MAYNOOTH UNIVERSITY

DOCTORAL THESIS

Expanding the Frame: Realising Engagement Through an Interactive, Visualisation-Based Search in Digital Humanities Research Environments

Author:
Shane A. MCGARRY

Supervisor:
Dr. John G. KEATING

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Computer Science
Head of Department: Dr. Joseph TIMONEY

January 8, 2020

Funded by:

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Contents

MAYNOOTH UNIVERSITY

Abstract

Science and Engineering
Department of Computer Science

Doctor of Philosophy

Expanding the Frame: Realising Engagement Through an Interactive, Visualisation-Based Search in Digital Humanities Research Environments

by Shane A. McGarry

As the Digital Humanities expands its modes of interrogation, it has continued to develop new ways of researching and exploring text. As a result, visualisations have risen to prominence as scholars have begun to explore concepts behind Moretti's *Distant Reading* and Jocker's *Macroanalysis*. While the study of these types of visualisations has largely focused on their ability to provide higher-level insights, little exploration has been conducted concerning their effectiveness within the context of a learning or research environment. Drawing on discussions from fields of psychology (specifically the role of working memory), education (modes and frameworks of learning), and computer science (usability and interaction design), this thesis attempts to discover the effects of interactive visualisations compared to standard keyword search approaches on a user's engagement with the overall system, as well as the effect on learning as a direct result of engagement. Centred around an 18th century manuscript detailing the expenditures of the Royal Irish College at Alcalá de Heneres, this thesis presents the design and implementation of the *Alcalá Record Books* and discusses a case study that was conducted to explore the effects of the visualisation-based search. Ultimately, this thesis advocates for the inclusion of an interactive, visualisation-based search as a complement to existing keyword searches, highlighting the advantages such searches bring to engagement, learning, and overall satisfaction with the system as a whole.

"The public is more familiar with bad design than good design. It is, in effect, conditioned to prefer bad design, because that is what it lives with. The new becomes threatening, the old reassuring."

Paul Rand

Acknowledgements

I would like to begin by thanking my husband, Joshua D. Savage, who has stood by me in constant support throughout the last 5 years, agreeing to move halfway around the world to support this crazy dream of mine. I couldn't have done this without you. I'd also like to thank my parents, Keith and Linda McGarry, who instilled in me a lifelong love of learning, supported me in my every endeavour, and have always pushed me to reach for my dreams and never give up. Your love and support have meant the world to me. Meredith Dabek and Stephen McCarthy, thank you for offering friendship and support as fellow postgraduates, sharing in (and sometimes mocking) my pain. You were both always there for a laugh or a shoulder to cry on. And of course, Rick Ramirez, my best friend in the whole world, thank you for listening to me complain and gripe and freak out. It's so rare to find a friendship that can span more than a decade and thousands of miles. Thank you for always being there for me.

To the Graduate Studies department, and especially Prof Maria Pramaggiore, who helped me navigate some of the hurdles I encountered mid-way through my studies, thank you for endless support and guidance. It was a long and bumpy road, but I ended up exactly where I needed to be. I have you to thank for helping to make that transition a reality. Also to the staff of the Maynooth Arts and Humanities Institute (formerly An Foras Feasa), thank you for allowing me to explore new avenues and for exposing me to areas of research I would have never considered on my own. I'd also like to thank the funders of this research, without whom I never could have embarked on this journey: Maynooth University (through the John & Pat Hume Scholarship) and the Irish Research Council (through the Government of Ireland Postgraduate Scholarship).

Finally, I wish to thank my committee, Prof Thomas O'Connor and Prof Fionna Lyddy, and my supervisor, Dr John Keating. None of this would have been possible without your patience, your guidance, your wisdom, and your inspiration. Each of you has had a profound impact on my research career, and I will be forever grateful for your time and energy. I'd like to extend a special thank you to Dr John Keating for taking me on rather late into my PhD and helping me reformat my work into something of which I can truly be proud. You threw me a life preserver when I was sinking in a sea of uncertainty and discouragement. I will be forever grateful for your help and guidance.

Dedicated to my amazing and supportive husband, Joshua D. Savage. You are my everything, and there is no way I would be where I am today without you by my side. I love you. Always and Forever.

List of Figures

List of Tables

Acronyms

ACD Activity-Centered Design. 75

ANOVA Analysis of Variance. 134, 147

DRE Digital Research Environment. 3, 4, 12, 17, 21–24, 26–30, 35, 37, 73, 74, 76, 78, 82, 87–93, 97, 99, 101–106, 108, 109, 111, 112, 116, 117, 123, 143, 153–156, 159–164, 166, 168–170, 173, 174, 385

ED Emotional Design. xix, 79–82, 116, 117, 123, 124

GDD Goal-Directed Design. 76–79, 82, 88, 89, 116, 117, 123, 167

GUID Globally Unique Identifier. 129

HCD Human-Centered Design. 75

HCI Human-Computer Interaction. 24, 26

KWIC Keyword In Context. 121, 168

LTM Long-Term Memory. 10, 11

NFC Need For Cognition. 42, 43, 81, 82, 86, 117

NMF Non-negative Matrix Factorization. xiv, xvi, 61, 62, 374

OOP Object-Oriented Programming. 121

OPL On-Paper Learning. 39

OSL On-Screen Learning. 39

SAS Supervisory Attentional System. 9

SOJ Second-order Meta-cognitive Judgement. 39

SSL Secure Socket Layer. 26

STEM Science, Technology, Engineering, and Math. 20, 130, 146, 158, 160

TEI Text Encoding Initiative. 168

UEQ User Experience Questionnaire. 117, 126, 130, 140–142, 154, 155

UI User Interface. 99, 111, 112, 168

UxD User Experience Design. 71

List of Symbols

$\chi^2(k)$	chi square distribution where k is the degrees of freedom
$\sum$	sum

Chapter 1

Introduction: Laying the Groundwork for a Theory-Based Approach

Reading is not a natural process, but rather is the result of a learned practice which we have continued to adapt to the ever-evolving technological landscape of the "book" [1]. From early wax tablets to the scroll and later the codex, the human mind has learned to adopt new methods of engagement with a source text in order to increase understanding and comprehension [2], [3]. As the Internet has evolved over the last 30 years, many scholars have sought to leverage the power and reach of this dynamic medium in order to disseminate their work to a larger audience or (in the case of libraries and archives) to create a digital preservation of an artefact. These technological advances have caused an evolution in the mode of scholarship, which has solidified around the field of Digital Humanities.

However, with the onset of the digital age and the digital reproductions created by Digital Humanists, new challenges have arisen regarding learning and comprehension when reading in digital environments—challenges which have never existed in prior media and which few within the field have sought to address. While some scholars in the humanities have begun to investigate alternative modes of textual engagement—such as Moretti's *Distant Reading* approach discussed in more detail in Chapter 2—in order to overcome these challenges, the work being conducted is often viewed through a lens of novelty and curiosity. While such a lens certainly lends itself to further investigation and research, it does little to help us understand the problem to be solved and the application of the solution within the wider ecosystem. The work herein seeks to address this very issue.

1.1 An Introduction to the Current State of Digital Humanities

When looking at work that seeks to propose alternative methods of engagement outside the scope of "close reading",[1] two particular scholars rise to primacy: Franco Moretti and his *distant reading* approach [4] and Matthew Jockers and his *macroanalysis* approach [5]. Both argue that their respective approaches allow for larger scale investigation of text and, like close reading, rely upon subjective analysis and contextualisation in order to gain new insights. However, their work primarily focuses on the investigation of the corpus / text and fails to explore issues related to learning and comprehension (the latter of which is distinct from the former). Johanna Drucker investigates the role of visualisations themselves not only as a means of knowledge production, but also as a mechanism for understanding both objective (i.e. quantitative) and subjective (i.e. qualitative) data [6]. In fact, Drucker goes so far as to define four distinct levels of interpretation via the use of modelling: modelling through "phenomenological experience in the making of humanities (data as *capta*...[that is], the representation of temporal and spatial experience)" [6, p. 135], modelling through relationships among documents (such as dates of creation, etc.), modelling through narrative, and modelling through interpretation [6]. In her four levels, the first two advocate for more objective (read: quantifiable) data approaches while the latter two are more subjective (read: qualified) approaches. It is through this additional lens that Drucker holds up that this work seeks to both evaluate the benefits of and promote best practices in the use of data visualisations within the ecosystems created within the Digital Humanities. But this work is also about moving beyond simple representations of design rules, which tend to focus on activities and their connections. Rather it is a more holistic approach to the design of immersive research environments with the Digital Humanities.

1.1.1 Asking the Question

In order to establish such a framework, however, more must be understood about the nature of the environment, the goals of the user, and the goals of the creator. If the purpose of the environment is to truly engage the user and allow for further knowledge dissemination (whether through in-depth, professional research methodologies or through public engagement), the question must be asked: "What effect do data visualisations, specifically within the remit of Digital Humanities, have on engagement with the source material—and by extension learning—in digital research environments?"

This question bears investigation for a number of reasons, but primarily due to the growing use of data visualisations within the humanities. While interesting, many of these visualisations are created either to showcase the novelty of the

[1] See Chapter 2 for more discussion on close reading and proposed alternatives.

technology or as an attempt to prove a hypothesis (but with a distinct lack of transparency around the underlying data model). As a result, the reader can find herself making erroneous suppositions regarding the inter-relatedness of data displayed within the visualisation, especially if the reader does not properly understand the tools and methods at play and attempts to draw assertions that *appear* to be valid but in fact are not.[2] Additionally, interactive data visualisations provide the reader with new modes of engagement that can shift not only the discourse around the text but also the discourse around usability, engagement, and learning. Understanding how data visualisations can affect the user experience and what contributions they can make to the learning and research processes serves to shift both our engagement with and design of the DRE. Finally, these digital environments created by Digital Humanists continue to grow in both scope and breadth; as such, it is our responsibility as creators of this content to ensure environments are constructed in such a way so as to optimise the research and learning experience. Otherwise, we risk creating nothing more than a digital preservation of an artefact that has lost much in its translation from the analogue to the digital.

1.1.2 Contextualising the Work within a Research & Learning Environment

Before continuing with this discussion, the work herein must be heavily contextualised. There is much work occurring within the Digital Humanities, and not all of it will apply to the theories and frameworks discussed herein. Therefore, the reader should consider this discussion within a pre-defined context: that of work that is a digital reproduction of an analogue text (usually historical manuscripts) where the text is presented alongside a digital reproduction of the analogue artefact. These environments generally contain both the artefact and the transcription (and where applicable, a translation as well) and some type of search and/or browse interface. While these types of environments go by many names, each with slight variations to separate them from the whole,[3] for the purposes of context, the author has created a larger umbrella under which to discuss this work: the Digital Research Environment (DRE). Any environment which meets the former description and whose goal is to provide an artefact for the purposes of further historical research or dissemination of knowledge to the general public shall fall under this larger umbrella for the purposes herein.

1.2 Foundational Work

While later chapters delve more deeply into the core arguments of this work (such as the issue of reading in digital environments discussed in Chapter 2 or the importance

[2] This particular issue is explored more fully in Chapter 2.

[3] Such environments are often referred to as digital archives, thematic research collections, digital scholarly editions, and so on.

of interactivity and immersion with regard to engagement as discussed in Chapter 3), what follows here is a discussion of a number of concepts that lay the very foundations for the larger arguments as a whole. As this work is interdisciplinary—touching on aspects of psychology, history, education, and computer science—it would be impossible to weave the various threads together without first constructing the loom upon which they lay.

For the remainder of this chapter, a structure will be built by drawing on concepts inherent primarily in education and psychology. While these sections will not lead the reader to an immediate understanding of the state of the art, they are important here as to lay the groundwork for later arguments within the thesis. This section will open with a discussion around the concept of Working Memory—a core, foundational concept which will be drawn upon repeatedly throughout the work. Following on will be a discussion of learning theories and the role each plays in overall system design (which is, itself, pedagogical in nature) as well as a discussion centred around inductive versus deductive reasoning—a critical component for understanding the various challenges at play within the DRE. This will then conclude with a discussion of Activity Theory and the various components which comprise and influence its use within this work, focusing on the particular framework developed by Diana Laurillard which was the authority upon which much of the design of this work was based.

1.2.1 Working Memory

The way individuals learn is intimately connected to how humans process sensory information via attentive processes and how such information is stored and retrieved—a process more commonly referred to as *memory* [7]. The earliest models of memory involved the concept of long-term versus short-term memory. In 1968, Atkinson and Shiffrin proposed the *multi-store model* of memory which became the predominant model for understanding memory architecture [8], [9].[4] According to their model, memory is effectively broken into three primary components or *stores*: sensory (one store—also referred to as a register—for each of the five senses), short-term (which has a limited storage capacity and is meant to store information for only a short period of time), and long-term (which is nearly limitless in its capacity and can hold information over long periods of time) [8], as seen in Figure 1.1. Information could degrade from the sensory store via decay,[5] from short-term store

[4] It should be noted that Atkinson and Shiffrin's model is no longer the predominant model in cognitive psychology and has been replaced with more modern approaches such as those detailed herein.

[5] Sperling notes in his work that the visual sensory store, for example, has a decay rate of about 500 milliseconds [10]. If the short-term store is otherwise occupied, information from the sensory store(s) could be lost, thus leading to the concept of *decay*.

via interference,[6] and from the long-term store via loss of strength within the store itself.[7]

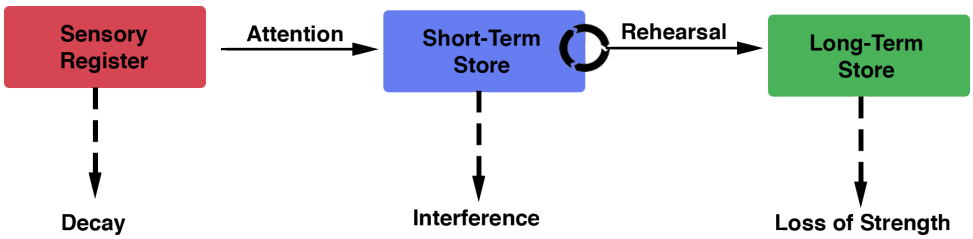


FIGURE 1.1: Schematic of Atkinson and Shiffrin's model of multi-store memory.

Atkinson and Shiffrin's model relied on earlier work by Sperling, who theorised that sensory stores are *pre-attentive*—meaning that they do not rely upon attention [8], [10]. Sperling additionally introduced the idea of *iconic (visual) memory* through his experiments in 1960. Here, Sperling provided study participants with a screen that flashed images of letters one at a time for a fraction of a second. Most study participants were able to recall some of the letters but typically no more than 4 or 5 [12]. Sperling theorised that this was due to the fleeting nature of the iconic memory store and its *pre-attentive* state. Baddeley and Hitch believed there was more involved with simple repository stores for memory, especially when considering the data from patients with brain damage whose ability to recall information did not fit neatly into the idea of a short-term vs long-term store. Baddeley and Hitch theorised there was a more robust system at play and sought to understand exactly what affected storage and recall. This led to the development of a new model for short-term memory, using the concept of working memory.

Baddeley's Model of Working Memory

In 1974, Baddeley and Hitch proposed a model for working memory that transformed the concept of short-term memory into a more robust, complex system [9], [13]. Their model stipulated that working memory[8] went beyond simple temporary storage for memory purposes, but rather acted as a type of workspace for tasks that were not memory-specific. Three primary components constituted the initial model: the visuospatial sketchpad, the phonological loop, and the central executive (see Figure 1.2 for a simple model of Baddeley and Hitch's early proposed system of working memory).

[6] Interference is caused when a new memory is found to be competing with resources during the commitment or retrieval of other memories. Interference theory is used to explain much of the nature of conflict within the working memory system, such as articulatory suppression, list length, and part-set cueing [11]

[7] Consider, here, examples of memory loss associated with ageing, trauma (such as amnesia or brain damage), illness (such as dementia), and (most predominantly) lack of use.

[8] Schacter, Gilbert, Wegner, et al. provide the following definition of working memory: "[working memory is the] active maintenance and manipulation of information in short-term storage" [7, p. 184].

FIGURE 1.2: Schematic of Baddeley and Hitch's early model for working memory.

According to Baddeley and Hitch's early theory, each of these components served a particular purpose: the visuospatial sketchpad was used for processing visual and spatial information, the phonological loop was for processing auditory information, and the central executive served as a type of processor: combining information from the other two components and integrating them into long-term memory. The central executive also assists with recall from long-term memory, passing off information to the visuospatial sketchpad and phonological loop as needed. A fourth component, known as the episodic buffer, was added in 2000 after Baddeley determined there was some type of cooperation between the visuospatial sketchpad and the phonological loop [14], [15]. Figure 1.3 describes the addition of the episodic buffer to the working memory model. As each component plays an integral role in the both the reading process and with comprehension as a whole, a short exploration of each component is necessary.

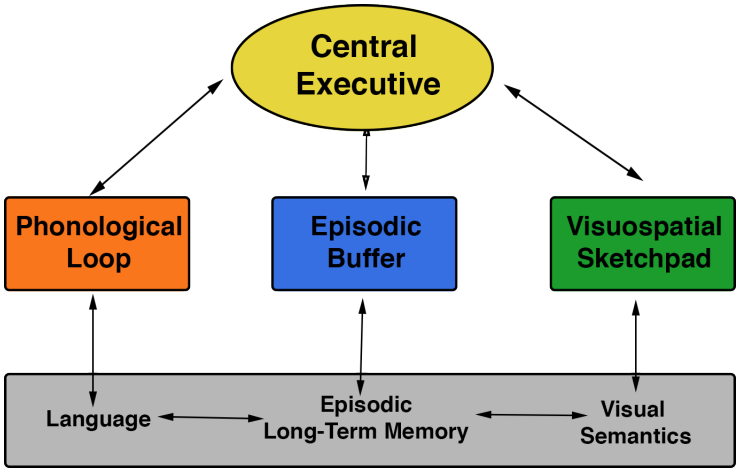


FIGURE 1.3: Schematic of Baddeley's model incorporating the episodic buffer.

Phonological Loop The phonological loop handles auditory information (primarily speech) in the working memory system; the most common function of the phonological loop is to maintain the order of words [9]. The phonological loop is also leveraged during *inner speech*—speech which is not audible but occurs within the mind—and as a result, the phonological loop plays a significant role in reading (whether silently or aloud) [16]. While the visuospatial sketchpad plays a vital role in the processing of text on screen, the phonological loop plays a significant role in the *retention* of information through its two subcomponents: the phonological store (or *inner ear*) and the articulatory control process (or *inner voice*) [13].

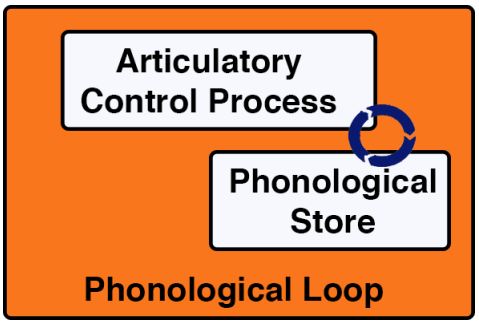


FIGURE 1.4: The phonological loop model as proposed by Baddeley and Hitch.

As seen in Figure 1.4, the articulatory control process acts as a rehearsal mechanism for information in the phonological store. As information enters the store, it is rehearsed in the articulatory control process and then eventually handed off to the central executive for processing (either through the episodic buffer or directly into long-term memory). It is this rehearsal mechanism which allows for comprehension during reading.

Visuospatial Sketchpad The visuospatial sketchpad is the area of the working memory system responsible for processing visual (*what*) and spatial (*where*) information [9]. Like the phonological loop, the visuospatial sketchpad plays an important role in the reading process by assisting with recall on the page through an orientation of the text within the page position.[9] Additionally, the visuospatial sketchpad can be used to interpret semiotics and other visual indicators (such as colour, shape, and pattern recognition). Recent research has also pointed to a reliance on the visuospatial sketchpad for the interpretation of haptic (i.e. touch) feedback [17].

Logie elaborated on the visuospatial sketchpad in 1995. In his work, Logie split the visuospatial sketchpad into two primary working components: the visual cache and the inner scribe [18]. Under Logie's model, the visual cache acts as a sort of store for visual information (such as colour or form). It is a passive component in and of itself—its purpose is only to store static information regarding visual patterns. The second component, dubbed the inner scribe, is an active component system that is responsible for the transfer of information from the visual cache to the long-term store [18].[10] Additionally, the inner scribe can store information regarding movement, sequence, and physical action.[11] Logie's model for the visuospatial sketchpad can be seen in Figure 1.5.

[9] More on the role of the visuospatial sketchpad in reading and digital environments can be found in Chapter 2.

[10] Transfers to the long-term store occur either via the central executive or via the episodic buffer and central executive.

[11] Because the inner scribe contains information regarding the control of physical action, Baddeley has theorised in his later work that the visuospatial sketchpad as a whole also serves as a storage and processing ground for haptic feedback [17], [18].

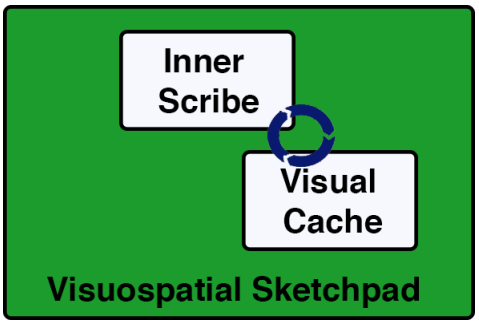


FIGURE 1.5: Logie's model of the visuospatial sketchpad.

Episodic Buffer The third component, added by Baddeley in 2000, is the episodic buffer, which acts as a sort of temporary store for holding *chunked* information—that is, it integrates input from the visuospatial sketchpad and the phonological loop. The idea of the episodic buffer was developed in an attempt to explain some inconsistencies observed by Baddeley in the original model. A set of experiments tasked readers with random key presses while attempting to read three different types of prose: a fairy tale ("easy" to comprehend), prose describing a medical condition ("medium" difficulty in terms of comprehension), and an excerpt of philosophical text (deemed "difficult" to comprehend, thus requiring more cognitive effort) [15]. The expectation of the experiment was that the key presses would be more random—thus indicating less cognitive effort—with the fairy tale than with the medical prose, and more random with the medical prose than with the philosophical text, the idea being that as cognitive effort rises (e.g. the comprehension of more *complicated* texts), there would be less available resources to continue to press keys at random. In fact, what Baddeley, Allen, and Hitch discovered was that the randomness of the key pressing did not change depending upon the readability of the text but rather remained consistent across all three texts, indicating a larger system at work [15].

Baddeley later expanded on this theory by examining retention rates of patients with amnesia; most of the patients he examined demonstrated great difficulty in both immediate and delayed recall of a piece of text they were tasked with reading. However, a small number of patients, who were determined to have "high intellectual capacity" [15], were able to demonstrate strong immediate recall of the text (but just as poor delayed recall as other patients). This suggested to Baddeley, Allen, and Hitch that these patients were able to leverage the central executive, largely thought to be an attentive process, to support recall; however, as this strategy seemed to only be effective in "high intellectual capacity" patients, it clearly was not a common strategy for the general populace. As a result, the episodic buffer was born.

In its initial concept, Baddeley theorised that the episodic buffer acted as both a passive store and an active process which served the role of integrating information between the phonological loop and visuospatial sketchpad, thus allowing the central executive to focus on attentional control [14]. Baddeley and his team set out to prove

this theory through a series of experiments designed to test both visual and verbal binding in memory.[12] However, the experiments produced no strong indication of the working memory system being a necessary component for verbal or visual binding. This led Baddeley, Allen, and Hitch to determine that the episodic buffer is not a system that *actively* integrates information but rather acts as a multidimensional storage mechanism which the central executive can use as a type of temporary store as it attempts to integrate information relayed from the phonological loop and the visuospatial sketchpad. In addition, the episodic buffer can also be used to store information retrieved from long-term memory, which will be integrated into working memory and used to assist the central executive with the parsing and integration of information from the phonological loop and visuospatial sketchpad. Thus the episodic buffer can be seen as a type of RAM storage within the working memory system: a temporary store designed to assist in both permanent storage and the integration of other multi-modal data for further analysis.

Central Executive The final component of Baddeley's model is the central executive. Referred to in his work as the "all-powerful homunculus" [17, pp. 223, 225], the central executive serves a multi-dimensional purpose—each of these purposes identified as *executive processes*. One of its primary roles is that of an attentional control mechanism with three primary executive processes: the ability to control (or maintain) attention, the ability to divide attention, and the ability to switch attention (also seen as the ability to switch between tasks) [17]. Much of this attentional model was based on Norman and Shallice's model of the executive function of attention known as the Supervisory Attentional System (SAS) [21]. The SAS theorises that the executive function of attention is composed of a series of learned actions that are then translated into scripts. As the system encounters situations which require attentional control, the SAS can leverage these scripts in order to make decisions related to attention [20]. Similarly, the central executive monitors attention and directs it to either:

(a) focus attention on a particular task

(b) split attention between multiple tasks

(c) switch attention to another task

This attentional switching plays a vital role in reading comprehension and encounters unique challenges with regard to digital environments.[13]

Another executive function crucial to the central executive is its ability to interface with long-term memory. This two-way interface allows the central executive

[12] These experiments are detailed in Baddeley, Allen, and Hitch 2010 paper entitled *Investigating the Episodic Buffer* [15] and in his follow up paper in 2011 entitled *Binding in Visual Memory: The Role of the Episodic Buffer* [19]

[13] For immediate reference, see [22]–[24]. For a more in-depth discussion of how attention affects reading and the unique challenges of attention in digital reading environments, see Chapter 2.

to write information to the Long-Term Memory (LTM) store, as well as retrieve information from LTM and store it in the episodic buffer in order to integrate it with other information in the working memory system.[14] Part of this integration also involves the *encoding* of information from the phonological loop and the visuospatial sketchpad[18]. This encoding allows for the integration of information stored in the episodic buffer which originated in the phonological loop and the visuospatial sketchpad and, when integrated, forms a more complete memory.

As an example of this phenomenon, consider the act of reading this thesis. As the reader scans the words on the page, the visuospatial sketchpad is using the visual cache and the inner scribe to encode and interpret the characters on the page into words. Simultaneously, the central executive is pulling information from LTM in order to assist with the comprehension of the words. As the words are articulated using the *inner voice*, the phonological loop is engaged to capture the reading in the working memory system. The data in both the phonological loop and the visuospatial sketchpad are stored in the episodic buffer, and the central executive is then engaged to integrate the multidimensional data into an array that can be encoded into memory, all the while keeping the attentional process engaged and focused on the task of reading.[15] But how exactly does the central executive maintain attention while integrating data from the episodic buffer? The answer can be found in Miyake, Friedman, Emerson, et al.'s Unity/Diversity Framework.

Unity/Diversity Framework

Miyake, Friedman, Emerson, et al. sought to categorise the executive functions of the central executive, which go beyond the four primary subsystems listed above, leading to the creation of the unity/diversity framework: a framework which describes executive functions as having common underlying functionality plus additional functionality that is unique to each function [25]. As part of their work, Miyake, Friedman, Emerson, et al. defined three standard functions (which subsume each of the 4 executive functions described by Baddeley). These functions are [9]:

1. *Inhibition Function* - the primary purpose of this function is to override prevailing responses or to prevent distraction. The most common example of the inhibition function is demonstrated by the *Stroop Task* where words for colours are written on a background of a conflicting colour (for example the word *red* is written on a blue background). When asked to name the colour of the card, the participant must leverage the inhibition function to stop themselves from reading the word printed on the card rather than stating the colour of the card as instructed

[14] This latter integration plays a crucial role in reading comprehension as discussed in Chapter 2.

[15] The process of reading and the differences between analogue and digital reading are discussed in depth in Chapter 2

2. *Shifting Function* - this function is the primary function for switching between tasks or shifting attention from one set to another

3. *Updating Function* - the updating function is responsible for both *adding* and *deleting* information in the working memory store. Logically, one can extend this to the function of the central executive in Baddeley's model that also retrieves and writes information into LTM

While each of these functions performs a specific task, they also contain a common executive function which Miyake and Friedman describes as "one's ability to actively maintain task goals and goal-related information and use this information to effectively bias lower-level processing" [26, p. 11]. Integrating Miyake, Friedman, Emerson, et al.'s work into Baddeley's model of the central executive (see Figure 1.6) provides a more holistic view of how the central executive manages attention and working memory as a whole—concepts crucial to understanding the problems with digital reading as discussed in Chapter 2.

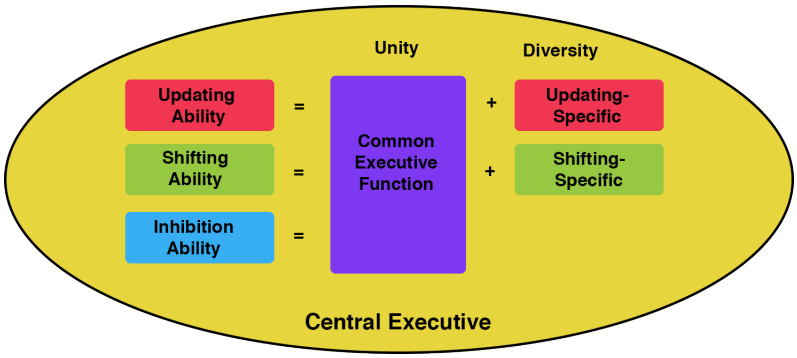


FIGURE 1.6: An integration of Miyake, Friedman, Emerson, et al.'s unity/diversity framework with Baddeley's central executive. Note that the *inhibition function* does not have its own specific diversity item as it is often seen as a parallel of the common executive function[9].

Working Memory: The Complete System

By integrating Miyake, Friedman, Emerson, et al.'s work with Baddeley's models as they evolved over time, a richer understanding of the working memory system is obtained. As information is initially captured by the appropriate sensory store (in the case of working memory, either the phonological loop or the visuospatial sketchpad), the central executive is able to pass information off to the episodic buffer for temporary storage (via the *updating function*), shift attention or resources to the integration of information within the episodic buffer (via the *shifting function*) or leveraging either the common executive function or (if necessary) the *inhibition function* to maintain attention and prevent distraction. This full model of working memory (see Figure 1.7) will be referenced throughout this thesis.

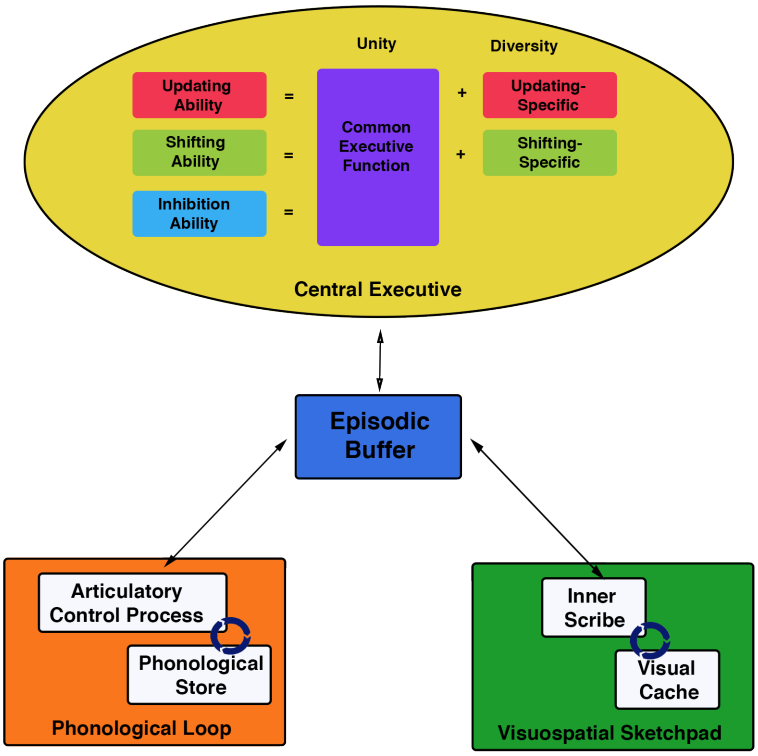


FIGURE 1.7: The full model of working memory integrating the work of Baddeley and Hitch, Logie, and Miyake, Friedman, Emerson, et al.

Working Memory and Reading

While reading will be explored in more depth in Chapter 2, it is worth discussing briefly here the role of working memory with regard to reading comprehension. The literature demonstrates a strong correlation between working memory and reading comprehension in both adults [27] and children [28]. While much of the literature discusses the role of the phonological loop in reading comprehension [29], [30], further discussions regarding the role of higher level-functions in the central executive highlight the importance of subsystems beyond that of just the phonological loop [31], [32]. Given the literature around the role of working memory as it relates to reading comprehension, an understanding of the working memory system will be essential for future discussions of reading (and other forms of knowledge acquisition) which occur throughout this thesis.

1.2.2 Theories of Learning: An Overview

Just as working memory plays a vital role in reading and comprehension, so too does learning theory, especially when considering the application of pedagogical principles to digital environments where learning and research are primary goals (as is the case with the DRE). When considering the notion of learning and how individuals acquire knowledge, numerous theories abound, and it is beyond the scope of this

work to attempt to discuss them all. However, there are three theories whose primacy in the subject of education and pedagogy is rarely questioned: behaviourism, cognitivism, and constructivism.[16] While there is no "one size fits all" approach to learning, there are definite pros and cons to each theory of learning and a general stance on one theory must be adopted for the purposes of this writing. What follows is a brief discussion of the three primary theories and a discussion of how the selected theory is applied to this work.

Behaviourism

The oldest of the three categories of learning discussed here, behaviourism developed from the work of Watson in 1913. In his seminal paper on the subject, Watson called for a new approach to how psychologists approached the understanding of the mind:

> Psychology as the behaviourist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behaviour. Introspection [the prevalent theory at the time, *ed. note*] forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. [36, p. 158]

By calling for this shift in approach, Watson established a new theoretical framework for studying the ways in which individuals learn by focusing on the *behaviours* of the learner. Under his definition, behaviourism has four basic principles:

- Only that which is observable should be the focus of study. The idea of objectivity, especially concerning what can be objectively measured, is the cornerstone of the behaviourist approach [37]
- Only behaviour can be objectively explained (not consciousness or thought) [37]
- Theories should be as simple as possible and should constrain themselves to the scientific method [37]
- The goal of any theory should be to break down behaviour into individual, "irreducible constructs" [37, p. 17]

This focus on behaviour translates into learning as an extension of *behaviour* acquisition, rather than *knowledge* acquisition [38]. To the behaviourist, only those traits which are directly observable—and thus by extension quantifiable under the scientific process—can truly be studied; all other data is too subjective and thus not

[16] While there are numerous other theories, such as connectivism [33], experiential learning [34], cultural-historical learning [35], etc., it is the author's belief that these other theories can reasonably be placed under one of the three broader categories discussed at length here. Thus, for the purposes of brevity, only behaviourism, cognitivism, and constructivism will be discussed.

empirical [39], [40]. Therefore, within the classroom setting, learning occurs through the construction of situations which reinforce the desired behaviours.

Behaviourism eventually fell out of favour and is now the least popular of the three aforementioned categories of learning [41]. However, it has had a lasting impact on theories of learning. For example, studies of language have moved away from philology (the history of language) to empirical studies of language use, such as linguistic analysis [38].

Cognitivism

More recent theories of learning have come to focus on how the structures of the mind (such as working memory) affect the learning experience. Cognitivisim attempts to understand and map the cognitive processes of the learner and determine the role these processes play in learning [41]. While behaviourism seeks to correlate man and animal (highlighting the similarities in adaptability of behaviours between man and animal test subjects), cognitivism seeks to set man apart based on the function of the brain [42]. Consider, for example, Anderson, Bothell, Byrne, et al.'s Adaptive Control of Thought-Rational (or ACT-R). ACT-R posits that there are multiple subsystems that comprise the mind, each with their own distinct role, and each controlled by a specific area of the brain. Of the various subsystems theorised, four are seen as critical to cognition: the retrieval module (located in the inferior ventrolateral prefrontal cortex), the imaginal module (located in the posterior parietal cortex), the goal module (located in the anterior cingulate cortex), and the procedural module (located in the caudate nucleus in the basal ganglia) [9], [43], [44]. Two additional modules—the manual module (located in the motor cortex) and the visual module (located in the fusiform gyrus)—concern themselves with fine motor control and the ability to see and interpret information respectively. A model of ACT-R can be seen in Figure 1.8.

ACT-R is discussed here, not so much for its application to the overall work discussed herein, but more to highlight the distinct differences between behaviourism and cognitivism. ACT-R is very strongly concerned with the internal structure of the brain and looks to these various structures to explain cognitive ability (or in the case of abnormalities, cognitive impairment). For example, the visual module attempts to explain visual processing and attention. As such, a cognitivist approach would attempt to explain how this structure supports reading and how digital reading may present certain challenges to underlying cognitive structures of the brain.

Another cognitive framework model is the one presented by Torey and Dennett. In *The Crucible of Consciousness*, Torey and Dennett refer to their model as the *mind-system* that is similar in nature to the work of Anderson, Fincham, Qin, et al., but less technical and slightly more abstract. The *mind-system* model focuses on the need for the physical structures of the brain (such as cross-hemispheric links in the corpus callosum) to give rise to reflective processes of thought, transcending that of the animal hind brain. The *mind-system* model posits that, from an evolutionary perspective,

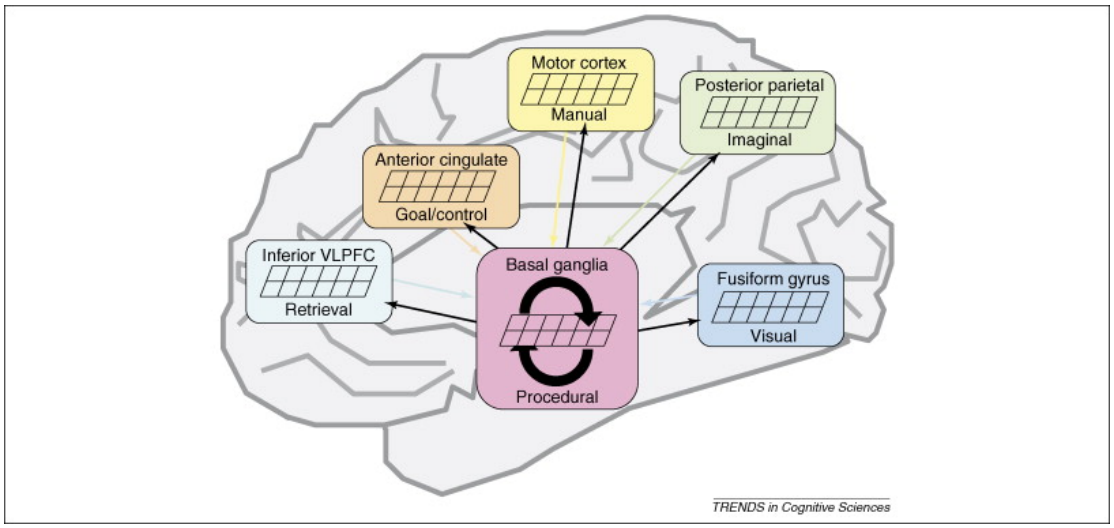


FIGURE 1.8: A representation of Anderson, Fincham, Qin, et al.'s ACT-R showing the main modules of the framework and their hypothetical locations within the brain. Reprinted with permission from Elsevier.

the human mind has developed beyond that of animal-like behaviour, providing us with levels of both conscious and unconscious thought that separate man from beast [45]. The work of cognitive scientists highlights the underlying differences between behaviourism and cognitivism by focusing on these structural differences within the brain that set man and beast apart.

Cognitivism plays a crucial role specifically within the field of computer science, as many of the modern metaphors used to understand and explain cognitivism are themselves drawn from the field of Computer Science, seeking to explain the cognitive structures of the brain as that of a computer system [42], [46].[17] The focus shifts to that of underlying component models (such as working memory, short term and long term memory, etc.) and that these individual systems can be isolated and studied independently: an idea that contradicts behaviourism, which holds that such processes cannot be independently isolated and studied [46]. Whereas behaviourism is focused on drawing connections between external stimuli (i.e. behaviours), cognitivism focuses on the how knowledge is connected to existing structures within the mind—an entirely internal process [47]–[49]. From an education and learning perspective, this means shifting the attention away from behaviour and towards those techniques, such as *chunking*, which derive the most utility from the underlying cognitivie sub-structures (e.g. attention, working memory, etc.).

Finally, cognitivism proposes an objective representation of the world or of the

[17] It should be noted that this is a circular reference as computers have always been designed within the constraints of the metaphor of the mind. Even more modern fields of study within computer science, such as Machine Learning and Artificial Intelligence, draw on our understanding of cognitive structures within the brain. Additionally, ACT-R (mentioned earlier) is one such example of a computational model of a cognitive architecture.

knowledge being acquired, through the construction of what is referred to as a *mental model* [41]. At its core, cognitivism revolves around the idea of how knowledge is constructed in the mind: "In cognitive learning theory, the key to learning and behaviour involves the individual's cognition, meaning a person's perception, thought, memory, and ways of processing and structuring information" [50, p. 130]. Supporters of this theory believe that the role of the teacher is to "create an accurate representation of the external world in our minds" [41, p. 10]. It is this representation of knowledge that is directly juxtaposed by constructivism, which seeks to integrate cognitivism with the lived experience of the learner.

Constructivism

While constructivism draws from cognitivism in that it also focuses on the underlying substructures of thought and memory, it moves beyond the notion of objective knowledge conveyed through a singular model and instead sees the learning process as knowledge filtered through the lens of experience [51]. The constructivist seeks to build upon additional scaffolds of knowledge and experiences of the learner, thereby allowing the learner to *construct* new knowledge. The constructivist believes that learning comes "from experience phenomena (objects, events, activities, processes), interpreting those experiences based on what we already know, reasoning about them, and reflecting on the experiences and the reasoning" [52, p. 2]. Under the constructivist paradigm, knowledge is heavily reliant upon notions of experience. Jonassen, Peck, and Wilson lay out five attributes which define their approach of "meaningful learning". In order for a learner to truly benefit from the experience, the activity must be:

- Active - humans are adaptable and learn through manipulation of an object or their environment, thus adapting their responses to environmental stimuli [52].
- Constructive - learners need to also reflect on their learning activity in order to construct mental models. This allows them to iterate on their learning, determine what additional data they need, etc [52].
- Intentional - since all human behaviour is goal directed [53], humans learn best when they have a clearly articulated goal or intention [52].[18]
- Authentic - we learn best when the information is properly contextualised. Stripping information of its context in an effort to "simplify" typically makes learning more difficult, not less [52].
- Cooperative - as social beings, humans learn best in collaborative environments [52].

[18] This attribute is later reflected in the Goal-Directed Design approach discussed in Chapter 3.

These attributes highlight the underlying foundation on which the arguments put forth in this work rest. It is through the construction of knowledge based upon both the learner's past experiences with the subject matter (including the learner's own enculturated context with the material) and the application of information in a same or similar context that new knowledge is born. While the author acknowledges the importance of understanding the underlying subroutines (such as working memory) which facilitate the learning process, the argument here cannot ignore the importance of cultural expectations and individual experience and their effects on the learning process. Thus a discussion with regard to learning must take into consideration not only the cognitive mechanisms through which knowledge is acquired, but also the context in which the learner applies this knowledge and the learner's enculturation within the sphere.

1.2.3 Pedagogical Approach to Reasoning: Examining Deductive versus Inductive Approaches to Learning

With the groundwork laid for a constructivist approach to learning (the underlying theme of which is carried through this work, especially within the context of Laurillard's *Conversational Framework* discussed further on), a further discussion around modes of reasoning inherent in the various fields of enquiry becomes relevant as it became an emergent trend throughout the course of this work (as discussed in Chapter 5). Additionally, it is important to consider that individuals are unique with their own preferences and approaches to reasoning. While each field of enquiry has different pedagogical approaches it leverages to train its scholars[19]—thus allowing assumptions to be made for how most scholars in a given field will approach the scaffolding process—each individual develops their own strategies for learning and applying knowledge.

However, when considering the role of the DRE in Digital Humanities as both a research environment *and* a learning tool, a pattern begins to emerge among users: those who follow an inductive reasoning approach (common to the Humanities) and those who follow a deductive reasoning approach (common to those in the Sciences).[20] As these different approaches to reasoning and learning have drastically different impacts on how the individual will leverage tools within an ecosystem—and furthermore the *types* of tools required by different reasoners—it is worth exploring the modes of instruction that privilege one type of reasoning over another and how these two types of reasoning inform the learning process.

[19]It should be noted that any discussion of *scholar* here is in the context of third-level education. While some of the work discussed within could also be extrapolated and applied to primary and secondary education contexts, this work concerns itself with tertiary education and beyond, and the methods employed to train scholars in their respective fields.

[20]This emergent theme is discussed in greater depth in Chapter 5.

The Inductive Reasoner

The inductive reasoning process focuses on moving from small details to a generalised conclusion that is the most likely probable solution to the problem given the underlying premises are true [55]. Due to its less definitive nature with regard to the conclusion, this type of reasoning is well suited to most humanities research: the humanities seeks out truth hidden between the layers. There is no one, single objective truth but instead interpretations of history, literature, art, music, etc. that can be seen as valid based on upon the evidence presented by the scholar. This is most readily seen in category-based induction where a scholar is able to make a larger generalisation about a group based on information observed in record [55].

In the Digital Humanities—and by extension the *traditional* humanities fields that inform much of its work, such as History, Literature, Archival Studies, etc.—this mode of inductive enquiry is at the core of its scholarly output. Jennifer Edmond highlights this use of inductive reasoning in her recent article entitled: "How Scholars Read Now: When the Signal *Is* the Noise" [56]. In her article, she discusses how humanities scholars leverage distraction in their reading and in their research as a method of gathering together the various bits of scholarly insight and building to a conclusive whole.

> Distraction is not a problem for scholarly reading in the humanities: it is, in fact, intrinsic to how we work. An openness to distraction facilitates knowledge creation in conversation with and between sources, bringing together disparate times and places, authors and forms of source material, frames of reference and layers of insight, harnessing peripheral vision as much as central focus to create the conditions for serendipitous discovery. This “loose hold” on focus results in a knowledgescape with a substantial tacit dimension, with the result that humanists most certainly know more than we can tell. [56, sec. "Abstract"].

Guillory further asserts this "distracted" or "messy" approach to reading, discussing the role of reading for keywords or dates and looking for various "signals" that call for a more focused reading.

> The technique of scanning can be organized, alert to keywords, names, dates, or other features of a text. This technique is a form of attention, even heightened attention, although the scholarly reader might ignore the continuous meaning of a text, deferring comprehension until some textual signal brings the scanning process to a temporary halt and initiates a more intensive reading. ... At some point reading must be decelerated for the purpose of a more analytic reading, which aims to correct distortions produced by scanning. [57, p. 13].

These discussions around reading actually illustrate why the inductive reasoning process is so inherent in most humanities fields of enquiry. As the historian or literary theorist seeks to build their argument, they must compile the evidence for their

assertions by drawing upon secondary literature within the field. The scholarship of humanities research is inherently "messy"—it must be built up using gems of insight in order to fashion a larger picture. It is akin to the construction of a mosaic: each individual coloured tile is a piece of the larger picture. Thus the humanist must build *up*, from small details to a larger narrative.

This mode of meaning-making is seen in how the humanities teaches its students to close read.[21] Kersh and Skalak mention the teaching of Dickens's *Oliver Twist* and the exploration of the "historical implications of industrialization, including background on the Poor Laws, mortality rates, and the general squalor of nineteenth-century London" [58, par. 5]. History, especially, applies a constructivist approach to its scholarship with a strong thread of inductive reasoning throughout. "[I]t is important to remember that the past is only ever *approximately* apprehended: one cannot simply 'go back,' but must rely on constructed memories, artifacts, written accounts, recordings, etc." [59, par. 1]. Furthermore, the use of paratexts (or texts which act as a frame for single work) is a common device in the humanities [59], and their use highlights this constructed, inductive reasoning approach to humanities scholarship. Again, it is the individual pieces which are pieced together to make the whole.[22]

Many of the tools being constructed in Digital Humanities are geared towards this process. Consider the rise of annotation software which allows for a type of "recursive reading" of the text [58]. Annotation allows for the reader to mark certain aspects of text and return to them for further in-depth review. This type of approach to teaching scholarship reinforces the transfer of knowledge and allows students to begin to apply knowledge from one contextualised space to another [58], [60]. Annotation can also support collaborative approaches to learning and scholarship, one of the key requirements of humanists with regard to the technologies they require to conduct their work [61]. Here again, this notion of annotation, which allows the reader to create small subsets of thoughts across a larger corpus and then compile such annotations into a larger body of work, is a prima facie case of inductive reasoning.

While it is worth noting that there are differences in every individual and not all individuals within the humanities are inductive reasoners (nor will even inductive reasoners rely *solely* on inductive reasoning approaches), the literature demonstrates that the inductive reasoning process is a core component of the pedagogy and methodology which underpins the scholarship of the humanities. The methods of interrogation and the process of meaning-making rely to a great extent on the inductive reasoning process. Thus, it is worth understanding how this process differs from the deductive reasoning process inherent in most science-based fields, as will be discussed next.

[21] Close reading is discussed in further detail in Chapter 2.
[22] This point will be contrasted with the deductive reasoning approach later on.

The Deductive Reasoner

Unlike inductive reasoning—which seeks the most probable solution by starting with the details and generalising out to a conclusion—deductive reasoning seeks to reach a conclusion through a gradual breakdown of knowledge to a singular truth. This is an approach that is generally common in STEM fields of enquiry where the researcher attempts to explain a phenomenon by tracing factual knowledge through various observations that can then be conclusively reproduced, given the same set of parameters. It is seen as the objective approach, as it leads to a singular interpretation of data based upon the facts known [62]. This so-called *objective truth* requires a different line of reasoning than the subjective form of humanities enquiry, as inductive reasoning suggests the most *probable* answer as opposed to the *only* answer. Additionally, the scaffolding process is inverted for the deductive reasoner, who starts with the general and breaks down to a single result.

The nature of most science-related scholarly work is observation-based. While the field also builds upon existing literature (as all scholarly work must), much of the work being conducted within STEM fields is based upon observed phenomena. Khan uses this observation-based approach to study the effects of computer simulations in science pedagogy [63]. Using the theory to inform a general hypothesis, Khan constructs a case study to test ideas and then break down the data into a formalised conclusion. This methodology is deductive by definition. Ali also demonstrates this deductive approach to pedagogy in their construction of computer science modules, relying upon constructivist theories of scaffolding:

> In my introductory C++ course, I successfully applied the approach of scaffolding...I assigned the following problem to my students for classroom discussion.
>
> Write a C++ program which will read a series of integers through the keyboard and identify (1) the largest integer, (2) the smallest integer, (3) only even integers, (4) only odd integers, and compute (5) the sum of all even integers, (6) the sum of all odd integers, and (7) the sum of all the integers. Print all the input integers, the largest integer, the smallest integer and the three sums on the screen. Create several user-defined functions (subprograms) to write this program. [64, pp. 246–257]

In this above exercise, Ali instructs the students to break down the problem and moves through the reasoning process that details how to effectively construct functions.

This approach is a common one in computer science even outside of education settings, as demonstrated in much of literature around software design [65]–[67]. A major component of the software engineer's job is to take requirements and distil them into multiple components. This, too, is reflective of the deductive reasoning approach where the larger item is broken down into a singular (or in this case multiple smaller) component(s). Even the construction of the component itself is then

further broken down into properties, methods, and events. There is a constant reinforcement of this deductive approach of moving from the larger, more generalised body of information to the smaller, more concise conclusion.

Additionally, this is not just inherent in computer science education or in the construction of software (although it is perhaps the most germane example given the subject of this work). The scientific method, utilised throughout the "hard sciences", is also a deductive approach to reasoning [68]. It begins with a hypothesis and then through observation, testing, and analysis, a conclusion is presented that is the only conclusion based upon the data collected and observed.

It is perhaps this difference that most sets apart the humanities from the sciences. While the sciences benefit from the ability to (generally) observe their data and follow a clear line from hypothesis to conclusion that leads to a singular answer, the humanities must construct their arguments based solely upon secondary literature. One cannot *observe* history, nor can the construction of an 18th century novel be *observed* by the modern-day scholar. The scholar must instead rely upon extant sources and paratexts in order to draw conclusions that are the *most probable* given the evidence at hand.

Reasoning and Digital Humanities

However, because Digital Humanities is an attempt to incorporate technology into the study of the humanities, it is perhaps uniquely poised to leverage *both* the inductive and deductive processes. As many of the scholars who construct DREs are historians or literary theorists, the research they conduct on the subject matter at the core of the DRE in question will likely follow an inductive reasoning approach. However, when proper system design is applied—as evidenced by the work discussed in Chapter 5 and in the conclusion in Chapter 6—the digital humanist is also employing a deductive reasoning approach. As the field broadens and begins to draw in scholars from non-humanities fields (such as computer scientists), it becomes more evident that the software design must also support diverse modes of reasoning and thinking. A better understanding of these needs can be gleaned from various design methodologies (discussed in more detail in Chapter 3); however the foundation for these methodologies and understanding how to leverage them to accommodate for different modes of reasoning can be found in Leont'ev's Activity Theory.

1.2.4 Activity Theory

Armed with this understanding of working memory, constructivist ideas regarding learning, and an understanding of the role reasoning plays in the learning process, the groundwork can now be laid for higher level theoretical concepts in which this work can be situated: the first of which is activity theory.

In 1934, Vygotsky first noted an inherent problem with approaches to learning—the lack of cohesion between the studies of intellect and affect [71]:

> [Intellect and affect]'s separation as subjects of study is a major weakness of traditional psychology, since it makes the thought process appear as an autonomous flow of "thoughts thinking themselves," segregated from the fullness of life, from the personal need and interests, the inclinations and impulses of the thinker. [70, p. 10]

Vygotsky sought to create a more holistic approach to education and psychology.[23] He proposed what became known as a "cultural-historical" theory of psychology wherein he posited that children develop better cognitive faculties when they are exposed to practical activities in a social or communal setting [70]—thus establishing a constructivist approach during the era of behaviourism. As a result, his theories were largely ignored by his contemporaries, and it was Alexei Leont'ev who incorporated Vygotsky's work into his own theory: known as activity theory.

Activity theory seeks to understand the relationship between the subject and the object via the activities that act as intermediaries [69], [71]. Rather than attempting to analyse and understand the subject and the object and then logically infer the activities, Leont'ev argued that the activity exists in primacy to the subject and object, and in fact, the various properties of the subject and object do not exist beyond the activity: it is the activity itself which brings these properties into view [72]. However, not just any entity can be considered a subject; a subject must have *agency*—the capability to act on an object and produce an effect. This translates into the Digital Research Environment (DRE) as the user (the subject) acting upon the object (the artefact) and, ideally, producing a boundary object,[24] thus highlighting the user's agency.

But agency goes beyond just the effect produced by the subject. To properly situate his work within Vygotsky's cultural-historical theory, Leont'ev argued that a subject is influenced by her position or place with the social hierarchy:

> It is self-evident that the activity of every individual man depends on his place in society, on the conditions that are his lot, and on how this lot is worked out in unique, individual circumstances. [69, p. 85]

Engeström later built on Leont'ev's notion of societal position in his development of what would come to be known as Scandinavian Activity Theory. In his seminal work, *Learning by Expanding*, Engeström discusses the role culture plays in the formation of activities, and the way in which these activities connect subjects and objects in meaningful ways. One of the exemplary mechanisms he leverages is that of gestures (building on the work of Leont'ev's own example):

[23] It is interesting to consider Vygotsky's comment within the context of modern-day thinking regarding education and technology, which also seeks to create a more holistic symbiosis between the two fields by integrating not only the psychology of learning but also the effect of user experience and Complex Network Theory.

[24] Diviacco and Busato defines boundary objects as "[Artefacts] that aim to bridge concurrent cognitive models through abstraction from all domains of partners" [73, p. 396]. These objects are used to represent constructed knowledge created by the user as a result of her interaction with the DRE and will be used within the scope of this work to discuss any object created by the user within the context of the DRE.

> [C]onscious gestures originated as people experienced that even when a work movement did not lead to its practical result for some reason or other, it was still capable of affecting others involved in production. It could, for example, draw them into the fulfilment of a given action...[gestures] are seen as derivative instruments of productive activity that necessarily [have] an interactive, communicative form. [74, pp. 42–43]

Engeström goes on to highlight the role that culture plays in both the use of gestures and the information they can communicate.[25] Without this cultural contextualisation, the gesture becomes meaningless and is unable to effectively communicate its message to the receiver.[26] This notion of contextualisation is applied to the DRE via the creation of the boundary object (and the user's understanding of the theory in which her constructed knowledge exists) as well as in the creation and use of the visualisation itself—for example, how does the user interpret the patterns she sees, what do the colours mean to her, etc. Much of this subjective experience is bound up in interpretation that is at least partially influenced by the user's enculturation.

Beetham extends this notion of culturally contextualised activity by reinforcing the role of activities within activity theory, in this case contextualising the activities within the culture of the system as a whole. Thus the same activity can take on a different meaning within disparate systems:

> We [discuss here] tools and resources in terms of how they *mediate* learning. This emphasizes that artefacts can have different meanings in different activity systems, and that the purposes for which artefacts are designed and the ends for which they are actually used need not be the same...artefacts are understood as transitional objects between activity systems. As outcomes of their own design process they reify certain purposes, roles, rules and divisions of labour; and as components of the learning activity process they mediate new, emergent outcomes and relationships. [Within Activity Theory], artefacts are never conceived as actors themselves but as mediators of human actors' intentions, through complex and interrelated activity systems. [75, p. 39]

These principles are easily applied to the design and creation of the DRE. When designing the types of tools the DRE makes available to the user, the "culture" of the target user base must be considered: after all, the cultural mindset of the historian is different from that of the textual scholar, just as the mindset of the undergraduate researcher is different from that of the post-doctoral scholar. As Beetham mentions, different communities will have different needs and expectations of the artefacts they interrogate. So one must consider these enculturated differences of the target

[25] One can of course see the links between Engeström and Leont'ev's work regarding the role of cultural experience and learning with that of the importance of cultural contextualisation within the reading process. Refer to Chapter 2 for further discussion.

[26] Much the same can be said for the affordances and metaphors leveraged within Interaction and User Experience Design as will be discussed in more detail in Chapter 3.

audience(s). This same concept also extends to data visualisations often produced within the DRE, which themselves must be not only contextualised within the source work (i.e. the corpus, manuscript, etc.) but also within the work of the system and the enculturated context of the creator. As the scholar behind the creation of the data visualisation has a certain enculturation (be it academic training, nationality, or even the idea of a contextualised research agenda), data visualisations cannot be viewed independently of their source environments.[27]

Much of the theory behind Leont'ev, Vygotsky, and Engeström's work lies with the concept of motivation: users (in the case of Human-Computer Interaction, hereafter referred to as HCI) only take part in actions that either meet an immediate need or that serve to fulfil a series of actions that, in their aggregate, meet a predefined need. Thus the notion of motivation becomes a critical linchpin in the development of any system interface and, by extension, that of the DRE.

Motivation & Maslow's Hierarchy of Needs

In digital environments, motivation is a key factor in determining exactly why a reader may utilise a particular resource. The *Oxford Dictionary of Psychology* defines motivation as:

> [a] driving force or forces responsible for the initiation, persistence, direction, and vigour of goal-directed behaviour. It includes the biological drives such as hunger, thirst, sex, and self-preservation, and also social forms of motivation such as need for achievement and need for affiliation. [76]

Schacter, Gilbert, Wegner, et al. simplify this definition by simply referring to motivation as "the purpose for or cause of an action" [7, p. 408].

Motivation, however, is far more complex. In 1943, A. H. Maslow sought to contextualise motivation within a framework, which he dubbed a "hierarchy of needs" [77]. Maslow's hierarchy identified five key factors, each building upon the former, that drive human motivation. In his model, if the more base needs are not met, they will override any higher-order needs (see Figure 1.9).

[27] This argument must be carried with the reader throughout the nature of this work as it serves as an underlying point to the nature of the work as a whole.

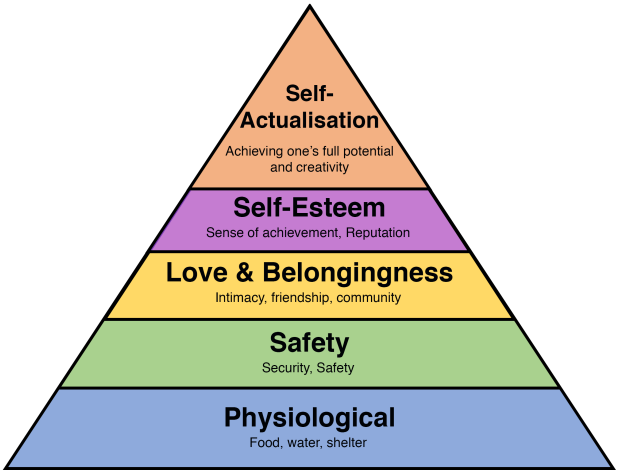


FIGURE 1.9: Recreation of Maslow's Hierarchy of Needs. Larger version can be found in Appendix E as Figure E.1

The first and most basic of these needs are known as physiological needs. These are items that no human can survive without, such as food, water, and shelter. As these needs are necessary to create a type of homoeostasis, no other needs can be addressed until the physiological needs are satisfied [78]. The second need which will drive motivation once the first need is met is safety and security. These needs can range from a basic feeling of safety to more abstract concepts such as rule of law, order, or limits [79]. Once these two base needs are met, individuals are then motivated to seek out relationships and community (known as "Love & Belongingness" needs) [80]. Motivation then builds upon these structures to feel a sense of achievement or establish a reputation (self-esteem) [81] which then directly leads on to self-actualisation, whereby the individual is motivated to achieve her full potential. As Zalenski and Raspa note:

> Living at this level [of self-actualisation] can lead to peak experiences and even transcendence—the experience of deep connection with others, nature, or God, and the perception of beauty, truth, goodness, and the sacred in the world. Such experiences become highly motivating and lead to feelings of being enlivened and enlightened. [82, p. 1121]

Rutledge's Theory of Social Needs & Connection

But some argue that Maslow's theory falls short because it fails to consider the underlying need for connection that is inherent at all levels of his hierarchy [83], [84]. Rutledge notes that especially for the three most basic levels of the hierarchy—physiological, safety, and emotional needs—all require social connection in order to be adequately met [84]. In fact, Rutledge goes so far as to state that needs themselves are not hierarchical in nature as Maslow contends: "Needs are not hierarchical. Life is messier than that. Needs are, like most other things in nature, an interactive, dynamic system, but they are anchored in our ability to make social connections." [84].

Rutledge's model removes the notion of hierarchy and instead proposes a model that more clearly demonstrates the type of interconnectedness and reliance upon the "belongingness" needs for which she argues, as can be seen in Figure 1.10.

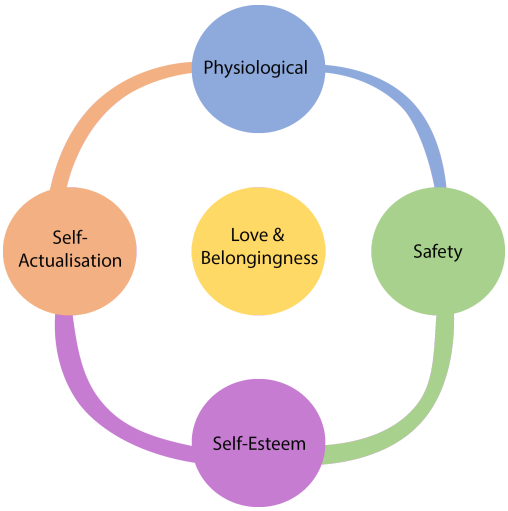


FIGURE 1.10: Recreation of Rutledge's model of Needs & Connection

When considered within the context of HCI, the idea of *belongingness* can be extrapolated into the idea of *connection*. What makes the application of Rutledge's work to HCI interesting is that this notion of connection can be understood within a broader context. Connection is no longer constrained to just the community of people; it is expanded to include the emotional connection to the software (via Norman's idea of emotional design)[28] or in the case of the DRE, the connection to the digital object. If considered within the scope of both Maslow's and Rutledge's work, a model of needs can be derived for the purposes of a discussion around HCI and motivation within the context of the DRE.

The first two aspects of Maslow's model (which are also included in Rutledge's work) are perhaps the least relevant as they pertain to this discussion. The physiological needs from Maslow can be interpreted as infrastructure needs within the concept of the software. These include things such as reliable hardware for hosting the software, reliable hardware for the client to access the software, a stable internet connection, etc. The notion of safety, then, can be re-imagined as data protection and privacy: secure connections (via mechanisms such as SSL), encryption of personally identifiable information (such as logins and search histories), and software free from malicious scripting. While both of these "needs" are less relevant to the discussion of HCI, it is interesting to note that both rely upon a notion of connection, albeit in this case connection is less about an emotional response from the user and more of a physical connection between endpoints.

The need for *Love & Belongingness* is articulated in much broader terms and, in fact, interweaves with Maslow's top two elements: self-esteem and self-actualisation.

[28] See *Emotional Design: Extending a System From Usable to Necessary* discussed later in Chapter 3.

Regarding the DRE—whose primary goal is to facilitate knowledge acquisition—self-esteem (redefined here as self-efficacy)[29] is highly important regarding knowledge acquisition, as the individual's perception regarding her ability to accomplish a task(s) has a major impact on her ability to learn underlying concepts [87]. It is not difficult to see how this extends into the realm of the DRE and the user's ability to both (a) navigate the system via search and browse techniques and (b) interrogate the artefact. If the user finds the system overly complex or feels she lacks the skillset necessary to interface with the system or digital object, she will likely abandon the system and thus fail in her quest of knowledge acquisition (at least with regard to the resource in question).

Self-actualisation then extends the notion of simple usability to that of true immersion with the object—or within the boundaries of Rutledge's model, connection with the object. While self-esteem concerns itself with the ability to connect to the object through utility (navigation, search, browse, inspection, etc.)—thus self-esteem is increased through the successful accomplishment of tasks—self-actualisation concerns itself with the engagement of the user with the object itself—that moment of true connection where the knowledge being sought is acquired. It is the "ah-ha" moment of true understanding.

Beyond Maslow's (and by extension Rutledge's) model, other scholars have sought to describe motivation. Perhaps most relevant to this discussion is that of Jang, Conradi, McKenna, et al., who describe motivation not as a single concept but as a collection of factors: attitude, interest, value, self-efficacy, self-concept, and goal [88]. Within the framework of an activity or task, attitude is used to understand the individual's feelings that affect engagement with a task (i.e. whether it is something to avoid or something enjoyable) [88]. Interest becomes an extension of attitude and is seen as a "positive orientation" toward that task [89]. Value, then, becomes an expression of the user's view of the intrinsic worth of the time investment in a particular activity (e.g. the time it takes to complete a task is seen as having a direct impact on the accomplishment of a goal, and thus worth the user's time and effort) [90].

Self-efficacy builds upon value by determining how successful the user is at accomplishing the task set before her. It is an extension of how the user views her innate ability to accomplish a given task. If she feels the task is too difficult, her self-efficacy will be lowered, with the inverse being true if the task is accomplished successfully. Self-efficacy plays a vital role not only in motivation but also in comprehension [91], a concept further explored later on.

Self-concept differs slightly from self-efficacy in that it focuses on a more holistic view of the self as it relates to the task at hand. In the case of reading (which is the focus of Jang, Conradi, McKenna, et al.'s work), it relates to whether the individual

[29] Bandura defines self-efficacy as an individual's belief in her ability to achieve a goal and notes that "it is not a fixed act or simply a matter of knowing what to do. Rather, it involves a generative capability in which component cognitive, social, and behavioral skills must be organized into integrated courses of action to serve innumerable purposes." [86, p. 122].

sees herself as a "reader" and is secure in her role as such. Unlike self-efficacy, which focuses on individual perceptions of the ability to accomplish a task, self-concept refers to a personal identification with the task itself—"Can I read?" versus "Am I a reader?" [88], [92]. This same concept can be applied to the research process that takes place within the DRE, especially around various aspects of the research tools provided. For those leveraging data visualisations, the question becomes "Am I a data analyst?" For those leveraging annotation tools, it is "Am I a historian or scholarly editor?" These are common questions that users of a DRE may ask themselves.

Finally, under Jang, Conradi, McKenna, et al.'s model, the goal is ultimately defined as the reason(s) the individual embarks on a particular task [93]. Goals can be shaped by multiple factors, and these factors can have considerable impact on some of the above-mentioned aspects of motivation. For example, Dweck, Mangels, and Good discuss two different types of goals: performance-based (or goals which involve validating one's ability against a predefined performance metric) and learning-based (or goals which seek to increase the overall learning experience) [94]. While performance-based goals tend to have a stronger short-term motivational impact (as these types of goals drive the individual's need for self-actualisation), they can also have negative side effects, such as placing unnecessary strain on attention as cognitive resources are allocated to constantly compare performance with other participants or creating negative feelings of self-worth which can lower self-efficacy or even self-concept [95]. However, the importance of establishing a goal,[30] especially with regard to motivation, cannot be ignored. Such an "objective-based" approach, as described by Beebe and Clark, necessitates the creation of an end-point before the steps can be undertaken to reach such an end-point. After all, the user must understand where she is going before she determines the steps she will take to get there.

All three of these theories of motivation (Maslow's, Rutledge's, and Jang et al's) contain different components to consider in the construction of the DRE—as do working memory, constructivism, and activity theory. As the ultimate goal of the DRE is the generalised concept of learning, it is thus logical to consider each of these components within a larger pedagogical framework within which the Digital Research Environment (DRE) can be constructed.[31] Enter Diana Laurillard's *Conversational Framework*.

1.2.5 Situating the Work Within a Framework

Each of the aforementioned items (working memory, learning theory, activity theory, and motivation), plays a vital role in the overall theory which underpins this

[30] For more on motivation and its relationship to system design via a Goal Directed Design approach, see Chapter 3

[31] For an application of the aforementioned theories as they apply to the construction of the DRE, see Chapter 3.

work. As the author advocates for a constructivist view of learning, both the structure of the mind (specifically the role of working memory in reading as discussed in depth in Chapter 2) and the user's past experiences and cultural expectations (as seen in discussions of activity theory and motivation, which play a large role in the arguments laid forth in Chapter 3) play a significant role in both understanding the inherent challenges of using a DRE and in crafting a solution that not only meets these challenges but extends the function of the DRE beyond that of knowledge acquisition (through close reading) and into the realm of engagement.[32]

But simply weaving together these foundational concepts is insufficient for the purposes of discussing learning and comprehension, especially when considering the future of the DRE and scholarship as a whole. As technology continues to evolve and becomes more ubiquitous in day to day life, the expectations of the populace change regarding format and delivery of content. These changes, especially in those considered "digital natives",[33] fundamentally alter not only how content is expected to be delivered, but also how knowledge is acquired, contextualised, and understood [98], [99]. Thus to truly extend the DRE into a tool that can be used for learning, the tool itself must also be situated within a theoretical learning framework, one that incorporates constructivist ideas and understands the needs of the digital learner.

Diana Laurillard sought to incorporate a constructivist approach with her "Conversational Framework", which postulates that students learn best in an environment where their constructed knowledge is balanced against the teacher's represented environment [54]. Laurillard took on Vygotsky's notion that information must be situated within a relatable context. This type of "situated learning", as put forth by Laurillard, focuses on an authentic learning experience, which not only conveys to students *how* to apply the knowledge they are gaining, but also *when* such knowledge is applicable [54]. To contextualise this argument, Laurillard uses the example of using a Swiss army knife to remove stones from the hooves of a limping horse. It is not enough for the student to know how to remove the stones from a horse's hoof without causing damage; the student must also be able to recognise the limping horse in order to apply this knowledge [54].

Laurillard recognises in her work, that because each student has different experiences (and thus filters knowledge through the lens of their experiences), a more customised approach must be taken when constructing knowledge environments—the goal should be to provide the student with an opportunity for "sense-making" [54]. This "sense-making" serves not only to assist with the contextualisation that is so important in activity theory, but also in achieving that state of self-actualisation

[32] The idea of engagement, its definition, and its role in the DRE are at the core of the discussion set forth in Chapter 3.

[33] As discussed in Chapter 2, digital natives are those individuals who were raised within the current technological ecosystem and thus never had to acquire "computer skills" in adulthood [97]. Typically, this term refers to the Millennial generation, Generation Z, and the latter portion of Generation X (depending upon how one categories each of these generations).

that is inherent in all three models of motivation. Thus Laurillard's work serves as the optimal framework under which to contextualise the design and development of the DRE.

Within a digital context, Laurillard's framework also rises to primacy as it facilitates many of the *soft skills*, such as innovative methods of communication that result from the use of social media [54], [98]–[100], that many digital natives possess. While other frameworks may provide a valid conceptual model under which this research could be positioned, Laurillard's *Conversational Framework* was specifically selected not only for its blending of many of the ideas and concepts addressed in the foundational theory of this work, but also due to its successful application within the Education Technology sector [101]–[103]. Thus a learning model, specifically proven within the realm of digital learning, serves as a solid foundation upon which to rest this work.

1.3 Format & Content of this Thesis

With the groundwork established, this thesis seeks to answer the question set forth earlier: "What effect do data visualisations, specifically within the remit of Digital Humanities, have on engagement with the source material—and by extension learning—in digital research environments?" The following chapter will discuss the most basic process that occurs within the DRE: reading. This discussion will focus on not only the importance of reading within the context of the DRE but also the inherent challenges that exist due to the nature of the medium (that is, the digital nature of the environment). Alternative modes of reading will be discussed, and a case will be made for how data visualisations can assist with the research process which incorporates reading.

Following on from this discussion of reading will be a discussion centred on the importance of immersion and engagement within a digital environment. This chapter will begin with a brief discussion of the selected design methodologies (*Goal Directed Design* and *Emotional Design*), touching on how the incorporation of these two methodologies addresses the concept of motivation discussed previously. This discussion then segues into a discussion around the role interactivity plays with regard to immersion and engagement, addressing not only why engagement matters but also how interactivity plays a role in the learning process via the mechanism of engagement.

The remainder of the thesis, then, concerns itself with the practical application of the aforementioned items, with an eye to answering the underlying research question. A case study will be presented, and a discussion will take place around the methodology behind the design of the software as well as the case study itself. This will follow on with an analysis of the results and a consideration of the impact of the findings of the case study and its application to future work within the field. Finally, the thesis will close with thoughts regarding the limitations of the work and a look

towards prospective applications of both the theory and practice of the thesis as a whole.

Chapter 2

Visualisations: Supporting Close Reading in a Digital Context

Humans are, by their very nature, adaptable creatures. Whether it be biological adaptation through evolutionary processes or cognitive adaptation to new technologies, human nature seeks to constantly grow and acclimate to change [104]–[107]. The process of reading is certainly no exception. As technology has evolved and developed new ways of presenting and distributing content, humans have adapted to these technological innovations by developing new methods of engagement with text. This ability to adapt is certainly nothing new, as humans have constantly been adapting to changing environs for millennia.[1] Within the last few decades, scholars have attempted to understand how individuals adapt their behaviours in digital environments. One of the theories they have developed to explain these behavioural adjustments is the theory of visual perceptual adaptation—that is, the ability of the brain to adjust to distortions or abnormalities within the visual field [108], [109]. While much debate continues in regard to this theory as to whether the observed adaptations are due to "cognitive correction" or "true adaptation",[2] it is apparent in the research that humans will modify behaviour in order to adjust to new influences and environments. The book itself (and by extension, the process of reading) is a prime example of this adaptability.

While reading itself is not a natural process,[3] how readers have interacted with text throughout the ages has evolved alongside the medium. The earliest books[4] were created on scrolls, and readers learned to read these by allowing the eye to

[1] As a recent example, consider the adaptation to cars and driving what is, essentially, a large, explosive devise (after all, the internal combustion engine—which is at the core of every automobile—is powered by small explosions whose energy is then harnessed to produce power for the rest of the machine). Piloting such a machine would have seemed like the fever dream of a madman only a few centuries prior but is now a common rite of passage for adolescents in modern western societies.

[2] The debate between cognitive correction and true adaptation is too extensive to be discussed here. For further information, see [110]. For more on perceptual adaptation, see [111], [112].

[3] Unlike speaking (and understanding spoken language)—which, thanks to evolution, has dedicated neural structures necessary to support the activity—reading is not natural and will not develop naturally without some type of systemic training [1].

[4] The use of the word "books" here refers to compiled sources of western knowledge, that followed on from the use of clay tablets. The history of the written word and its conveyance through various systems of writing is beyond the scope of this work, and so the notion of "books" here begins with the use of the paper or vellum scroll form.

follow the text down the length of the document.[5] The technology as such enforced a continuous reading approach by forcing the reader to engage with the content in a continuous fashion, from beginning to end [2]. With the development and adoption of the codex form, readers learned to read across the page, allowing the eye to flow from the top-left corner to the bottom-right corner (in the case of western manuscripts). With the development of the printing press, various affordances—such as the footnote, table of contents, etc.—eventually gained prominence and became commonplace over the years [3]. In addition, the codex form contrasted the scroll by allowing readers to approach the text in a discontinuous manner, supporting the capability to read page 42, and then page 8, followed by page 180 if the reader so desired [2], a process which would be nigh impossible with the scroll due to a lack of some of the affordances of the codex, such as page numbers. Much of the codex form's lifespan has been an encouragement of this kind of discontinuous reading, developing, at least partially, out of a need to break from the continuous method of reading enforced by the scroll. This disconnected approach allowed readers to embrace the text very differently than they could when reading from a scroll. Consider, for example, the Bible (which dominated the codex form for much of its history), wherein readers were encouraged to engage in a disconnected fashion with the source material (thus the development of "books", "chapters" and "verses" as reference points), an approach which would have been extremely difficult, if not nearly impossible, using the scroll. It has only been in the last few centuries that the novel, which encourages a more continuous form of reading, has come to be the dominant implementation of the form. Thus, "[t]he novel has only been a brilliantly perverse interlude in the long history of discontinuous reading" [2, p. 47].

With the advent of the digital age, text could now be moved from the physical page to the digital screen. No longer bound by the limitations of the physical medium, technology began to introduce new modes of presentation, while conversely reverting back to older forms of textual presentation (that of the scroll). New technologies, however, have allowed the scroll form to be utilised in different ways—one of the most influential of these technologies being hypertext [113]. And with these new technologies, readers were once again presented with new modes of interactivity, and thus began to adapt their reading techniques to the new opportunities presented by the technology. However, these new methods have not really changed the way we read, but rather simply provided us a different *mode of transportation*.

This chapter seeks to explore the evolution of reading specifically within a digital context. The scope of the argument will be constrained to those reading environments generally produced within Digital Humanities: the digital scholarly edition, thematic research collections, digital archives, and similar digital collections of text

[5] Unless otherwise stated, all text (and implementations of text) discussed here are concerning Latin script, created for a western audience.

meant to be distributed for a scholarly purpose (which will be referred to as a Digital Research Environment or DRE). Due to this scope, considerations of device are somewhat irrelevant as few of these types of projects produce different versions for multiple devices. Furthermore, when viewed on multiple different devices (whether they be laptops, mobile devices, or tablets), the experience is largely the same from a cognitive perspective, as the challenges are inherent to the digital screen as opposed to the paper medium [114], [115].[6] Further, most of these types of projects also employ the traditional web interface of scrolling text and eschew the "page-turn" metaphor that is inherent in the traditional e-pub format popularised by devices such as Amazon's Kindle, Apple's iPad, and Barnes & Noble's Nook, thus encouraging a type of continuous reading. However, because of the implementation of hypertext (which permits the reader to jump to various sections of the text), these projects encourage a discontinuous reading approach as well, thereby blending the traditional codex and scroll forms of the book.

These new modes of engagement (i.e. discontinuous reading via hypertext, also known as hyperreading) will be juxtaposed against a more traditional close reading approach. Due to the nature of the digital environment, both modes present new cognitive challenges to the reader—challenges not generally present in analogue environments—and these challenges will also be discussed. Following on, Moretti's concept of *distant reading* will be explored, where the argument will be made that, despite its radically different method of engagement with text, distant reading is actually closely aligned to more traditional close reading techniques, given its reliance upon cultural contextualisation.[7] The chapter will close with an exploration of the challenges from a cognitive and visual perception standpoint and a brief analysis of a few of the projects that utilise data visualisations to facilitate both a close reading and distant reading approach.

2.1 Close Reading

2.1.1 Examining the Cultural Contextualisation of Close Reading

Reading[8] can best be described as "the process of constructing meaning from written texts. It is a complex skill requiring the coordination of a number of interrelated

[6] Notions of designing for different devices is outside the scope of this current chapter, which is specifically concerned with the cognitive effects of reading in digital environments, and how various theories within the realm of cognitive psychology should affect our approach to the use and implementation of data visualisations.

[7] Neuroscientist Stanislas Dehaene argues that the brain does not evolve to accommodate new cultural phenomenon, but rather the inverse, that culture has adapted itself to work within the confines of cognitive structures: "...the human brain is a much more constrained organ than we think...it places strong limits on the range of possible cultural forms. Essentially, the brain did not evolve for culture, but culture evolved to be learnable by the brain" [116, par. 7]. Under Dehaene's paradigm, this chapter contends that distant reading is not so much a new modality of reading but rather a recycling of close reading approaches using a different mechanism.

[8] Portions of this section are reused in an upcoming journal article entitled "Engagement through Visualisation: A Case Study on the Alcalá Record Books" in the *IEEE Transactions on Learning Technology*.

sources of information" [117, p. 389]. Reading is a process that is grounded in both cultural and historical practice [118], [119] and is further defined by the way in which individuals interact with text in "natural, meaningful, and functional" ways ([120, p. 269], [121, pp. 3–4]).[9]

Most modern-day readers approach content from a holistic standpoint, processing each word as a whole within the context of the sentence and then within subsequent larger groups (such as paragraphs, sections, and chapters). Within an academic context, this is typically referred to as *close reading*, defined as "[the] instructional practice that makes complex texts accessible using repeated reading, cognitive scaffolding,[10] and discussion" [124, p. 35]. Frederico refers to close reading as the kind of reading that engages the reader in such a way as to open new or deeper understandings of a work, thereby enhancing the overall *experience* of the reader [125]. Birkerts—who uses the term *deep reading* in place of close reading—extends Frederico's metaphor of experience to that of immersion, describing how immersion is a necessary component of the overall adventure [126].[11]

Nicholson adds additional context to the definition of close reading by tying it to more than just literary text, but also to that of cultural objects: "Close reading can reveal the tensions and contradictions in a text, and it can illuminate moments of experience by placing them in the context of other cultural, artistic or social practices" [127, p. 184]. Here, Nicholson is advocating for the application of close reading techniques in more performance-based humanities arenas with the intent of using these techniques to critically analyse the underlying "ideas, debates, politics and practices that are inherent within [them]" [127, p. 184]. It is not the goal to uncover some kind of *universal truth*,[12] but rather to apply a careful analysis, situated within the culture of the object itself, to a medium that is traditionally experienced rather than read.

While Nicholson was largely concerned with applying critical analysis to theatre, his argument can be applied back to traditional takes on close reading. By encouraging a reading that is situated within the culture of the object itself, close reading allows the reader to draw inference, theme, metaphor, etc. by situating the reader's understanding of the text within the cultural context of the work itself. Consider Austen's *Pride & Prejudice* [128], set in Georgian England. Much of the novel

[9] Understanding that reading is also culturally informed is important to consider, as the literature in this chapter demonstrates. Much of the work being done with data visualisations (and distant reading as a whole) is specifically situated within western cultural standards of semiotics and visual interpretation.

[10] Cognitive scaffolding is used in reference to how the brain builds temporary *structures* (or *scaffolding*) in order to support working memory and recall. Baddeley explores the concepts of working memory in much of his work (see [122], [123]).

[11] Birkerts's slightly more colourful definition of *deep reading* is defined as "the slow and meditative possession of a book. We don't just read the words, we dream our lives in their vicinity. The printed page becomes a kind of wrought-iron fence we crawl through, returning, once we have wandered, to the very place we have started ... we hear the words in the theater of our auditory inwardness" [126, p. 201].

[12] Close reading has traditionally been associated with the work of F.R. Leavis, who sought to analyse literature with the goal of producing "universal values" or truths [127, p. 183]. This particular view was eventually seen as problematic as critics in recent decades have argued these "universal truths" are in fact reflective of white, western, male privilege and are, therefore, not truly *universal* [127].

is geared towards cultural commentary around the roles of marriage and class in Georgian society, and to truly understand the themes and narrative Austen is communicating, the novel must be read within that cultural context. Attempting to read it outside of the cultural context and situating it within an entirely different set of cultural expectations and social mores removes aspects of the overall tone and message of the narrative. That is not to say that only Austen's peers could successfully understand her work; certainly not. However, understanding the expectations of women in Georgian England and the importance class played in such endeavours provides the reader with a greater understanding of the work as a whole. A recurring theme in Austen's work is that of the preoccupation with marriage as a means of financial stability (while such a thing still occurs in modern day settings, the necessity of such an arrangement is much more clearly understood when situated within a Georgian context). This is perhaps best illustrated in another of Austen's works, *Sense and Sensibility* [129], when considering the fate of the Dashwood women after the death of their father. Throughout the book, notions of money, class, and financial stability plague the various relationships of many of the characters (for example, Lucy, Edward, and Eilnor or Willoughby and Marianne). By devoting effort to the understanding of the cultural mores which surround these narratives, the reader increases her immersion with the text, further enhancing the overall experience. This notion of contextualisation becomes increasingly more apparent in the latter half of this work, which concerns itself with the creation of a DRE with an 18th century Spanish manuscript detailing the expenses of the Royal Irish College of Alcalá as its source.[13]

This method of reading, which can be applied to any text regardless of its length or technical complexity, continues to be seen as the "gold standard" with regard to literacy and reading education, as evidenced by the US Common Core standards, which notes:

> Students who meet the Standards readily undertake the close, attentive reading that is at the heart of understanding and enjoying complex works of literature. They habitually perform the critical reading necessary to pick carefully through the staggering amount of information available today. [130, p. 3]

Hinchman and Moore note the importance of close reading, specifically highlighting the skills which develop as a result of its practice: skills such as meticulous analysis of language patterns, pattern combination, and analysis of irony, symbolism, and metaphor, without which one cannot truly comprehend the text at hand [131].

While the general association with this kind of attentive reading is to that of literary analysis of complex works, close reading is involved in the consumption

[13] More information regarding the explored source text for the experimental portions of this work can be found in Chapters 4 and 5.

of nearly every facet of the traditional narrative (i.e. linear texts). Whether one is reading a work of fiction, a short story, a poem, a letter from a loved one, or even a periodical (such as a newspaper or magazine), the skills and processes of close reading are inherently intertwined with the consumption of the narrative. When reading a periodical, for example, the reader will often re-read sections of the article in question to parse fact from opinion or to facilitate the construction of her own opinion based upon facets of information relayed within the article itself. Another example is the reading of a letter from a loved one: a reader may carefully parse the text of particular sections, subconsciously looking for clues as to the author's state of being or thoughts related to a particular situation being discussed within the letter. The methods utilised in these two examples are a small sampling, but are representative of how close reading can be applied across multiple types of objects. Upon turning one's attention to larger works (whether poetry, prose, drama, or other works typically studied within a humanities scope), these practices become even more obvious, with the reader analysing motive, theme, or plot in an effort to better understand the characters or the larger arc of the supporting narrative.

2.1.2 Obstacles & Challenges of Close Reading in Digital Environments

While close reading techniques have served readers well throughout the print age, numerous studies have noted that the nature of text in a digital environment presents several problems for those engaging in close reading, problems which either do not exist or are not as problematic in analogue environments. One of the most prevalent problems is the challenge of comprehension. A study conducted in 2013 by Mangen, Walgermo, and Brønnick compared rates of reading comprehension between a group of readers reading two versions of the same text: one print and one digital [132]. The reading itself consisted of a narrative text as well as an expository text. Participants were then asked a series of questions designed to assess the comprehension of the source material. The study showed that those who read the material via print had a greater comprehension rate than those who read the same material in a digital format.

The study proposed a number of possible factors contributing to the lower rate of comprehension for those reading digitally. The first is an issue of navigation within the text [132]. The entire text of the digital article did not fit on the screen, thus it was necessary for the reader to scroll through the text. This scrolling has proven to be problematic with regard to reading as it produces a kind of "spatial instability" [132, p. 65]. Piolat, Roussey, and Thunin note that reading is clearly informed by positional placement of text, based on various eye tracking experiments which highlight how the gaze of the eye is repositioned during the consumption of texts in digital environments [133]. As continuous scrolling forces the eyes to constantly reposition, there exists a logical inference that this instability of the content on-screen has an effect on the reader's ability to comprehend the text.

Several studies show that the position of text in physical space is often used by readers when attempting to recall particular information [133]–[135]. The rigid nature of the printed text provides a structure upon which the reader can build a mental map—one that is formed without regard to text length. Mangen, Walgermo, and Brønnick note that even in shorter texts of less than 4 pages, this cognitive framework played a crucial role in comprehension [132]. By building these mental maps of the text, the reader is able to rely upon them to aid in recall of specific aspects of the text, thus furthering comprehension and, by extension, cognition. However, the digital space lacks such rigidity. And while one could argue the screen of the device itself provides boundaries—because of the *spatial instability* of a scrolling text on screen (as mentioned by Mangen, Walgermo, and Brønnick [132])—the physical boundaries of the device screen do not provide the same type of tangible boundaries provided by print. Thus, in a digital environment, the reader is deprived of a mechanism that has a strong correlation to recall.

Another potential contributing factor is that of metacognition, defined as "[k]nowledge and beliefs about one's own cognitive processes" [76, par. 1]. Ackerman and Goldsmith conducted a study in 2011 to evaluate the role metacognition played with regard to On-Screen Learning (OSL) versus On-Paper Learning (OPL) [136]. The results showed that OSL readers performed statistically more poorly than OPL readers primarily due to two reasons: overconfidence of ability and an inability to properly self-regulate. One possible reason for overconfidence is the assumption by the reader of the level of difficulty involved in OSL, a phenomenon which can be traced back to observations made more than 30 years ago. In 1982, Lichtenstein, Fischhoff, and Phillips observed a phenomenon whereby individuals showed a level of overconfidence when answering difficult questions but under-confidence when answering easy questions. Dubbed the "hard/easy effect", the underlying idea is that as a task increases in difficulty, humans react with overconfidence in their assessment of their performance of the task; on the contrary, when a task is perceived as relatively simple, individuals will second-guess themselves as to whether they completed the task correctly, thus showing a level of under-confidence in execution [137], [138]. Ackerman and Goldsmith assert that, because of the perceived difficulty of OSL (due to an increase in cognitive effort), most readers are overconfident in their ability to read and comprehend text on-screen. As such, readers do not expend the additional effort needed to gain the same levels of comprehension in OSL, thus actually leading to a decrease in comprehension [136].

Regarding the inability to self-regulate, Ackerman and Goldsmith also note an apparent bias towards OSL wherein learners feel their ability to study on screen is less reliable than on paper [136]. This bias then translates into a meta-meta-cognitive process (known as a Second-order Meta-cognitive Judgement, or SOJ), which has a stronger, overriding meta-cognitive effect. There is an additional bias in association that paper-based learning is a more *effortful* task as opposed to screen learning, which is perceived as being geared towards more shallow and faster readings of

text. These meta-cognitive processes have a high correlation when factoring in reading comprehension [136], and therefore should not be ignored when considering how individuals read in digital environments.

It is interesting to note that even digital natives[14] have higher levels of comprehension in digital environments when previously exposed to print. Duncan, McGeown, Griffiths, et al. note in their study that both children and adolescents who have a history with reading print material tend to have higher levels of comprehension than those who do not [142]. And while Duncan, McGeown, Griffiths, et al. mentions several scholars who argue that the use of social media and text messaging provide exposure to a type of literary text, they note that others argue this type of exposure does not, in fact, translate to online reading comprehension [142].

Each of these items (spatial instability, lack of mental map, meta-cognition, and the inability to self-regulate) all foster difficulties related to comprehension, thereby increasing the cognitive effort required to closely read text in digital environments. As a result, humans began to adapt their reading practices to the screen in an effort to reduce the cognitive load and consume the text in a manner more befitting the medium. With the advent of hypertext, readers discovered a new way of approaching reading in digital environments: hyperreading.

2.2 Hyperreading

2.2.1 Digital Reading through Hypertext

Partially due to the advent of the internet (and, more specifically, hypertext) and partially as a response to the challenges listed above, readers have begun to adapt their reading with new methods of engagement beyond that of printed material [143]–[145]. While some have posited fears regarding the downfall of the traditional book[15] or the decline of close reading skills,[16] readers continue to engage with close reading but often seek to enhance the experience through additional means, adapting their reading methods to the technology and the modes of interface provided. These new methods of engagement do not replace close reading, but rather serve to augment the experience or provide new insights through alternative engagements with the text.

[14] The term *digital native* is typically used to refer to those individuals born or raised after the advent of digital technology and thus never needed to acquire digital skills in adulthood [97]. However Prensky's model on which this assumption was based has since been debunked [139] and studies have shown that age plays little (if any) role in adoption of technology (which is evident in the case study discussed in Chapter 5). White, Silipigni Connaway, Lanclos, et al. developed a more robust theory of *digital residents* and *digital visitors* [140] that more concisely explains the relationship individuals often have to technology. However, for our purposes here, the term *digital native* will be used both in a broader sense than its original definition (by also including those *digital immigrants* who gained digital fluency in adulthood) as well as for its familiarity in a colloquial sense. It should be noted that wherever this term is mentioned, it should be subsumed under White, Garton, White, et al.'s paradigm.

[15] See [126, pp. 201–204]

[16] See [146, p. 56]

While there is no one, unified definition of the term *digital reading*, for the purposes of this discourse, the term is defined as the type of reading that takes place in any digital environment (within the remit of Digital Humanities as previously defined) and that does not solely employ traditional *close reading* techniques but also offers alternative, non-traditional modes of engagement. As the scope of work laid out has already been constrained to those types of projects generated by Digital Humanities, the term will not refer to device-specific issues (such as the reading of e-pub formatted books on a Kindle or the effects of the *page turn* mechanism) but rather will consider common implementations of these *digital reading* mechanics across devices, as most of these types of projects have singular implementations—that is, they typically do not exist in one form for the web and another form for mobile devices or tablets. As most of these types of projects generated within the context of Digital Humanities are themselves *digitally native*,[17] they are uniquely positioned to take advantage of alternative methods of engagement offered by digital reading approaches.

Digital reading, as defined above, truly began with the development and implementation of hypertext. Numerous definitions of hypertext exist, but it is George Landow who provides the most comprehensive definition, stating that "[hypertext is] text composed of blocks of words (or images) linked electronically by multiple paths, chains, or trails, in an open-ended, perpetually unfinished textuality described by the terms *link*, *node*, *network*, *web*, and *path*" [113, p. 2]. Specifically, it is the notion of text linked by "multiple paths" that is "open-ended" and "perpetually unfinished" [113, p. 2] which sets the stage for radical new ways of approaching textual engagement in digital formats. Whereas the traditional print form forced engagement through linear means,[18] the nature of hypertext specifically allows the reader to chart her own path through the text, as opposed to print reading, which is generally hierarchical by design. The structure of the printed text is traditionally meant to flow from sentence to paragraph to page. Pages are organised into sections and then chapters, which are compiled into a book. These conventions were developed according to two disciplines: the Outline, where information is organised by its relative importance to the overarching theme or themes, and Syllogism, where information is organised according to the "narrative structure of the argument" [22, p. 106]. These conventions hold true today regarding printed text, and analogue narratives are still largely constructed following these two disciplines. And while Stallybrass contends the codex form was originally created to encourage discontinuous reading (the analogue form of hyperreading), many readers today approach print with a linear (i.e. continuous reading) mindset [2].

[17] Baron defines projects considered digitally native as resources which are "designed to capitalize on the special possibilities digital technology offers" (see [147, ch 1, sec. "Definitions"]). These types of resources typically contain a form of user interaction, such as visualisations, connecting to other outside, electronic sources, or implementing a social affordance such as a forum.

[18] While printed text can be read in a non-linear fashion via the use of devices such as an index or table of contents, for the purposes of this analysis and the definition used herein, the focus will be on reading materials in a digital (or hypertext) environment.

By contrast, however, hypertext allows for both linear and lateral navigation through the text due in large part to the lack of physical boundaries of the digital medium. This lateral movement largely breaks the metaphor of the traditional, structured text and organisation of ideas. Burbules notes that as a result of this lateral movement through the text, hypertext can actually create a more nuanced and complex relationship between multiple narratives [22]. Such *lateral thinking* is shown to be highly effective in developing critical thinking skills as well as learning how to problem-solve, as it forces the reader to associate seemingly disconnected nodes to form a whole [148]. Developing these types of skills is highly beneficial as the brain does not naturally think in non-linear (i.e. lateral) ways, but rather in logical, pre-determined paths that are built up over time based on past experiences [149]. By harnessing the power of lateral thinking, the reader can actually learn to visualise the information in a different sequence or structure, thereby changing her understanding of the subject as a whole [150].

This non-linear engagement with hypertext is known as hyperreading [151]. Due to its lateral movement and associative approach to narrative construction, hyperreading can promote a type of "thinking by association" [151, p. 24]. This type of thinking plays a strong role in the development of reading skills due to an increase in cognitive effort on the part of the reader to associate multiple, variable paths through the narrative, as afforded by the environment [152]. These skills position the reader to better understand and consolidate information from a variety of sources:

> Reading hypertext is a naturally dynamic, recursive, and integrated process, one that provides multiple opportunities for students to acquire, test and re-frame knowledge through cognitive reconstruction of text, inter-textual analysis and exposure to varied perspectives on issues. Thus, hypertext may promote increased comprehension through the elaboration and integration of new information into the existing knowledge network as readers create and expand the cognitive map that guides their construction of meaning. [153, par. 5]

By allowing the reader to engage with the content in this associative manner, the text is meeting a very real Need For Cognition, otherwise known as NFC, defined by Cohen, Stotland, and Wolfe as the "need to structure relevant situations in meaningful, integrated ways. It is a need to understand and make reasonable the experimental world" [154, p. 291]. Also defined as "people's tendency to engage in and enjoy effortful cognitive activity" [155, p. 197], NFC is highly relevant to understanding information seeking behaviour and how readers consume digital resources, largely due to the *experimental nature* of the digital world. Many digital readers utilise information seeking as a means of meeting this Need For Cognition [156].

Various modes of research have shown that NFC is tied to success in academia.[19]

[19] For further discussion on the effects of NFC in academia, see [157] (for discussion on the interrelation of intelligence, personality, and interests, much of which serves as a background for NFC); [158]

Given that most projects created within a Digital Humanities context are geared towards the dissemination of research and aimed at an audience of academics, there is likely a higher NFC among the user base as intelligence (and by extension, intellectual curiosity) is one of the contributing factors to a higher level of NFC [158], [159], [161]. This understanding of how NFC can affect behaviour in these types of digital environments, therefore, informs both the design of the hypertext as well as the intentions of the creator in how the various narratives are linked together and what additional tools may be necessary to supply to the reader.

The very nature of hypertext (and by extension hyperreading) provides a facility for information seeking, a task which is highly correlated with NFC [156]. Furthermore, the interactivity of hyperreading (in this case the selection of content through personal choice via the use of hyperlinks) actually allows for those who rank lower on the NFC scale to engage with content more effectively [162]. Thus hyperreading provides not only an alternative method for exploring textual content, but it can also actually facilitate further engagement with a wider audience by providing a level of interactivity aimed at drawing in low NFC readers.

2.2.2 The Challenges of Hyperreading: Cognitive Overload, Working Memory, and Attention

Hyperreading is, however, not without its disadvantages, as noted by Uso-Juan and Ruiz-Madrid [152]. Some of these disadvantages—such as a lack of clear contextual or navigational cues, the connection to other content nodes, or screen-related issues such as resolution, glare, or eye strain—are addressable via either information architecture principles or via improvements to hardware designs, which are constantly evolving as new technology emerges. However, the most significant problem faced by hyperreaders is that of cognitive overload, as noted by Kasper [153]. Cognitive Load Theory plays an integral role in understanding how hypertext specifically can lead to issues of cognitive overload. When first proposed in 1998, Sweller, Merrienboer, and Paas sought to understand the role of working memory, defining Cognitive Load Theory by offering the following definition:

> Cognitive Load Theory has been designed to provide guidelines intended to assist in the presentation of information in a manner that encourages learner activities that optimize intellectual performance. The theory assumes a limited capacity working memory that includes partially independent subcomponents to deal with auditory/verbal material and visual/2- or 3-dimensional information as well as an effectively unlimited long-term memory, holding schemas that vary in their degree of automation. These structures and functions of human cognitive architecture have been used to design a variety of novel instructional procedures

(for discussion on the role of personality and NFC in academic performance at the third level); [159] (for discussion on how NFC affects intellectual investment); and [160] (for discussion on how NFC affects motivation in higher learning environments).

> based on the assumption that working memory load should be reduced and schema construction encouraged. [163, p. 251]

The theory hinges specifically on an understanding of working memory,[20] which is the area of the cognitive system which stores small amounts of data for short periods of time [164]. DeStefano and LeFevre further define working memory as a "set of mental resources people use to encode, activate, store, and manipulate information while they perform cognitive tasks" [165, p. 1618].

Numerous studies regarding Cognitive Load Theory have focused on second language acquisition, specifically within multimedia environments. Baddeley discusses the link between working memory and language, addressing the role working memory plays in language acquisition. In his research, Baddeley highlights four different subsystems which make up working memory;[21] however, it is the visuospatial sketchpad that is perhaps most salient to the discussion of distant reading, specifically due to its role in maintaining page layouts and facilitating eye tracking [122]. However, given that page layouts no longer exist in digital environments (at least not in terms of being able to construct a stable mental map of the page) and that eye tracking is in a constant unstable state, the working memory system is forced to adjust the processing of digital text from the visuospatial sketchpad to the central executive. This can actually lead to a reduction in comprehension, a theory which is supported in a number of studies [168]–[170].

In keeping with this theory, Plass, Chun, Mayer, et al. conducted a study which examined the role of cognitive load during foreign language learning, focusing on the role visual cues played in reading comprehension [171]. Much of their work was based not only on Baddeley's model of working memory but also on Mayer's generative theory of multimedia learning, which holds that individuals learn better from the use of both the verbal and the visual [172]. As part of the experiment, Plass, Chun, Mayer, et al. had students read digitally a portion of text in a foreign language. Some students were then presented with hypertextual annotations for certain words that contained both visual and verbal cues as to the word's definition (a picture followed by an English Language description), while other students were presented with text which either contained hypertextual annotations with verbal cues (such as the actual written definition) or no hypertextual annotations at all. Their findings stated that vocabulary acquisition was higher where the use of both visual and verbal cues were present [171]. However, they noted that in *low-ability* learners, parsing both visual and verbal cues led to poorer reading comprehension

[20] See Chapter 1 for an in-depth discussion of working memory.

[21] The four subsystems of working memory are: the phonological loop (responsible for verbal / auditory input), the episodic buffer (responsible for the integration of multiple types of information in preparation for storage in long-term memory), the visuospatial sketchpad (responsible for visual / spatial input) and the central executive (responsible for managing attention within working memory). Each of these plays a different role in the abilities of working memory. For more information on these subsystems, see: [122]; [14]; [166]; [21]; [167]; [19]. Also, refer to Chapter 1 for an in-depth discussion on working memory and its various components.

and an increase in cognitive overload [171]. Conversely, this was not the case in *high-ability* learners. They observed that individual learning preferences played a major role in comprehension, concluding that those with higher spatial-learning abilities were able to process the text more easily than those with more verbal-focused learning abilities [171]. It can be concluded that offering readers a choice in how they are presented with contextual cues can have a strong impact on reading comprehension as it allows them to tailor the experience to best fit their cognitive abilities, thus decreasing the effort of the overall cognitive load of the process.

The study conducted by Plass, Chun, Mayer, et al. is particularly intriguing when considered within the framework of digital reading. When considering their results within the context of Baddeley's model for working memory, those who struggle with strictly visual cues are likely experiencing a type of cognitive overload due to the lack of strong boundaries within a digital environment. Unless the reader has a highly developed visuospatial sketchpad, the central executive will be forced to offset the additional workload, thus making it more difficult to make the conversion to long-term memory. Additionally, due to the central executive's role in moderating attention, these individuals will also struggle with maintaining proper attention to comprehend the text. However, in instances where both the visual and the verbal are utilised, the episodic buffer (the fourth component in Baddeley's model) is capable of taking on some of the additional workload to process and convert data from both the phonological loop as well as the visuospatial sketchpad, thus offloading some of the work to convert to long-term memory. Plass, Chun, Mayer, et al.'s study provides strong evidence for more of a *mixed media* approach to reading within digital environments.

Kasper notes a different type of cognitive overload regarding hypertext, as her work focuses on structural organisation rather than visual versus verbal cues [153]. In her study, Kasper notes that students reported a higher success rate for reading comprehension when working with *controlled hypertexts* (or hypertexts which have some, but limited, branching nodes) as opposed to *free hypertexts* (or hypertexts with no editorial, structural controls in place and that allow the reader to move in a completely unstructured, non-linear way throughout the text with no editorial intervention). Kasper notes that *free hypertexts* lead to cognitive overload due to their lack of structure and organisation, thus inhibiting the reader's ability to comprehend the totality of the text in question [153].

The other major challenge faced by hyperreaders is that of attention. Best described by William James in the latter part of the 19th Century, attention can be defined as the "taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought" [173, p. 403]. This understanding has changed little in the intervening century as proven by Buschman and Kastner in 2015: "the selective prioritization of the neural representations that are most relevant to one's current behavioral goals" [174, p. 127]. With working memory having a limited capacity, attention becomes a necessary

component in order to prioritise the storage and processing of information. Just as digital content has conditioned the reader to screen information more efficiently [175], it has also led to a reduction in her ability to sustain attention for any significant length of time [22]. Furthermore, attention facilitates information-seeking behaviour in large part by determining the reward associated with the behaviour: is the effort involved in seeking the information in a particular environment likely to produce a sufficient reward [176]? While attention is widely studied, this discussion will concern itself primarily with the theory of selective attention, focusing on visual attention, defined as attention which "refers to a set of cognitive operations that mediate the selection of relevant information and the filtering out of irrelevant information from cluttered visual scenes" [177, sec. "Definition"].

There are two primary methods by which visual attention can be focused. The first is through signal detection, otherwise known as visual spatial orientation. Signal detection relies upon a type of visual cue that draws the eye to the intended target [178]. These visual signals cause attention to orient to the physical location of the stimuli. Within digital environments, visual spatial orientation can be demonstrated via the use of coloured hyperlinks. By leveraging a colour which contrasts against the colour of the body text as well as that of the background, attention is drawn to the link, giving the reader some type of visual indicator that said text is different from the main body. In the case of hyperlinks, this signal detection is used to provide a clue as to additional information or functionality contained within the text and is useful for scanning a hypertext document for linked information (see Figure 2.1).

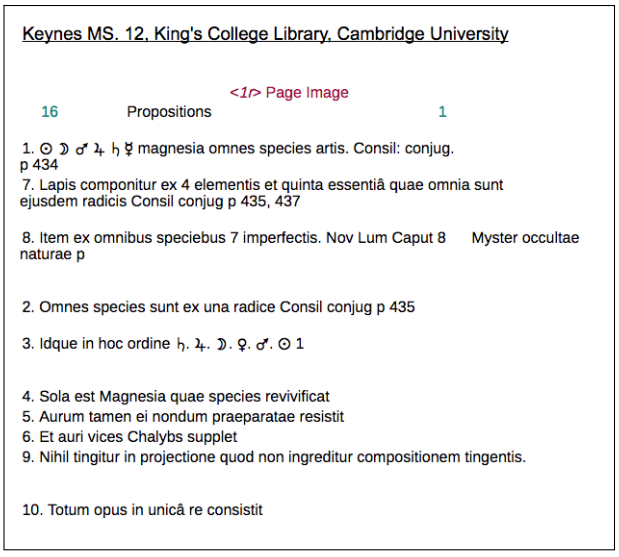


FIGURE 2.1: Use of signal detection (via the use of colour) to indicate links and additional information as seen in *The Chymistry of Isaac Newton* [179].

The example from Figure 2.1 highlights a particular problem with this method of signal detection: accessibility. In the above example, the numbers "16" and "1"

appear in green text, and the "<1r> Page Image" text is in red text. The green items provide modal overlays with additional information or contextual cues, whereas the red text provides a hyperlink to an image resource. However, because of the use of "red" and "green" colours, those who experience red/green colour-blindness may not be able to distinguish a difference in colour at all from the main body, let alone a difference in colour between the two different "types" of visual artefacts. Thus, when relying upon colour for signal detection, it is imperative to consider the limitations of the colour palette being used, as it may prove entirely ineffective for certain users.

The second method of visual attention is that of visual search, as proposed and discussed by Treisman and Gelade. Visual search concerns itself with the method of focusing attention in order to isolate a particular parameter among a morass of visual distraction or clutter [180]. Tying back to the earlier hyperlink example, a visual search may take place to locate the one "visited" hyperlink on the page (which is generally coded differently via the use of a secondary colour) amidst a sea of "unvisited" hyperlinks. The hyperlinks themselves (visited or otherwise) will stand out from the body of the hypertext due to their colour coding; however, the status of the link (visited versus unvisited) will be further highlighted via colour distinction. As the use of two colours (that are often in related colour families such as blue and purple) increases the visual clutter on the page, the visual search method is utilised to focus attention on the desired information.

One final aspect of attention which works in conjunction with Cognitive Load Theory and which can have a strong impact upon digital reading is that of the *split-attention effect*. Schnotz and Kürschner define this effect as the "[occurrence] when the learner's attention must be split between multiple sources of visual information that have to be integrated for comprehension, because the individual sources cannot be understood in isolation" [181, p. 471]. They specifically discuss the increase in cognitive load when attempting to integrate sources of material which are related but spatially separated. When readers must move between multiple sources or jump between spatial locations in order to access related information, the split in attention leads to an increased cognitive load in order to integrate the pieces of information; however, it is important to note the split-attention effect is only applicable to sources of information which are difficult or impossible to understand unless they are integrated with their related counterparts [181]. When sources of information are redundant, hiding the secondary sources (or spatially separating them) actually decreases cognitive load, thus leading to an increase in comprehension rates [24]. It is therefore essential to consider the relative importance of information relationships when designing hypertext readings, as these relationships can obviously play a vital role in the reader's ultimate comprehension.

An example of the dual nature of the split-attention effect can be seen in Figures 2.2 and 2.3. Here, both websites have been queried to show similar types of data: murders which occurred during a specific time period in a specific place (London,

UK and Harlem, NY USA respectively). Both allow the user to visualise the occurrence of crime and perhaps identify problematic areas via clustering. However, Figure 2.2 gives no indication as to the type of crime.[22] All occurrences of data are marked using the same icon—the only difference being that of size, which indicates the number of instances. Furthermore, no other information about the occurrence is available (on::hover, each map point simply indicates the number of occurrences related to that point with a link to view detailed results). In order for the reader to truly understand the scope and nature of the occurrence, her attention must be split to another screen entirely.

FIGURE 2.2: Map of murders in London between 1674 and 1819 as visualised in *Locating London's Past* [182]. A larger version is available in Appendix E in Figure E.2

However, Figure 2.3 takes an entirely different approach. First, this visualisation provides a visual indicator for each individual type of crime (each notated by a different icon on the map). The legend that describes each icon is visually separated but, as most of the icons are intuitive, it does not cause the split-attention effect as the legend is largely redundant. Furthermore, each occurrence is plotted individually and contains summary information regarding the occurrence, which allows the reader—on initial glance via the on::hover function—to view pertinent information and decide if the occurrence is worth further investigation without having to manoeuvre to an additional screen.

[22] It should be noted that while the only type of crime shown for both websites is that of murder, both websites allow for multiple crime types in queries, meaning that more than one type of crime could potentially be visualised at the same time.

FIGURE 2.3: Map of murders in Harlem between 1915 and 1930 as visualised in *Digital Harlem* [141]. A larger version is available in Appendix E in Figure E.3

Given that attention is finite and technology has conditioned us to expend less attention on digital text [147], [183], [184], the effort involved in selective attention methods is important to understand. Leveraging *real world* approaches to visual search (such as search for collections of objects in a scene such as *cars*, *people*, etc.) demands significantly less attention and cognitive load than the hyperlink example discussed above—otherwise known as a conjunction search [174]. Due in part to the elastic nature of hypertext and in part to cognitive overload that results from an abundance of choice with little clear direction, hyperreading has made close reading more difficult (via the creation of rabbit hole scenarios) and by a reduction in our ability to sustain attention in digital environments. These issues have led to the exploration of other alternatives that can mix some of the strategies for attention mentioned here (such as visual search) with multimedia learning, allowing for a more robust approach to comprehension in online environments. Enter Moretti's concept of *distant reading*.

2.3 Distant Reading: A Digital Evolution of Close Reading

While[23] hyperreading has served as a novel approach to textual engagement, it was unable to offer the same kind of deep, immersive experience that close reading could offer. As technology evolved, scholars began looking at using computer algorithms to *read* text and gain potentially new insights. One of the major advantages of this approach is that a machine could analyse millions of texts at a time, something physically impossible for any human being to undertake in a single lifetime.

Thus the concept of *distant reading* was born. In an attempt to answer the question 'How can you a read million books?' [185], Franco Moretti noted the greatest

[23]Portions of this section are reused in an upcoming journal article entitled "Engagement through Visualisation: A Case Study on the Alcalá Record Books" in the *IEEE Transactions on Learning Technology*.

shortcoming of a close reading approach: a dependence upon a small corpus [4]. In his work with world literature, Moretti felt it necessary to understand the genre as a whole, and in order to do so, felt it necessary to migrate away from a small cannon to an impossibly large one, thus gaining a greater understanding of the entire picture of the genre: "...[close reading is] a theological exercise—very solemn treatment of very few texts taken very seriously—whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how *not* to read them" [4, p. 48]. Matthew Jockers expanded upon Moretti's approach, asserting that these new methods of discovery via big data—such as Moretti's distant reading—allow the reader to pursue new avenues of discovery and understanding [5]. Close reading still plays a role at the micro-level, but distant reading allows the reader to begin her investigation at the macro-level. In the modern world of literary analysis, both the micro and macro approaches are essential components, given the volume of data available and the need to analyse both the overall view, as well as the insights available at a more granular level:

> Today's student of literature must be adept at reading and gathering evidence from individual texts and equally adept at accessing and mining digital-text repositories. And *mining* here really is the key word in context. Literary scholars must learn to go beyond search. [5, p. 9]

Close reading allows the reader to gain insight into "nuggets" of truth while data mining allows the reader to probe the "deeper veins" of knowledge buried beneath the detritus [5, p. 9]. This macro-level approach, or Moretti's "little pact with the devil" Moretti, p. 48, forms the basis of the theory of distant reading.

Distant reading can be defined as the ability to "focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems" [4, pp. 48–49]. Jockers, however, uses the word 'macroanalysis' in juxtaposition to distant reading, primarily because, as he argues, the word *reading* implies a broad, interpretative approach, whereas *analysis* implies a more concentrated, objective approach [5]. He does note, however, that "interpretation is fueled by observation, and as a method of evidence gathering, observation—both in the sciences and in the humanities—is flawed" [5, p. 6]. Even in the most objective of experiments, subjectivity still plays a role in the interpretation of results. True objectivity is not attainable [5].

Objectivity, however, is not absolutely necessary. After all, literary analysis (and, for that matter, most other types of humanities data) is by definition a more subjective field of inquiry and relies more on subjective interpretation and informed opinion than on hard, numerical data. Just as close reading requires the user to situate their analysis of the work within the historical and cultural underpinnings of the setting, so too does distant reading require this more *subjective* approach to analysis. Moretti attempted to apply such an analysis to his study of British novels written between 1740 - 1850. In his analysis, he noted a trend in the length of the title: moving

from an average of 15 - 20 words in 1740 to an average of 6 words by end of the era of study [4]. He found this curious, and through investigation discovered that the frequency of publications drastically rose: from a few books per year at the start of the era to around 100 per year by the close of the period [4]. With the influx of books, various magazines began to publish reviews of the books which provided summaries of the novels themselves. Prior to review publications, the title itself needed to serve as a summary of the novel—thus the longer title was necessary. But with the rise in reviews, which served the purpose of providing summaries to readers regarding the books in question, longer titles were considered superfluous and thus began to shorten [4]. Further, as more novels were published, readers became more familiar with the concept, and longer titles were not seen as necessary for *guidance* to the reader [4]. This, coupled with other constraints, such as library catalogues needing to constrain titles for space reasons, as well as the rise of advertising in the literary space (which could create flyers to advertise more content regarding the book), led to the decrease in the length of the title. Moretti's conclusions regarding the length of titles and their evolution was extrapolated using a distant reading approach, but could not be fully understood without this historical and cultural contextualisation. Thus, like close reading, this distant reading approach was culturally contextualised in order to properly understand the information.[24]

Thus, the assertion here is that distant reading is not that different from close reading, merely a different way of assessing the information at hand. As previously noted, close reading is most effective when situated within a particular context, a context that is generally culturally specific. Whether it be understanding the cultural references in a literary work, the underlying economic factors which might have driven a particular historical event, or the portrayal of certain characters in a painting due to perceptions of wealth, social class, or nationality, the notion of cultural hegemony[25] and effects of cultural on humanities data plays a vital role in understanding the content at its core.

> Broad attempts to generalize about a period or about a genre by reading and synthesizing a series of texts are just another sort of microanalysis. This is simply close reading, selective sampling, of multiple "cases"; individual texts are digested, and then generalizations are drawn. [5, p. 25]

[24] Moretti's examination of title lengths is but one example among many, too numerous to list here. For further examples, see [186] which details multiple examples of situating literary data analysis via historical context. Jockers also discusses this in his examples of Fanning's analysis of the decline of publication rates of Irish American authors or situating the use of language within particular cultural contexts, such as the difference in the use of definitive articles between American and British English [5].

[25] Cultural hegemony is defined as a situation wherein the ruling class manipulates a society's culture in order to impose their worldview, thus making it the accepted cultural norm [187], [188]. While it is usually portrayed in a negative sense (see Marxist philosophy), various cultural movements through the ages (in which works of art, history, music, and literature are situated) are by their definition a cultural hegemony: the ruling class of the time has shaped the culture to enforce a norm. The artists creating these works that we study are typically seeking to either further the established cultural norm or challenge it by providing commentary or criticism. Therefore, it is through these cultural norms that context is provided in the study of humanities data.

And while Jockers also states that such a macroscopic investigation is "contextualization on an unprecedented scale" [5, p. 2], utilising such an approach helps the reader better situate the material within the proper context. This situates both distant reading (as defined by Moretti) and macroanalysis (as defined by Jockers) as close reading approaches at their core, as they provide context in the way that close reading provides context—through their application of subjective analysis of other contributing factors (such as historical context, economics, etc.).

But distant reading has its own inherent problems. As a relatively nascent theory, perfecting the visualisations (which are a necessary component) is still a work in progress. One of the largest issues with these visualisations is their implementation (or sometimes lack thereof) of gestalt principles of visual perception.

2.3.1 Distant Reading: Leveraging Gestalt Principles of Visual Perception

While distant reading does provide an advantage from a cognitive standpoint (given that the brain is naturally primed to interpret patterns and images as opposed to reading, which requires training the mind to interpret words and derive meaning), many of the visualisations implemented in distant reading approaches fail to consider the ways in which the human brain understands patterns and interprets visual information. An understanding of how the brain interprets visual information allows the creator to provide a more nuanced analysis of the use of visualisations within textual environments. The basis of current understandings of pattern recognition is founded on the Gestalt Principles of Visual Perception.

Born out of gestalt theory (which was developed in the 1920s by German psychologists Max Wertheimer, Wolfgang Kohler, and Kurt Koffka), the gestalt principles focus on how individuals perceive objects based primarily on six factors: proximity, similarity, symmetry, closure, continuity, and figure-ground [1], [189], [190]. Each of these factors is used to describe how a user may perceive an object and the importance which may be placed upon the object. In addition, these factors influence how patterns are recognised and how information is *read* in a visualisation.

The first of the principles, the principle of proximity, specifically deals with the distance between objects and how the user perceives the objects to be organised. Objects which are placed closer together are seen as being *grouped* together, whereas objects with distance between them are seen to be separate [1]. This principle is demonstrated in Figure 2.4. In the left group, the diamonds are arranged closer horizontally than they are vertically, giving the appearance of 3 distinct rows. In the right group, the arrangement is reversed, giving the appearance of 3 distinct columns instead.

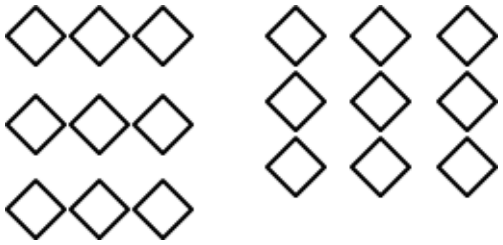


FIGURE 2.4: Use of proximity to group objects

The principle of proximity can also be used to show relative importance [189]. In Figure 2.5, the placement of the dot within the left frame indicates a place of priority, as it is centred. However, the dot placement within the right frame is off to the side, as though it is an afterthought. As a result, the brain ascribes the dot in (b) less importance than it does to the dot in (a).

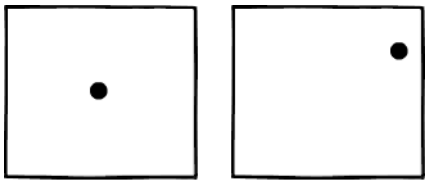


FIGURE 2.5: Use of proximity to show importance as demonstrated by Schlatter and Levinson [189]

The principle of proximity perhaps plays one of the most important roles when considering how data visualisations are read. Many of the visualisations seen within Digital Humanities deal with textual analysis. When such an analysis is represented as a visual item, assumptions are immediately made by the reader concerning the proximity of words in the visualisation relative to others (meaning those which are closer in position are *more related* than those which are further apart). However, this is not always the case depending upon the methodology employed. In a data visualisation generated using the *RoSE* tool created by University of CA, Santa Barbara (see Figure 2.6), the reader can easily see that some authors, who are positioned more closely to William Shakespeare at the centre of the visualisation, have a closer (in most cases, direct) tie to Shakespeare himself; whereas those authors positioned further away from Shakespeare lack this closer tie and are only tangentially related to him (perhaps via the notion of *6 degrees of separation*).

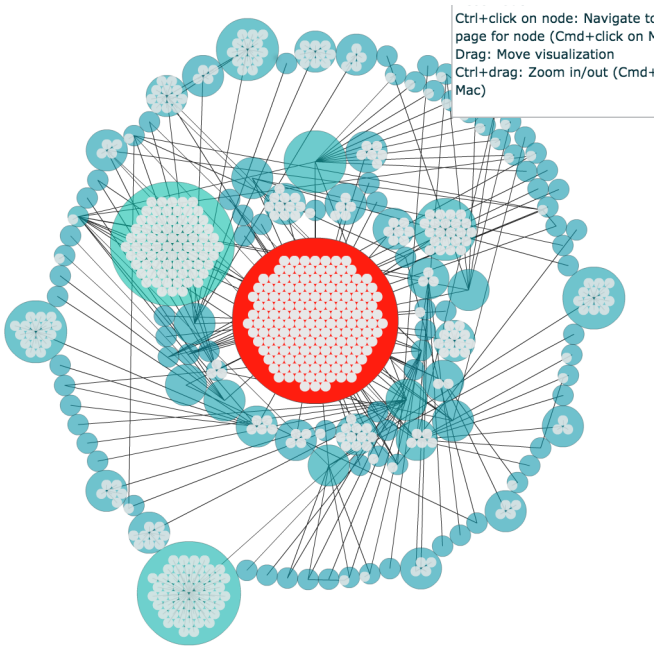


FIGURE 2.6: Use of proximity in a data visualisation created by *RoSE* [191] which shows groupings of authors and relationships

However, Figure 2.7 highlights the problems that can arise from unintended use of the principle of proximity in a data visualisation. Here, in a data visualisation generated by the *Letters of 1916*, cosine similarity is used to model the closeness of words to the keyword *bread*. This visualisation implies that there are 9 keywords (the 9 circles that all cluster around the centre keyword *bread*) which have a closer semantic relationship than all the other keywords. However, in this case, that relationship is spurious at best. This particular methodology does not visualise the differences in the *distance* of the relationship between keywords; it simply visualises *all* keywords which fall within the specific cosine similarity. These nodes simply happen to cluster around the keyword at random. If the visualisation is run again, these nodes will arrange themselves differently around the keyword each time as this *clustering* is simply due to the software package's (in this case R's visNetwork) rendering of nodes and edges. As a result, a relationship is implied where none exists. This is an inherent problem with the method of vector space analysis (more of which will be discussed later).

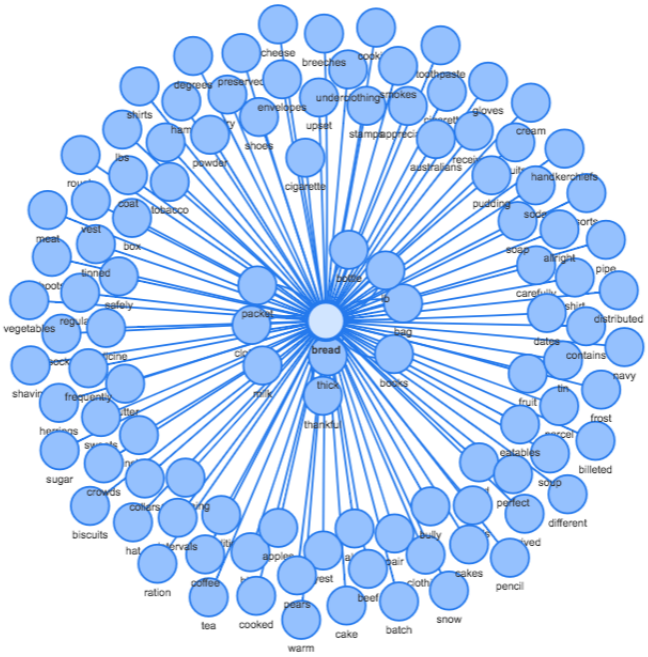


FIGURE 2.7: Use of proximity in a data visualisation in the *Letters of 1916* which implies unintended relationships between keywords [192]. A larger version of this image is available in Appendix E in Figure E.4

Like proximity, the principle of similarity can also be used to affect perception regarding object grouping. This principle states objects appearing similar to each other will appear to be grouped [1]. In Figure 2.8, the solid colour diamonds appear to be related to each other in a way that they are not to the transparent diamonds.

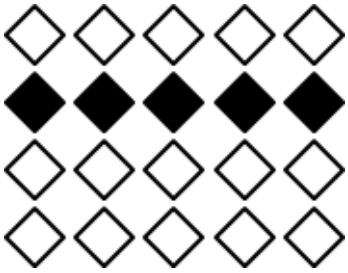


FIGURE 2.8: Use of similarity to group objects

Here, too, is another principle commonly leveraged within Digital Humanities visualisations. Colour is often used within these types of visualisations, which leads the reader to assume connections that may or may not exist. Because these coloured items are seen as *similar* (e.g. all the blue items are related and all the green items are related), the reader is able to intuitively and quickly draw connections in the dataset in a manner that is not as easily reproducible using a standard search or browse approach.

A strong example of the use of the principle of similarity in data visualisations is that found in the *Letters of 1916* (see Figure 2.9).[26] Here, the letters are grouped into topics, each of which is assigned a colour, immediately drawing the eye to various collections and signalling to the brain that the letters within the coloured groups are somehow related. In some places the letters overlap more than one colour, signalling multiple relationships among these letters—a valid interpretation as letters may belong to more than one topic [193].

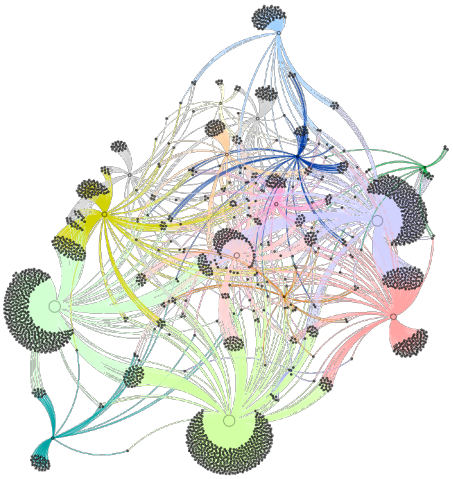


FIGURE 2.9: Visualisation of letters dispersed into topics by colour as seen in the *Letters of 1916* [193]

The application of the principle of similarity here greatly assists the reader in identifying groups of letters which may be of interest in a manner that is less cognitively intensive than scanning a list of topics. Consider, by contrast, a list of letters by category as seen in Figure 2.10. While both Figures 2.9 and 2.10 allow the user to see similar information (letters divided into categories and an understanding of the amount of letters in each category), Figure 2.9 is viscerally more appealing, less cognitively difficult for many readers to parse (as it is much easier to immediately recognise groupings of objects by colour than it is to scan a list of text and numbers), and additionally allows the reader to identify areas of overlap between topics that are not at all apparent in Figure 2.10.

[26] The *Letters of 1916* is a public humanities project developed at Maynooth University. It offered a collection of letters—both governmental and personal correspondence—written to or from Ireland (originally from November 1915 to October 1916, but later expanded to 1923), with the initial goal of exploring contextualisation around the Easter Rising.

FIGURE 2.10: List of letters by category as seen on the original *Letters of 1916* website [192]. A larger version of this image is available in Appendix E in Figure E.5

The remaining four principles are less obviously related to data visualisations but are important to understand as they serve the purpose of explaining how the brain resolves visual ambiguity. The third principle, known as symmetry, concerns itself with how the brain attempts to reduce more complex images into simpler ones. While a particularly complex scene may have multiple interpretations, the brain will automatically look for the simplest solution first, providing it symmetry [1]. In Figure 2.11, the brain interprets the original figure on the left as two interconnected circles, not as two half-moons stacked atop each other. This demonstrates the brain's ability to reduce the more complex object on the left to the simpler conjunction of two circles (which is less complex than seeing it as two half-moons or even as a possible third or fourth interpretation of shape).

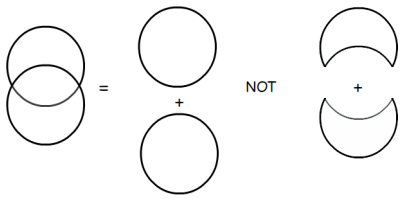


FIGURE 2.11: Use of symmetry

While not as obviously utilised in data visualisations, especially when compared to the principles of proximity and similarity, the principle of symmetry should still be considered if only to understand how users might recognise patterns within the visualisations themselves. While it has already been noted that overly complex visualisations can create cognitive overload, if the visualisation is arranged in such a way that symmetry exists, the reader will find it easier to recognise patterns and draw conclusions. The application of the principle of symmetry is far more subjective here than those of similarity and proximity, but it is an important consideration

nonetheless.

The fourth principle is that of closure, which is the principle used to describe the brain's tendency to see objects as part of a whole, rather than separate parts. Johnson uses a common example of the closure principle, as seen in Figure 2.12 [1]. Although the image itself consists solely of 3 open faced triangles and 3 'Pac-Man'-like shapes, the object is perceived as consisting of 2 triangles and 3 circles.

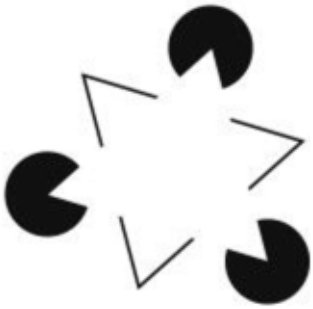


FIGURE 2.12: Use of closure as seen in [1]

Perhaps the most common implementation of the closure principle is seen in many modern software applications where a collection of objects (often a series of messages or documents) is represented by what appears to be multiple objects of the same type stacked atop each other (see Figure 2.13). The image itself is merely the one object with additional lines giving the appearance of objects *behind* the primary object. The closure principle describes how the brain sees this object as a single collection. Each of the three icons illustrated in Figure 2.13 use the closure principle to illustrate common functions in software applications.

FIGURE 2.13: Icons using the closure principle

Like closure and symmetry, the principle of continuity is used to assist the individual with the resolution of visual ambiguity. According to Johnson, the principle of continuity "states that our visual perception is biased to perceive continuous forms rather than disconnected segments." [1, p. 15]. The mind is conditioned to automatically fill in aspects that are visually missing in order to ascertain a complete object. In Figure 2.14, the brain sees both a complete square and a complete circle. While it could be argued that the aspect of the circle underneath the square may not be complete or could be misshapen or construed, the general consensus would be the existence of a complete circle.

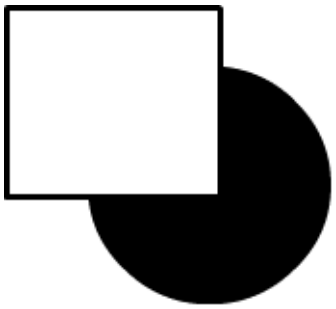


FIGURE 2.14: Use of the continuity principle

The logo used by International Business Machines (IBM) is also a classic example of the continuity principle (see Figure 2.15). While the logo itself is made up of solid horizontal lines with spacing in between, the brain forms a complete picture and allows for the word "IBM" to be brought to the forefront.

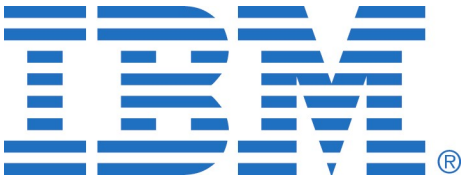


FIGURE 2.15: The IBM logo which demonstrates the continuity principle. Reprinted under IBM's terms of use [194]

The final gestalt principle is that of figure-ground. The figure-ground principle is used to focus attention, allowing for objects placed into the figure (i.e. the *foreground*) to be given a higher place of focus and attention than those objects placed into the ground (i.e. the *background*) [1]. Artists often use the figure-ground principle to play with perspective and draw the eye or to allow for double meaning in a work (see Figure 2.16).

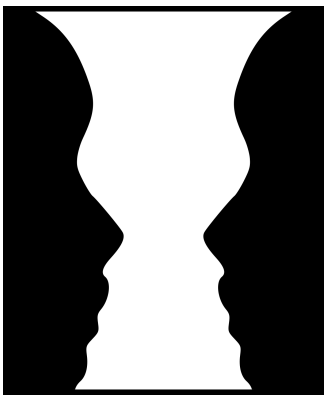


FIGURE 2.16: Shows the use of figure-ground. Is the foreground image a *vase* or *two faces*? Both interpretations are valid depending on how one interprets foreground versus background

Figure-ground is also used commonly in many software systems. The use of the modal overlay to display temporary information (such as in *Borchward's Journey*

as seen in Figures 2.17 and 2.18) is a common implementation of the figure-ground principle [195]. In this particular scenario, the footnotes are stacked on top of the next in order to make room for additional text (as in Figure 2.17). However, utilising an on::hover action with the mouse, the user can place the cursor of the mouse over the desired footnote which then moves to the figure, drawing the eye and allowing the user to view the content (see Figure 2.18). This content is then also seen as temporarily being *more important* than other footnotes as it is given visual prominence in the figure, while other content is relegated to the ground. This type of interaction technique allows the user to gather new information without breaking the interaction within the current figure.

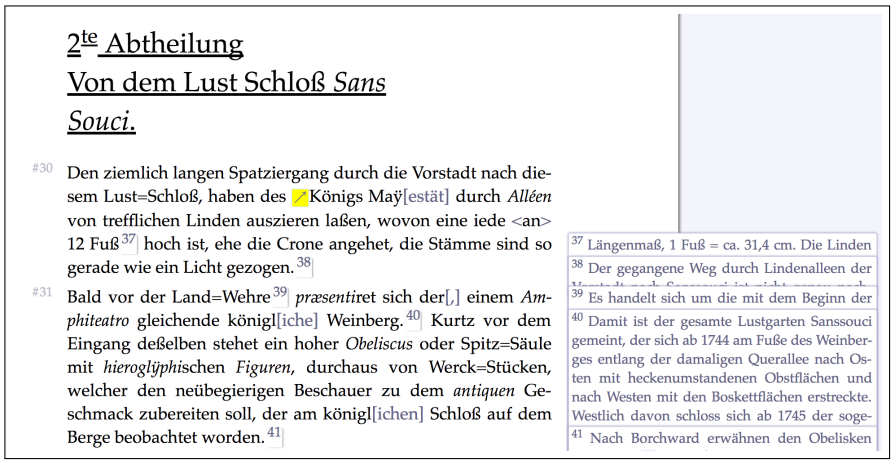


FIGURE 2.17: Footnotes in *Borchward's Journey* stacked atop each other [195]

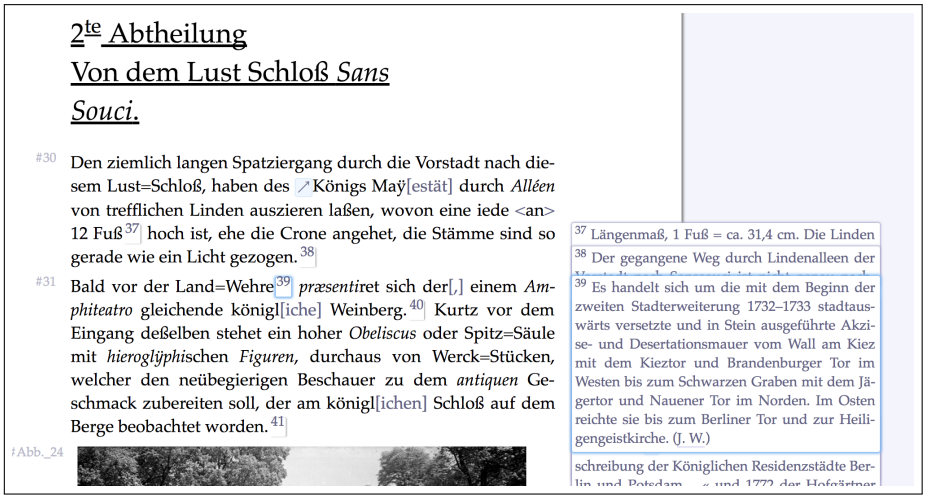


FIGURE 2.18: Footnote 39 in *Borchward's Journey* brought to figure in order, elevating its visual importance [195]

There is a seventh gestalt principle known as common fate, which, unlike the previous six discussed (which deal with static images), concerns itself with moving images. This principle states that those objects that are perceived as moving together are somehow interrelated or grouped [1]. Take, for example, a series of

squares lined up across numerous rows. If certain squares were seen to *wiggle* in the animation, then those squares would be perceived as being grouped together, even if the squares were not in proximity to each other or visually grouped in any other way. Since they move together, they are seen having a *common fate* and thus are related. Because this principle deals with moving images, it is not discussed in this chapter in detail (as much of the work here is concerned with more static content); however, it is noted here as it may have implications in how certain data visualisations may be designed within Digital Humanities (specifically if these visualisations incorporate motion).

Each of these principles mentioned above work in concert, not in isolation. Used correctly, designers can easily draw the reader's eye to a particular area of the visualisation or provide visual clues to assist with pattern recognition and understanding of the dataset as a whole. However, due to their subconscious nature, gestalt principles can easily imply visual relationships which may not be intended, thus leading to confusion or other unintended behaviours. As demonstrated by some of the examples here, ignoring the impact of gestalt principle of perception can lead the reader to draw inaccurate conclusions with regard to relationships among sets of data or keywords, thus lowering the usefulness of the visualisation as a whole.

2.4 Discussion: Applying Explored Concepts Within a Critical Analysis

As computation techniques continue to evolve and hardware becomes more cost-effective, data visualisations continue to rise into prominence. Already, projects such as *London Smells* [196], *Personæ* [197], *Visualising Emancipation* [198], *Letters of 1916* [192], and *The Diplomatic Correspondence of Thomas Bodley* [199] have begun to incorporate the use of data visualisations to convey deeper meaning. Many of these projects attempt to visualise textual analysis in an effort to identify patterns in text. Some of the methods used include topic modelling, vector-space analysis, sentiment analysis, and network analysis. Examples of each are presented below for the sake of discussion.

2.4.1 Topic Modelling

The first example is also the most prolifically used: topic modelling. Topic modelling is one of the most common means of approaching textual analysis (especially as an underlying data model for visualisation of text) and encompasses a wide range of different approaches. In Figure 2.19, Emily Barry presents a topic model-based visualisation using Non-negative Matrix Factorization (NMF)[27] in order to visualise

[27] Originally proposed by Lee and Seung in 1999 as an algorithmic method for parsing image data [200], NMF has since been adopted by textual scholars to explore the distribution of topics within document clusters [201]–[203]

the distribution of topics in United States Supreme Court cases over the last two centuries.

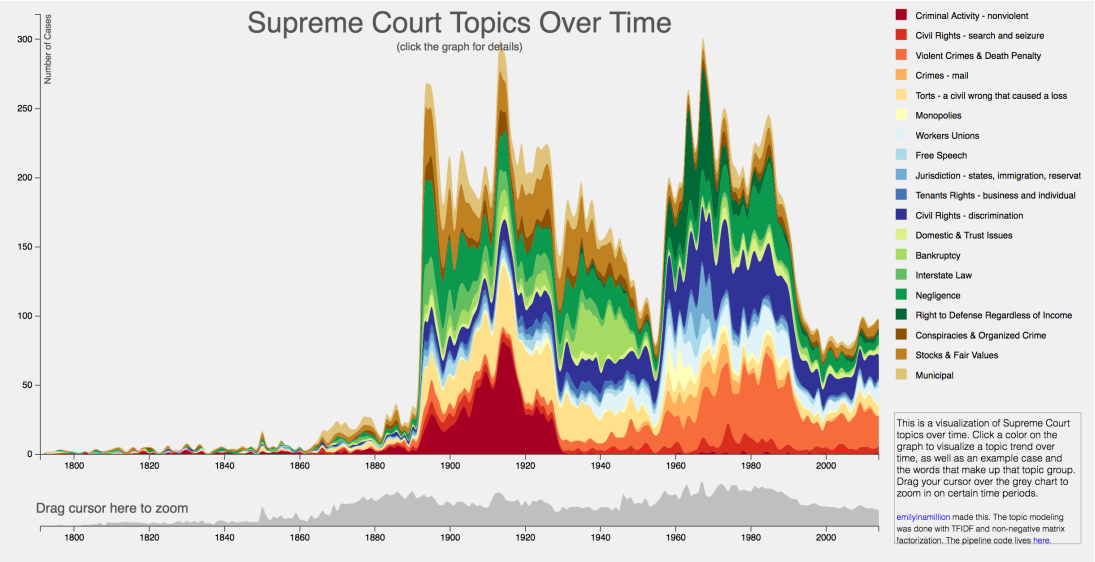


FIGURE 2.19: Topic model leveraging Non-negative Matrix Factorization showing the distribution of topics in U.S. Supreme Court cases over time [204]. A larger version of this figure is available in Appendix E in Figure E.6

An initial inspection of this visualisation allows the reader to quickly ascertain if there is a particular time period on which she should focus her research if she is interested in a particular theme or category. Additionally, it allows her to quickly identify thematic trends in cases over time (acting also as a type of temporal visualisation). If attempting a solely close reading of these cases, it would be difficult (as well as time consuming) to notice a trend in theme. As an example of this ability to identify patterns and trends, consider, Barry's NMF model in Figure 2.19. When examining the visualisation between the period of 1950-2000, we see an up-tick in cases related to civil rights (discrimination in blue and search and seizure in orange). Considering the social upheaval of this period in American history, these thematic trends are not surprising. During this time period, America experienced the rise of the Civil Rights Movement, followed closely by the LGBT movement. Both of these movements revolved around cases of discrimination and illegal search and seizure, a common practice of state intimidation of these marginalised communities in the U.S. at the time. This visualisation provides easy insight into some of the larger issues shaping American politics during this time period. Knowledge of the history of this era allows such insight to be situated within its proper context, just as a close reading of the individual cases would also allow.

From a Cognitive Load and Visual Perception perspective, Barry's visualisation provides a solid example of best practice. The use of colour to differentiate the categories leverages the gestalt principle of similarity as well as that of symmetry. The reader is able to clearly distinguish between the topics as groups (similarity) and is

also able to view each subsection as its own unique item (symmetry). This approach also aids attention via the use of visual search, allowing the reader to quickly scan the page for information relevant to her query. With the implementation of the colour-coded legend alongside the visualisation, the split-attention effect is minimised, thus reducing the cognitive load on the reader and making it easier to identify areas of interest for further inspection.

2.4.2 Vector-Space Analysis

Another example of the types of data visualisations being leveraged within Digital Humanities is that of vector-space analysis. Ryan Heuser leverages vector-space analysis to analyse the text of "every significant English-language and foreign-language title printed in the United Kingdom during the 18th century, along with thousands of important works from America" [205], [206]. In Figure 2.20, Heuser shows the distribution of words in these works, with the different colours representing each word's syntactic meaning (e.g. noun, verb, adjective, etc.).

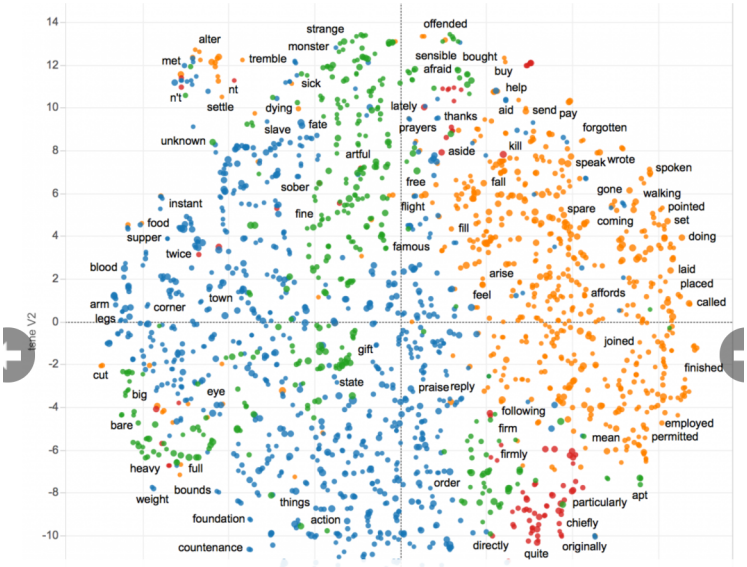


FIGURE 2.20: Use of Word Vector Analysis to show syntatic relationships between words [206]. A larger version of this image is available in Appendix E in Figure E.7

It is interesting to note here that, as in Figure 2.7, Heuser mentions that these types of visualisations may confuse the reader who would (likely incorrectly) assume the distance between each node is a semantic difference, meaning that there is a correlation between words and distance (the closer the words exist to each other, the more closely they would be related in the text). For example, the word *state* and *gift* must have some kind of semantic relationship because they appear close to each other on the graph; however because of the type of vector space analysis that was leveraged, the distance between nodes in Figure 2.20 represents a *syntactic* difference rather than a semantic one [206, sec. "1B"]. So in the aforementioned example, there

is no true semantic relationship between these two words; they are simply syntactically related in their use as nouns.

Without careful contextualisation, readers would likely draw inaccurate assumptions regarding these texts based solely on the visualisation. The issue here then is not only one of understanding the applied gestalt principle of proximity but also providing additional contextual information to the reader, so she understands both *how* the graph was constructed as well as the *context* in which the data has been plotted. Due to the size of the dataset and the disparity in the data (every significant literary work in 100 years produced across cultures and languages), it would also be nearly impossible to provide any kind of historical or cultural context given the sheer size of the sample set.

In a focus group study conducted in late 2016 (see Appendix: A),[28] numerous individuals mentioned the need for context when leveraging data visualisations. One participant noted:

> So...a lot of the time [I] might go into these visualisations or these websites and find myself thinking "That's a fantastic resource. I don't know how to use it to advance my work. How does this benefit my research." So...and then I think images without context don't have meaning. They're just an image. [207, t.s. 00:17:34]

From the perspective of the reader, this lack of context is problematic. However, when context regarding the *method* behind the visualisation is presented (which includes an explanation as to the role of vector-space analysis), the reader can gain new (and more accurate) insights into the text. Although it remains pertinent to note that this type of visualisation completely abstracts any kind of non-methodological context from the data, making a true correlation between close reading and distant reading here less plausible.

2.4.3 Sentiment Analysis

Other types of methods for generating textual visualisations find themselves less fraught with problems. Sentiment analysis, for example, is commonly utilised with social data (for example, when analysing tweets from Twitter or comments from YouTube). However, it can certainly be applied to the type of textual analysis common within Digital Humanities. *London Smells*, as seen in Figure 2.21, provides an example of the use of sentiment analysis on a number of medical health reports.

[28] See Chapter 4 for further information on the design and analysis of this focus group.

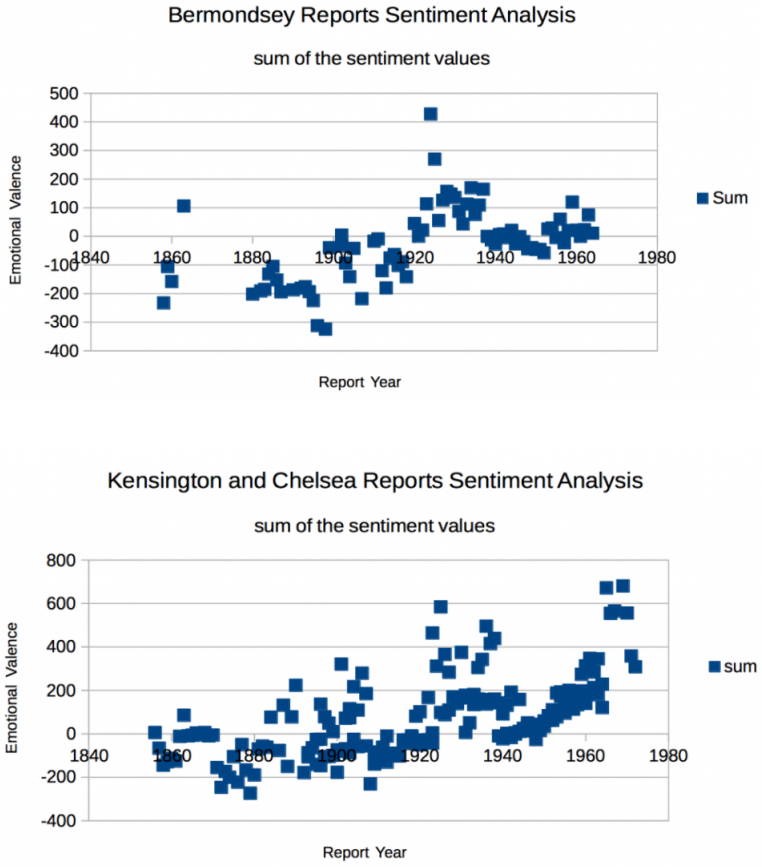


FIGURE 2.21: Use of sentiment analysis in the *London Smells* project [196].

Here, *London Smells* uses the Syuzhet package (written by Matthew Jockers) in R to analyse the general sentiment of the language used in the reports. This package, which works in conjunction with Standford's coreNLP package, analyses a text to determine if the text in question uses language that would be seen as *positive* (i.e. any number which is greater than zero), *negative* (i.e. any number which is less than zero), or *neutral* (i.e. any number which is zero) [208]. While *London Smells* presents these data visualisations without context (a problem in and of itself), it can be deduced from the data and the methodology that various reports in a given time frame are analysed using this sentiment analysis method, and the sum of their sentiment is then plotted according to the *emotional valence*. This allows the reader to ascertain if there are certain years where there is a higher preponderance of positive or negative sentiment which may be worth investigating. Additionally, this information can be situated within an historical context due to the narrowed field of enquiry (understanding sanitation issues in historical London, for example). Finally, due to the gestalt principle of proximity, readers are able to quickly ascertain certain areas where sentiment is clustering at various times, an analysis which would not be easily deduced through close reading. All this combines to provide the reader with a different method of reading these medical reports without necessarily needing to parse

complex medical terminology or understand particular cultural jargon, which could skew the sentiment to one side of the analysis or another. This type of analysis allows the text to be accessible to the reader in a manner similar to (albeit functionally different from) a close reading approach.

2.4.4 Network Analysis

Finally, the *Personæ* project at NUI Galway utilises network analysis to investigate the relationships of characters in Shakespeare's *A Comedy of Errors* [197]. Pictured in Figures 2.22 and 2.23, *Personæ* attempts to visualise the number of speeches by a particular character, as well as how other characters are referenced by a particular character in the text. The outer ring indicates a character, who is highlighted via the use of colour and by toggling a checkbox in the legend along the left-hand side. This ring identifies the frequency of speech by a particular character. The inner ring then shows who the character is speaking about, allowing the reader to form a type of network regarding the characters. Such a visualisation assists the reader with constructing a cognitive scaffold (or mental map) of how various characters interact with each other.

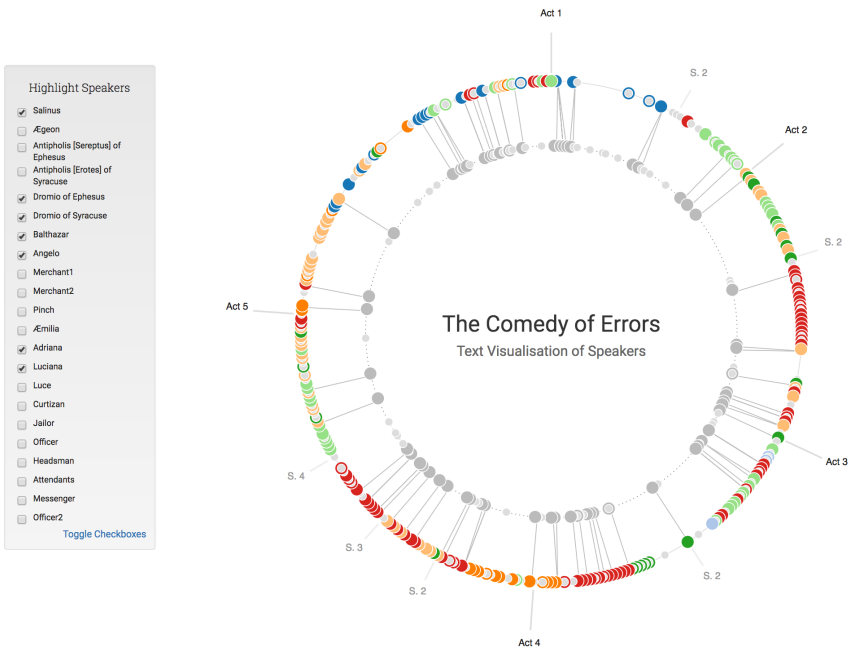


FIGURE 2.22: Use of network analysis in *Personæ* [197].

Unfortunately, while this visualisation does allow the reader to see who each character is discussing individually, it does not build a true network visualisation as the reader cannot easily see connections between all of the various characters at once. Doing so requires a careful reading and analysis of each individual node. However, because of the use of the gestalt principle of similarity (e.g. the use of colour) and the gestalt principle of figure-ground (specifically seen in Figure 2.23), the reader is able to build a small conceptual graph in memory and use said graph

for further research and exploration. In this case, the visualisation is acting as an augmentation for close reading—a sort of reference guide to the material (rather than a true immersive experience, which would transmute the visualisation from a research tool into a true engagement with the source text).

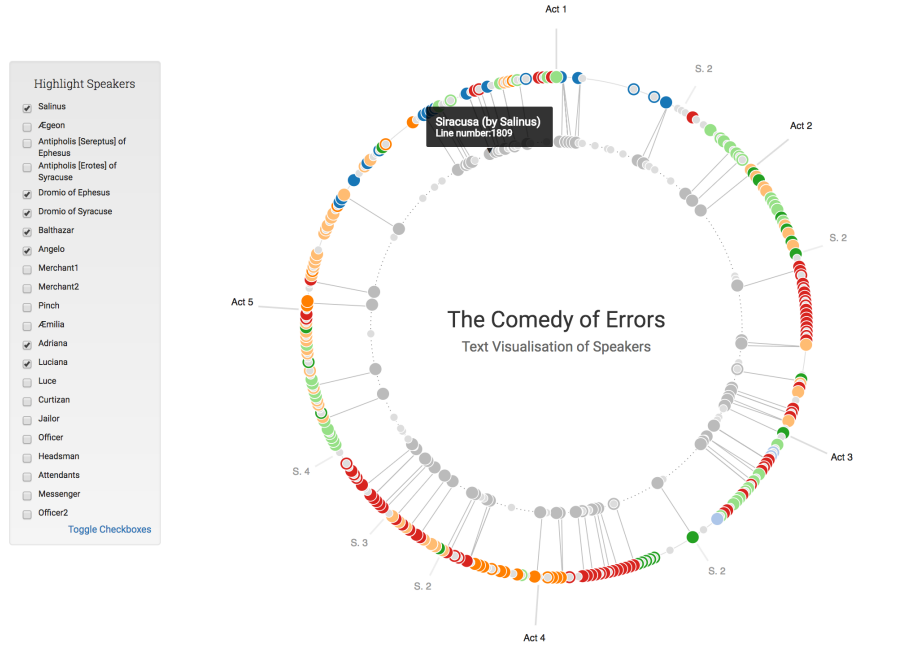


FIGURE 2.23: Highlights the use of the modal overlay on::hover in *Personæ* to showcase the character being discussed in the network visualisation [197].

2.4.5 Other Types of Visualisation

Textual analysis, however, is not the only tool in the digital humanist's tool chest. While data visualisations can be applied to textual sources via the use of various text mining algorithms or can assist in the identification of patterns and phenomena, visualisations can also be used to augment the search and browse experience of the user, thereby augmenting a discontinuous, close reading of the object. Consider the implementation of the interactive timeline in the *The Diplomatic Correspondence of Thomas Bodley* as seen in Figure 2.24 [199].

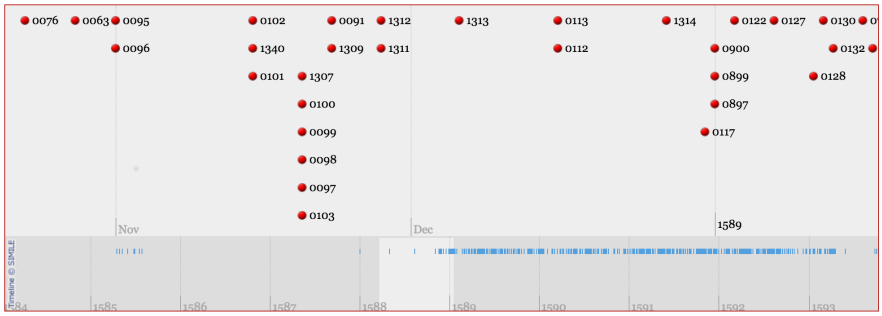


FIGURE 2.24: Timeline visualisation used to facilitated alternatives to search and browse mechanic, as seen in *The Diplomatic Correspondence of Thomas Bodley* [199].

Here, Adams has leveraged a temporal visualisation in order to allow the user to visually understand the movement of Bodley's letters over time. With each node *clickable* into its related object of correspondence, the reader is provided with an alternative, non-textual approach to the traditional search and browse technique. Search and browse, while ubiquitous in digital environments, tend to be very text heavy and often rely upon the reader having at least a cursory knowledge of the dataset in order to input the proper terms and receive valid search results. The type of search and browse implemented in Figure 2.24 may impose less cognitive load on the reader by engaging the visuospatial sketchpad and allow her to more quickly (and perhaps accurately) browse the collection. Furthermore, it does not rely upon an understanding of the data within the corpus itself, thus allowing for a wider audience for dissemination. By incorporating such a visualisation, *The Diplomatic Correspondence of Thomas Bodley* is providing another form of access into their corpus, catering to those readers who wish to experience the data in a more visual manner. This technique serves to facilitate discontinuous reading without the cognitive overload inherent in hyperreading.

Another example of a data visualisation used to facilitate search and browse is that of the geo-spatial visualisation. *Visualising Emancipation* is an example of the use of geo-spatial data transformed into a mapped visualisation. This map, much like Figure 2.24 above, provides an alternative implementation of search and browse (see Figures 2.25 and 2.26).

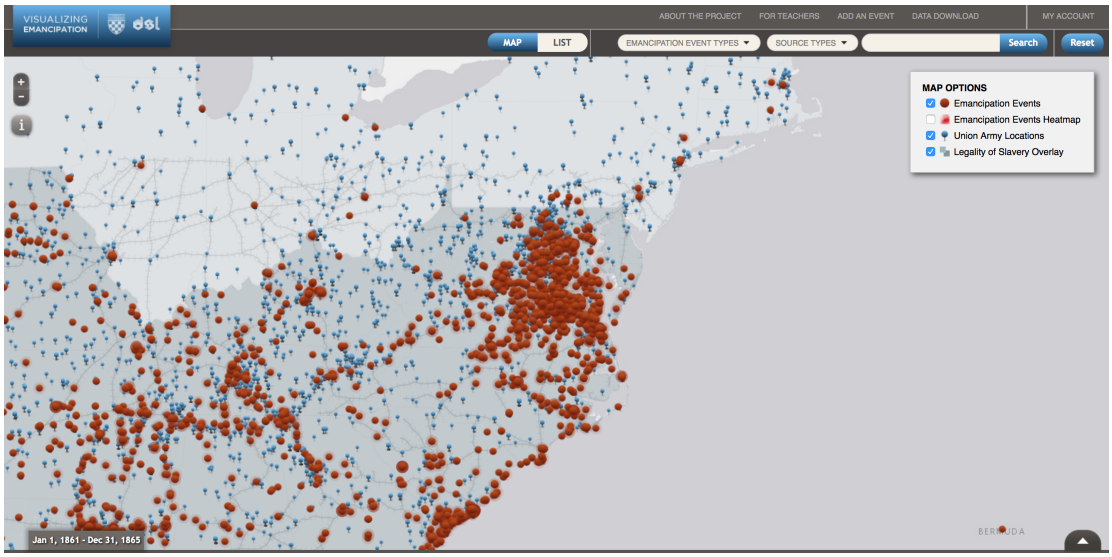


FIGURE 2.25: Use of geo-spatial mapping techniques to facilitate search and browse as seen in *Visualizing Emancipation* [198].

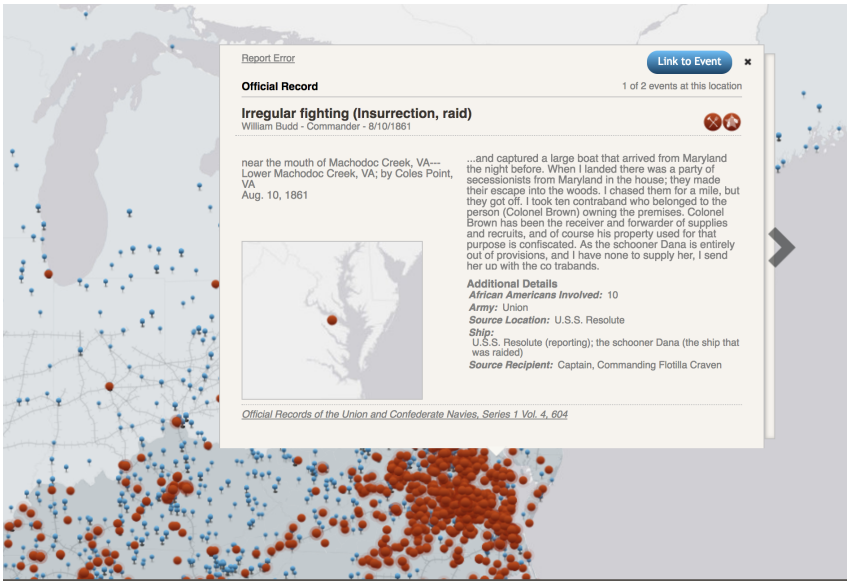
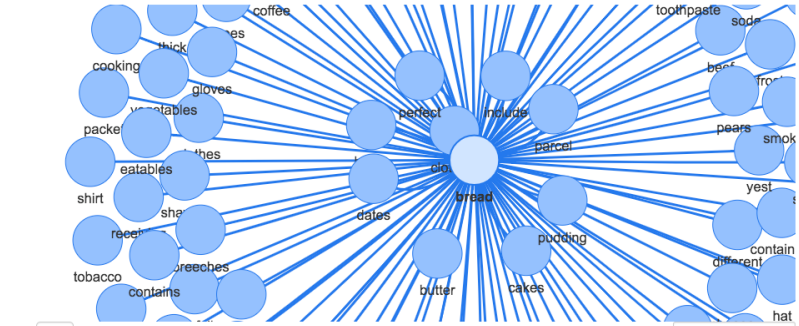


FIGURE 2.26: Additional object information presented by *Visualizing Emancipation*, allowing the reader to click into a particular record from the geo-spatial visualisation [198].

The use of the spatial visualisation in *Visualizing Emancipation* allows the reader to identify *geographical* areas of interest, as well as to identify patterns of data as related to the physical location of the event occurrence. The reader can then click into a specific node on the map to view additional information about the event (see Figure 2.26); this *modal overlay* annotation of the node also contains a hyperlink that links to the actual record of the object, providing the reader with the same data they might expect from a more traditional *search and browse* approach and additionally engaging the reader's brain through the use of visual patterns and cues. Here again, this type of search pattern engages the visuospatial sketchpad and offloads some of the

cognitive effort from the central executive, thereby decreasing the effort involved in searching the corpus. This mode of search also assists with the split-attention effect because it hides unrelated information until it is necessary, allowing the reader to discover larger patterns before examining the minutiae of the data. The application of numerous gestalt principles is readily apparent here in the clustering of nodes (proximity), the use of colour to differentiate between node types (similarity), and the use of *pop-ups* to convey detail (figure-ground). Overall, this implementation still allows for a *traditional* close reading approach, while offloading some of the cognitive effort inherent in close reading in digital environments.

The earlier example from the *Letters of 1916* (see Figure 2.7) is also an example of a search and browse replacement. While this visualisation serves as an analytical tool for identifying patterns and relationships within content nodes, the visualisation also contains an element of interactivity, as seen in Figure 2.27.



Letter	Left Context	Keyword	Right Context
101	parcels one containing a large loaf of	bread	and the other containing tea sugar
1031	we had a little of the prison	bread	when taking lunch with the priest
118	nearly all the men here is receiving	bread	from switzerland every week except myself
123	would send me some and also some	bread	as i cannot get any from

FIGURE 2.27: Interactive view of a vector-space analysis model in the *Letters of 1916*.

By providing interactivity through a mouse click action on a content node, the visualisation reveals each instance of the selected keyword *in context* within the letter. The reader then is given the option to click further into the letter object itself, thus allowing for a different approach to search and browse. This type of interactivity (along with the other mentions of search and browse techniques here) has an immersive effect upon the reader, which in and of itself provides cognitive benefits.[29]

Each of these examples highlights different approaches to distant reading. Some (such as Figures 2.19 or 2.21) can provide the same types of insights as a more traditional close reading (albeit in very different, functional ways). They allow for the data to be culturally or historically contextualised, assist the reader with constructing cognitive scaffolding or providing deeper understanding of the work, thereby

[29] The concept of interactivity and immersion and their effects upon cognition are discussed in further depth in Chapter 3.

enhancing the experience of the reader, all of which are aspects of the various definitions of close reading. Others (such as Figures 2.24 or 2.26) act as augmentations to a traditional close reading by enhancing the experience of a discontinuous reading approach (without the inherent difficulties associated with hyperreading) but still rely upon the source text for the user to truly understand inherent meaning. Finally, some distract from close reading altogether by failing to provide proper context or by ignoring inherent gestalt principles of visual perception. As with any tool or method, a failure to understand its scope and place can lead to misunderstandings or false interpretations.

2.5 Conclusion

While reading is an important aspect of any textual analysis, digital environments provide the reader with more options beyond that of a traditional close reading. And while hyperreading can augment the experience of the reader by allowing her to move through a text in a non-traditional manner (specifically non-linearly), these methods are fraught with their own challenges, such as spatial instability, metacognitive bias, cognitive overload, and a lack of physical markers to aid with recall.

Data visualisations, however, can assist the reader with parsing more complex concepts, as well as offer alternative implementations of the traditional search and browse technique. By engaging the visual centres of the brain—such as the spatial cognition region which is responsible for drawing context and meaning [209] and the visuospatial sketchpad in working memory—the reader is able to draw new context and meaning that is not possible through the application of traditional reading modalities (or even newer, digital modalities such as hyperreading). Furthermore, it is still necessary to apply contextualisation and outside analysis in order to gain deeper insight into the text via the visualisation. This specifically situates distant reading as not simply a *new* method of reading but rather a *reinvention* of close reading, a mode of adaptation of a practice given new environmental changes. Just as the codex form allowed readers to engage with text in new ways, so too does distant reading.

However, creators of these visualisations must understand how the brain draws context and meaning from visual images (such as the impact of Gestalt Principles of Visual Perception or the role of working memory and Cognitive Load Theory); without this understanding, the creators may allow for inaccurate conclusions to be drawn, which can then call into the question the veracity of the work at large and its underlying methods (despite there being no errors in these underlying methods). These types of visualisations must be approached with some level of caution and understanding of many principles inherent to field of User Experience Design.

Perhaps the greatest boon of these visualisations, though, is the introduction of new methods of search and browse through the application of interactivity. In Chapter 3, the role of interactivity and its impact on the brain is discussed, as are common

techniques of interaction. This chapter, coupled with the Chapter 3, will serve to provide a strong basis for how data visualisations can truly transform digital scholarship.

Chapter 3

Interactivity, Immersion, and Engagement: Examining the Importance of Interactivity & Immersion in Digital Research Environments

While visualisations can be used to augment the reading experience in digital environments,[1] more than a static image is required to truly engage the cognitive faculties of the brain. If the goal of a digital project is to disseminate knowledge and further research, then the visualisation must employ an interactive component in order to fully immerse the reader in the experience. An understanding of how interactivity influences immersion, which in turn influences engagement, is essential in positioning the role of the interactive visualisation in the Digital Research Environment (DRE).

Studies have shown that online environments contribute to an increase in visual intelligence—that is, a type of "non-verbal reasoning" that "requires individuals to manipulate aspects of problem space using visual mechanisms for example, size, difference, and identification of relationships amongst objects or remembering/recognizing sequences" [210, p. 253].[2] It is a natural extension of this statement that visualisations better serve those digital natives than traditional narrative text as they are able to more quickly decode information contained within visual datasets due to this higher visual intelligence. Further research by Small, Moody, Siddarth, et

[1] See Chapter 2 for further discussion

[2] In discussing the work of Johnson and Johnson [211] and Subrahmanyam, Greenfield, Kraut, et al. [212], Hadlington notes that individuals who report being "frequently online" have a higher incidence of visual intelligence than those who report being online less frequently [210, p. 253]. It is concluded that these *digital natives* had higher visual intelligence due to the nature of the digital environment and its need for non-verbal reasoning (for further information on how the term *digital natives* is utilised in this thesis, please refer to Chapter 2). Hadlington does note, however, that this conclusion should be approached with caution as this increase in non-verbal reasoning could be more influenced by personality than by any kind of quantitative study [210].

al. discusses how digital natives (referred to as *internet savvy participants*) have increased activation in the area of the brain that is responsible for decision-making and the ability to process complex information and synthesise it in order to extrapolate information [213]. This study indicates that the task of searching is actually more cognitively stimulating than reading itself, which leads to the assertion that providing innovative search and browse techniques is essential to producing an experience that is both useful and cognitively fulfilling. Johanna Drucker reinforces this need for robust search and browse, stating:

> The standard theory of interface, based on the 'user experience,' is reductively mechanistic. Its goal is to design an environment to maximise efficient accomplishment of tasks—whether these are instrumental, analytic, or research oriented—by individuals who are imagined as autonomous agents whose behaviours can be constrained in a mechanical feedback loop. Challenges to that conception arise from within the information studies community—where interface is embedded in the motivations of an embodied user engaged in some activity that may or may not be goal oriented, highly structured, and/or driven by an outcome—but might equally be the diversionary experience of wandering, browsing, meandering, or prolonging engagement for the purpose of pleasure or an even lower level notion like keeping boredom at bay or idle distraction and time squandering. [6, p. 151]

Data visualisations are one mechanism that can lead to this more robust search and browse experience, but only through the application of interactivity. Interactivity offers the reader the opportunity to truly engage with the data by immersing her in the *experience* of the data. This immersion leads not only to further understanding and comprehension but also increases her self-efficacy, which has been shown to provide a boost to cognition when attempting to learn in a digital environment [87].

The following chapter attempts to explore the notions of interactivity and immersion, tying them to the benefits these concepts can provide to visualisations and reading as a whole. Drawing on the work of Leont'ev as discussed in Chapter 1, this chapter will open with a discussion centred on Cooper, Reiman, and Cronin's concept of *Goal Directed Design*, laying the groundwork for understanding behaviour in digital environments and the motivations that often drive users of the types of Digital Research Environments that are the primary concern of this work. Norman's idea of *Emotional Design* will also be explored as a central component for designing effective systems. This then leads to an exploration of the subject of interactivity as it relates to immersion and how immersion truly affects the user experience on not only an emotional level, but also on a cognitive one. This will then conclude with a discussion around the application of these principles to visualisations already in use and an analysis of how these principles may or may not be applied in many of the resources within Digital Humanities that offer data visualisations.

3.1 Goal Directed Design: Designing for the End Game

Understanding what drives the user—that is, the motivations behind her actions—is a major component of building an immersive experience. The principles of gestalt psychology[3] form the foundation for a discussion of understanding user behaviour and how to design for those behaviours. Gestalt psychology explains how the user draws connections and infers relationships between objects on screen. This mode of interaction then affects how the user will interpret visual information on screen in order to accomplish a task. Every interaction a user has with a digital resource involves a task or series of tasks—whether that task is as simple as reading a page or a more complex interaction such as conducting a search or analysing a network—an assertion that has its roots in activity theory. By understanding that software systems (whether they be common, everyday computer programs or scholarly resources such as Thematic Research Collections, Digital Archives, or Digital Scholarly Editions) are meant to assist the user with accomplishing some kind of task, Norman developed his theory of Activity-Centered Design, which created a hierarchical relationship between Tasks, Activities, Actions, and Operations (wherein tasks contain activities, which contain actions, which contain operations) and based much of his theory on the work of Leont'ev [215]. Since humans have an innate desire to accomplish tasks, Norman's theory situates itself within the proper psychological framework regarding motivation and system design.[4]

Norman defines Activity-Centered Design (ACD) as an extension of the more commonly known Human-Centered Design (HCD) approach, which itself seeks to extend the notion of *user-friendliness* by focusing on the needs and desires of the users and stakeholders from the outset and continuing to involve said individuals through the entire design and development process, not just during end-stage testing [215], [216]. ACD, however, differs from HCD by extending the metaphors inherent in HCD—the focus on understanding the needs and desires of the users—to the technology, tools, and activities the user may leverage in their use of a software system. Norman begins by focusing on activities and crafts a hierarchical approach to how those activities may be accomplished through a series of sub-activities, tasks, and sub-tasks.

While ACD extends the concepts of HCD, Cooper, Reiman, and Cronin believe this extension does not reach far enough. They argue that ACD may be useful in understanding *what* the user is attempting to accomplish, but it fails to answer the question of *why* [214]. By understanding the motivation behind a user's action, the designer can create more appropriate, and ultimately more satisfying, systems. By appealing to the emotional impact of the experience (in this case by increasing satisfaction levels), the system gains both credibility and trustworthiness [217]—attributes highly valued in academically focused systems. These satisfaction levels

[3] For further discussion on gestalt psychology, see Chapter 2
[4] See Chapter 1 for further information motivation.

are further enhanced when meeting the needs and motivations of the user, as discussed in Chapter 1. As the system increases the connections between the system and the community, the system and the data, and the system and the interface, the user's self-efficacy and self-actualisation rise in conjunction with satisfaction—a necessary component to both the user's feeling of success and her ability to learn and comprehend new information.[5]

By understanding the goals a user may possess, the designer can understand the user's motivations, as well as her expectations. This understanding contextualises the activities and tasks designed for a given system. Such an understanding would, logically, influence the design of any software system. Within this theoretical framework, determining *what* aspects of a system to make interactive and *how* that interactivity should function is largely accomplished through an understanding of the user's goals and motivations.

From a neuro-biological standpoint, the notion of a goal-based approach is even more strongly reinforced. Current research suggests that dopamine[6] significantly impacts reward-motivated behaviour [219]. In the case of Goal-Directed Design (GDD), this reward-motivated behaviour is that of accomplishing the goals determined by the user, where the *reward* is the successful accomplishment of the goal(s). When the reward is achieved, the brain releases additional dopamine, leading to a sense of satisfaction [220, sec. "Dopamine: 'The Reward Molecule'"]. It is therefore a logical conclusion that a user who is provided with a resource that allows her to accomplish her goals (read: motivation for leveraging the resource) will walk away from her experience with the resource feeling happy and satisfied. This is one of the primary goals of user experience design, from which much of the following work is derived.

Thus, Goal-Directed Design is based on the above-stated premise. Because it leads to ultimately higher levels of satisfaction (which in turn leads to an increase in credibility and trustworthiness), leveraging a GDD approach can translate into possible repeat uses of the software, as well as recommending the software to others [214]. While some may argue that certain systems may be instrumental to a particular field regardless of the *satisfaction level* these systems produce, this does not mean they cannot be improved upon in order to provide a more meaningful experience.

While Cooper, Reiman, and Cronin primarily gear their work towards the private sector, the same principles translate well to the resources generated by Digital Humanities within the academic community (such as the DRE). Successful implementation of a digital resource is more than just a launch of software that is functional at its base level: the resource should also be usable, promote scholarship, and ultimately serve the purpose of knowledge dissemination (which is often a major

[5] See Chapter 1 and the section on Motivation for further discussion.

[6] Tobler, Florillo, and Schultz discuss the effects reward has on neurological responses in the brain, specifically as it is related to dopamine receptors. When the "neural representation" of reward is strongly understood, it can "[reinforce] learning and decision-making" [218, p. 1645].

component of modern day grant applications). Satisfied users who can successfully accomplish their goals are thus a major component of a successful academic resource, as their satisfaction can directly translate into further knowledge dissemination via recommendation and supplementary publication. In a space that, while narrow, still competes for a digital "mind-share",[7] any resource (whether scholarly or otherwise) must find a way to set itself apart and claim its place.

In discussing Goal-Directed Design, Cooper, Reiman, and Cronin lay out three specific criteria to which a software system should adhere: the primary motivation and needs of the user must be understood, especially within the context of how the user will work; the domain within which the software will function and its requisite constraints and requirements must also be understood; and finally, the knowledge gained from an understanding of the first two items should be used to create systems which are "useful, usable, and desirable" [214, p. 4] while also adhering to the limitations imposed by economic viability and technical feasibility [214]. Each of these three principles is immediately applicable to the resources created within DH.

When designing for these environments, the needs and motivations of the audience (typically researchers in a particular field) must be understood in order to effectively present information. Often, numerous types of users (also known as *personas*)[8] may exist and each of these types must be considered, as they may have different goals. For example, consider the *Letters of 1916*, wherein one of the types of users would be that of the general public. The common goals of this user may be to contribute a letter from a relative or perhaps search for letters from a particular relative who was living through the events of the time. In order to accomplish the latter goal, this user needs to be able to run a search by name and find any letters which reference this name. However, this same user may also be interested in understanding the different connections made via letters by a particular ancestor. Leveraging a visualisation to build a network graph would provide a clearer picture of the various individuals to which the individual wrote than a close reading of every single letter. Understanding this goal allows for the construction of a useful visualisation.

Another type of user of the *Letters of 1916* is the historian. This user's needs and motivations likely differ drastically from those of general public. While the general public is likely more interested in a casual experience that facilitates a simple search, the historian is likely more interested in identifying previously unknown themes

[7] Ruecker, Radzikowska, and Sinclair discuss the role of interface design in digital scholarship, pointing to the importance design plays in marketing and attraction. As most content on the internet is designed to be attractive and memorable, academic resources must also compete to stand-out in the user's mind and carve out their own piece of digital landscape within the virtual "mind-share" [221, p. 172]

[8] Proposed by Cooper, Reiman, and Cronin as a method of ascertaining user needs and requirements, personas have been mentioned by numerous other critics within the field as playing an important role in the design process [189], [222]. Personas seek to define an archetype of users placed within the context of how the user will interact with the system in question. They serve to provide context for the designer, allowing them to understand not only the users' needs and motivations, but also their limitations.

or drawing connections. She may require the ability to see information tagged according to a broader theme or view data visualisations of topic models or a network analysis. By providing additional functionality that meets the needs of the historian (which may not be readily utilised or even understood by the general public user), the historian is able to more quickly accomplish her goals and gain the information that she seeks.

Similarly, the requirements and constraints of the field (presentation of artefacts, for example) must be taken into consideration. When dealing with collections of literature, an original scan of the source text may be deemed necessary, depending upon the field. Users of numismatic collections find it necessary to be able to view detailed images of the individual coins in order to identify various markings, etc. A failure to understand the general requirements of the field can lead to a failure in providing for the users' needs and motivations. Furthermore, in the case of scholarly resources, such failings can ultimately degrade the perceived *scholarliness* of the resource. This ties directly to data visualisations as well, as they are often contextualised within the resource as a whole. Data visualisations that are not given sufficient context either through descriptors or through relevant base data[9] available on the object itself will be seen as less trustworthy (or in some situations, less useful) than those with sufficient context or the ability to drill-through to the underlying details.

All of this, however, does provide a valuable approach for designing Digital Research Environments for Digital Humanities. All of these areas—consideration of user needs, motivations, and requirements in addition to the constraints of the field—are strongly present in these types of resources and important factors in determining not only the types of tasks and activities to construct but also the mode(s) of interaction. Motivation, in particular, is tightly tied to an individual's emotional state. While some might argue that emotion should play no role in the consideration of motivation, goals, and usability as a whole, it is impossible to separate the emotional experience from practical application of a software system. As Valacich, Parboteeah, and Wells note, users seek out three primary factors in their evaluation of a software system: "structural firmness" (the stability of the website and its performance, which correlates to the base needs under Maslow's paradigm[10]), "functional convenience" (or the user's perception of how easy the system is to use), and "representational delight" (described as the emotional impact or *wow factor* evoked by the software) [223, p. 86]. There exists an *ordo essendi* in these three items (as listed here) whereby each builds upon the former. While much research exists in the literature around the first two items, it is the emotional component that also needs to be explored within the context of motivation and Goal-Directed Design. Afterall, humans are emotional creatures; thus a discussion around the concept of what Norman

[9] Base data refers to the underlying detail-level data that serves as the foundation of any visualisation. Providing the user with the ability to *drill-down* to this base-level data is an essential aspect of interactivity.

[10] See Chapter 1 for further discussion on Maslow's theory of motivation.

coined as *Emotional Design* seeks to further inform the ideas behind GDD. It is the "rich, entertaining experience" [223, p. 88] which provides the emotional connection necessary for an immersive environment.

3.2 Emotional Design: Extending a System From Usable to Necessary

Goal-Directed Design offers valuable insight into how the user interacts with a given system and can provide a strong framework for developing usable systems. However, there is more to a software system than its perceived usability, and in fact, user perception of usability is influenced by both objective measures (such as goal accomplishment) and subjective ones. It is this latter, subjective experience that is often overlooked when discussing Interaction Design—a concept that Norman sought to rectify with his theory of Emotional Design. While a successful interface must be usable in order to be successful, it must also contain an emotional appeal. As Shih and Liu note, "Users are no longer satisfied with efficiency and effectiveness; they are also looking for emotional satisfaction" [224, p. 203].

Shih and Liu's appeal to an emotional connection to system design is supported by the literature in psychology, which discusses the importance of emotion and logic and how tightly bound these are to the reasoning process, as discussed by Schacter, Gilbert, Wegner, et al. [7]. To illustrate this connection, Schacter, Gilbert, Wegner, et al. note the complications encountered by those with Capgras syndrome.[11] Patients who suffer from Capgras syndrome have a neurological disconnect between the temporal lobe (which assists with recall of information, specifically concerning familiar objects) and the limbic system (which is responsible for general emotional reactions). This disconnect inhibits the individual from connecting the logical recognition of a person or place (such as recognising the face of a spouse) with the emotional reaction that should accompany the recognition (such as a feeling of warmth, love, or a sense of home). Without the connections between these two areas, patients who suffer from Capgras syndrome feel a profound sense of disconnect and often view the people or places that were once familiar as "impostors" (due to the lack of emotional recognition). Even though logic dictates that the persons or places in question look as expected, because the emotional recognition is absent, patients with Capgras assume these items are fake. Studies have shown that those individuals who lack the ability to apply emotion to decisions (typically due to brain injury), often are either unable to make decisions because they are unable to draw an emotional reaction between choices [225] or may be more prone to risk taking activities (such as gambling or investing) because they have no anxiety—an emotional reaction—to the risk they may be undertaking [226].

[11] Capgras syndrome is a disorder in which the individual believes a loved one has been replaced by an imposter. It is often, but not always, the result of neurological damage [7].

In order to rectify this connection between emotion and logic, Norman developed his theory of Emotional Design, elaborating on the importance of the emotional connection to the system in order to facilitate usability. In his book *Emotional Design: Why We Love (or Hate) Everyday Things*, Norman states that there are three important components of design: usability, aesthetics, and practicality [85]; two of these elements—usability and practicality—have already been addressed in this chapter. However, it is this third aspect, that of aesthetics, which is often ignored, much to the detriment of the end user. "But what many people don't realize is that there is also a strong emotional component to how products are designed and put to use...the emotional side of design may be more critical to a [system's] success than its practical elements" [85, p. 5].[12]

The importance of emotion regarding cognition and decision making is readily apparent. Norman cites an anecdote from the 1980s during the rise of colour monitors for computers. At the time, the use of colour served no logical purpose as machines were primarily text based; yet many businesses were spending extra money in an effort to secure colour monitors at great expense. Norman himself, sceptical of the appeal of the colour monitor, acquired one as well to test his hypothesis that it would serve no discernible scientific purpose that could not be served by a standard black and white monitor. While his hypothesis was correct—no discernible, scientific difference between the black & white and colour monitors existed—he found he could not part with his colour monitor either as he had developed an emotional attachment, which influenced his experience with the device as a whole [85]. He goes on to note the neurochemical reactions that occur within the brain, affecting various decision making processes: "One of the ways by which emotions work is through neurochemicals that bathe particular brain centres and modify perception, decision making, and behaviour. These neurochemicals change the parameters of thought" [85, p. 10].

Norman is not the only scholar to note the effects of emotion regarding usability. Japanese researchers Masaaki Kurosu and Kaori Kashimura conducted an experiment in the early 1990s that highlighted similar emotional reactions, as recorded by Norman's above anecdote. In their article, Masaaki and Kashimura created twenty-six versions of an ATM interface (screen and keypad layout) and asked more than two hundred participants to evaluate the interfaces based on usability and aesthetics [227]. Their research showed a strong correlation between the aesthetic nature of the design and a system's perceived usability. They concluded that there is strong evidence to support the idea that the emotional impact of aesthetics plays an important role in the user's response to the inherent usability of a system [227]. These results were later replicated and confirmed by Israeli scientist Noam Tractinsky, noting that cultural differences—originally theorised by Tractinsky as the major factor in Kurosu and Kashimura's work, meaning they would not translate the same in other cultures—were found to have minimal impact on the overall findings [228].

[12] These ideas are borne up by the data collected and discussed in Chapters 5 and 6.

In an attempt to explain the impact of emotions on usability, Norman discusses three important aspects of design: the visceral, the behavioural, and the reflective [85]. The visceral aspect of design deals with the initial impact a software system has upon the user: that of the first impression, "its appearance, touch, and feel" [85, p. 37]. The visceral tends to focus on the appearance of the system. The behavioural concerns itself more with the experience of the system: its function (what the system can do), its performance (how efficiently the system can perform said functions), and its usability (how easily the user can leverage said functions in order to accomplish her goals) [85]. Finally, the reflective concerns itself with the interpretation of the system by the user, the "interpretation, understanding, and reasoning" that takes place in the user's mind in evaluating the system as a whole [85, p. 38]. The reflective level is the level that is most vulnerable to external factors such as culture, education level, class, etc. See Table 3.1 for a simplified break down of these three levels.

TABLE 3.1: Norman's 3 Aspects of Emotional Design

Aspect	**Description**
visceral	appearance
behavioural	pleasure, effectiveness of use
reflective	self-image, personal satisfaction, memories

Each of these aspects meets a very real psychological need: the Need For Cognition (NFC). As Attrill states, there is a "desire to both employ cognitive effort and to enjoy the rewards of that effort" [229, p. 39]. This NFC can strongly influence a number of behaviours in a digital environment—behaviours such as information seeking [230] and attitude formation [231]. Additionally, it also influences *how* the user will leverage a system. Those who tend towards the analytical will often evaluate a software system from a logical (or behavioural) aspect, whereas those users who tend towards the creative will evaluate a product on its aesthetic nature (thus relying upon the visceral) [229]. Thus, these aspects of personality must be considered during the design process, as they can prove to be a strong predictor of the effectiveness of a system for a user. However, it is important to note that the behavioural and the visceral aspects of emotional design concern themselves with the *now*; they have an immediate impact upon the user. In contrast, the reflective is more long term, as it often involves cognitive processes which continually take place over a period time. Therefore, a reflective design seeks to engage a long term relationship with the user and invoke feelings of satisfaction, especially regarding ownership, use, and display [85]. As the reflective can often override that of the visceral and the behavioural, it can have the strongest impact on the emotional experience of the system [85].

While also discussing the importance of the need for cognition, Gangadharbatla notes the role of "internet self-efficacy" [232, p. 6], defined by Daughtery, Eastin, and Gangadharbatla as a user's "confidence in their ability to successfully understand, navigate, and evaluate content online" [233, p. 71]. Gangadharbatla applies the concept of internet self-efficacy to not only its importance to the overall user experience

but also, more importantly, the likelihood of adoption. If a user experiences a low level of self-efficacy, she is unlikely to return to the software system to meet future needs. This concept also ties in closely to Goal-Directed Design, as the easier it is for the user to accomplish her goals within a given ecosystem, the higher her internet self-efficacy will be [232]. This also ties back to the both Maslow's and Rutledge's needs models,[13] where self-efficacy can be understood as *self-esteem*. And as previously stated, self-efficacy plays not only an important role in adoption (as stipulated here) but also in knowledge acquisition itself [87].

Another aspect of emotional design that ties closely to a Goal-Directed Design approach is that of flow theory. As described by Attrill, flow theory:

> represents the idea of achieving a psychologically optimal experience that occurs when people become totally immersed in a current activity, an immersion that sees a person as completely absorbed by what to them is a positive experience characterized by a clear objective, feedback, concentration, a sense of control and a loss of self-consciousness. [229, p. 40]

While flow theory was originally developed by Hungarian psychologist Mihaly Csikszentmihalyi in 1975, it was Hoffman and Novak who adopted his theory to describe user behaviours in online environments. They theorised that users achieved a state of flow when they were able to accomplish specific goals within a software system [234]. When applied to Digital Research Environments, it can also be concluded that achieving flow within a goal-directed state would meet a user's NFC as defined by both Attrill and Gangadharbatla.

But how can this flow be achieved? What steps can be taken by the designer to facilitate this psychological state? The answer to both of these questions is found in the idea of immersion.

3.3 Immersion in the DRE: Understanding the Emotional and Cognitive Benefits

One[14] of the greatest benefits of Goal-Directed Design and Emotional Design is that of immersion. But immersion is a nebulous animal at best when discussed within the context of Interaction Design and even more so within the context of the Digital Research Environment (DRE), especially as it is a term most often associated with narrative media such as games, film, and narrative texts. The idea of immersion proves highly beneficial within the DRE, as it allows for increased cognition and higher levels of satisfaction and self-efficacy—the latter two of which also provide a cognitive advantage.

[13] See Chapter 1 for further discussion.

[14] Portions of this section are reused in an upcoming journal article entitled "Engagement through Visualisation: A Case Study on the Alcalá Record Books" in the *IEEE Transactions on Learning Technology*.

3.3.1 Defining Immersion

In her 1997 book *Hamlet on the Holodeck* (later updated in 2017), Murray gives perhaps the most evocative, metaphoric definition of immersion:

> Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or a swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus. We enjoy the movement out of our familiar world, the feeling of alertness that comes from being in this new place, and the delight that comes from learning to move within it. Immersion can entail a mere flooding of the mind with sensation, the overflow of sensory stimulation. [235, p. 124]

Numerous other scholars build upon Murray's work through an elaboration on what is (by Murray's defintion) a subjective experience. These scholars focus on the sensory experience associated with immersion, whether through an extension of Murray's definition of being completely enveloped by the environment [236], or through a multi-sensory experience of the virtual environment, often through artificial stimuli [237]–[239]. These experiences are what Gerrig describes as *transportative immersion*, or the experience of being *transported* into a crafted world, typically as part of a narrative text or via the technology into a constructed virtual world [240].

Slater, utilising the metaphor of transportative immersion, attempts to define it, then, as that of an objective (rather than subjective) experience created by the hardware:

> Let's reserve the term "immersion" to stand simply for what the technology delivers from an objective point of view. The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is "immersive". This is something that can be objectively assessed, and relates to different issues than how it is perceived by humans. [241, sec. "Immersion & Presence", par. 1]

Here, Slater draws a distinction between the idea of immersion—an objective measurement by his account—and presence—the subjective experience of the user. He makes this distinction via the analogous metaphor of colour, where immersion is the wavelength distribution of light and colour itself is the perception of the wavelength distribution by the individual. Slater's definition, then, can be interpreted as immersion as the *property* of a system, as opposed to an *experience* that exists within the mind of the user (which Slater refers to as *presence*). To his mind, immersion is an objective measure of the quality of the display, the feedback from the peripherals, etc. It is an extension of the hardware, rather than an esoteric emotional experience.

Slater's definition is problematic, however, as it assumes that immersion can be objectively measured. While he believes that immersion is an extension of what the technology can deliver in terms that "preserves fidelity in relation to their equivalent real-world sensory modalities" [241, sec. "Immersion & Presence", par. 1], this assumes that every individual's sensory experience is objective. However, how colour blind individuals experience the world through sight is different from that of non-colour blind individuals. Under Slater's definition, a 4K monitor would provide an objectively more *immersive* experience than a CRT monitor, but when considering the sensory experience of the colour blind individual compared to that of the non-colour blind individual, the level of *immersion* is not necessarily the same. Furthermore, the difference between an LCD or LED screen and a 4K screen may or may not be noticeable to an individual due to the quality of the individual's eyesight. Such examples demonstrate that the experience is subjective regardless of the objective measures we apply and cannot be separated from *objective* measures of technology—at least not without reinforcing hegemonic language around ableism. While the differentiation between immersion and presence can be useful (as they can be qualified through different affective responses, specifically when dealing with virtual and mixed realities), the distinction for the purposes of discussion here is irrelevant, as it dismisses the subjective aspect of the immersive experience, which is crucial to understanding why immersion plays such an important role in the learning process.

Returning then to Gerrig's work on immersion, Gerrig also highlights a second type of metaphor for immersion—that of performative immersion, or immersion that takes place due to the actions of the reader [240]. Other scholars have built off this other metaphorical definition of immersion by defining it as a response to systemic challenges presented to the user [237], [239], [242]–[244]. While each defines immersion through a slightly different lens, they all revolve around the idea of immersion as a response to absorption in a task or a desire to complete a challenge or task set before the user. This builds upon Gerrig's idea of performative immersion—where he focuses on the reader's psychological performance in constructing the world as part of the active reading process—by looking at the unique challenges presented to the reader depending upon the medium:

- System immersion, or immersion as a result of cognitive challenges that match (but do not exceed) a user's capabilities, which results in a type of hyperattentional focus on overcoming the challenge to achieve a sense of satisfaction [237]

- Engagement, or immersion (specifically within games) that is derived from focused attention on completing a challenge or *winning*[15] [238]

[15]It is interesting to note that, of the various types of immersion discussed here, only McMahan refers to immersion as engagement, thus creating a bit of conflation between the two terms, which will be clarified below.

- Strategic immersion, or immersion that results from focused attention on planning or devising strategies to overcome obstacles or challenges [242]
- Challenged-centric immersion, or immersion that results from focused attention on meeting challenges (specifically challenges that demand advanced mental acuity or fine motor control)[16] [239]
- Playful immersion (also termed "ludic immersion" by Nilsson, Nordahl, and Serafin[245, p. 111]), or immersion as a result of intense, focused attention on a single task or activity in order to achieve completion [244]

It is interesting to note that the literature surrounding the idea of immersion as a response to a challenge is less about the subjective, emotional *experience* and more about the absorption in a task or series of tasks—a state that could be redefined as attention; in fact, each of the above-noted works specifically discusses the idea of focused attention as part of the immersion process. Viewing immersion as a response to a challenge, then, involves the recognition that, at least in part, immersion is about the ability of a system to focus the attention of the user effectively and efficiently.

As demonstrated, defining immersion is no simple task, especially when considering the ideas of presence and engagement. Scholars often conflate the terms or make distinctions between one and not the other. Further complicating this idea is that notions of immersion, presence, and engagement are typically only discussed within the context of interactive, narrative media—such as video games, virtual reality, or mixed reality—or with regard to overall narrative—such as arguments found in literature, electronic or otherwise.

Such a disparate discourse around the subject of immersion necessitates clear definitions of the context of the term for the purposes of this work. Here, immersion will largely concern itself with the subjective experience of the user, relying upon Murray's metaphoric idea of submersion in water. However, as complete immersion would be impossible under this paradigm due to the nature of the system and the work involved, the definition of immersion will constrain itself to that of immersion within a set of tasks and activities contextualised by an overarching goal, thus incorporating the ideas of immersion as an experience of attention. Therefore, immersion is seen as the subjective experience of the user within a software ecosystem wherein she becomes absorbed in a set of tasks and activities that ultimately lead to the satisfactory completion of a goal. Under this paradigm, then, presence is a part of immersion, as it is part of the subjective experience. Thus, whenever the term immersion is used, presence should be considered within the same context. Engagement, however, becomes a function of immersion under this paradigm.

[16] Ermi and Mäyrä's definition differs somewhat from the others as it focuses on the attention required on the *process* of overcoming the challenge, rather than the attention focused on the challenge itself.

Where immersion describes the subjective experience of the user, engagement describes the mechanism through which the user experiences this immersion—most often accomplished through the design of the user interface.

3.3.2 Conflating Attention and Immersion

Immersion, then, has two components: an emotional component (or the subjective experience), and an attention-based component, where the former is a response to the success of the latter. Discussing the relationship between immersion and attention, then, becomes a necessary component in understanding how the emotional (or *affective*) response plays a role with regard to cognition.

As discussed in Chapter 1, attention is the system through which the individual is able to shift cognitive processing resources to focus on a particular task or tasks. While attention is born partially from merely cognitive, biological processes within the brain,[17] it is also an affective response to challenge. Humans are, by their nature, driven by the need to overcome challenges and obtain new knowledge [229], [246]. There is a very real Need For Cognition (NFC) within the human psyche and when the perception exists that a goal or system will meet such a need, our underlying attentive processes (such as those discussed in working memory) become part of the response to rise up and overcome the challenge that will fulfil this need.

When discussing the role of immersion within the context of knowledge acquisition or cognition, it then becomes about the idea of attention (specifically focused attention) and the mechanisms provided by a system through which the user's attention becomes focused. Treisman (basing her working off an earlier paper she published in 1980) discussed the role attention plays in the process of binding—the ability of the brain to integrate various visual aspects into a coherent whole—and how attention influences the ability of the brain to conduct visual search. "...[A]ttention is needed to bind features together, and that without attention, the only information recorded is the presence of separate parts and properties" [23, p. 1304]. Under Treisman's model, the brain is not able to properly bind various objects together into a cohesive image from which the mind can process information unless attention is focused enough to bind disparate objects together. This *feature integration theory* (originally proposed by Treisman and Gelade in 1980 [180]) can produce *illusionary conjunctions*, or combinations of features in a seemingly random pattern that make no cognitive sense to the user [9].

While much of the work discussed here is focused on notions of visual search and integration with working memory, the theory can be extrapolated to apply to notions of immersion. As previously demonstrated, scholars who argue for the idea of immersion as a response to challenge all discuss ideas of focused attention or mental absorption as playing a key role in the experience. This focused attention results in a narrowing of the visual field to a particular aspect of the system, thus removing

[17] See Chapter 1 for further discussion on working memory and the role of attention.

distractors and allowing for an increased capacity for feature binding. This increase, then, allows for better integration of information, which ultimately leads to higher levels of cognition and knowledge acquisition within the context of the DRE. By focusing the user's attention on a task or set of tasks that assist her with completing a goal, she is more immersed in the task to be completed, thus leading to higher levels of binding and, ultimately, higher levels of cognition.

Another benefit to focused attention brought on by immersion is seen through the analysis of the theory of threaded cognition. Threaded cognition is a theory put forth by Salvucci and Taatgen that seeks to explain how multitasking functions within the brain. Under this theory, cognition is treated much like that of a multi-threaded computer system where resources become locked by a particular *thread* (or attentionally focused goal in the context of this argument) and when the resources are no longer needed, they are immediately released [247]. This theory is primarily used to describe the function of multitasking—a common phenomenon in the digital age, as users are continually encouraged to engage with multiple tasks, ultimately leading to a more shallow experience of each [248]. Working memory further supports this theory when considering the strain multi-tasking places on the central executive. Not only must the central executive manage multiple attentional threads, it must also work to constantly integrate disparate sources of information and commit everything into long-term memory. Eventually, working memory resources can become deadlocked, and cognitive overload occurs.

Focused attention through immersion, however, reduces the amount of multitasking taking place (i.e. reducing the number of *threads* in play at any one time) thus reducing (or all together removing) the likelihood of deadlocking resources that are needed to integrate information and thus increase cognition. Newport builds on this idea in his book *Digital Minimalism* where he discusses the importance of removing digital distraction [249]. While his work is more of a life philosophy, the parallels regarding focused attention can certainly be easily extrapolated. By removing distraction and focusing in on a task (or set of tasks), the user becomes hyper-focused on the work at hand. This focused attention becomes an immersive experience that ultimately leads to a more cognitively fulfilling result [249].

3.3.3 The Affective Response

While the above has focused on the idea of immersion as a type of focused attention, it is more than that; it is also the affective response[18] that plays a considerable role in understanding its cognitive benefits. This is the *other side of the coin* with regard to the earlier definition laid out for immersion.

Emotions are complex, and from a psychological perspective, defined in multiple terms. *Emotions* are generally described as:

[18] Defined as "the emotional response to a situation" [250, par. 1]

> specific and consistent responses to a significant internal or external event that are brief in duration and consist of a particular pattern of physiological responses...[they] are object oriented, in that they are generated by discrete events and have typically been studied in experiments designed to elicit a particular emotional response. [7, p. 393]

Emotions are often public events; meaning emotional response is often displayed outwardly. This definition of emotion is then contrasted against the definitions of *feelings*—which are "private conscious thoughts" that result from emotions [7, p. 393]—and *moods*—which are more generalised states, often where the emotion has been diffused, but lasts over a longer period of time than the initial emotional response [7]. Both feelings and moods tend to be more private, internalised reactions, which may or may not be object oriented but are often tied to an emotional event. For the purposes of the discussion here, the term emotion will encompass all three separate definitions provided above. It will be used interchangeably with the term *affect* or *affective response* to indicate the emotional state or response of the user to immersion and the effects of this response on cognition.

The primary affective benefit of immersion within the context of the DRE is that of satisfaction. Studies have shown that users who experience deeper levels of immersion, specifically with regard to activity completion and goal setting, also experience higher levels of overall satisfaction: both with regard to expectation (as in a satisfactory feeling of goal completion) and a feeling of satisfaction with the software itself [251]. This satisfaction is situated within the context of authenticity—the experience of the user where there is a clear parallel between the activities being performed and the perceived value of the outcome of such activities [252]. The experience of authenticity then leads to increased cognition and motivation [253].[19] When considered within the context of Goal-Directed Design, the notion of authenticity strongly shapes the design decisions made around task construction, asking the questions:

1. How does the task translate into knowledge that has an application to the stated goal?

2. How does the environment assist with the translation of the analogue experience to the digital, specifically within the enculturated context of the user?

While the former question is one often considered within any design leveraging the GDD methodology, it is the latter question that is of greater importance when considered within the DRE. Each user group (or persona[20] as referred to in Cooper, Reiman, and Cronin's work) has certain expectations of a software ecosystem based

[19] Here, motivation is used to refer to the user's desire to continue to engage as opposed to what drives the initial engagement as discussed earlier.

[20] Personas are defined as "detailed, composite user archetypes that represent distinct groupings of behaviors, attitudes, aptitudes, goals, and motivations observed and identified during the Research phase [of the design process]" [214, p. 21].

on their own cultural background. Within the DRE, this *cultural background* generally refers to the ideas, expectations, and training received within a user's particular field of study. For example, historians will have a set of expectations and approaches to leveraging the various tools and tasks associated with a DRE different from those of the paleographer, who will in turn have expectations and needs different from those of a statistician. In order to ensure an *authentic* experience (as defined previously), the design of the software must consider these enculturated differences among user groups, as they have a considerable impact on the level of immersion (and by extension, authenticity) experienced by the user [254].

By ensuring authenticity, the system provides three distinctive benefits: cognitive, metacognitive, and motivational [251]–[253]. The cognitive benefits can be explained in part by the work of Cahill and McGaugh who demonstrated a causal link between "emotional arousal" [255, p. 295] (in this context, a sense of satisfaction) and long-term memory, as heightened emotions stimulate both hormones and various regions of the brain that control and regulate access to long-term memory [255] in addition to helping sustain "active and continuous learning" [256], [257, p. 465].[21] The impact of the emotional experience is one that has been thoroughly documented and researched [255], [258]–[263] and thus must be considered within the context of GDD, as GDD itself heightens the emotional response to a system through (ideally) successful completion of goals.

In addition to the positive contributions to cognitive systems, emotional satisfaction also plays an important role in metacognition (those processes which influence the cognitive process as a whole), specifically around the ideas of self-regulation and self-efficacy. As noted in Chapter 2, one of the problems with employing a close reading approach within the DRE is self-regulation, due largely to metacognitive bias.[22] By incorporating a goal-directed design approach, there is a challenge set before the reader, thereby engaging the metacognitive processes and removing the biased perception that leads to shallow engagement. This allows for a deeper sense of immersion, thus leading the user to more effectively self-regulate.

In addition, the user's sense of self-efficacy is also improved. As positive emotional valence within a software system occurs, levels of trust begin to rise. This is most readily seen in designs that are properly situated within a particular cultural context (otherwise known as *localisation*) [264]. In the context of the DRE, this cultural context is focused upon the culture of the user group or persona (as previously discussed). The user, already familiar with various cultural cues or practices, gains trust in a system that signposts or mimics these cues or practices. In addition, trust also increases when a website provides ease of task accomplishment or goal completion [265]–[267]. The greater the ease with which an individual can accomplish a task, the greater the trust the user places in the system.

[21] For further discussion on the role of long-term memory in the learning process, see Chapter 1.

[22] Chapter 2 discusses the role of metacognitive bias in digital environments under the paradigm of the *hard-easy effect*.

Naturally, as trust rises, the user also becomes increasingly more confident in her ability to perform tasks within the system, as she trusts the system to guide her and provide her with the tools necessary to accomplish tasks or goals set before her. This confidence in ability, referred to here as *self-efficacy*,[23] has not only a lasting effect on the motivation to complete a task (or to continue leveraging a particular system) [269], [270], but also a strong impact on the user's cognition [87].

Hamilton, Kaltcheva, and Rohm reinforces much of the above theory regarding the importance of immersion and satisfaction by examining these concepts within the context of social media branding and marketing of products. While not a direct correlation to software built for research environments, certain parallels exist. In their research, Hamilton, Kaltcheva, and Rohm evaluate three specific factors:

- Customer Lifetime Value (CLV) - defined as the total amount of financial contribution (such as purchases) made by the consumer over their "lifetime association" with a product [271, p. 125]. This includes repeat purchases.
- Customer Influencer Value (CIV) - defined as the total influence a customer has on other customers or potential customers, i.e. the customer's willingness to recommend the product or brand [271].
- Customer Knowledge Value (CKV) - defined as the propensity of the customer to provide feedback to the brand related to "innovations and improvements" [271, p. 125].

Hamilton, Kaltcheva, and Rohm's work evaluated these metrics under three different strategies: immersion only, satisfaction only, and a hybrid strategy of immersion-satisfaction that combined the previous two. They discovered that the hybrid strategy of immersion-satisfaction had the greatest impact across all three metrics (whereas the other two only impacted 2 out of 3 metrics, immersion-satisfaction impacted all three). Thus one can logically conclude from this study that focusing on both the immersive and affective responses of the social media interactions will influence the overall success of the brand.

While perhaps not immediately obvious, this theory can be applied to the construction of the DRE. CLV can be reinterpreted as the likelihood of the user to return to the system as a trusted source of knowledge. CIV, then, evolves from an idea of "product promotion" to one of knowledge dissemination—that is the willingness of the user to leverage the system for original research and to cite the software as part of that research process. CKV has perhaps the least applicable correlation, as the DRE as it is currently defined tends to be a rather static object: once developed, it rarely

[23] Defined as "confidence in one's ability to succeed at a task and [a] primary influence in motivating a person. This construct determines various aspects of an individual's behaviour toward a task, including the individual's thoughts, motivations, and overall performance, especially when the individual faces a difficult task." [268, par. 1].

undergoes an iterative process of improvement that would rely upon the feedback loop inherent in the definition of CKV.[24]

3.4 Discussion

When considering the application of the aforementioned principles, there are many avenues the DRE can take in order to adequately fulfil notions of immersion, emotional satisfaction, and goal-focused design. The most common interface techniques are the implementation of search and browse mechanics. These are often implemented through a faceted search environment (see Figure 3.1) or through a google-like search metaphor (see Figure 3.2), which may or may not incorporate advanced search features (see Figure 3.3). While these interface techniques often serve to fulfil the goals of most users on a base level, they fail to achieve both a strong visceral appeal as well as the truly immersive experience that leads to a cognitively fulfilling experience.

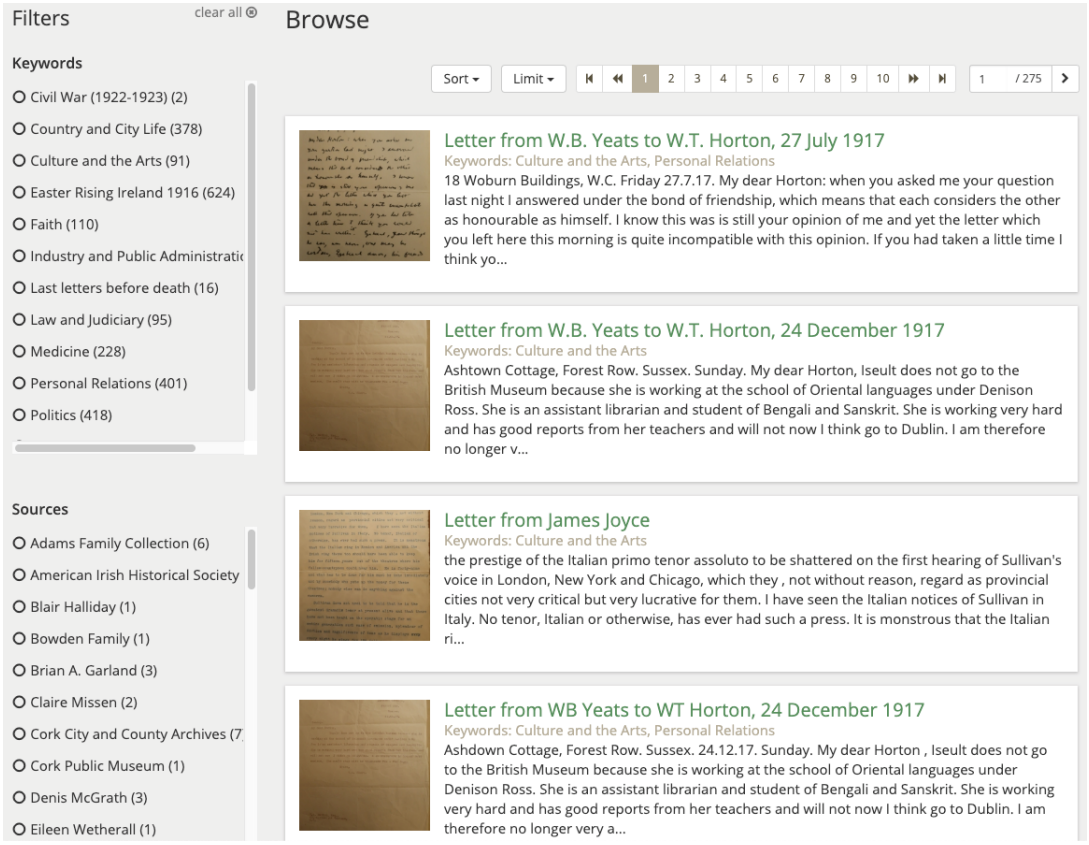


FIGURE 3.1: Implementation of browse through faceted search in *Letters of 1916* [192]

[24]It is worth noting, however, that some DREs position themselves as *public humanities* projects that seek to engage the public, often to assist with more menial tasks such as translation, transcription, or categorisation. In such cases, a stronger feedback loop does tend to take place between the software creators and the larger community, and occasionally the funding exists to make improvements based on CKV. In such circumstances, these systems gain additional advantage from a focus on immersion and satisfaction.

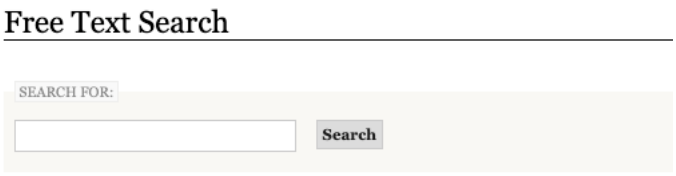


FIGURE 3.2: Implementation of Google-like search in *Jane Austen's Fiction Manuscripts* [272]

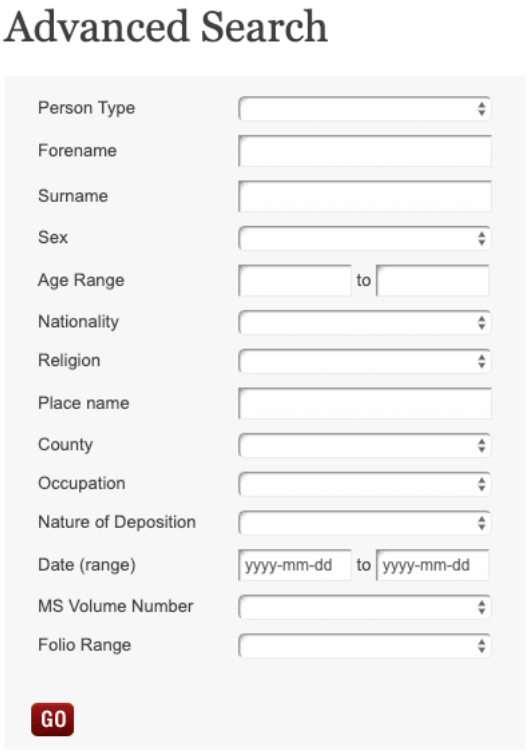


FIGURE 3.3: Implementation of advanced search in *1641 Depositions* [273]

In order to create a cognitively immersive environment, an element of interactivity is required [274]. While this interactivity can differ depending upon the system in question, within the context of the DRE, interactivity essentially involves creating an environment that allows for a conversation between the user and the data, as seen in the definition of interactivity put forth by Rafaeli: "[interactivity is] an expression of the extent that in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions" [275, p. 111]. In the context of the DRE, this can be applied to the information received from one object that can be applied to the next (or refer back to a previous object). Humanities research is about drawing connections, so this type of information contingency must inform the interactions created within the software ecosystem. In fact, this use of the metaphor

of interaction between objects is well-reflected in the literature on education pedagogy, especially within the framework of Laurillard's conversational framework as discussed in Chapter 1. In her work, Laurillard discusses the importance of the establishment of a conversation between the teacher's constructed environment and the student's interpreted environment, mediated through the lens of educational technology [54]. When these theories are then applied to the DRE, it can be extrapolated that a conversation must occur between the user and object (or source data) via an intermediary. While search and browse can offer some extension of this interactivity, the levels of immersion presented in such metaphors are insufficient. Thus it is theorised here that data visualisations fill a niche within the DRE that can bring about a truly successful, cognitively fulfilling experience.

However, the data visualisations themselves, in order to be truly immersive experiences, must be more than static images, for the type of immersive experience laid out in this chapter cannot be achieved without an interactive element. Take, for example, the static visualisations leveraged in the *Letters of 1916* (see Figures 3.4 and 3.5).

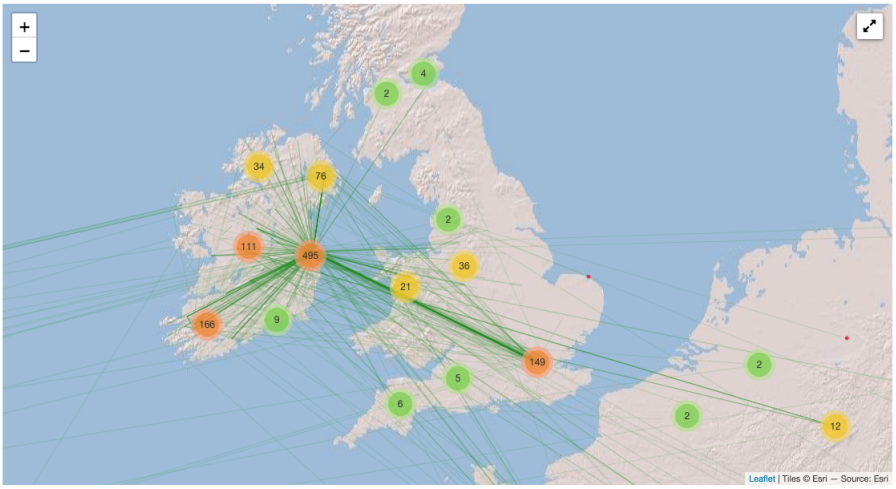


FIGURE 3.4: Static visualisation of letters by location in the *Letters of 1916* [192]

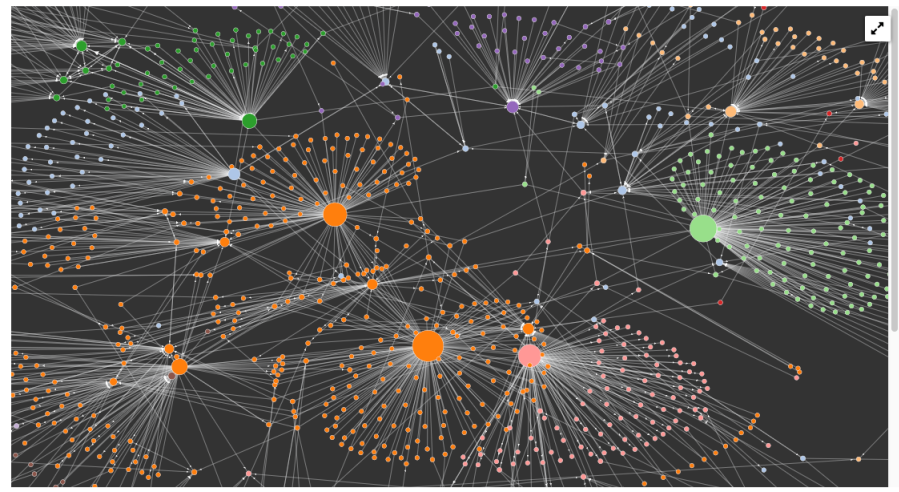


FIGURE 3.5: Static visualisation of networks of people in the *Letters of 1916* [192]

Both of these visualisations highlight interesting information. The map seen in Figure 3.4 allows the researcher to view where the concentration of letters originated and where the most frequent correspondence was taking place, thus viewing connections between places. The visualisation in Figure 3.5 similarly allows the researcher to view connections between people by highlighting major contributors and the various nodes (i.e. people) they connect to in their writing. The researcher can then follow these connections to other nodes to establish a type of *6 degrees of separation* between authors.

The use of these types of static visualisations is quite common. *The Diplomatic Correspondence of Thomas Bodley (DCTB)* leverages a number of static visualisations in order to highlight places mentioned (see Figure 3.6) or connections between correspondents (see Figure 3.7).[25] As in the above mentioned examples, these visualisations can allow the researcher to explore the data from a different view point, allowing patterns to emerge; however the utility becomes somewhat limited.

[25]Numerous other examples of these types of static visualisations exist and the websites presented here should by no means be considered an exhaustive list. Further examples include *Mapping the Republic of Letters* [276] (which has a mixture of static and interactive visualisations), *1641 Depositions* [273], and the *Earlier Latin Manuscripts (ELM)* project [277].

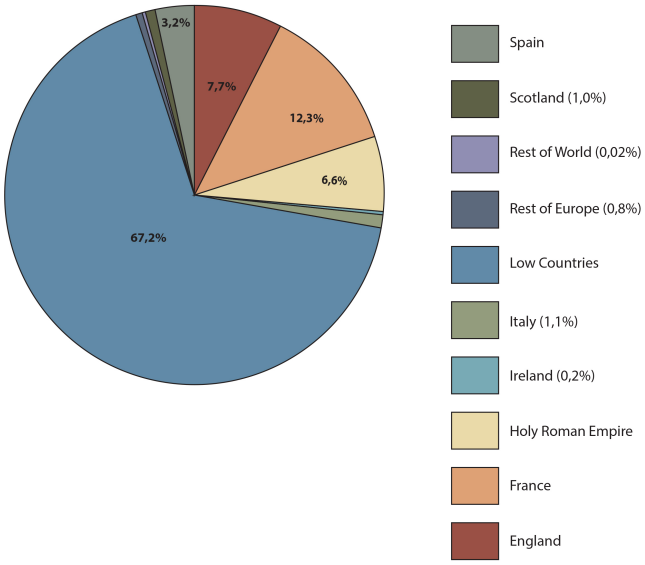


FIGURE 3.6: Pie graph of places mentioned in *The Diplomatic Correspondence of Thomas Bodley* [199]

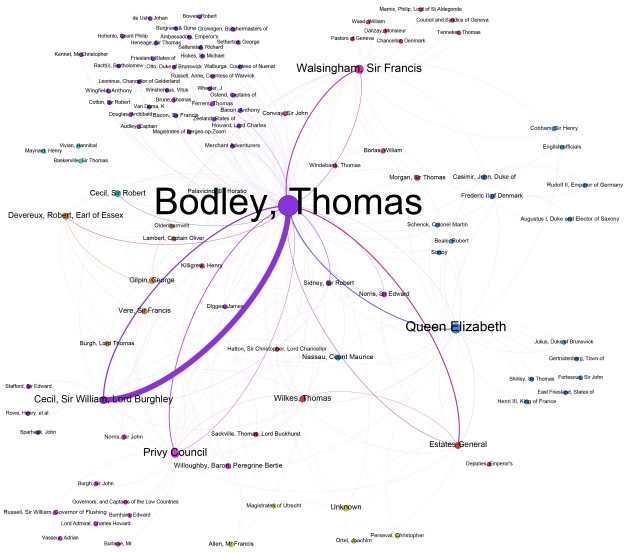


FIGURE 3.7: Network graph of people mentioned in *The Diplomatic Correspondence of Thomas Bodley* [199]. A larger version of this image can be seen in Appendix E in Figure E.8

None of these visualisations can offer a true sense of immersion, however, as they are missing the critical element of interactivity. While some basic interactivity exists in that the user can zoom in on the map or click on a node in the network diagram to view a person's name,[26] neither of these visualisations interact with the underlying data model to move the researcher into engagement with the source material and the creation of the boundary object(s). If the researcher notices an interesting pattern

[26]While the *Letters* visualisations do have this basic interactivity, the visualisations shown from *The Diplomatic Correspondence of Thomas Bodley* collection are entirely static and offer no form of interactivity beyond viewing a larger version of the image.

or discovers an avenue of exploration, she must navigate away from the visualisation and engage with other facilities (such as the aforementioned search or browse techniques) in order to locate the source data that leads to the pattern. Depending upon the nature of the data and the interfaces themselves, this data may become difficult to locate or identify as part of the underlying pattern, thus leading to frustration rather than a sense of satisfaction. This lack of interactivity within the data visualisation thus breaks the immersion within the system.

By contrast, consider the example of the *Hestia* project [278]. Here both a map view and frequency graph are leveraged to allow the reader to visualise the locations mentioned in the *Histories of Herodotus* (see Figure 3.8).

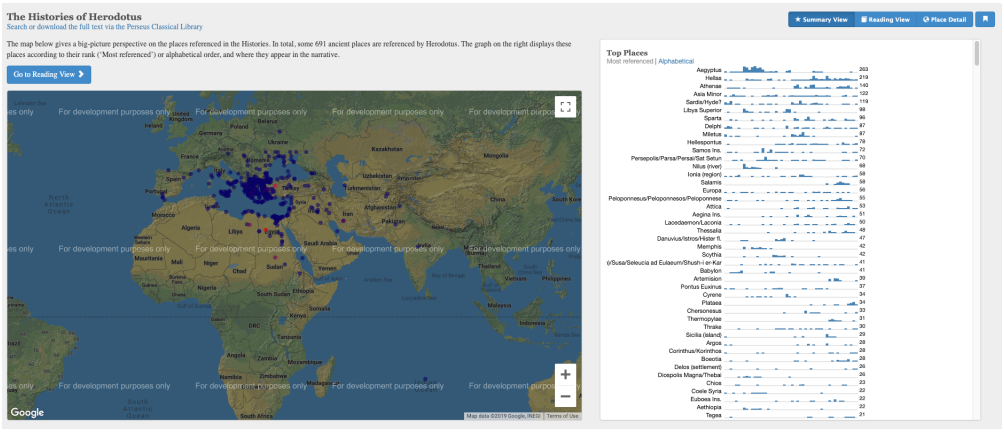


FIGURE 3.8: Interactive visualisation of places in the *Hestia* project [278]. A larger version of this image can be seen in Appendix E in Figure E.9

Like the map visualisations in *Letters* or *Mapping the Republic* or some of the various graphs presented in *DCTB* or *ELM*, these visualisations provide that higher level view of the data that can aid in pattern recognition and cognition, as discussed in Chapter 2. However, the visualisations in *Hestia* extend this functionality through the application of interactivity between the visualisation and the data source (see 3.9). This interactivity thus furthers the engagement of the user with both the visualisation and the source material by allowing her to seamlessly move to the underlying source. This type of engagement not only leads to a deeper sense of immersion but also raises the trustworthiness of the visualisation itself, as it allows the user to understand the underlying source from which it is derived. The user is thus able to verify the authenticity of the system, leading to a replicable state.

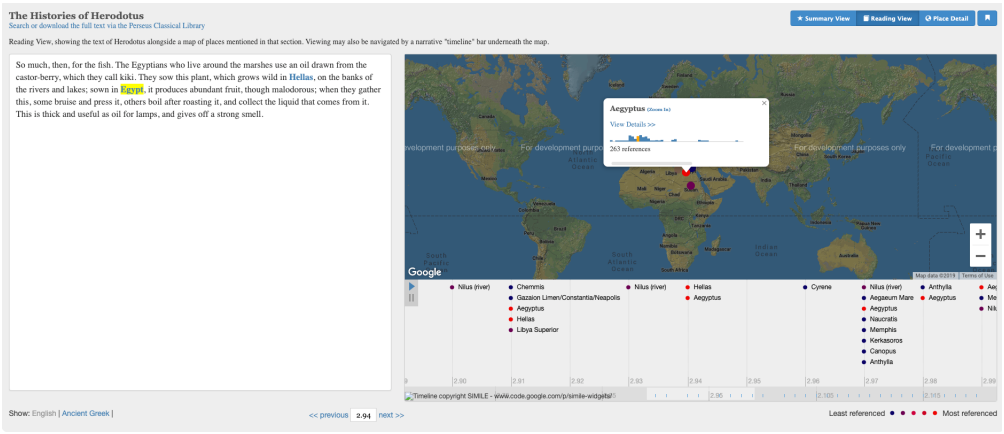


FIGURE 3.9: Viewing the underlying data associated with a visualisation in the *Hestia* project [278]. A larger version of this image can be seen in Appendix E in Figure E.10

3.5 Conclusion

Immersion and interactivity play a significant role in creating a truly satisfying and cognitively fulfilling experience. As the focus of the Digital Research Environment must be the ultimate goal of knowledge acquisition through research, tools must be provided that facilitate both immersion and interaction. However, in order to create these immersive, interactive environments, the goals and motivations of the end-user must be considered in order to craft an experience that can successfully meet their individual needs. By understanding how motivation drives interaction, as well as how a goal-directed approach to design can facilitate immersion and interactivity, the creators of the DRE can craft an experience that moves the system beyond the concept of *usable* and into *valuable*. This subjective perception of the system is highly influenced by the emotional reaction, which itself is guided by the aforementioned principles of motivation and goal fulfilment.

While there are many avenues the DRE can explore in creating an immersive experience, it is perhaps data visualisations, which employ interactivity and facilitate a type of communication between various data objects and the end user, that can seek to offer the most immersive—and by extension most rewarding—experience. The work laid out in the previous three chapters has laid the theoretical groundwork for why data visualisations play such a vital role in the DRE: from learning (in Chapter 1) to reading (see Chapter 2) to immersion (as discussed in this chapter). The remainder of this work, then, seeks to build upon the theoretical groundwork laid thus far through the lens of a practical application of the oeuvre. In the following chapter, the methodology behind the design and construction of the *Alcalá Record Books* will be presented, and later chapters will discuss the case study used to provide the data necessary to answer the questions at this core of this thesis.

Chapter 4

The Alcalá Record Books: Methodology & Design

The previous chapters have served to lay a theoretical foundation upon which the remainder of this thesis is built. As stated in Chapter 2, this work hypothesises that the application of data visualisations as an interactive search mechanic will provide an answer to the underlying research question: "What effect do data visualisations, specifically within the remit of Digital Humanities, have on engagement with the source material—and by extension learning—in digital research environments?" In order to test this hypothesis, a Digital Research Environment (DRE) was constructed using the account book manuscripts from the Royal Irish College of Alcalá de Henares—a set of 18th century manuscripts currently archived at the Russell Library at Maynooth University.

The following pages serve to describe the methodology behind the practical application of the theoretical work conducted for this thesis. It will open with a brief discussion of the *Alcalá Record Books*, providing an overview of the history of the manuscripts and why they are important. Following on will be an analysis of the design of the software application, from the initial data collection and discovery that informed some of the design choices, to the architecture of the UI and the system as a whole. The chapter will then close with a brief summary of the overall goals of the software, laying the groundwork for the case study discussed in Chapter 5.

4.1 Introducing the Alcalá Record Books: Background and History

The *Alcalá Record Books* are a collection of manuscripts dating back to the late 18th century (specifically 1774-1781 for the manuscripts used as part of this project). The manuscripts are a detailed list of accounts for the Royal Irish College of Alcalá de Heneres, established in 1649 as part of a larger network of Irish colleges abroad. These colleges were established in part as a response to the penal laws passed in

the 16th and 17th centuries that penalised the practice of Catholicism [279].[1] These colleges were established all over Continental Europe with a significant portion concentrated in Spain. The Royal Irish College of Alcalá de Henares operated from 1649 until it was merged with the Royal Irish College of Salamanca in 1785 [281]. The College at Salamanca continued to operate until its closure in the mid-twentieth century, whereupon the archives were transferred to St. Patrick's College, Maynooth in 1951 [282]. The manuscripts presented as part of this project hail from the college's account books—entitled *Libros de gastos del colegio de Alcalá*—from 1774 to 1781 (with the exclusion of the year 1780, which is missing from the records). The information contained within details finances of the daily running of the College and includes information on expenditures related to food, *viaticums*,[2] maintenance, and other administrative matters. In total, there exist 324 pages, detailing more than 1,000 financial transactions over the course of the 7 year span.

These manuscripts were originally digitised in 2007 with the aim of creating a Digital Edition of the texts [282]. Unfortunately, this original edition is no longer online, due in part to the underlying technology used to construct the edition—Adobe Flash—being deprecated and no longer supported by most modern-day browsers. An archive of the original software and the data—as created by Keating, Teehan, Gallagher, et al.—remained on a server held by The Arts & Humanities Institute at Maynooth University (formerly known as An Foras Feasa). As such, all of the images and their XML-based transcriptions and translations were still available.

In addition to their accessibility as described above, the manuscripts from the Salamanca archive that form the *Alcalá Record Books* offers some unique opportunities regarding the application of the theory discussed previously. There are unique challenges in understanding given the multi-cutural bridge that exists at its core. While the rector and most of the staff are Irish, they were residing in a foreign country and conducting their daily lives in a foreign tongue.[3] The archive provides historians of the period with a unique view into concepts of globalisation, allowing them to investigate not only the assimilation of the Irish into Spanish culture in the 18th century but also how the Irish maintained their own customs and cultural touch points while residing abroad. Such information also translates well to the classroom and discussions around historical concepts of globalisation. The *Alcalá Record Books* allows lecturers to demonstrate to students in a tangible way that globalisation is not a *new phenomenon* and has, in fact, been taking place for centuries (albeit at a much slower pace than might be observed today).

[1] It is worth noting that the penal laws are not the sole factor in the establishment of the continental colleges. Additional factors—such as migration, the growing importance of a broader education, and the need for a more educated populace to assist the mercantile networks of port towns in Ireland—also played an important historical role in the formation, but such a discussion is outside the scope of this work. For further information, see [280].

[2] *Viaticums* were travel stipends issued to students upon completion of their studies in order to return to Ireland [279]

[3] In fact, the manuscript itself is in Spanish and was translated into English as part of the 2007 digitisation.

The text also presents unique opportunities for digital research. As it is a financial record, it allows for different types of data visualisations to be generated. The data itself is a time series of financial transactions. Unlike most DREs which may be collections of letters, papers, or books, the *Alcalá Record Books* contains data that can be visualised in ways not constrained to topic models or network graphs: it provides an opportunity to view the flow of money over time, which itself can tell its own story.

It is this story through spending that is perhaps the most engaging. A traditional close reading of the work would present many of the cognitive challenges discussed in Chapter 2 (many of which are inherent to all digital texts). However, understanding the trends in spending or discovering discrepancies in the data would be more difficult if leveraging close reading techniques. The data visualisations allow the various narratives inherent in the work to emerge. Trends are more easily identified and investigated and discrepancies are immediately visible. Thus the nature of the manuscript itself uniquely positions the use of data visualisations for the purposes of research and for the application of distant reading.

Finally, the multilingual nature of the manuscript allows for discourse around translation and interpretation. Translation is largely an interpretive act. While directly translating a word from Spanish to English (such as *Dios* to *God*) is a relatively exact science,[4] the translation of a full text becomes interpretive. The meaning of words can change depending upon their context within the larger whole. It would seem at initial glance that the *Alcalá Record Books* would be relatively straightforward in its translation given that there is little context to consider when translating line expenditures in an account ledger. However, there still exists the larger cultural contextualisation to consider of the Irish as non-native speakers of Spanish in addition to the inclusion of descriptions around extraordinary expenses and the sign offs by the rector and the college's administrative staff. Thus, the multi-lingual nature of the archive allows for discussions to emerge between linguists and historians.

The type of data available in the archive, coupled with the already completed digitisation of the manuscripts (and their subsequent translations and transcriptions), provide a unique opportunity to focus on the digital presentation of the manuscripts and apply new search methods to explore the aforementioned hypothesis. Thus the *Alcalá Record Books* project (available at [283]) was born.

[4] It should be noted that even direct translation of a single word can occasionally be problematic as various words in many languages can have multiple meanings. The intent here is to generalise the translation of single words.

4.2 Initial Discovery: Utilising Focus Groups to Identify User Needs & Motivations

The idea of the focus group was an early emergent research item when the focus of the thesis was directed in a more theoretical manner.[5] Thus, originally, these focus groups were conducted with the intent of gaining an understanding for how users leverage these types of resources, as well as identify some of the key problems or sources of frustration experienced by users.

As the work of the thesis evolved away from a theoretical approach and towards a more practical one, the data from the focus groups was re-used in an effort to identify specific areas of focus for the design of the DRE that would eventually become the *Alcalá Record Books*. In addition, the data from the focus groups allowed for a focused application of the various theories presented in Chapters 2 and 3.

4.2.1 The Reasons Behind a Focus Group

In any software design practice, there is always a phase of requirements gathering and analysis of needs [284]. In a corporate setting, this can be undertaken in a variety of ways: through meetings with various business units, conducting market surveys, gathering feedback from a particular user community via feedback forms, etc. However, one of the more effective tools for understanding the needs and desires of a user community[6] is the focus group. The focus group allows for the researcher to obtain a small sample of a larger community and, if properly conducted, provide an interactive forum in which participants can be observed as well as heard. Thus it was the author's decision to employ the focus group (a mechanism with which they are intimately familiar through many years of use) as the primary tool for understanding the needs of those who leverage the DRE as a research tool.

4.2.2 Focus Group Design

Two focus groups were conducted in late 2016. A call for participants was sent out using social media and various networks within Maynooth University for any participants who utilise DREs such as *The Letters of 1916* [192], *The Woodman Diary* [285], and *Vincent Van Gogh The Letters* [286].[7] Participants were asked to fill out an online form providing basic demographic and contact details, as well as information regarding availability for a list of potential dates for the focus group. Participants

[5] It is worth noting that the original thesis proposal did not include a practical application such as the design of a DRE. Originally, the idea was to build a theoretical framework for the evaluation of user experience mechanics. As most research does, this idea evolved into a more practical approach, focused on testing a hypothesis about user interaction and the role of data visualisations.

[6] Authors note: while the literature offers numerous options without giving specific weight to one option over the other, it is the author's contention, based on personal experience in the corporate world as a software designer and technical architect, that focus groups tend to provide the most robust data for understanding the needs of a user community and the problems they face.

[7] This list is by no means exhaustive and was used simply to illustrate to the participants the types of systems being discussed.

were constrained to those living in or around the Dublin area who were at least 18 years of age and had experience leveraging DREs (then referred to as Digital Humanities Projects, or DHPs, in the literature describing the focus groups) in a research context. Ethical approval was gained on the grounds that all participants would be over the age of 18, their participation in the study was purely optional, all of their data would be anonymised, and they could opt out of the study at any point prior to the anonymisation of their data. Ethical approval for the study was obtained through Maynooth University's Social Research Subcommittee, and all participants were provided with a list of focus group rules, an information sheet regarding the study, and a consent form. All of this information is detailed in Appendix B. Participants were randomly placed into a focus group based upon their confirmed availability (as provided in the application form).

Two separate focus groups were conducted (with no overlap between the participants). Each focus group consisted of 6-7 participants in keeping with best practice [287]. The gender makeup skewed somewhat female with 61.5% of participants identifying as female compared to 38.5% identifying as male (see Figure 4.1). Most participants (61.5%) indicated a Masters degree as their highest level of education (see Figure 4.2) and the vast majority of participants indicated an age between 18 and 40 (see Figure 4.3).

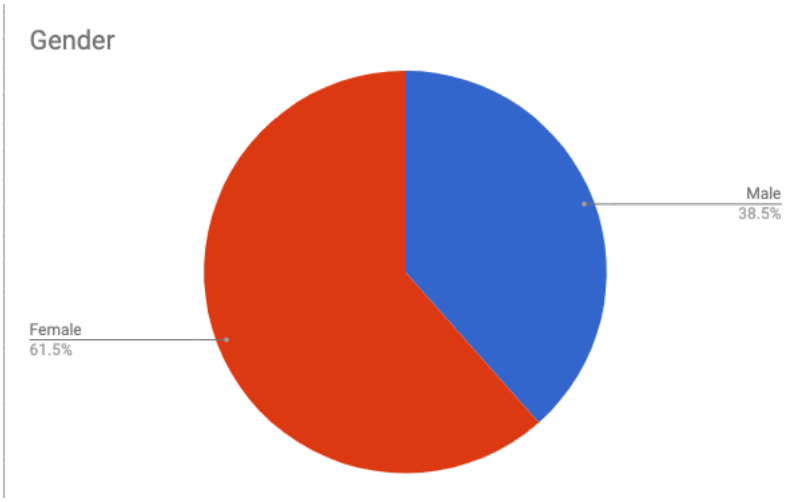


FIGURE 4.1: Gender makeup of participants across both focus groups

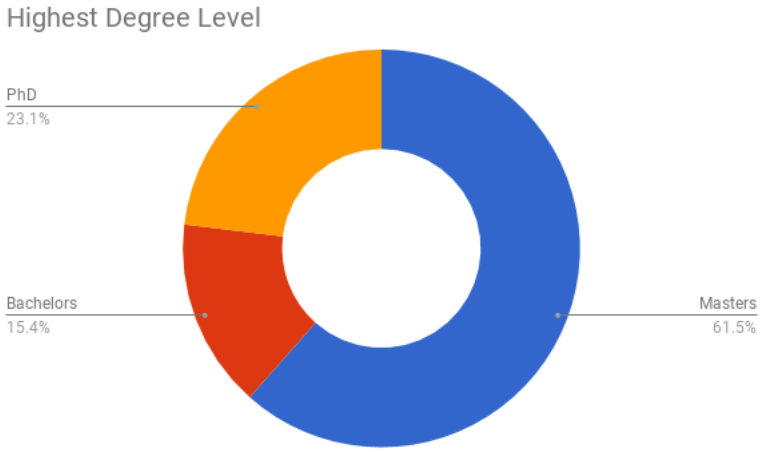


FIGURE 4.2: Breakdown of highest level of education as indicated by participants across both groups

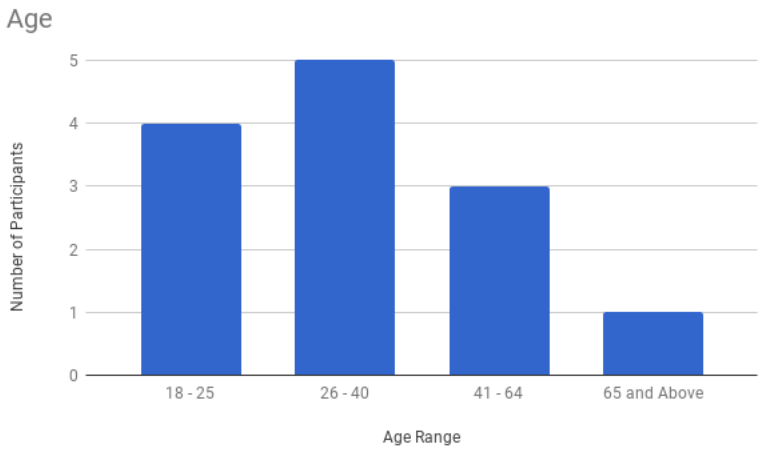


FIGURE 4.3: Bar chart showcasing the age range of participants across both groups.

The role of moderator was fulfilled by the author of this work (hereafter referred to as the researcher) with an additional research assistant conducting participant observation. Each focus group followed a carefully planned schedule prepared in advance in order to ensure a similar experience between the groups. The following details a brief outline of the focus group and its component activities:

1. Introduction (10 minutes) - begun by the researcher, this activity served both to set the tone for the focus group through a discussion of the rules and remind the participants of the aims of the focus group, as well as to be an icebreaker to give each member an opportunity to speak. This was accomplished through an activity known as *Draw a Card* wherein each person was given an index card on which they should write out a word or phrase that comes to mind when referring to a DRE. When finished, all participants placed the cards in a pile

that was then shuffled, and each participant drew a card and read aloud its contents. Discussion was then facilitated around the items on the card.

2. General Discussion (15 minutes) - a short group discussion, moderated and led by the researcher that aimed to generate discourse around some of the following topics:

   (a) Why do you use DREs?

   (b) Are visualisations in existing DREs helpful?

   (c) If there were no constraints, what would you like to see DREs accomplish?

   Discussion was allowed to naturally evolve with as little input and direction from the researcher as possible (with interjections only to clarify points or prompt further discussion).

3. Open Card Sorting (10 minutes) - a common focus group activity, this activity involved instructing each participant to list out all of the tasks or goals they would like to accomplish when using a DRE. Each task or goal was to be written on a separate sticky note and, upon completion, all sticky notes were brought together. The group was then charged with working together to organise all of the notes (not just their own) in order of importance (specifically indicating whether a task or goal was *very important*, *somewhat important*, or *don't know/care*).

4. How / Now / Wow (10 minutes) - this activity involved utilising the same cards generated in the previous activity that were marked as *very important* and sorting them along two axes: the x axis ranges from a *normal* idea on the left to an *original* idea on the right and the y axis ranges from *easy* at the bottom to *impossible* at the top. This causes the cards to be sorted into one of four quadrants: those ideas which are *normal* and *easy* fall into the *Now* category; those ideas which are *normal* and *impossible* are sorted into a *discarded* category, as they are not seen as being worth the effort for an unoriginal idea given the level of effort involved; those ideas which are *original* and *easy* are sorted into the *Wow* category, as they are seen as an easy win with high value for minimal effort; those deemed *original* but *impossible* are sorted into the *How* category, as they are seen as good ideas that bear further thought and investigation. As the x and y axes are scales (not absolutes), items can be visually sorted in order to illustrate their respective levels of difficulty and originality, allowing participants to further prioritise. An example of the grid system used to plot cards can be seen in Figure 4.4. A capture of actual output by the one of those focus groups using this technique can be seen in Figure 4.5.

5. 6-Part Story Method (40 minutes) - developed and detailed below, this activity involves each participant drawing images into 6 boxes that represent different

aspects of a narrative (the protagonist, the goal, a hindrance, a tool, the climax, and the ending). Participants then shared their stories with the group.

6. The Anti-Problem (20 minutes) - the group was given the question 'How do we make it impossible to read in DREs?' They were then given 5 minutes to draft a number of different answers to this problem and then 5 minutes to devise solutions to the answers they gave (a rebuttal). Each participant then shared their lists with the group.

7. Love Letter / Breakup Letter (15 minutes) - this activity involved each participant crafting a letter which could either be a love letter to a particular DRE they utilise, extolling its virtues and why they love it, *or* they could craft a breakup letter to a particular DRE, highlighting the problems inherent in the system and why they are choosing to no longer utilise the software. Each participant was then given the opportunity to read their letter aloud to the group.

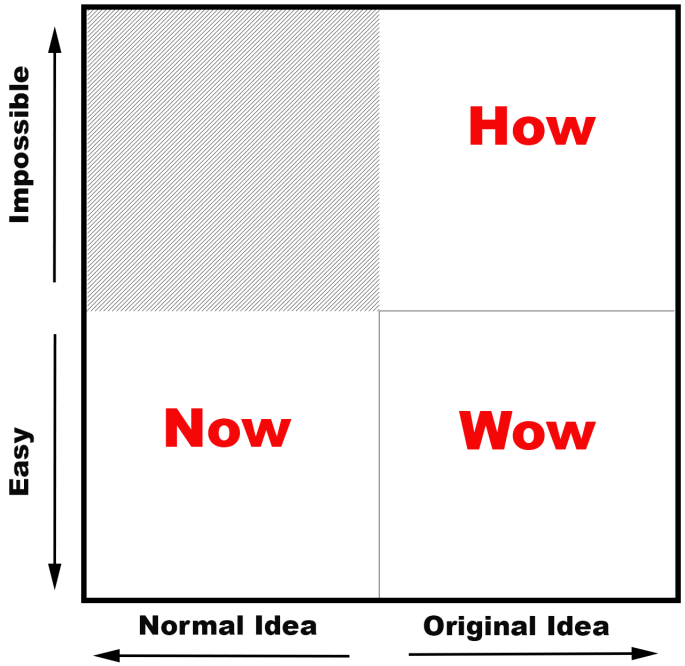


FIGURE 4.4: Example quadrant grid for plotting cards in the *How Wow Now* focus group activity.

FIGURE 4.5: Actual quadrant grid from one of the *How Wow Now* focus group activities.

The 6 Part Story Method (6PSM)

The 6 Part Story Method (or 6PSM) allows for a unique approach to system design by focusing on the user's individual story. The purpose is to allow the user to "tell a story" about how she uses a resource, providing insights for designers. One of the benefits of the approach is that it draws out user motivation and self-efficacy, the latter of which can have a major impact on how successful a user perceives a system to be [87]. Each of the stories created by various users can be easily translated into personas[8] and then analysed to determine the needs of each user group, by uncovering unconscious motivation or bias that would influence not only the way in which the user utilises the system but also objects or interactions that serve as a hindrance to use or adoption by the user community.

The 6PSM proposed here is an adaptation of drama therapist Alida Gersie's 6 Part Story Method. Under her original model, patients would be provided with a framework to craft a story that often tied into some type of subconscious meaning [289]. These stories could then be used to assist the therapist with uncovering deep-seated issues that require further therapeutic intervention [290]. In order to craft these stories, patients are given a piece of paper divided into six boxes and instructed to draw a picture in each box representing the following six ideas: the protagonist of the story, the goal or objective of the protagonist, the obstacle that keeps the protagonist for reaching the goal, an item or character that helps the protagonist during the story, the critical turning point, and the end of the story.

This method was adapted to a focus group setting for the purposes of understanding *how* and *why* users leverage a particular system, as well as uncover particular frustrations they may have encountered in the past. Each participant was given

[8] Pruitt and Adlin note the importance of defining personas which provide insight beyond that of a *user* and give the designer a sense of real-life individuals who are considered the target audience for the software, allowing the designer to more effectively engage the user and bring the designer's model closer to the user's mental model [189], [214], [288].

a piece of paper divided into 6 squares (3 across the top and 3 across the bottom). Participants were told they would be crafting a story about how a particular individual might set about accomplishing a goal or task using one of the aforementioned types of systems. Participants were then instructed to begin by drawing in the first box (the upper left-hand corner), a picture representing the main character of their story. Once completed, they were instructed to move one box right (to the middle box in the top row) and draw a picture of the goal or task the main character is attempting to accomplish. Participants continued the exercise by drawing individual images in the remaining boxes representing each of the following: an obstacle that stood in the way of the main character accomplishing their task, an object (i.e. a tool, person, etc.) that helps the main character in their story, the critical turning point in the story (also referred to as the *Eureka!* moment), and the final outcome of the story (i.e. the ending). Participants were then instructed to share their story with the group, using the images as a visual guide to illustrate the story to the wider group.

4.2.3 Focus Group: Analysis

Both focus groups were filmed and an additional audio recording was made as backup. The recordings were then transcribed with the participants anonymised in the transcription in keeping with Maynooth University's ethical practices for research [291]. Upon completion of the transcription, the transcripts were loaded into MAXQDA,[9] and the transcripts were coded using an open coding methodology [292]. Rather than breaking down the analysis line by line, it was decided to break down the coding by speaker:[10] so each time the speaker changes, that is considered a new segment of analysis (unless there is overlap among speakers with multiple people speaking at the same time, in which case these are treated as a single segment, but such instances are rare). As the codes developed, three primary themes began to emerge:

1. *Problems* - items under this theme referred to issues the participants encountered when using DREs. This was then subdivided into individual codes in order to identify recurrent problems, such as an inability to locate information, problems reading the content, technical issues, lack of context around the data being presented, etc.

2. *Reasons for Use* - items under this theme referred to why the participants had chosen to leverage the systems in question. This theme was subdivided into individual codes to identify specific uses, such as discovery, confirmation of ideas, public engagement, collaboration, and access (as in the ability to see the resource online as opposed to having to travel to a physical location).

[9] MAXQDA is a software toolset used for qualitative analysis of textual documents.

[10] This method of breaking down the text by a reasonable section as opposed to line by line is advocated by Corbin and Strauss. See Chapter 5 of [293].

3. *Emotive Reactions* - items under this theme attempted to gauge general emotional reactions (based on the language used) to whatever topic was being discussed. Responses here fell into one of three categories: Neutral, Positive, and Negative. It should be noted that this was meant to simply gauge an emotional response at the time. For example, there is no *Emotive Reaction* under the *Problems* theme or any of its subcodes. The idea was to simply gauge the overall emotional response of the participant.

Further codes were also added based on particular research interests. As one of the early underlying ideas of the research project had been the idea of escaping the book metaphor in online environments, a code was added to look for positive (approve of the use of the metaphor) and negative (disapprove of the use of the metaphor) reactions to the topic. However, very few instances of the code emerged (only 4 in total). A second code group was added to refer to the activity 'The 6 Part Story Method' (6PSM). This method, developed as part of this research, was specifically geared towards understanding how participants engaged with DREs. As a result, this code was used to capture any information related to the 6PSM for the purposes of analysis to prove the development of the theory (as detailed below).

Three additional codes were created as *catch-alls* for particular themes: Emotive Desire (the participant's desire for an emotional reaction to the system), Visualisation (used for anything related to data visualisations in general), and Safety (any discussions related to online safety and security).

4.2.4 Analysis & Discussion: Emergent Themes

To begin, the total code coverage once the coding was complete came to 46% coverage across both documents—that is, of all of the characters present in both transcripts, 46% of those characters were contained by at least one coded segment. The first step taken in the analysis of the data was to examine the most recurring codes. There were a total of 343 instances of codes across both focus groups. The top 5 codes are listed below: Individually, each of these codes accounted for roughly 5%

TABLE 4.1: Top 5 emergent codes showing # of occurrences and the % of occurrence overall.

Code	#	%
Discovery	42	12%
Technical Limitations	36	10%
Readability	28	8%
Visualisations	28	8%
Lack of Context	22	6%

code coverage across both transcripts, and thus cumulatively accounted for approximately 25% of all code coverage. The only code with a higher percentage of coverage was the 6PSM code (which applied to an entire activity). Upon closer inspection,

three of the five codes were related to specific problems participants had encountered (Technical Limitations, Readability, and Lack of Context), one was related to *Reasons for Use*, and the final code was related to data visualisations in general. The *Discovery* code also consisted of three subcodes: Analysis (using the system to facilitate some form of analysis of information, such as a comparative analysis), Learning (using the resource as an active learning tool), and Pattern Recognition (using the resource to discover unseen patterns in the data). This latter code (Pattern Recognition) was strongly correlated to the Visualisations code.

Each of these codes was used to inform the overall design of Alcalá. Discovery was clearly an important factor for many of the participants and thus one of the aims of the *Alcalá Record Books* was to find ways to facilitate that discovery process. Further analysis of the subcodes showed a marked interest in the use of data visualisations as a method of discovery. At one point, participants in one of the groups discussed the *Trans-Atlantic Slave Trade Database* [294] and the impact the data visualisations themselves had on understanding the data:

> Yeah the transatlantic life trade, that database, there's some very good data visualisations of that but when you speed it up you see the enormous impact it had. You're just seeing ships. You're seeing people. But you're seeing data as a...everyone was accounted for as one unit of cargo. Important people. So when you see that amount of cargo moving, you know, around the triangle, but when speeded up, you don't see it until you see it over time, 20, 30, 40 years, impact. I think that one was quite good. [207, 00:14:40]

This same individual noted that they did not truly understand the impact of the data until it was visualised. Another participant also noted how "mind blowing" the experience of visualising this type of data can be: "But I think when you see the data moving as an action over a period of time, like spatial history, you know and across a geographical space, it's definitely, it's mind-blowing." [207, 00:16:13]. In a separate discussion, another participant noted the need to be able to switch between the visual search and the traditional search:

> You know sometimes you can go to a site and you can search by text or you can explore the graph or something. But what I want to do is like a single thing to go between those things. Because my mind does not go to like 'oh let's do it by text or let's do it by picture'. It wants to go where it wants to go. [207, 00:42:26].

Some of the needs around discovery were further illuminated by an examination of the *Visualisations* code. Participants repeatedly mentioned using data visualisations as a prompt or a window into the data. They were effectively replacing the traditional browse mechanism with the use of a data visualisation. Some were focused on discovering specific information to support an idea or hypothesis, but an

equal number were leveraging data visualisations as a means of exploration. Such behaviour holds true with observations made by Johanna Drucker, who notes the importance of building mechanisms into a system that are beyond that of the efficient accomplishment of tasks and can facilitate novelty and enjoyment [6].[11]

The coded segments around technical limitations showed either issues with access (such as the lack of high speed internet) or basic problems with either UI or System Design. These include broken links, the inability to close modal overlays (or the lack of an intuitive close button), fidelity of images, etc. While important to recognise (as they are clearly recurring issues for participants), most of these items can be addressed through standard best practices in software development. Some of the problems, however, were related to access to the actual data: the ability to download the source object in a specified format (i.e. standard XML, TEI, METS, etc.) or the ability to download the data behind a visualisation to reproduce it locally. While ultimately considered outside the scope for the *Alcalá Records Books* (which is considered a prototype), these items were taken into consideration in the construction of the boundary objects and poster interfaces (to be discussed later).

Having perhaps the most significant impact were the codes around Readability and Lack of Context, as both related back to the Discovery code and the use of data visualisations to facilitate discovery. More than one participant mentioned the overwhelming volume of data contained in many of these DREs and the difficulty presented in attempting to read every item. In these cases, data visualisations were mentioned as ways of exploring the data at a higher level and allowing for a more focused approach to consuming the material present in the DRE. Multiple users also noted the importance of contextualising the data contained within the visualisations, as too often these types of visualisations are presented with little to no context regarding their construction or implementation. Both of these factors were strongly considered in the later design of the *Alcalá Record Books*.

It became clear through the analysis that the vast majority of participants were driven to DREs with the goal of discovery. Despite the common belief that most users seek out DREs in order to read the content, it is clear from the sample set mentioned here that users are in fact driven to discover and explore; reading is a by-product.

> Users of these systems demand alternative methods of engagement that assist with findability, accessibility, and discovery, as it is clear that their use of these systems is not simple engagement with the source material; the users are instead driven by a need to find and discover. Thus, the primary goal of the user is not to read but rather to explore. [295]

Armed with the information collected from the focus groups, this work set out to design Alcalá with the idea of discovery and exploration of the manuscript as the foremost goal with learning through annotation as an important secondary goal.

[11] For the full quote by Drucker regarding the importance of moving beyond the "reductively mechanistic" [6, p. 151] theory of interface, see Chapter 3.

4.3 System and UI Design Methodology

Taking into account the data present in the manuscripts coupled with the technical resources available and the results of the focus group, the goal of the work was to design a prototype of what could become a much larger system in-time. The following section details the overall system and UI design and describes the decisions made.

4.3.1 Personas: Crafting the User

Leveraging the data from the focus groups (primarily the various users crafted as part of the 6PSM), three distinct user personas were developed: the researcher, the learner, and the general user. Abbreviated versions of these personas can be found below and the full write up of all three can be read in Appendix C.

The researcher is referred to as Laura, a post-doctoral scholar in her 30s who is interested in the early Irish diaspora to continental Europe. She is comfortable with technology and has used various DREs throughout her graduate and post-doctoral work. She faces frustrations, however, with most of the DREs she has leveraged because the keyword searches do not always provide her with accurate results and she has difficulty identifying themes inherent in the manuscripts. She wants a system that will provide her with a better way of exploring the manuscript (beyond standard keyword searches) and would additionally like to be able to create a replicable result set that she could publish with her work.

Jenny is the learner. She is a student at Maynooth University working on her undergraduate degree in History. She prefers to use digital tools over the traditional pen and paper. As part of her Digital History class, she has been given an assignment to explore a DRE and create a presentation on the movement of human capital and culture in the 17th or 18th centuries. She needs to find a system that will allow her to annotate the information she finds and then compile everything into a format that will facilitate her presentation as part of her assignment. She is worried about finding a system that will have the information she needs for her presentation and how she can present that information as part of her assignment.

Matthew is the general user: a computer programmer based in Dublin who has an interest in his family history. As a computer programmer, he is extremely comfortable with the use of web-based technologies but has little understanding of how to leverage the content in DREs to locate information about his ancestor who attended Alcalá in the late 18th century. Matthew has no background in history and is thus unfamiliar with many of the lexical terms that may help him locate more information about his ancestor. He hopes to be able to find information that he can download and add to his collection for his family tree project, which he is building out on his own.

All three of these personas have different motivations but at the core, possess the same goals: discovery. While the reasons may differ, they all have concerns with

standard search and browse techniques and seek a system that offers an alternative that will facilitate locating and finding information. This theme of discovery carries throughout the Interaction Design of the project. In addition, the learner persona adds an additional layer of complexity by requiring the system to provide her with an interface for annotating and compiling the information she locates.

4.3.2 Interface Design: Crafting an Experience

Given the core of the experience needed to focus on discovery, it was decided to incorporate both a standard keyword search feature (with advanced search options) as well as a visualisation-based search feature. The latter would provide a data visualisation as its search output (as opposed to table of results which is standard practice for a keyword search). This data visualisation would clearly demonstrate the patterns in the data and would also contain an interactive element that would allow the user to *click through* to an underlying data model for a relevant aspect of the visualisation. This would allow the user to effectively create a smaller subset of results that could then be individually read and analysed. Such a system would—at least in theory—support notions of interactive search, furthering immersion in the data and engagement with the source object. It also would act as a facilitator for close reading (as opposed to a replacement), as it would create a narrow set of documents to be read by the user.

In addition to facilitating discovery, the design also sought to address some of the concerns with replicability and access to the underlying data. While wholesale exports of the encoded text and images were not possible given copyright restrictions, the decision was made to implement a metaphor that would not only further learning and engagement, but also create a reproducible, static object that could be downloaded and re-used. The goal here, then, was to enable a feeling of data reliability, as well as to craft an environment for the learner to leverage in an attempt to facilitate the overall learning experience. This goal was accomplished through the use of boundary object[12] and the poster metaphor, the design of which drew inspiration from Laurillard's *Conversational Framework*. By allowing the learner to annotate information and construct her own knowledge of the system, the system is forming the basis for what could be developed into a larger conversational model, allowing the teacher and the learner to communicate learning objectives and outcomes. It was within this larger context that the boundary object design was implemented.

The boundary objects exist in three forms: a keyword search (see Figure 4.6), a visualisation search (see Figure 4.7), and a page view (see Figure 4.8). The user can opt to save any keyword search, visualisation search, or individual page and annotate the saved object with their own comments. Relying on a notecard metaphor, the idea here was to create an annotated object, using the user's own words that could be reliably and repeatedly replicated with complete accuracy. This was accomplished

[12]For discussion and a definition of boundary object, see Chapter 1.

by storing the exact parameters (in the case of searches), visual configurations (in the case of the visualisation search), or unique identifiers (in the case of page views) along with the text of the annotated object. Each object is tied to a user and can also be referred back to a specific search log entry (in the case of search-related boundary objects) based on the parameters and creation time of the object. Thus this boundary object furthers engagement with the source material by forcing the user to *chunk*[13] information into a smaller subset and provide annotation in their own words for why the boundary object matters. Such a design seeks to offload some of the cognitive load of working memory through *chunking*, as well as providing the user with the space to more deeply engage with the manuscript.

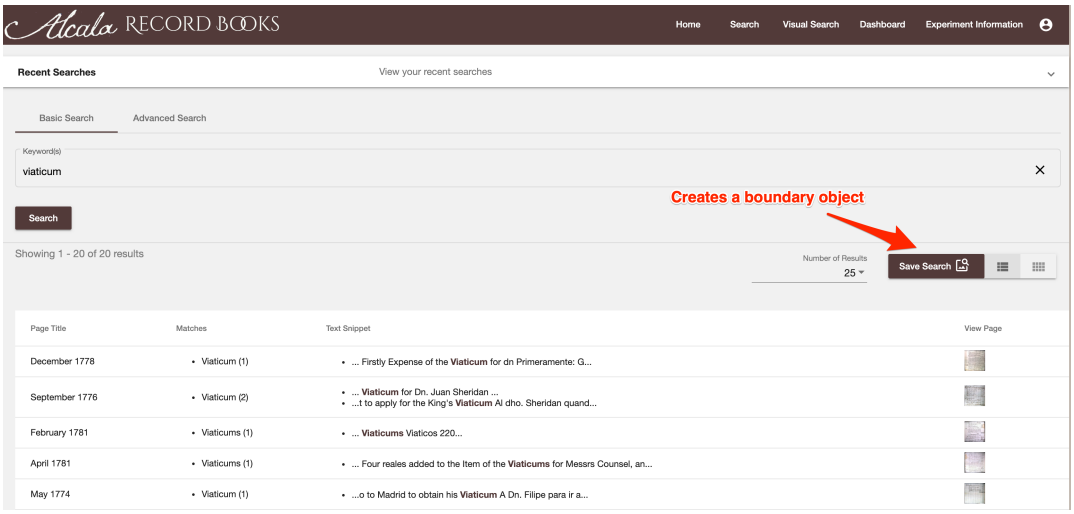


FIGURE 4.6: Annotated screen grab highlighting the ability to create a boundary object from a keyword search. A larger version is available in Appendix E in Figure E.11

[13] For more information on *chunking* and how it relates to working memory, see Chapter 1.

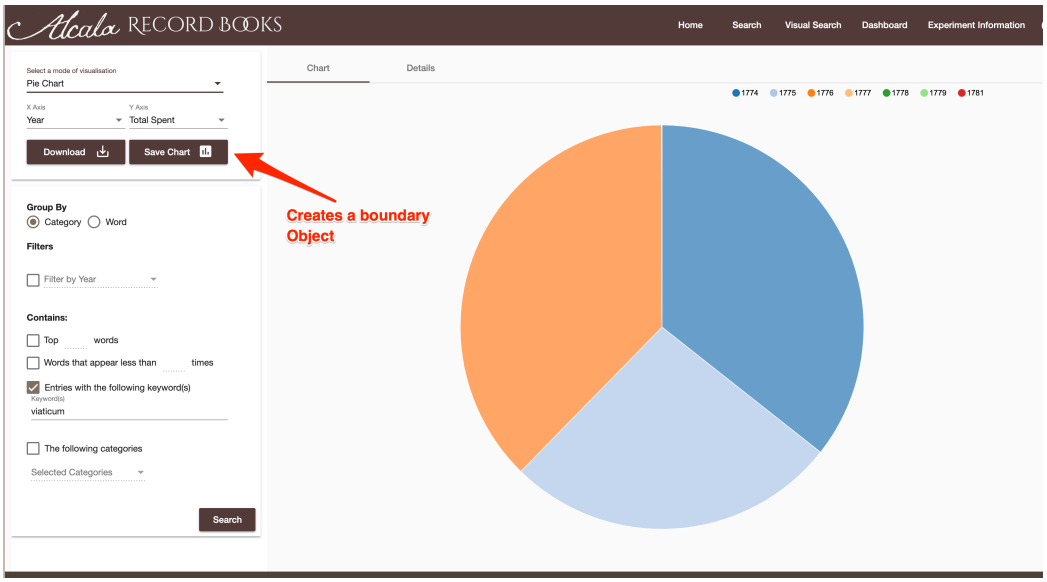


FIGURE 4.7: Annotated screen grab highlighting the ability to create a boundary object from a visualisation search. A larger version is available in Appendix E in E.12

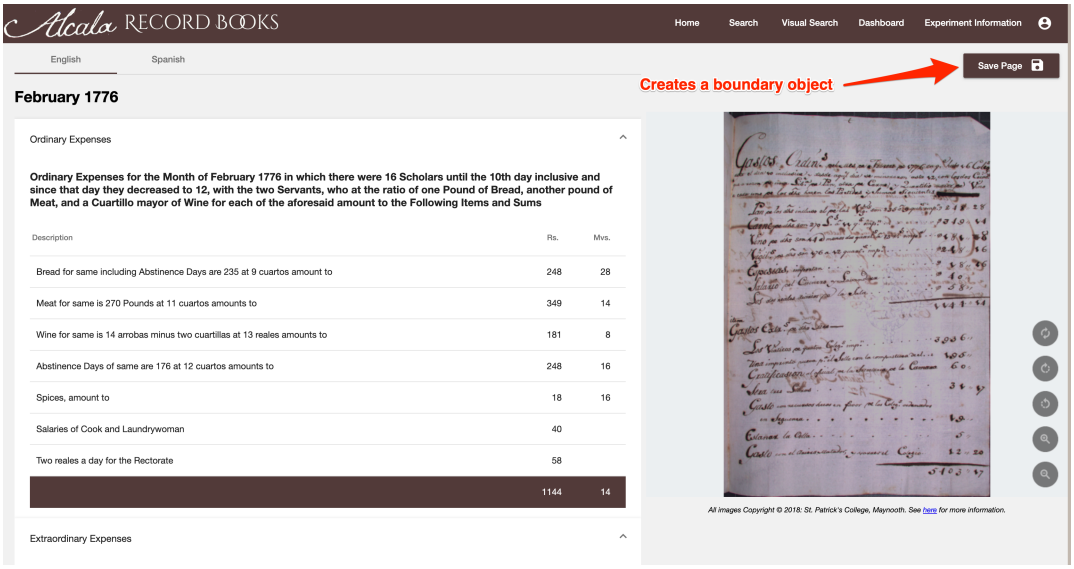


FIGURE 4.8: Annotated screen grab highlighting the ability to create a boundary object from a page view. A larger version is available in Appendix E in Figure E.13

Initially, there was an included dashboard that highlighted various metrics in the system (such as number of Boundary Objects created, Keyword with the most matches, etc.), as well as other user specific metrics such as recent searches, lists of boundary objects, etc. (early designs of these components can be seen in the wire-frames described in Appendix C). During initial beta testing (which occurred using volunteers from the Computer Science undergraduate programme), many of these metrics were found to be confusing, and the purpose of the dashboard was not clear. Thus much of the dashboard was scrapped in favour of a simplified approach that

allowed users to see all of their created boundary objects and then compile these boundary objects into a poster.

The poster metaphor (originally referred to in the initial designs as a story) sought to offer a compilation mechanism for the boundary objects and create a static, downloadable object that could be re-used. Drawing on the metaphor of the academic conference poster, the interface provides the user with the ability to describe the overall narrative of the work and then place selected boundary objects which support the narrative into the poster. This poster could then be exported to *.png* and downloaded so it could be re-used.

Additionally, at any time a user can log into the system and view a dashboard which will display all of their created boundary objects as well as all of their user-generated posters. This functionality allows the user to manage their boundary objects (editing or deleting as needed), as well as their poster objects. Originally, the idea of the dashboard was more sophisticated, allowing the user to track particular aspects of the system they found intriguing (such as most frequent word, or the number of transactions in a given category), allowing the user to build out a top-level custom view into the data. The undertaking proved to be confusing in initial beta tests and was later dropped for simplicity's sake.

The initial wireframes highlighting each of these items described above can be seen in Appendix C.

4.3.3 Relating the Design to the Literature

The overall design of the experience as discussed in the previous section was strongly influenced by both Goal-Directed Design (GDD) and Emotional Design (ED) philosophies as described in Chapter 3. Reinforcing these design principles as an inherent part of the overall creation of the *Alcalá Record Books* is an essential component in establishing the system's overall validity as a research output.

Applying a Goal Directed Design Approach

One of the primary goals of all three of the described personas is that of discovery, thus, under the GDD paradigm, the system must facilitate modes of discovery in its approach. This goal is accomplished both through the keyword search *and* the visualisation search, with the former assisting inductive reasoners and the latter assisting deductive reasoners, as is discussed in great detail in Chapters 5 and 6. Through an understanding of the primary motivation for using the DRE as a mode of discovery,[14] the implementation of multiple search modalities provides the end-user with numerous paths through which they may accomplish their goal. Additionally, by leveraging simple financial graphs to show spending across categories or words, the contextualisation of the graph is inherent (bar charts, line charts, and pie graphs

[14] See the previous section that discusses the findings of the Focus Group.

are readily understood by most readers); this increase in contextualisation also contributes to an increase in readability of the data visualisation, as it is easier to understand the purpose of the visualisation and what the data is highlighting—ideas borne up by the data collected and analysed in Chapter 5.

A secondary goal mentioned in the personas is that of knowledge dissemination. Each user persona sought a way to compile their thoughts and findings and share this compilation in a format for others to consume. The poster and boundary object functionalities fulfil this goal by both allowing the user to craft references to data within the system—contextualised through their own lens of experience—and compile these references into a narrative device for distribution.

The overall design of these components accurately reflects the three primary criteria of the GDD approach:[15]

1. Understand the primary needs and motivations of the user
2. Understand the domain in which the software resides (in this case, the Digital Humanities Digital Research Environment)
3. Applying knowledge gained from the first two items to create a system which is "useful, usable, and desirable" [214, p. 4]

The first two items in the list are demonstrated in the previous sections through the data gathered as part of the focus groups. The third item can be seen in not only the design as it is laid out here but, more importantly, through the analysis of the User Experience Questionnaire that is part of the case study detailed in Chapter 5.

Applying an Emotional Design Approach

An Emotional Design approach is reflected through multiple facets of the overall design. As demonstrated in the previous section, each user persona's goals were incorporated into the design of the various system components. As mentioned in Chapter 3, the completion of a goal meets a Need For Cognition (NFC) and increases satisfaction levels in the user. Furthermore, goal completion leads to increase immersion (as defined in Chapter 3), which ultimately leads to a stronger emotional connection to the system as a whole.

However, the accomplishment of a goal is not the only factor that contributes to the ED approach. The implementation of the boundary objects (which is operationalised as a metric for engagement[16]) also increases immersion with the system through playful immersion and challenged-centric immersion.[17] The creation of a boundary object becomes a challenge in and of itself, as it requires focused attention on an aspect of the source object (in this case, a keyword search, visualisation search, or page) and then applying careful thought and analysis to the data in order

[15] See Chapter 3 for further discussion.
[16] See Chapter 5.
[17] See Chapter 3 for more on the various types of immersion contained within this work.

to re-contextualise it through their own lens of experience. This aids not only in a feeling of immersion, but also aids working memory as a whole, through the implementation of *chunking*. By forcing the user to stop reading and apply careful though to the object, the working memory system is given time to process and integrate the information into long-term memory.

All this translates into an emotionally more satisfying experience. The user does not feel overwhelmed by the data in question because there is a decreased likelihood of cognitive overload (as explained above by the *chunking* process). In addition, the user experiences a much more novel approach to discovery through the incorporation of the visualisation-based search and is left with a stronger, overall reflective experience. The data discussed in Chapter 5 supports these conclusions.

4.3.4 System Architecture: Crafting an Ecosystem

As mentioned earlier, the original translation and transcriptions (along with digital copies of the individual manuscript pages) were provided in advance, as they were part of an earlier project (see [282]). Due to time constraints, the original encoding of the transcriptions was kept intact, and thus some of the underlying design decisions were informed by the assumptions made in the original data model—primarily the use of the account ledger metaphor inherent in the data model. The standard model was defined as follows:

1. Year
   - 1.1. Month
     - 1.1.1. Ordinary Expenses
     - 1.1.2. Extraordinary Expenses
     - 1.1.3. Adjustments
     - 1.1.4. Summation
   - 1.2. Adjustments
   - 1.3. Summation
2. Sign-offs

It should be noted that while the page is an important aspect of the model (in order to tie the data to an analogue counterpart in the manuscript), it is somewhat irrelevant within the model itself, as page breaks can happen at any point within the subsequent model—that is to say not every page starts with 'Year', some may start with Ordinary Expenses, Monthly Adjustments, Year Summations or even Sign-offs, thus it is not included as part of the data model above; however within the XML itself, <*page*> is always the top level node.

As the original data was built on the eXist-db database, the decision was made to leverage this as part of the underlying database implementation. This proved effective for displaying the individual pages; however, it lacked some of the analytical power that would later be needed to process the data visualisations. As a result, an additional NoSQL database using mongodb was implemented in order to process

all data visualisation searches, as well as storing any user generated content (such as system logins, search logs, boundary objects, and posters).

In order to begin construction of the data visualisations, the data first needed to be scrubbed[18] and loaded into the mongodb instance. This involved a Python script that began by loading all of the data from the eXist-db instance (using the pyexistdb package) into the mongodb instance (using the pymongo package). The format of the mongodb data was considerably flattened as part of the scrubbing process. This flattened model can be seen in Figure 4.2.

TABLE 4.2: Structure of an AnalysisItem (scrubbed data format from eXist-db to mongodb).

Field Name	**Data Type**	**Description**
_id	ObjectID	unique identifier generated by mongodb
categories	string[]	Categories assigned to the transaction (e.g. Travel, Food, etc.)
maravedises	int	One of the numerical values related to amount (approximately 34 maravedises to 1 real)
reales	int	One of the numerical values related to amount (singular is real)
month	int	The month in which the transaction took placed, expressed as an integer (e.g. 6)
monthName	string	The string representation of month (e.g. a value of 6 in month would give monthName a value of 'June')
pageid	string	The unique page identifier that references a unique key in the eXist-db instance
words	sring[]	A bag of words that makes up the description of the transaction. This is all of the words in the description minus stop words (such as 'a', 'the', etc.)
Year	int	the year of the transaction expressed as an integer (e.g. 1775)

Once this scrubbing was accomplished, 3 random training sets of 100 transactions each were generated and then imported into their own document collection in mongodb. An interface was then built around this training data which allowed a properly authenticated user to go through each transaction, view the page on which the transaction originated, and *tag* the transaction with one or more categories. Prof. Thomas O'Connor (Professor of History at Maynooth University) and Dr John Keating (Senior Lecturer in Computer Science) were tasked with handling the document classification of this training data as both had been involved with the previous iteration of the project and were knowledgeable on the subject matter.

[18]Here the term *scrubbed* is referred to as a reformatting of the data to fit into the necessary JSON object model inherent in all mongodb collections.

Once all training data was tagged, the transactions were then run through a Document classifier using the *sklearn* package. This classifier used the built-in *CountVectorizer()* from *sklearn* and TF-IDF as the transformer. Once the predicted categories were determined, they were then appended to their relevant transactions through a standard mongodb update. The full code (minus the wrapper classes which read and write data from the datastore) can be found in Appendix C.

The system itself is designed with a RESTful Flask API written in Python acting as the middleware between the two databases (eXist-db and mongodb) and the website, which is written in Angular 6.0. Additionally, a site was created for hosting all of the image files in order to allow these files to be relocated to another server (if necessary), requiring only a minor configuration tweak to the website. A high-level diagram of the system can be seen in Figure 4.9.

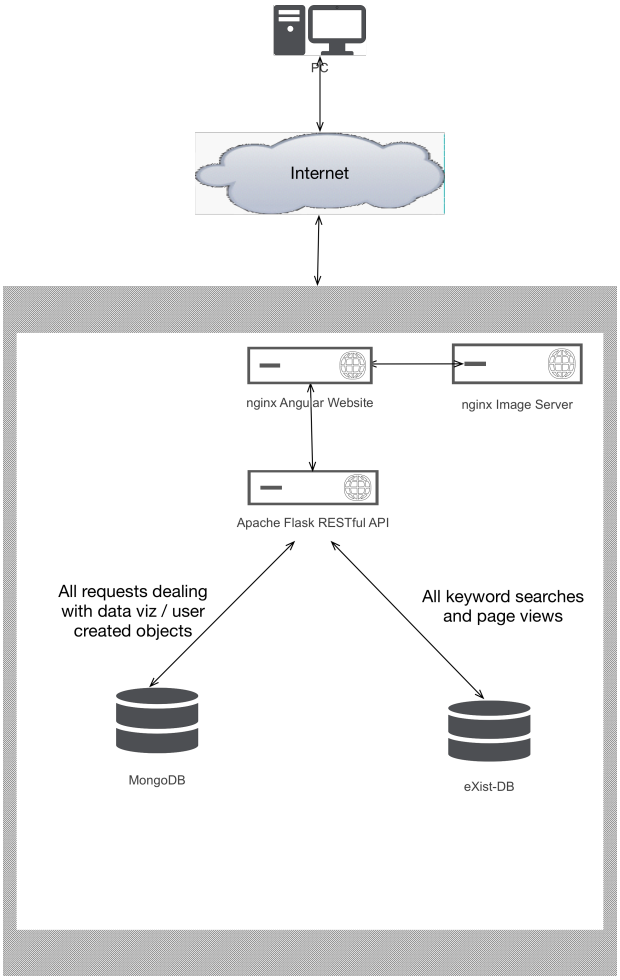


FIGURE 4.9: High level diagram of the system architecture for Alcalá

Requests received from the client to the RESTful API may take one of two paths:

1. All keyword searches[19] or requests for individual manuscript page views are

[19]It should be noted that while keyword searches are routed through the custom-built ExistData python module, all searches regardless of type are logged through the custom-built MongoData python module.

routed through a python module (ExistData), which leverages the *pyexistdb* package to communicate with the eXist-db database. The wrapper then loads the data into custom objects (which contain the XML mappings) and serialised to JSON for the response. The eXist-db implements a lucene index and the Keyword In Context (KWIC) module that allows for keyword searches to return the context in which the keyword appears in each search result. An example query leveraging the lucene index and the KWIC module is seen below:

```
1        import module namespace kwic="http://exist-db.org/xquery/
     kwic";
2        let $hits :=
3            for $hit in doc("alcala/books/ledger.xml")//pages/page
4            where $hit//textContent[ft:query(.,"viaticum*")]
5            order by ft:score($hit) descending
6            return
7                $hit
8        let $total-hits := count($hits)
9        for $hit in $hits
10       return
11           <result>
12               {$hit}
13               <matches>
14                   {kwic:summarize($hit, <config width="40"/>)}
15               </matches>
16           </result>
17
```

LISTING 4.1: Sample eXistDB Query as executed by ExistData python Module

2. All visualisation searches, as well as any user-specific functions (such as authentication, search logging, creation of boundary objects, etc.), are routed through a python module (MongoData), which leverages the *pymongo* package to communicate with the mongodb database. The user-specific functions all have JSON models that match their expected website implementation stored in mongodb and thus do not require a mapping object to translate between systems (although mapping objects were created to maintain a consistent Object-Oriented Programming (OOP) design). The data required for the data visualisations undergoes significant transformation in order to work with the visualisation framework selected for the project. The design decision was made to format the data on the server rather than sending only the raw data to the client and formatting it via javascript (which could become memory intensive for the user's browser).

The code for the two server modules mentioned (ExistData and MongoData) can be found in Appendix C minus the code for the object mappers, which are somewhat superfluous to the underlying implementation demonstrated here.

As mentioned previously, the website code is written in Angular 6.0 and used a standard Component architecture. Each aspect of the interface was split into components with related services that handled the communication with the Flask API and the Image Server. In addition to the standard implementation, the website also utilised the *Angular-nvd3* charting component, which is an Angular 2.0+ implementation of the *nvd3* reusable chart library. As most of the data contained within the manuscript is financial data, the decision was made to focus on financial charts that would be familiar to a modern web user (things like bar charts, pie charts, line charts, etc.). It was theorised that these charts would be easier for most users to interpret, as they rely upon metaphors many users encounter in their daily lives. A deliberate decision was made to specifically avoid things like topic modelling and vector space analysis partly due to their tendency to be confusing without heavy contextualisation and partly due to their inability to provide perfectly replicable results (as these approaches are based on statistical models).

The chart aspect of the website has two specific aspects worth mentioning. The first is the data manipulation that takes place on the server. While the underlying raw data is also sent to the client when a user initialises a data search, this raw data is also reformatted on the server, which provides numerous summaries of the data. The complete DataPackage object returned from the server contains:

1. summary - contains the total amount of reales, maravedises, a grand total, and a total count of transactions
    1.1. timeSummary - an array of objects that contain summaries of reales, maravedises, total amount, and transaction count grouped by a particular time value (year or month depending upon the request).
2. data - an array of objects which contain the sum of data grouped first by the allowable y axis groups and then by a time component (year or month depending upon the request)
3. rawData - an array of transactions that match the specified search criteria and were used by the server to create the various summary packages

The creation of this data package can be found in Appendix C under the section entitled *MongoData Module*.

The second aspect of the charting component is the creation of the chart itself. The aforementioned data package is only created when a new search is issued (that is, when the user clicks the *search* button); however, there was a desire to provide the user with the ability to change their view of a single search result without having to refresh the page and constantly reload the data. In order to accomplish this, an Abstract Factory Pattern[20] was implemented in order to determine the proper *nvd3* component to instantiate in the browser. This provides the user with the ability to constantly change their view of the data by trying out different charts and changing

[20]The Abstract Factory Pattern is a software design pattern that facilitates the implementation of various concrete classes via a Factory method. The benefit of the Abstract Factory is that it can implement an n number of concrete cases provided each case implements the required interface [296].

the x and y axis groupings. Such an approach allows the user to customise the visualisation to meet their individual needs or preferences. Figure 4.10 demonstrates the ability of the user to modify the axis and change their view of the data. The code for the Abstract Factory can be found in Appendix C.

FIGURE 4.10: Screen grab from the Alcalá Record Books annotated to highlight the options for changing the x and y axis of the chart. A larger version can be seen in Appendix E in Figure E.14

The other advantage to the implementation of the Abstract Factory Pattern is the ability to swap out the charting library should the need arise. A number of different charting frameworks (such as *HighChartsJS* and *ChartJS*) were experimented with in the early stages of development. As this evaluation was occurring, it was imperative that the system be designed in such a way that various frameworks could be swapped in or out with minimal implementation changes. Should *nvd3* no longer be supported or should another framework be developed that better suits the needs of the software, the implementation cost is low as it requires implementing only the specific charts required and activating them in the factory.

4.4 Conclusion

Understanding the needs and motivations of the community is imperative in any software design. By leveraging multiple focus groups, new insights were gained into what drives the users to these DREs and understanding their motivations and frustrations. From this data, personas could be constructed to drive the design, as well as to gain deeper insight into the goals of the user base. Facilitating a Goal-Directed Design approach, coupled with aspects of Emotional Design,[21] a software system was crafted that sought to provide:

[21] See Chapter 3 for more discussion on these concepts

- alternative methods of discovery - accomplished through an interactive visualisation search, this additional mode of discovery not only seeks to appeal to the visceral aspect of an *Emotional Design* approach[22] but also aims to offset some of the cognitive overload that occurs when attempting to read in digital environments (as discussed in Chapter 2). Additionally, the interactive visualisation provides an activity which is heavily contextualised by the user's enculturated experience with the source material and, theoretically, assists with the accomplishment of a goal.

- an attempt to further engage with the source material - accomplished through the creation of boundary objects, this engagement seeks to appeal to both the behavioural and the reflective nature of Emotional Design by heightening the learning experience. Through the use of *chunking*, the load on working memory is lessened, and the user experiences a greater sense of immersion with the source text as they stop to engage with a particular section through the creation of the boundary object.

These pieces of functionality (coupled with the more standard keyword search and page view implementations common to these environments) provide a practical implementation of the theoretical underpinnings of this work.

However, to truly explore the underlying theory, the system must be tested with a user community that represents all three personas. The following chapter details a case study conducted in December 2018 and March 2019, which sought to ascertain the success of the implementation and provide data for analysis and discussion.

[22] See Chapter 3.

Chapter 5

Case Study: Assessment of the Impact of Data Visualisations on Learning & Engagement

In[1] order to evaluate the usefulness of a visualisation-based search—and thus provide data to support the hypothesis that a visualisation-based search enhances not only the overall experience of the system but also increases engagement, thus facilitating the learning and research process—there must be a formal evaluation and testing of the software. As a result, two separate case studies were carried out over the course of a three-month period.

The first case study drew participants from volunteers across Maynooth University who participated in a workshop hosted by the Arts & Humanities Institute that featured discussions around the Irish in Europe. This case study drew a total of 19 participants across a range of disciplines. The second case study was part of a seminar series on Usability and User Experience Design conducted as part of a module on UI/UX for the Masters in Computer Science at Maynooth University. The particular seminar that was the case study, which contained a total of 16 participants, was open to both postgraduate (Masters) and undergraduate students in Computer Science.

In order to obtain ethical approval, two separate rounds were sought. The first round was simplified as all users were volunteer attendees of a workshop hosted by the Arts & Humanities Research Institute at Maynooth (thus there were no power imbalances to consider). All participants were required to be over the age of 18 and consent forms and information sheets were provided. Similar to the Focus Group discussed in Chapter 4, participants were ensured anonymity and could opt out of the study at any point up until the anonymisation of the data.

While the structure of the study remained consistent across both groups (as described below), a second round of ethical approval for the seminar was sought due to issues around power dynamics. Unlike the workshop, the seminar participants were

[1] Portions of this chapter are reused in an upcoming journal article entitled "Engagement through Visualisation: A Case Study on the Alcalá Record Books" in the *IEEE Transactions on Learning Technology*.

required to participate, as the seminar was part of a module. However, students were given the opportunity to opt out and not have their data collected, thereby allowing them to participate in the seminar but voluntarily participate in the study. This was detailed in the information sheet and consent form in order to to ensure there were no power dynamics at play in the collection of the data. As with the workshop, all data was anonymised, and students were given the option to opt out at any time until the anonymisation of the data. All documentation related to ethical approval can be found in Appendix D.

5.1 Study Design

Both studies were conducted using the same rules. Prior to the start of the case study, all participants were asked to fill out a questionnaire that captured demographic data (such as education level, gender, age, and area of research), as well as their familiarity with technology: the latter of which was accomplished through a modified use of the *Media and Technology Attitude and Usage Scale* [297].[2] The complete pre-test questionnaire can be found in Appendix D. Prior to the start of the case study, participants were randomly sorted into two groups: Group A and Group B. The reason for the grouping split was to control for the order effect[3] as demonstrated below.

During the case study, participants were given a basic overview of the system and the historical background of the *Alcalá Record Books*. Participants were then tasked with using either the keyword search only (for participants in Group A) or the visualisation search only (for participants in Group B) to locate information related to food. They were instructed to create boundary objects as they located items of interest and to generate a poster of their findings (time permitted). Participants were given 30 minutes to conduct their search with a further 15 minutes to complete a post-test questionnaire. This questionnaire, referred to as a User Experience Questionnaire (UEQ), attempted to ascertain the overall user experience using a series of questions developed by Laugwitz, Held, and Schrepp [299], as well as provide space for free-text comments and feedback. Upon completion, participants were then asked to switch their search method (with Group A now using the visualisation search and Group B using the keyword search) and complete the same task as before. After 30 minutes, they were then asked to fill out another version of the post-test questionnaire based on their most recent experience with the system. This final version of the post-test questionnaire contained an additional section that allowed for the participant to indicate a preference in search style, as well as final

[2] It should be noted that only an adapted portion of the scales was used as some of the items were not relevant to case study.

[3] The order effect is a concept in psychology that refers to how participants respond to items in a test based on the order in which the material is presented [298]. In this case, the experiment attempted to control for the order effect by randomly splitting participants into groups to ensure that there was an even distribution of those starting with the keyword search as opposed to the visualisation search and vice versa.

feedback comments. The full post-test questionnaire (including the *Final Thoughts* section which allows the participant to indicate their preference between the two search mechanisms) is included in Appendix D.

5.2 Participant Demographics

In total, there were 35 participants[4] split across the two case studies. A total of 15 female participants and 20 male participants were split across 5 research areas: Computer Science (16 [45%]), History (7 [20%]), Other Humanities (4 [11%]), Library Studies (6 [17%]), and Other Sciences (2 [6%]).[5] See Figures 5.1 and 5.2.

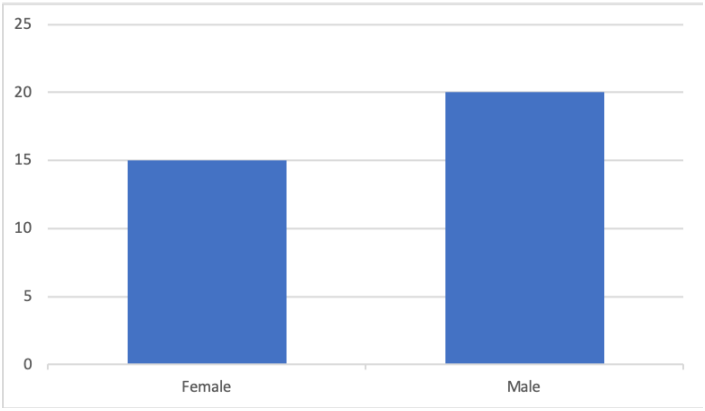


FIGURE 5.1: Gender makeup of participants across both case studies

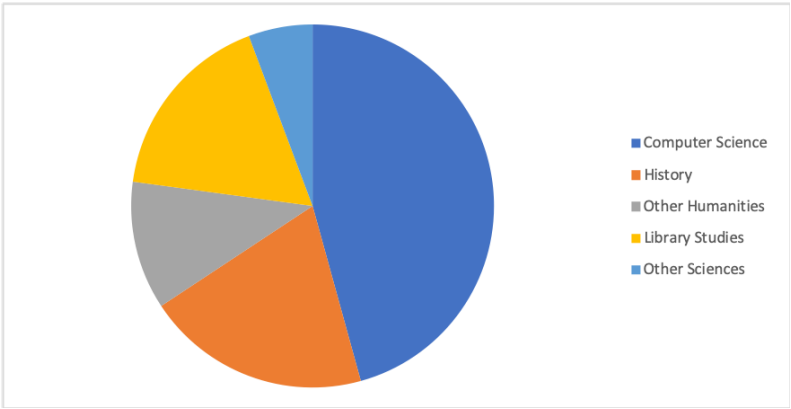


FIGURE 5.2: Specified research area of participants across both case studies

Most participants possessed a Masters degree (86%) or higher with the majority of participants ranging in age from 22 - 40 (71%). See Figures 5.3 and 5.4.

[4]For the purposes of the data analysis, both groups are analysed together. Unless otherwise noted, $n = 35$.

[5]These groupings were created based on free form feedback from the user concerning their research area. The "Other Humanities" and "Other Sciences" categorisations were made to ensure there were no categories with a population of only 1, which would thus exclude them from statistical analysis of means.

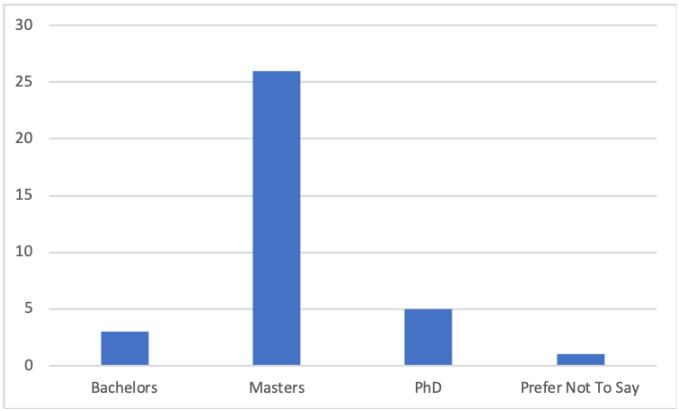


FIGURE 5.3: Highest education level of participants across both case studies

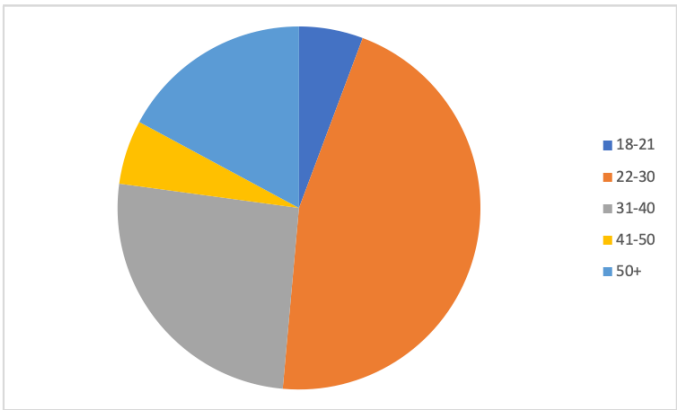


FIGURE 5.4: Age range of participants across both case studies

The median usage score (on a scale of 1 to 10, which maps to the scales in the pre-test questionnaire) was generally highest among traditional STEM participants with both Computer Science and Other Science participants having the highest median usage score of 9. A similar correlation was also seen in the median attitude score (on a scale of 1 to 5, which maps to the scales in the pre-test questionnaire) with Computer Science and Data Science having identical median scores of 4.[6] See Figures 5.5 and 5.6.

[6]As these scales are Lickert scales, a median value is more statistically correct than the mean (or average); thus these numbers represent the median of the median for each respective research area.

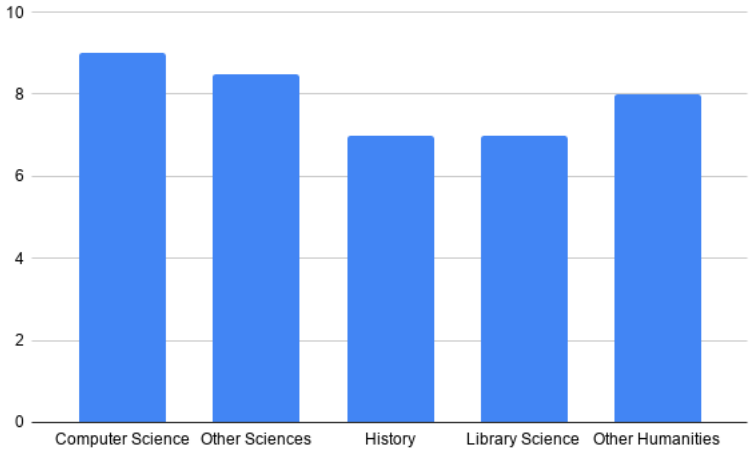


FIGURE 5.5: Median usage (scale of 1 - 10) of technology of participants by research area across both case studies

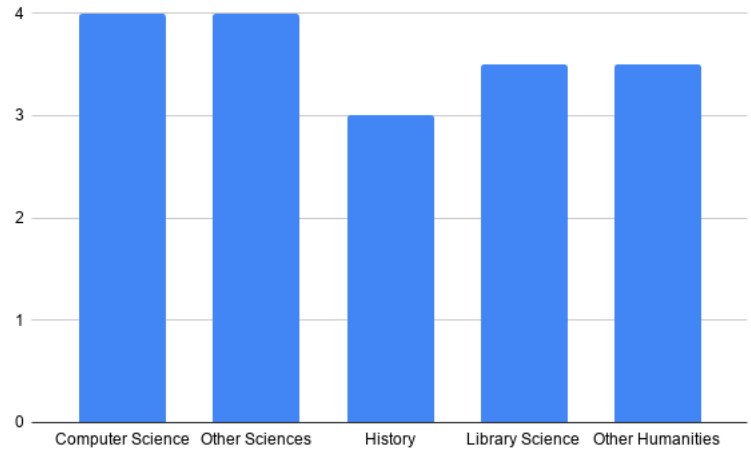


FIGURE 5.6: Median attitude (scale of 1 - 5) of participants by research area across both case studies

All participants were provided with an information sheet and a consent form, which can be found, along with the ethical approval for the study, in Appendix D. Upon completion of the case study, all participant data was made anonymous with names replaced with a system generated Globally Unique Identifier (GUID), and other identifying information (such as email addresses) were removed from the system.

5.3 Analysis of Results

To begin the initial analysis, two different dependent variables were investigated:

1. the selection of a search preference that occurs at the end of the post-test questionnaire

2. the number of boundary objects created by each participant

The selection of a search preference acts as a gauge as it relates the reflective reaction to the system.[7] The number of boundary objects was operationalised in order to gauge the level of engagement with the system. It was determined that these two variables would give an initial insight into the data, highlighting additional paths for further, in-depth analysis, especially concerning the analysis of the UEQ and the qualitative analysis of the free-text comments.

5.3.1 Analysis of the Search Preference

Overall, the visualisation search was the preferred method of interaction by 21 participants (60%), compared to 14 (40%) who preferred the keyword search. Utilising IBM's SPSS programme, either a Pearson's Chi Square test or a Fisher's Exact test—the latter was selected if the independent variable had only two levels, otherwise the former was used [300], [301]—was used to begin to look for correlations in the data. The first test was to confirm if the control for the order effect mentioned earlier was effective. When comparing the starting group of the participants with the final selection of their preferred method, 60% of Group A (9 participants) and 60% of Group B (12 participants) indicated a preference for the visualisation search. A Fisher's Exact Test shows no statistical significance between starting group and preferred method selection ($p = 1.000$), thus the control was effective.

From here, other correlations were investigated. A Fisher's exact test shows that while 75% of male-identifying participants preferred the data visualisation search method as compared to only 40% of women, there was no statistical significance between gender and preferred method selection ($p = .08$). A final Fisher's Exact test was run to determine if the participation in a particular case study (workshop versus seminar) bore any statistical significance, and the test did highlight an effect ($p = .036$).

Most of the Pearson Chi-Square tests proved no correlation. For example, age range played no role in the selection of a search method ($p = .776$) and neither did highest level of education ($p = .775$). Research area, however, did show a statistical correlation ($p = .022$) Additionally, it was decided to break the down the participants into a further grouping by type of training: the deductive reasoning common in STEM training and inductive reasoning common in humanities training. The reasoning behind this breakdown was twofold:

1. to explain the statistical difference between study types (where one consisted solely of STEM participants, which was suspected of skewing the results).

2. to determine more conclusively if the type of training background typical to these fields might have a broader influence.

The research areas were mapped as follows:

[7] For more about the reflective reaction and its role in *Emotional Design* techniques, see Chapter 3.

TABLE 5.1: Mapping of research areas to training type

Research Area	Training Type
Computer Science	Deductive
Other Sciences	Deductive
Other Humanities	Inductive
History	Inductive
Library Studies	Inductive

Utilising this mapping, the groups were nearly identically split, with 17 participants grouped as inductively-trained and 18 participants grouped as deductively-trained. See Figure 5.7.

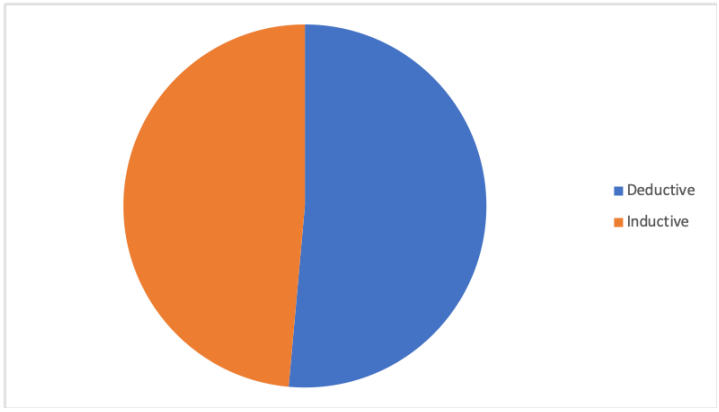


FIGURE 5.7: Type of training of participants across both case studies

This mapping is supported by an analysis of attitudes and usage. When re-grouping participants into the these two groups and comparing the median usage and attitude scores to the same scores noted above by research area, those with a deductive-trained background score higher on the usage scale (9 compared to 7) and the attitude scale (4 compared to 3). See Figures 5.8 and 5.9.

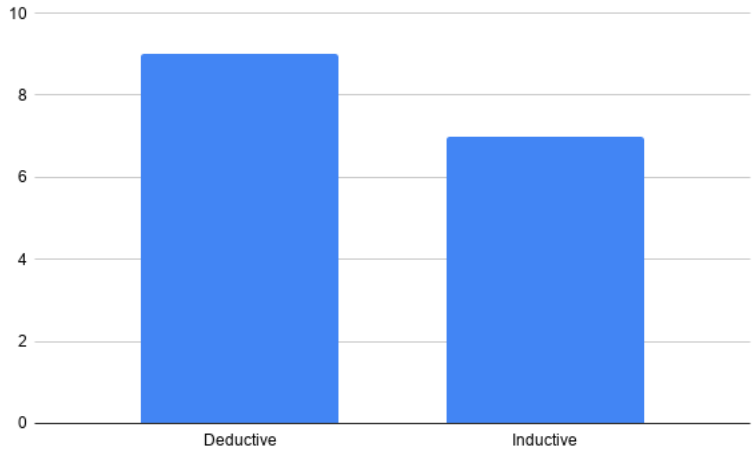


FIGURE 5.8: Median usage (scale of 1 - 10) of technology of participants by type of training across both case studies

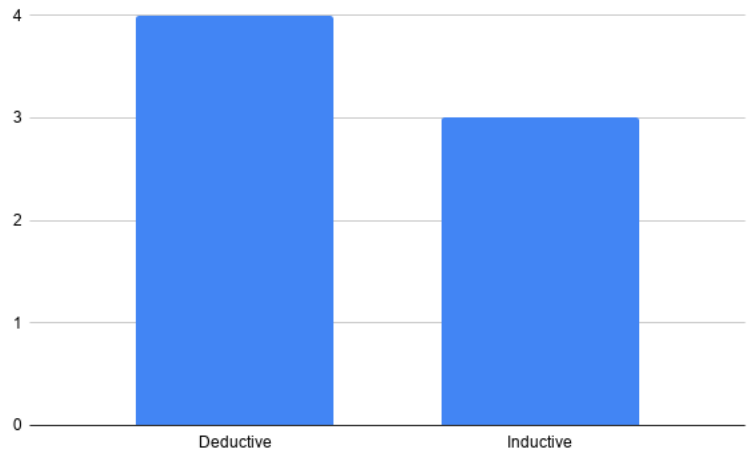


FIGURE 5.9: Median attitude (scale of 1 - 5) of participants by type of training across both case studies

Only 35.5% (6 of 17) of inductive-trained participants selected visualisations as their preferred method compared to 83.3% (15 of 18) of deductive-trained participants. A Fisher's exact test indicates this is a statistically significant comparison ($p = .006$).

Based on the above statistical analysis, it would appear the independent variable that had the most significant impact (and thus bore further investigation) was the research area and—by extension—the type of training associated with the research area.[8]

With these findings in mind, further exploration was needed to see if there was perhaps any correlation between the above and the statistical analysis of the individual's engagement with the system, which was operationalised as the number of boundary objects created.

5.3.2 Analysis of Engagement

The decision to use the number of boundary objects created as the measurement for engagement was based primarily on two factors:

1. nearly all participants created at least 1 boundary object during each test in a given case study.

2. boundary objects highlight an attempt by the participant to critically engage with the source material and enlist various aspects of working memory to facilitate the knowledge acquisition process.

While the number of posters created could also potentially be operationalised, this metric was ruled out due to the small number of posters actually created (only 30% of participants attempted to create one). Given the short time frame allowed during

[8]While the type of study in which the user participated also had a statistical significance, it is not considered here due to its correlation with the reasoning type, as discussed earlier.

each test (a total of 30 minutes), it can be theorised that many participants chose not to create the poster, as it was a more time-intensive process.

Because of the small sample size ($n = 35$), a non-parametric test was required to look for statistical correlations. Relying upon the same independent variables in this analysis as in the analysis of the preferred search method, an inquiry into the participants' preferred method of search was deemed necessary to determine if it had any correlation to the number of bboundary objects created. A Kruskal-Wallis H test [302] was run and distributions of the total number of boundary objects were similar for all groups based on a visual inspection of a box plot. The median number of boundary objects created was not statistically significant among groups: $\chi^2(1) = 1.205, p = .272$. A Kruskal-Wallis H test also revealed that there was no statistically significant correlation between age and median boundary objects created, although it should be noted that this lack of correlation was barely above the threshold: $\chi^2(4) = 9.424, p = .051$. Additionally, there was no statistically significant correlation with gender: $\chi^2(1) = 1.006, p = .316$. A final check of the control put in a place to account for the order effect also showed no statistical correlation with the median number of boundary objects created: $\chi^2(1) = .091, p = .763$.

All other independent variables revealed some type of correlation using a Kruskal-Wallis H test. A test between highest education level and total number of boundary objects created showed a significance of $\chi^2(3) = 7.860, p = .049$ although it should be noted that a visual inspection of a box plot showed the groups were dissimilar. A further pairwise comparison showed the only significant difference was that between those with a Bachelors and those with a PhD ($p = .037$). Given the vast difference in research training between these two groups, such a significance is to be expected, and thus can be somewhat ignored for the purposes of this study.

Upon comparing the research area to the total number of boundary objects, a statistical significance was also observed ($\chi^2(4) = 19.747, p = .001$) and a visual inspection of a box plot shows a similar distribution among groups. The two groups with a statistically significant difference were Computer Science and Library Studies ($p = .013$) and Computer Science and History ($p = .003$). These findings were further reinforced when comparing training type (where a visual inspection of the box plot showed similar distributions across groups), which highlighted a statistical significance: $\chi^2(1) = 15.260, p < .0005$. As training type has only two levels, there is no pairwise test to conduct, but given the extremely low p value, it would appear that those with the strongest deductive training (e.g. Computer Science) produced fewer boundary objects than those with the strongest inductive training (e.g. History).

The final Kruskal-Wallis H test conducted was between the particular research study (workshop versus seminar) and the total number of boundary objects created. Distributions were similar for both groups based on a visual inspection of a box plot, and a statistically significant correlation can be seen: $\chi^2(1) = 18.931, p < .0005$. This test reinforces the decision to group research areas into inductive versus deductive as the seminar group consisted solely of deductively-trained participants, thus proving

a correlation between training and engagement.

The findings of the non-parametric K tests support the conclusions of the analysis of the preferred search method: that the research area—and by extension the type of training associated with the research area—has a strong correlation with engagement in addition to the preferred method of search.[9]

The above findings support the notion that training (and to some extent the specific area of research) have an impact both on preferred method of use and on engagement with the system via the creation of boundary objects. However, further analysis is required to understand the impact of these items.

5.3.3 Further Statistical Analysis

In order to gain further insight into the research question and the hypothesis, a comparison between the different types of search and their requisite outcomes was required. As research area—and by extension training background—had the most statistical impact, these variables became the focus of further analysis. All search-related metrics (boundary objects, the number of searches, and the average time spent between searches) were divided into two groups: keyword and visualisation. As every search was logged in the system and the type recorded, and as only one type of search was conducted during a particular time frame, this data was easily generated by the system.

The first type of test conducted was a Wilcoxon signed rank test [303] to determine if there was a statistical significance in the search metrics. Using keyword search as the control and visualisation search as the experiment (for the purposes of the test), there was no statistical significance found in the number of boundary objects created ($p = .133$) nor in the average time between searches ($p = .098$). There was, however, a statistical significance in the number of searches conducted. Visualisation searches saw a statistically significant increase in the number of searches generated ($p = .008$).

While no statistical significance existed when comparing the means of the two experiment groups, a statistical significance *does* exist when running a non-parametric ANOVA. When comparing the number of boundary objects generated during the keyword search across research areas, a Kruskal-Wallis H test reveals a statistical significance: $\chi^2(4) = 13.961, p = .007$. This significance becomes even more pronounced when examining the number of boundary objects generated during the visualisation search: $\chi^2(4) = 16.906, p = .002$.[10] In a pairwise comparison of the keyword search, the statistically significant comparisons are between Computer Science–History ($p = .030$) and Computer Science–Library Studies ($p = .036$). In the visualisation search, a pairwise comparison showed only a statistical significance between Computer Science–History ($p = .013$).

[9] As noted in the search preference analysis, study type is ignored due to its likely association with training.

[10] In both scenarios, a visual inspection of the box plot shows a normal distribution of means.

When considering the grouping by training type (as noted above), the results were somewhat similar. Again, a Kruskal-Wallis H test reveals a statistical significance in the number of boundary objects created across type of training: $\chi^2(1) = 12.788, p < .0005$. In this instance the keyword search was actually slightly more statistically relevant than the visualisation search using a Kruskal-Wallis H test, which had a result of $\chi^2(1) = 10.262, p = .001$. While both had normal distributions, it should be noted that the visualisation search in this case contained a couple of minor outliers in the data. As they were minor, they were not discarded.

Comparing the other search metrics studied in the Wilcoxon signed rank test, there appears to be no statistical significance in variance among research area. See Table 5.2 for results. However when looking at the variance of the same metrics by training type, statistically significant evidence arises for how individuals engage with search. With the exception of the amount of time spent on the keyword search, all search metrics showed a statistically significant variance. See Table 5.3 for results.

TABLE 5.2: Kruskal-Wallis H tests for search metrics by research area.

Search Metric	**KW Result**	**Sig.**
Average time (keyword)	$\chi^2(4) = 3.777$	.437
Average time (visualisation)	$\chi^2(4) = 7.675$	.104
Number of searches (keyword)	$\chi^2(4) = 7.406$	.116
Number of searches (visualisation)	$\chi^2(4) = 8.634$	.079

TABLE 5.3: Kruskal-Wallis H tests for search metrics by training background.

Search Metric	**KW Result**	**Sig.**
Average time (keyword)	$\chi^2(1) = 2.948$	.086
Average time (visualisation)	$\chi^2(1) = 6.376$	.012
Number of searches (keyword)	$\chi^2(1) = 6729$	.009
Number of searches (visualisation)	$\chi^2(1) = 4.682$	.030

Two separate binomial regressions were run in order to determine if any of the above metrics had an impact on the user's reflective experience (i.e. their choice of preferred search method). In both models, preferred method was treated as a dichotomous variable, where 0 represents keyword search as the preferred method and 1 indicates visualisation as the preferred method.

The first regression looked at the effect of the number of searches generated during each type of search and the average time spent between searches in each type of search (for a total 4 metrics). The preferred method was used as the dependent variable. Linearity of the continuous variables was confirmed via the logit to the dependent variable as assessed by a Box-Tidwell procedure [304], using a Bonferroni correction [305] applied to all terms, resulting in statistical significance at $p < .0056$. There were no standardized residuals that fell more than ± 2.5 standard deviations from the mean, thus there was no need for correction. The model fit was considered statistically significant: $\chi^2(4) = 10.815, p = .029$, and the model explained 35.9%

of variance (using Nagelkerke R^2) and correctly classified 77.1% of cases. See Table 5.4 for the values for sensitivity, specificity, and positive and negative predictive values. The area under the ROC curve (see Figure 5.10) was .803 (with a 95% CI of .654 to .951), which is considered an excellent level of discrimination according to Hosmer, Lemeshow, and Sturdivant [306]. However, of the 5 predictor variables, only 1 proved to be statistically significant: average number of seconds between searches using the visualisation search (see Table 5.5) with it having only a slight impact (odds of 1.014) of increasing the likelihood to select the visualisation search.

TABLE 5.4: Sensitivity, Specificity, and Predictive Values for the Binommial Regression using Search Metrics.

Metric	**%**
Sensitivity	81.0
Specificity	71.4
Positive Predictive Value	80.9
Negative Predictive Value	71.4

TABLE 5.5: Logistic regression predicting the likelihood of selecting Visualisation as the preferred method based on number of keyword searches, number of visualisation searches, average time between keyword searches, and average time between visualisation searches.

Variable	**B**	**SE**	**Wald**	**df**	**p**	**Odds**	**Lower**	**Upper**
# of Searches (Kw)	-.039	.040	.965	1	.326	.962	.890	1.039
# of Searches (Viz)	.028	.027	1.074	1	.300	1.028	.975	1.084
Avg. Time (Kw)	.001	.007	.018	1	.894	1.001	.988	1.014
Avg. Time (Viz)	.014	.007	.969	1	.046	1.014	1.000	1.028
Constant	-1.687	1.705	.979	1	.322	.185		

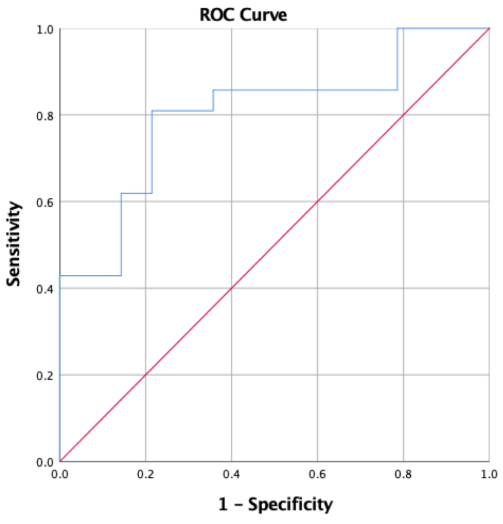


FIGURE 5.10: ROC curve for binomial regression of search metrics to predict preferred method

The second regression looked at the effect of the number of boundary objects created for each search (for a total 2 metrics). The preferred method was used as the dependent variable. Linearity of the continuous variables was confirmed via the logit to the dependent variable as assessed by a Box-Tidwell procedure [304], using a Bonferroni correction [305] applied to all terms, resulting in statistical significance at $p < .01$. There were no standardized residuals that fell more than ± 2.5 standard deviations from the mean, thus there was no need for correction. The model fit was not considered statistically significant: $\chi^2(2) = 1.835, p = .399$; however, a Hosmer and Lemeshow test did not result in a statistically significant result: $\chi^2(7) = 10.045, p = .186$, which suggests the model is still a good fit [306] (see Table 5.6 for the contingencies).

TABLE 5.6: Contingency Table for Hosmer and Lemeshow Test.

Step	Observed (Kw)	Expected (Kw)	Observed (Viz)	Expected (Viz)	Total
1	2	2.576	2	1.424	4
2	2	2.063	2	1.937	4
3	4	1.723	0	2.277	4
4	1	1.578	3	2.422	4
5	2	1.429	2	2.571	4
6	1	1.600	4	3.400	5
7	0	1.255	4	2.745	4
8	2	1.210	2	2.790	4
9	0	.565	2	1.435	2

Additionally, the model explained 6.9% of variance (using Nagelkerke R^2) and correctly classified 62.9% of cases. However, the area under the ROC curve (see Figure 5.11) was .675 (with a 95% CI of .485 to .865), which is considered a poor level of discrimination according to Hosmer, Lemeshow, and Sturdivant [306]. None of the variables were statistically significant in the prediction, thus this regression is not considered applicable to prediction of preferred method. As such, it is not relevant to further analysis.

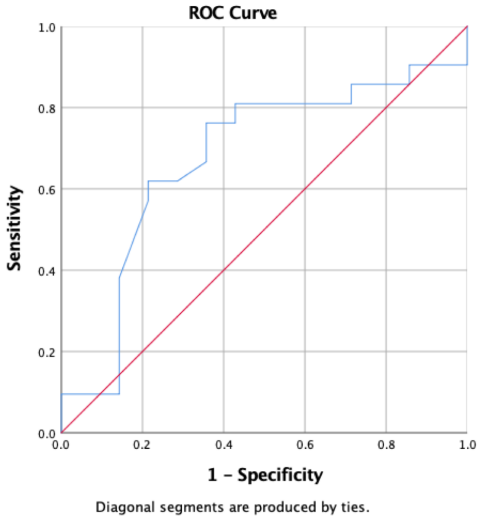


FIGURE 5.11: ROC curve for binomial regression of boundary objects to predict preferred method

Finally, two additional binomial regressions were run to determine if any of the same metrics referenced above could be used to predict whether the participant fell into the "inductive" training group or the *deductive* training group. The first regression tested the number of boundary objects created between the two types of search (for a total of 2 covariates). Linearity of the continuous variables was confirmed via the logit to the dependent variable as assessed by a Box-Tidwell procedure, using a Bonferroni correction applied to all terms, resulting in statistical significance at $p < .01$. There were no standardized residuals that fell more than ± 2.5 standard deviations from the mean, thus there was no need for correction. The model fit was considered statistically significant: $\chi^2(2) = 19.418, p < .005$, and the model explained 56.8% of variance (using Nagelkerke R^2) and correctly classified 70.6% of cases. See Table 5.7 for the values for sensitivity, specificity, and positive and negative predictive values. The area under the ROC curve (see Figure 5.12) was .884 (with a 95% CI of .777 to .991), which is considered an excellent level of discrimination according to Hosmer, Lemeshow, and Sturdivant [306]. Of the 2 predictor variables, only the number of boundary objects created during the keyword search proved to be statistically significant (see Table 5.8) with each new boundary object created having a likelihood to predict an inductive participant by 1.544.

TABLE 5.7: Sensitivity, Specificity, and Predictive Values for the Binommial Regression using Created Boundary Objects.

Metric	**%**
Sensitivity	70.6
Specificity	77.8
Positive Predictive Value	75
Negative Predictive Value	73.7

TABLE 5.8: Logistic regression predicting the likelihood of determining if the user is an inductive reasoner based on the number of boundary objects created in either the keyword or visualisation search.

Variable	B	SE	Wald	df	p	Odds	Lower	Upper
Boundary Objects Created (Kw)	.434	.183	5.652	1	.017	1.544	1.079	2.208
Boundary Objects Created (Viz)	.194	.120	2.610	1	.106	1.214	.960	1.536
Constant	-2.334	.841	7.710	1	.005	.097		

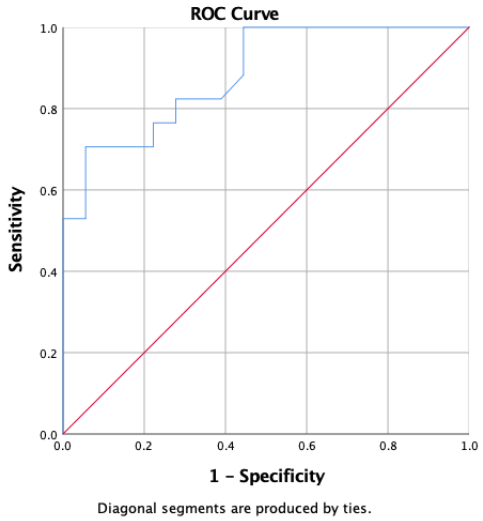


FIGURE 5.12: ROC curve for binomial regression of boundary objects to predict training type

The second regression tested the number of searches conducted and the average amount of time between searches across the two types of search (for a total of 4 covariates). Linearity of the continuous variables was confirmed via the logit to the dependent variable as assessed by a Box-Tidwell procedure, using a Bonferroni correction applied to all terms, resulting in statistical significance at $p < .0056$. There were no standardized residuals that fell more than ± 2.5 standard deviations from the mean, thus there was no need for correction. The model fit was considered statistically significant: $\chi^2(4) = 16.210, p = .003$, and the model explained 49.4% of variance (using Nagelkerke R^2) and correctly classified 76.5% of cases. See Table 5.9 for the values for sensitivity, specificity, and positive and negative predictive values. The area under the ROC curve (see Figure 5.13) was .837 (with a 95% CI of .706 to .968), which is considered an excellent level of discrimination according to Hosmer, Lemeshow, and Sturdivant [306]. Of the 4 predictor variables, only the average amount of time spent between visualisation searches proved to be statistically significant (see Table 5.10) with each additional second having a likelihood to predict an inductive participant by .982.

TABLE 5.9: Sensitivity, Specificity, and Predictive Values for the Binommial Regression using Search Metrics.

Metric	%
Sensitivity	76.5
Specificity	72.2
Positive Predictive Value	83.3
Negative Predictive Value	76.5

TABLE 5.10: Logistic regression predicting the likelihood of determining if the user is an inductive reasoner based on number of keyword searches, number of visualisation searches, average time between keyword searches, and average time between visualisation searches.

Variable	B	SE	Wald	df	p	Odds	Lower	Upper
# of Searches (Kw)	.084	.045	3.55	1	.059	1.088	.997	1.187
# of Searches (Viz)	-.048	.031	2.432	1	.119	.954	.898	1.012
Avg. Time (Kw)	.002	.006	.061	1 .	.804	1.002	.989	1.014
Avg. Time (Viz)	-.018	.008	5.069	1	.024	.982	.967	.998
Constant	2.034	1.859	1.197	1	.274	7.646		

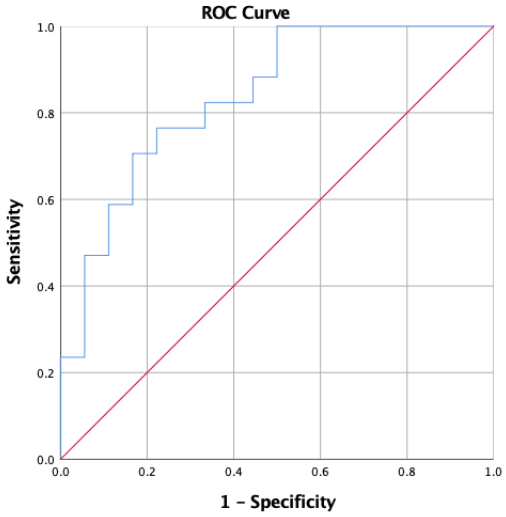


FIGURE 5.13: ROC curve for binomial regression of search metrics to predict training type

A further investigation of the impact of this data is conducted in the discussion section of this chapter.

5.3.4 UEQ Analysis

As part of the post-test questionnaire, the UEQ was incorporated in order to ascertain the user experience of the particular type of search used. Each participant completed the post-test questionnaire twice (once for the visualisation search and once for the keyword search). The test itself asks 26 questions—all answers are

ranked on a Likert scale of 1 to 7—and attempts to understand the overall user experience based on six (6) areas: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [299]. As the system evolved, Schrepp, Hinderks, and Thomaschewski grouped 5 of these 6 categories into super categories: pragmatic (which includes perspicuity, efficiency, and dependability) and hedonic (which includes stimulation and novelty); as a result, attractiveness is then seen as a valence dimension of which the pragmatic and hedonic are part [307]. It is worth noting here that these aforementioned categories, which constitute attractiveness, map to two of Norman's aspects of *Emotional Design*: the behaviour aspect (which maps to the pragmatic) and the visceral aspect (which maps to the hedonic).[11] A further exploration of the application of Norman's *Emotional Design* to this analysis will be discussed further on.

As part of the analysis, the UEQ is packaged with a data analysis tool built in Excel, which provides a number of metrics and benchmarks. The data from both studies was combined across two separate analyses: one for visualisation search and one for keyword search.[12] Each analysis is provided with results which calculated the means across the scores of the 26 areas (where each question is mapped to one of the 6 aforementioned categories). Overall, the visualisation search slightly outperformed the keyword search when looking at a summation of the mean benchmark scores for each area (visualisation search: $\sum = 4.45$, keyword search: $\sum = 4.39$).[13] When evaluating the attractiveness and hedonic qualities, the visualisation search out-performed the keyword search; however, the keyword search was seen to be the more pragmatic. See Table 5.11.[14]

TABLE 5.11: The Attractiveness, Pragmatic, and Hedonic Scores for both types of search.

Search Mechanism	Attractiveness	Pragmatic Quality	Hedonic Quality
Keyword	.68	2.73	.98
Visualisation	.95	1.54	1.95

In an attempt to look for correlations in scores between different research areas or training backgrounds, separate statistical tests had to be constructed. Both a Kruskal-Wallis H test and a One-Way MANOVA [308] (Hotelling's T^2 test in the case of training background as it was dichotomous [309]) were executed using either the research area or the training background as the independent variable and

[11] For more on *Emotional Design*, see Chapter 3.

[12] $n = 35$ for both.

[13] It should be noted that these scores should not be considered formal KPIs, as the UEQ does not calculate KPIs unless additional data is implemented, which was not considered for this study. This is merely a summation of the mean scores for each of the 6 areas.

[14] Values above .8 can be seen as a positive valuation. Values between -.8 and .8 can be seen as neutral. Anything below -.8 is seen as a negative valuation.

the attractiveness, pragmatic, and hedonic scores[15] for each type of search (for a total of 6 dependent variables). All four tests (each of the two tests run twice: once for research area and once for training background) did not produce statistically significant results. Thus it can be assumed that neither research area nor training background had any influence on the overall user experience.

5.3.5 Qualitative Analysis of User Feedback

In order to analyse the written comments that were part of the post-test questionnaires, the transcriptions of these written responses were broken into 3 documents: keyword feedback, visualisation feedback, and final thoughts. Each of these transcripts correspond to a section in the post-test questionnaire—questions 32-36 for the keyword feedback, questions 38-43 for the visualisation feedback, and questions 47-50 for the final thoughts.[16] The transcripts can be found in Appendix D. Each question was treated as a codeable segment (as most responses were only a sentence long). In no situation does a coded segment span more than one question. The actual coding of the documents leveraged the same open coding technique detailed in Chapter 4. Additionally, if the participant made the same (or similar) comment in multiple questions (such as "I didn't like x" or "I really enjoyed y"), it was only coded once so as to avoid artificial inflation of the codes.

The code coverage across all three documents was 44%. The codes primarily broke down into 2 main themes: positive and negative feedback. Each of these themes then broke down into 3 further categories:

- relevance - feedback regarding the relevance of the data returned via a search result
- clarity - feedback regarding how easily the participant understood the data
- usability - feedback related to how usable the system was from a functional standpoint

Other codes not part of these two themes were codes for tracking:

- a preference for visualisation search over keyword search
- a preference for keyword search over visualisation search
- cognitive overload (which relates to any comments by the user of feeling "overwhelmed" by the system or the data in their result sets)

[15] It should be noted that these tests were somewhat spurious from the outset as the calculations for attractiveness, pragmatic, and hedonic scores in the UEQ workbook rely on means; however, since the questions themselves are based on a Likert scale, they should be leveraging the median rather than the mean, as it is ordinal rather than interval data. As the data in the workbook is already based on peer-reviewed and accepted work, the calculations in the workbook were not adjusted for this flaw, and thus these scores were imported as means and treated as continuous interval data for the additional statistical analysis in order to remain consistent with the accepted work.

[16] All questions for the post-test questionnaire are found in Appendix D.

- learning (which was used to track any references to new information the user gleamed as a direct result of the using the system)

Table 5.12 details the percentage of each code as a reflection of all coded segments, cross-tabulated by the transcription document in which they appear.

TABLE 5.12: Cross-tabulation of the % of codes as a reflection of all coded segments across transcriptions.

Code	Final Response	Keyword	Visualisation	Total
Keyword Preference	2%	3%	1%	2%
Visualisation Preference	6%	3%	3%	3%
Cognitive Overload	1%	0%	2%	1%
Learning		19%	20%	17%
Positive (General)	5%	5%	7%	6%
Relevance	43%	7%	1%	8%
Clarity	27%	11%	8%	11%
Usability	14%	12%	11%	12%
Negative (General)		11%	3%	6%
Relevance	8%	5%	2%	4%
Clarity		6%	29%	17%
Usability	1%	35%	22%	25%

Generally speaking, the qualitative analysis supports the findings from the quantitative analysis.[17] Most users found the data in the visualisation search more relevant to their questions but had a more difficult time understanding the data (i.e. it took more cognitive effort to parse the visualisation). This point supports the ideas in Chapter 2 that visualisations are treated as more cognitively demanding. The keyword search, however, was seen as much easier to understand but made it more difficult to find relevant information. Additionally, while the keyword search had slightly better positive usability comments (12% versus 11% for visualisation), it had considerably more negative comments. Many of these were related to not being able to order search results beyond the default relevance sort, and some wanted a more in-depth default search.[18] In addition, there is a slight edge to what users stated they learned during the visualisation search as opposed to the keyword search (19% versus 14%). Again, this ties back to the comments in Chapter 2, which discusses the importance of visual information and the advantage it provides in assisting individuals with data extrapolation.

There also appeared to be a strong correlation between participant comments about their mode of thinking and their ultimate choice of their preferred method of search. The participants who specifically stated they preferred to start with the micro-analysis of the text through the keyword search before looking for larger

[17] A more in-depth discussion of this analysis will be conducted further on.

[18] It should be noted that a basic default keyword search and advanced search were implemented to mimic many of the common functions inherent in most DREs. Therefore, many of the comments related to a desire for more robust search are comments which should be considered beyond the scope of just this work and should be applicable to all DREs.

trends via the visualisation search tended to prefer keyword searches. Conversely, those who preferred to conduct a macro-analysis first and use that analysis to drill into more detailed results preferred the visualisation search. Additionally, numerous participants noted how well the two search mechanisms complement each other and that they are much better suited to the overall goals of the user when used in concert, as opposed to standing on their own. This is perhaps best summed up by participant 5c0a5fdc9e02299a92d16b52:

> I started to identify the data around the key pieces of information I was looking for. For example, I wanted to use the visualisation tool to get a sense of when more wine was purchased. Then, subconsciously, I was using the keyword search tool to find out more information around mentions of wine in the archive. That's when I began to see other patterns and identify other questions, such as what about abstinence days? How can I incorporate the cheaper rates into my previous visualisations...It reminded me that subconsciously I was using the two tools to address the same theme but my expectations of what answers I would get out of the results modified my research questions to meet the search/exploration tools. [310]

Finally, as part of the overarching research question also concerned itself with aspects of learning, the analysis of the data from both case studies also attempted to determine if there were any statistically significant effects on learning. While there were no direct evaluations of learning outcomes,[19] there was data captured in the post-test questionnaire that specifically asked what the individual learned as part of a particular search method. If a user directly stated a fact or piece of data which was gleaned during a search in response to any of the questions in the post-test questionnaire, this was recorded as an indication of learning. Each participant was assigned a value of either 0 or 1 for each search indicating whether they acquired new knowledge, thus two dichotomous variables were created to indicate that either the participant learned new information during either the keyword search or the visualsiation search, or they did not. If the user was vague in their response to the question—such as "I learned about diet" or "There were 88 results related to 'spice'"—these instances were discarded and coded as 0 (or no knowledge acquired) as the statement was too vague to indicate that there was any real knowledge gained. When adding these variables to the larger dataset and running both Fisher's Exact Tests as well as binomial regressions for preferred method, study type, primacy group, and training type, there appeared to be no statistical significance between these variables and whether the user obtained new knowledge during the visualisation search or the keyword search.

[19] The case study focused on the engagement aspects, due in part to the restrictions around the gathering of participants and an inability to directly leverage the system within a structured classroom setting.

5.4 Discussion: The Impact of the Data on the Research

"Data are just summaries of thousands of stories—tell a few of those stories to help make the data meaningful" [311, qtd. in Chapter 3 opening]. In the previous section, the data collected from the case studies was presented and the methodology behind the analysis was defined. But data needs a discussion in order to be born into relevance. Without contextualisation, the data is spurious at best and meaningless at worst, and so the question remains: what stories do the data tell?

5.4.1 The Story of Search Preference

Perhaps the first and most obvious story within the data is the selection of the preferred method. When looking at the overall selection across both case studies, visualisation is the predominant choice: 60% (21 participants) preferred the visualisation search to the keyword search (40% or 14 participants). However, this breakdown becomes increasingly more interesting when looking at it across research area and study type—the latter of which also reinforces the earlier mentioned classification of participants into inductive vs deductive reasoning types. When isolating the results from the first case study, which contained primarily humanities-based participants, the results are nearly flipped with 58% (11 participants) preferring the keyword search as compared to 42% (8 participants) preferring the visualisation search. When breaking down the preference by research area, History and Library Sciences remain nearly evenly split in their decision; however, Other Humanities (which consists of Digital Humanities, Digital Media, and Classics) and Other Sciences (which consists of Sociology and Data Science) are unanimous in their choices, with the former choosing the keyword search and the latter choosing the visualisation search (see Table 5.13).

TABLE 5.13: Search Method Preference for the Workshop Case Study

Research Area	**#**
Keyword Preference	11
History	4
Library Studies	3
Other Humanities	4
Visualisation Preference	**8**
History	3
Library Science	3
Other Sciences	2

When considering the second case study (which consisted of only Computer Science students as a part of a seminar series), the numbers overwhelmingly skew in favour of the visualisation search, with 81% (13 participants) preferring the visualisation search as compared to only 19% (3 participants) preferring the keyword search. As the analysis suggests, there is a clear correlation between the type of educational training and the preference in search methodology. When considered within

a learning context, while both Science, Technology, Engineering, and Math (STEM) and Humanities generally rely on constructivist views of learning (as discussed in Chapter 1), STEM fields tend to have a stronger reliance upon deductive reasoning—that is, reasoning that derives from general observations and moves through a series of logical deductions to a specific conclusion [312].[20] Visualisations lend themselves well to this type of reasoning as they allow the reader to begin with a high-level view of the data—in other words, a general view of the data—and through a series of iterations within the visualisation itself, reduce the data set down to specific data points which support a logical conclusion.

The Humanities, however, tend to focus on a more inductive approach—that is, moving from specific data points to a more generalised conclusion [313]. In this case, the keyword search more closely mimics the inductive process of reasoning, where a reader is moving from one search result to the next and building a knowledge base that will lead to a more generalised conclusion. Both methods rely upon the *construction* of knowledge, but the inductive mode of reasoning does not expect an absolute truth to evolve from the data—merely a conclusion that is *probable* given the premises of the argument are true. The deductive method seeks more definitive conclusions, thus relying upon a different type of scaffolding during the learning process. Moving from the generalised view of the data provided by the visualisation allows the deductive reasoner to look for patterns and follow those patterns to a logical conclusion.

5.4.2 The Story of Engagement

The analysis of the engagement with the system via the creation of boundary objects also highlights the reasoning differences noted above. When comparing the number of boundary objects created between the inductively-trained participants and deductively-trained participants, the inductive group was nearly 5 times (4.92) more likely to produce a boundary object during the keyword search and nearly 3 times (2.96) more likely to produce a boundary object during the visualisation search.[21] The use of annotation as a learning tool—which is essentially the function of boundary object creation—is more closely aligned to the inductive mode of reasoning as it relies upon making small observations and building to a larger conclusion. While such a tool can prove useful for humanities-trained scholars (or for teaching new scholars how to conduct humanities scholarship), it is less effective for those who

[20] The author acknowledges that this is not a universal fact and that some STEM-related fields may not always rely upon deductive reasoning; however, the general consensus among many of the STEM-related disciplines is a reliance upon the deductive reasoning process inherent in the scientific method and conceptualisation of experimental design and interpretation of results. See Chapter 1 for further discussion.

[21] It should be noted that, while the binomial regression model used to predict the number of boundary objects created was considered significant, only the keyword search proved to be statistically significant in its prediction, $p = .017$. So while the average number of boundary objects created during the visualisation search was nearly 3 times more in the inductive group versus the deductive group, it cannot be said with complete confidence that this is a statistically significant indicator without further testing.

are trained deductively, as it runs contrary to the typical reasoning process of the deductively-trained reasoner.

Additionally, while the keyword search was seen as the preferred option for the inductive reasoners, that is not to say that the visualisation search was unhelpful. Indeed, nearly every single research area group saw an increase in the number of boundary objects created during the visualisation search as opposed to during the keyword search as see in Figure 5.14.

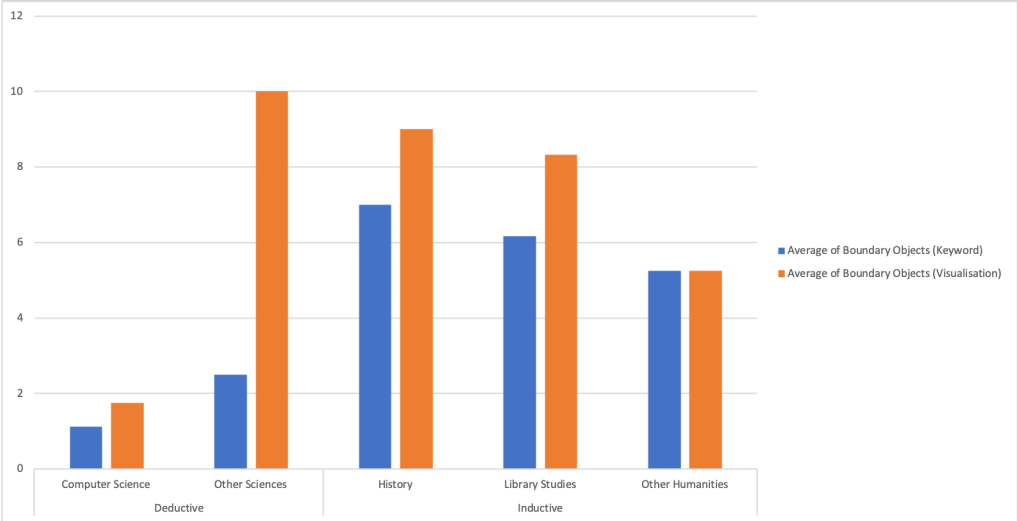


FIGURE 5.14: Graph showing the change in the average number of boundary objects created during each search type across research area and training type. A larger version can be seen in Appendix E in Figure E.15

While a Wilcoxon signed rank test did not determine a statistical significance in this comparison ($p = .133$, see the earlier analysis section for more detail), the change is worth noting. While possible that the lack of statistical significance here could be legitimate and the increases are merely coincidental, it could also be due to the small sample size. As the p value itself is not overly large, a larger study might reveal a stronger correlation. This is further supported when considering the results of the non-parametric ANOVAs detailed above wherein a statistical significance in the variance across means in the number of boundary objects created was found to be statistically significant for *both* types of search across both research area and training. These results bear up the supposition that visualisations can increase engagement with the source text. Therefore the incorporation of this type of search mechanism, in conjunction with the keyword search, can facilitate the learning process irrespective of research area or educational background.

Finally, these types of additional tools can serve as a measurement to check against. As one inductive participant noted:

> When I did the keyword search, I noticed wine appeared more often than bread and meat. But the visualisations showed that while wine was mentioned more, meat and bread accounted for a larger percentage of costs

> - so my perception of wine being the biggest expense from the keyword search was wrong when I realised the visualisations were showing actual total cost. [314]

Other participants noted in their responses the importance of starting with the keyword search and then using the visualisation search to either confirm their hypotheses or to gain a broader understanding of the data after a closer inspection. In this way, the keyword search could be more closely related to a type of close reading approach discussed in Chapter 2. As such, it would explain why inductive participants (many of whom were already familiar with the history of the era or had a background in archival studies or history) found the keyword search more fulfilling: they already possessed the contextual knowledge in which to situate the data received from the keyword search. Conversely, the deductive participants were largely unaware of the historical and cultural context of the archive. Although a talk was provided beforehand in which to help situate the *Alcalá Record Books* within a proper context, such a short contextualisation is likely insufficient for a group of participants to accurately engage at a close reading level. Thus the use of a distant reading approach to gain a higher level of understanding before moving into a more contextualised understanding at the detail level more closely served their understanding and contextualised experience.

5.4.3 The Story of Learning

While the analysis of the learning mechanisms did not show a direct statistical correlation between the training background (or preferred method) and whether knowledge was gained during each type of search, it is interesting to consider the data as a whole outside of its statistical significance. When examining the total number of participants who indicated some kind of learning, the visualisation search saw an increase of 26% for deductive reasoners but a decrease of 18% for inductive reasoners (see Figure 5.15).

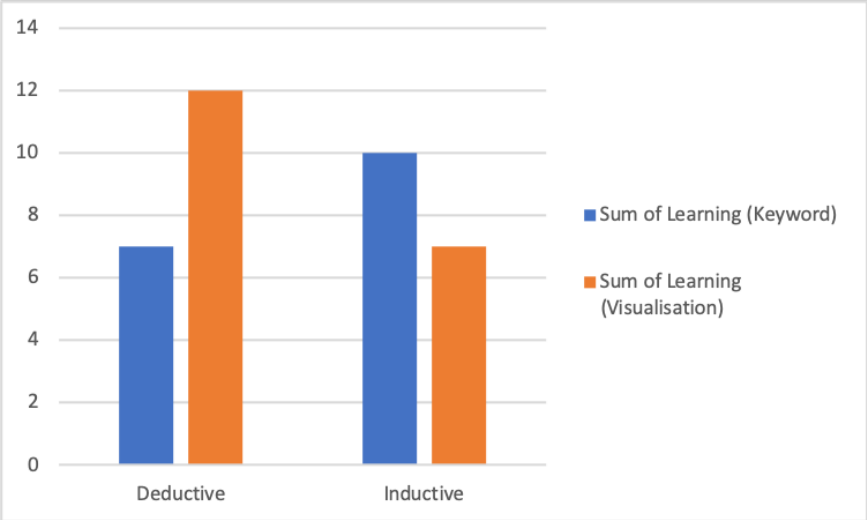


FIGURE 5.15: Bar chart showing the number of participants who indicated acquiring knowledge through each type of search intersected by training type

Interestingly, a nearly identical result can be seen when comparing the same learning metrics with the preferred method selected by the user. Among those who selected visualisation as their preferred method, there was a 27% increase in the total number of participants who acquired new knowledge during the visualisation search as opposed to the keyword search. Among those who selected the keyword search, there was a 14% decrease in the total number of participants who acquired new knowledge during the visualisation search as opposed to the keyword search. See Figure 5.16.

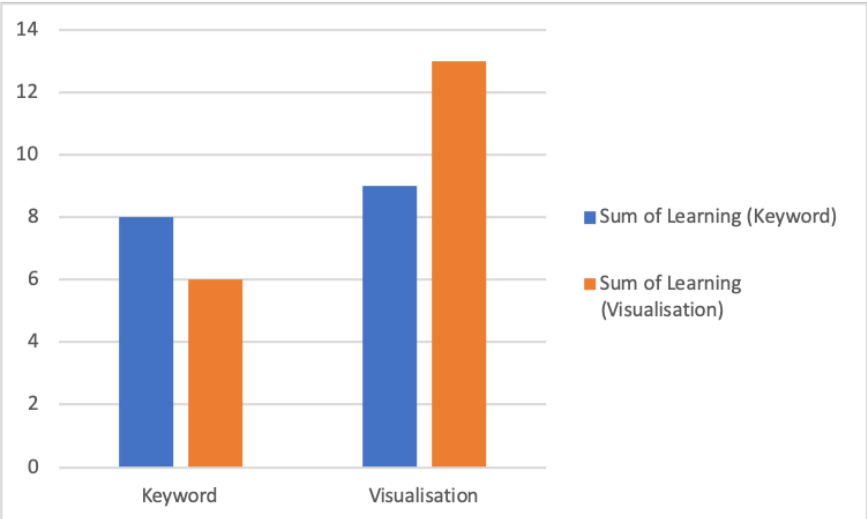


FIGURE 5.16: Bar chart showing the number of participants who indicated acquiring knowledge through each type of search intersected with preferred method

While neither of these metrics are born up by any statistical significance tests, the correlation between the two separate figures nonetheless remains intriguing. At

minimum, these two figures would seem to imply some type of relationship between training type, selection of preferred method, and the effect on learning based on search mechanism. Further research is needed (as discussed in Chapter 6).

5.4.4 Crafting a Truly Interdisciplinary Approach

In its earliest inception, Digital Humanities was born from the desire of Humanities scholars to publish their work in a digital environment. Early pioneers in the field foresaw the revolution that technology would bring to learning and academia as a whole and understood the new medium of the Internet could provide them with new and unique ways to present (and preserve) their data. As the field has grown, it has (as a whole) recognised the importance of interdisciplinary work and collaboration with the sciences (specifically with Computer Science). However, the field remains dominated by Humanities scholars who have dipped their toe in the pool of Computer Science—there is under-representation in the field of actual Computer Scientists doing the work of Digital Humanities. The collaboration and interchange of knowledge has historically been a one-way street: Computer Scientists are being leveraged by Humanists in the essentials of software design and methodology but rarely are Humanists engaging Computer Scientists with regard to the type of inductive, methodological work involved in the deconstruction of a historical manuscript, artwork, works of literature, etc. Conversely, this leads to misunderstandings in the design and implementation of much of the software. Since Computer Scientists are trained to think deductively, they construct websites that facilitate deductive reasoning—a problem which leads to less usable, less immersive experiences for Humanities scholars.

In addition, acknowledging and accommodating these differences in learning is essential when training new Digital Humanities scholars, especially in programmes that are either taught as interdisciplinary at the undergraduate or postgraduate levels, where students may be hailing from diverse backgrounds that may include training in a more deductive format (as opposed to the presumed inductive training inherent in the Humanities). No longer can the field assume only a Humanities-based training on the part of its student population. Additionally, the construction of the tools used must inevitably involve those versed in the Computer Sciences. As a result, if the Computer Scientists involved do not understand the inductive reasoning process—and by extension the needs of those conducting this type of reasoning/research—then the tools built will fail to assist the user in accomplishing her goals. This failure will result in a loss of satisfaction, which will decrease immersion, usability, and both the behavioural and reflective responses to the overall design of the system.

5.5 Conclusion

As digital tools become more ubiquitous in the classroom (and in scholarship), the need for true collaboration becomes increasingly more apparent. Humanists must embrace the methodologies inherent in Computer Science that lead to solid software design (including iterative testing, case studies, and focus group research), including in their grant proposals the time and materials necessary to incorporate this type of methodology. Additionally, Computer Scientists must embrace new ways of approaching their scholarship when applying it to Humanities; it is imperative they remain open to the differences inherent in how Humanities-trained scholars reach their conclusions. If Digital Humanities is to thrive as its own unique field,[22] it must embrace not just the *Humanities* scholarship but also the *Digital* scholarship. While the field has made an effort to be more collaborative and open than most, it needs to recognise that the attempt to marry two divergent types of reasoning is a challenge, and both sides must better understand the other in order to successfully collaborate and create a truly interdisciplinary environment.[23]

The following (and final) chapter of this work seeks to extend this discussion into the wider scope of the research question. While this chapter has focused solely on the case study design, analysis, and discussion of the data, Chapter 6 will discuss the broader impact and propose possible solutions to address the inherent problems highlighted in this chapter.

[22] Awareness exists of the ongoing debate of whether Digital Humanities is a field or merely a set of tools, but such arguments are beyond the scope of this work. As most Digital Humanities scholars consider Digital Humanities a field, this work does not seek to interrogate this particular question and instead chooses to side with the prevailing opinion.

[23] A more detailed plan for achieving this type of interdisciplinary collaboration is laid out in Chapter 6.

Chapter 6

Conclusion: Engaging the Work Beyond the Thesis

The overarching theme of this work is perhaps best encapsulated by Johanna Drucker in her 2014 work, *Graphesis*:

> As a scholarly act, interpretation has almost always been textual, based on close reading, and intimately bound to the graphic form of the work to which it attaches. None of this is exclusively true any longer. [6, p. 180]

While Drucker's observation was intended to address notions of interface and usability design within the Humanities, it also applies to the concept of learning and engagement in the Digital Research Environment (DRE). This is especially apparent when considering the interdisciplinary nature of Digital Humanities, and the inevitable draw of divergent methods of reasoning students and scholars alike bring to the field as it grows and seeks to expand its interdisciplinary reach. No longer can the creators of the DRE assume a user base of exclusively humanities-trained scholars (or even scholars who leverage similar modes of reasoning). As Digital Humanities begins to expand and attracts scholars from fields outside the scope of the traditional Humanities umbrella, the needs of these researchers must be addressed in how the creators of the DRE choose to design tools that facilitate the research and learning process. Such an expansion not only strengthens the contributions of the field as a whole, but can also only occur if the proper tools are provided that meet the needs of a divergent demographic.

The following chapter acts as a final summation of the work. It will open with addressing the question at the very heart of the thesis: "What effect do data visualisations, specifically within the remit of Digital Humanities, have on engagement with the source material—and by extension learning—in digital research environments?" This question will be broken down into its subsequent parts, and each part will be analysed and answered given the data collected and discussed in Chapter 5 and contextualised by the underlying theory discussed in Chapters 1, 2, and 3. Following this analysis, the thesis puts forth a new model for the construction of the DRE and grounds this model in the need for methodological system design and an iterative design approach in order to facilitate continuous feedback through public

engagement. The chapter will close with a look towards the future, listing potential avenues for future exploration and describing some of the alternative paths this work could have taken and the reason(s) these paths were discarded.

6.1 Revisiting and Answering the Research Question

"What effect do data visualisations, specifically within the remit of Digital Humanities, have on engagement with the source material—and by extension learning—in digital research environments?" This question has remained the crux of every chapter of this work. At its core, there are actually three questions to be answered:

1. "What effect do data visualisations have on engagement?"
2. "What effect do data visualisations have on learning?"
3. "What effect do the previous two questions have on Digital Humanities, specifically as it relates to the construction of the Digital Research Environment (DRE)?"

Each of these questions must be answered in order to bring this work to a conclusion.

6.1.1 Visualisations and Engagement

The first aspect of the question focuses on the use of data visualisations and their effect on engagement with the source material. As discussed in Chapter 3, engagement with a system is the direct result of an individual's immersion. While operationalised as the number of boundary objects created for the purposes of the case study,[1] engagement is difficult to measure, as it is largely a subjective response steeped in the reflective aspect of emotional design. However, leveraging both the results of the UEQ tests and the tests for engagement detailed in Chapter 5, a better understanding of engagement for both inductive and deductive reaosners can be theorised.

As mentioned in Chapter 5, the keyword search was the preferred method of search for the inductive reasoners. However, in terms of engagement (using the metric of number of boundary objects created), the visualisation search was the more engaging search mechanism for *both* inductive and deductive reasoners. In both instances, the average number of boundary objects increased when leveraging the visualisation search over the keyword search. This suggests that while the keyword search may have been the more *preferable* mode of search for the inductive reasoner, the visualisation search was the more *effective* mode of search within the context of engagement.

This idea is borne up by the results of the UEQ analysis also detailed in Chapter 5. While the keyword search had the higher pragmatic score (thus implying it was

[1] See Chapter 5 for more information on the operationalisation of engagement.

the easier search mechanism to use), the visualisation search surpassed the keyword search in both attractiveness and hedonic quality, suggesting an overall higher reflective experience for the user, as well as stronger visceral response. As a number of statistical tests indicated no significance in terms of the UEQ analysis and reasoning type, it can be concluded that the use of the visualisation raises the reflective response of the user *universally*, regardless of the type of training or background of the user or their mode of enquiry. It can also be theorised based on this data that the visceral and reflective responses of the user have a stronger impact on the perceived usability of and immersion and engagement with a system than that of the behavioural/pragmatic. This holds with Shih and Liu's assertion that users require an emotional component to the systems they leverage [224].

It also holds with the view of psychology regarding the role emotion plays in the reasoning process. As explained in Chapter 3, those with Capgras Syndrome[2] often have difficulty reconciling the logical outcome of a situation because the emotional connection to a person or place is missing (often to due to some form of brain trauma). Leveraging this example and re-contextualising it within the notion of engagement within the DRE, the relationship between between data (i.e., logic) and engagement (i.e., the emotional response) becomes readily apparent. Even if the data is presented in a usable format, the lack of an emotional connection to the data will lead to a lack of understanding and extrapolation. The data collected from the case study and discussed in Chapter 5 clearly articulates that visualisations increase this emotional impact and thus facilitate engagement with the underlying system. As Chapter 3 discusses, this facilitated engagement also feeds into the scaffolding process with regard to learning as the user becomes more immersed and invested in the system—a concept explored further on. Additionally, the increased engagement also increases the user's perceived self-efficacy, which is directly tied to learning.[3] It can be conclusively stated that visualisations have a direct positive effect on the user's ability to engage with the software.

6.1.2 Visualisations and Learning

As the above section indicates, engagement has a direct effect on the user's learning process, which is why the learning process is described within the context of the overarching research question as an *extension* of engagement. While specific tests for learning were not conducted for the purposes of this work,[4] the post-test questionnaire of the case study did ask each participant to indicate any specific knowledge acquired during each of the searches. While these tests showed no statistically significant correlation, an analysis of the data as a whole did reveal some interesting parallels between preferred method, training, and the total number of participants

[2] See Chapter 3 for further information on Capgras Syndrome and the role it plays in emotion and logic.

[3] See Chapters 1 and 3 for information on the effect of self-efficacy and learning.

[4] See Chapter 5.

who indicated some type of learning during a specific search process. While the data here cannot be used to definitively prove the effect visualisations have on learning given the constraints of the case study (as detailed in Chapter 5), it can be used to theorise the possible implications that the visualisation search can have on the learning process, specifically with regard to the type of training (inductive versus deductive).

In order to further investigate the link between training, learning, and data visualisations, it is useful to consider the working memory system discussed in Chapter 1. In looking back, a discussion of working memory focused on the primary differences between the phonological loop and visuospatial sketchpad, and the role of the central executive in the learning process. When the working memory system is considered in light of the discussions in Chapter 5—specifically with regard to the differences between the deductive and inductive reasoner—a much larger and more relevant discussion becomes apparent. If there are differences in the working memory system for deductive reasoners as opposed to inductive reasoners, then these considerations must be taken into account when designing tools to facilitate learning. Thus this question seeks to address these differences.

The Inductive Reasoner

As discussed in Chapter 1, inductive reasoning relies upon the collection of smaller bits of information and building to a larger conclusion and is a common reasoning approach in the humanities. How the inductive reasoner reaches their conclusions and leverages working memory is an important distinction, as this process informs the construction of the tools needed to conduct their work.

In some of the work conducted around the manuscripts that form the basis for the *Alcalá Record Books* project (as well as the parent archive from the Royal Irish College of Salamanca), historians have extrapolated information about the impact of Irish Diaspora on both Continental Europe and Irish culture during the 17th and 18th centuries [279], [280]. This type of subjective extrapolation of information is the centre-point for humanities research, and an inductive approach to research further facilitates this process.

From a working memory perspective, this type of reasoning tends to rely heavily on both the visuospatial sketchpad and the phonological loop. Often this type of reasoning is born from close reading of a work and is facilitated through *chunking* to scaffold ideas. As described in Chapters 1 and 2, both the visuospatial sketchpad and the phonological loop are used in the close reading process, and the information is integrated by the central executive and then stored in the episodic buffer until it can be committed to long-term memory. In the case of the inductive reasoner, all of the subsystems are (ideally) leveraged to scaffold the learning process.[5]

[5]Obviously, these subsystems are not always optimally leveraged; otherwise there would be no point to this work or any work concerning itself with aspects of cognitive overload or attention in DREs. The point here is to highlight the *ideal* situation where all 4 subsystems are leveraged effectively.

The Deductive Reasoner

The literature shows a marked difference in how deductive reasoners leverage the working memory system. Deductive reasoning relies primarily upon the phonological loop and the central executive to process information, as evidenced by numerous studies that highlight a reliance on the phonological loop as part of the rehearsal mechanism to scaffold the learning process [315]–[318]. While the visuospatial sketchpad still plays a role, the literature suggests a stronger reliance upon the phonological loop in deductive reasoning and scaffolding than upon the visuospatial sketchpad, especially in areas of inspection that are very maths-centric (of which Computer Science is one). This would suggest that the episodic buffer is not as strongly utilised in the deductive reasoning process (as the episodic buffer is primarily meant to store integrated information from the visuospatial sketchpad and the phonological loop). As a result, fewer resources are available in working memory, as the central executive is unable to rely upon the episodic buffer as an additional memory store. The selection of the visualisation method of search by the majority of deductive reasoners in the case study is therefore a logical conclusion: the visualisation engages the visuospatial sketchpad in conjunction with the phonological loop, thus allowing the central executive to leverage the episodic buffer to offload some of the material in the phonological loop once it has been integrated with the data from the visuospatial sketchpad—a notion that also supports the idea put forth in Chapter 2 that the visualisation reduces the cognitive load on the central executive. In both inductive and deductive reasoning, the visualisation assists with cognitive load but in subtly different ways.

Learning in the DRE

When considering the effects each mode of reasoning has on the working memory system, it is easier to understand the larger effect on learning when contextualising this data within a pedagogical framework, such as Laurillard's *Conversational Framework*.[6] Laurillard advocates for the construction of knowledge by building upon the student's pre-existing knowledge and specifically allowing the student to scaffold new knowledge under their own enculturated context (or experience) [54], a notion also supported by the work of Vygotsky [70]. Laurillard's framework concisely explains exactly why the inductive reasoners not only generally preferred the keyword search over the visualisation search but also demonstrated more propensity for acquiring new knowledge during the keyword search. For the inductive reasoner, the keyword search (and the annotation function of the boundary object creation) mimics the close reading process that is core to the training of most fields of Humanities study. Close reading relies upon a scaffolding of detailed facts into a larger generalised conclusion, much as the keyword search relies upon the gathering of smaller portions of data and generalising outward. Similarly, Humanities scholars

[6] See Chapter 1 for more information.

are trained to *chunk* portions of the text they read while also relating these *chunks* to the larger theme of the work. With regard to working memory, such a mode of reasoning would rely heavily on not just the visuospatial sketchpad and the phonological loop, but also on the episodic buffer (which would integrate the information with the help of the central executive), which would act as additional storage space for information to be integrated either directly with long-term memory or with other data currently residing within working memory. When examining the results of the learning evaluation within Laurillard's framework, the keyword search most closely mirrors the inductive reasoners' enculturated experience of learning.

Conversely, the deductive reasoner works from a larger dataset in an effort to reduce results to an observable phenomenon. As a result, the working memory system relies primarily upon the phonological loop and largely eschews the benefits of the additional memory store that is the episodic buffer. This results in more resources being dedicated to temporary storage and offloading on the part of the central executive. Most systems designed for deductive reasoners generally accommodate this restriction through the highly structured nature of most STEM data. Humanities data, however, are much less tightly structured and thus generally require a more robust approach. The visualisation search, therefore, assists these types of reasoners by providing not only a visual diagram from which to further infer information (thus engaging the visuospatial sketchpad and, by extension, allowing the episodic buffer to be engaged as information is integrated from the visuospatial sketchpad and phonological loop) but also a familiar experience given their deductive training. The visualisation environment feels more *familiar* to the deductive reasoner, as it more adequately facilitates the deductive reasoning process. Under Laurillard's framework, the deductive reasoner is able to more easily grasp concepts and gain new knowledge because the mode of presentation is situated within the context of the user's training background and enculturated experience.

The above conclusions regarding the different experiences of the individual are further supported in the literature. Shen notes the differences between a group of Computational Data Visualisation students and Visual Communication Design students [319].[7] As part of a collaborative class focusing on visualisation design, students were asked to rank the quality of communication and state reasons for their score. One of the reasons highlighted by Shen was a "communication gap [stemming] from different problem-solving processes and the overall learning goals of each discipline" [319, par. 2]. Shen's assessment is further supported by other scholars who have sought to incorporate cross-disciplinary models into their pedagogy, many of whom have noted not only the benefits of such models [320], but also the challenges inherent in such due to the differences in reasoning or problem-solving

[7] While Shen does not categorise these students as inductive or deductive, the same methodology behind the categorisation in this work could be applied. In this case, the Computational Data Visualisation students—likely trained in STEM—would be categorised as deductive reasoners whereas the Visual Communication Design students—trained as artists and thus part of the broader Humanities umbrella—would be categorised as inductive reasoners.

approaches [321] and the importance of contextualisation within an enculturated experience [322], [323]. This latter point of enculturation can also have a strong impact on the reading process, particularly regarding how the data is structured and the ability of the reader to comprehend the work, as demonstrated by Fukuoka and Spyridakis [324]. Thus, allowing for data structured in a more deductively-produced manner (such as that of a data visualisation) can actually have an impact on the ability of the reader to comprehend the overall text.

Based on both the data and the supporting literature, it can be theorised that visualisations do in fact have an effect on the learning process; however, the level to which the visualisations can facilitate learning and comprehension is most strongly affected by the individual's mode of reasoning.

6.1.3 Visualisations and the Digital Research Environment

When considering the answers to the first two parts of the question, the final aspect of the question becomes clear: What impact do these visualisations have on the construction of the DRE within the remit of Digital Humanities? Both answers to the previous two questions highlighted the roles visualisations play in learning (for both inductive and deductive reasoners) and engagement (through the emotional and reflective experience of the user). It therefore stands to reason that the incorporation of visualisations is of tantamount importance in the construction of the DRE. From a learning perspective, the incorporation of visualisations may have a stronger impact on those who are classified as traditional deductive reasoners; however, as visualisations have a positive impact on engagement *universally* (irrespective of the individual's mode of reasoning), the incorporation of data visualisations brings a net positive to any DRE.[8]

Furthermore, the type of immersive experience provided by interactive data visualisations is shown to have a positive effect on the learning process [325], [326]. While many of the DREs constructed within the remit of Digital Humanities may not necessarily be geared towards pedagogy (or used within a pedagogical context), they are most often used for research, of which learning is an essential component. Therefore, even solely *research-focused* DREs must consider the benefits of immersive learning through engagement with interactive visualisations, as these benefits lead to a richer, more satisfying research experience.

However, as Chapter 2 notes, not all visualisations are created equal, and the incorporation of data visualisations merely for illustrative purposes lacks the impact necessary to facilitate the user's goals (as was uncovered during the initial discovery phase of this work detailed in Chapter 4). Rather, these visualisations must move beyond the *illustrative* and into the realm of *discovery*. They should serve as an alternative mode of engagement with the source text—an extension and reinterpretation

[8] It should be noted that this work emphatically holds that visualisations should in no way replace the standard keyword search, as the keyword search still proves essential to the learning process, especially for inductive reasoners.

of the traditional search and browse mechanisms inherent in modern day systems. It is the essential component of discovery through interactivity that lends the visualisation credence within both the *ordo essendi* (engagement) and *ordo cognoscendi* (learning).

The role of visualisations and their impact within the remit of the DRE becomes even more relevant when considering the DRE outside the immediate scope of Digital Humanities and the field's scholars, and broadening the scope to the sphere of public engagement. Increasingly, funding bodies—such as the Arts and Humanities Research Council, Humanities in the European Research Area, the European Research Council, and the Irish Research Council—have championed the notion of public engagement and have foregrounded this goal as a requirement for funding. This raises the need for tools that engage the deductive reasoner, as this mode of reasoning is given preference in primary and secondary education settings, primarily due to the rising prominence of STEM and the devaluation of humanities education by the wider public [327]. As a result, the general public is more apt to rely on deductive reasoning techniques than inductive. If Digital Humanities wishes to engage the public, the use of visualisations will not only raise the reflective response of the user due to the visual aesthetics of the medium, but will also assist them with deeper levels of immersion through tools that facilitate their acquired mode of reasoning.

By contrast, if the system is designed with solely inductive reasoners in mind, it will ultimately prove difficult for deductive reasoners to engage. As the deductive reasoner moves deeper into the system, she will likely find herself less and less immersed in the work, as her satisfaction levels with the results will be lower due in large part to a failure of the system to provide her with the tools necessary to accomplish her goals. As the goals fail to produce satisfaction, the emotional response to the overall design of the system decreases across the board: the system is unlikely to provide a strong visceral appeal, the behavioural aspect fails to provide the necessary feeling of satisfaction and goal-completion, and both of these then feed into the final reflective response, which typically leaves the user with the feeling that the system is incapable of providing the answers she seeks. While the conclusion is likely untrue—the system likely does possess the necessary data needed to obtain the knowledge sought—because it fails to leverage tools that assist a different type of reasoning, the system is seen as a failure. This not only effects the individual system, but also has an impact on the wider perception of the humanities. By building systems that enable both inductive and deductive work, these DREs begin to build the cultural caché of the Humanities as a whole within the public sphere. Therefore, the inclusion of data visualisations is of vast importance to not only the notions of engagement and learning but also to the overall success of the DRE.

6.1.4 Summarising the Results

Ultimately, the incorporation of an interactive visualisation-based search has a positive effect for the Digital Humanities and the construction of the DRE. Regarding

engagement, the visualisation-based search has universally applicable benefits as it increases overall engagement with the source material—as defined not only through the creation of boundary objects but also through the reflective and visceral responses detailed in the analysis. In regard to learning, the visualisation-based search has a positive effect on deductive reasoners, providing them with a better cognitive scaffold that more closely aligns with their enculturated experience. While the effect of the visualisation-based search for inductive reasoners does not demonstrate a net positive result, it does not demonstrate a negative result either, and many inductive participants indicated the visualisation-based search was useful as a secondary tool. Thus, this work does not advocate for the *replacement* of the keyword search with a visualisation-based search, but rather advocates for it as a *complement* to the keyword search. Thus the learning of the inductive reasoner is not impeded (and in some cases, augmented through the use of the additional tool), and the deductive reasoner is greatly aided by this additional search mechanism. Finally the DRE as a whole benefits from the inclusion of the visualisation-based search because it widens the audience—specifically with remit to the general public—that can effectively engage with the software. An interactive, visualisation-based search must be considered as an essential component going forward and should be part of any DRE design.

6.1.5 Relating the Results to the Overall Design

When considering the results as they relate back to the original design documentation,[9] each of the goals of the primary user personas are aptly met through the inclusion of the visualisation-based search. The research persona[10] is given a tool which allows her to explore the corpus in a wider manner without reliance upon particular phrases. She can use the visualisation-based search to validate her keyword search results and determine if there is any additional data missing from her analysis (one of the key concerns detailed in her persona). The learner persona[11] is aided by learning how to apply different modes of thinking to the exploration of a corpus. As a student of Digital Humanities, she is gaining knowledge in how the study of historical artefacts can be employed in different, unique ways from that of the traditional scholarship model. Finally, the general user[12] is greatly aided by the visualisation-based search as it is catered more towards his enculturated reasoning process of deductive thinking (as a computer programmer). Additionally, as he is not intimately familiar with the text or time period in a way the researcher or learner might be, he is not reliant upon foreknowledge of the corpus in order to properly explore and search the contents; rather, he can visually explore the corpus and begin to narrow his search through further refinement of the visualisation.

[9] See Chapter 4
[10] See Appendix C
[11] See Appendix C
[12] See Appendix C

6.2 A New Model for the Construction of the DRE

The breakdown of the research question in the previous sections highlights the important role visualisations play, not only in the learning and engagement processes, but also their importance within the wider construction of the DRE. Visualisations also play into the aesthetic appeal of the overall system and sit somewhere between the experienced-based and the techno-futuristic approaches to interface categorisation, as detailed in the work of Udsen and Jørgensen [328]. Thus the notion of the interactive visualisation is not only essential to the research question presented as part of this work, but it is further supported by the literature and the data presented by this work as a method of pushing forward the boundaries of digital research.

However, the design of such systems is still nascent despite their being a mode of enquiry and interest engendered within the Digital Humanities over the last decade. Both Unsworth and Palmer, Teffeau, and Pirmann have noted the importance of establishing basic sets of functionality inherent to the form (referred to as "scholarly primitives" by Unsworth [329, par. 2] and "scholarly activities" by Palmer, Teffeau, and Pirmann [330, p. 4]). Hughes, Constantopoulos, and Dallas note the importance of incorporation of digital methods in the construction of a DRE, as does Warwick, Nyhan, Terras, et al. [331], [332].

Yet despite this awareness and discussion of the incorporation of design activities and a general awareness of standard software development principles and practices with regard to system design *in theory*, rarely are these principles applied *in practice*. A quick review of the Digital Humanities Awards for 2018 illustrates this point. Only the winner of the "Best DH Data Visualization" award, a numismatic-based website visualising coin collections, demonstrated through their documentation an attempt at standard design practices [333]. *Derrida's Library* [334], a nominee for "Best DH Tools or Suites", detailed technical documentation regarding system architecture, but offered no insights into a research-based design approach (or rather none was available from the project's website). The remainder of the relevant projects fell short with regard to demonstrating any methodology behind their design decisions. While these projects may have in fact undertaken such an approach, there is no documentation linked to the project to offer evidence of such. Thus it can be inferred that most of these projects were undertaken in the vein that (in fairness) many software development projects are undertaken (even in the commercial sector)—with a particular goal or idea in mind that may have strong academic merit, but with little research to support the design decisions underpinning the interface.

It is impossible to state (at least within the scope of this work) whether such a seeming lack of design methodology is due to a lack of funding [332], a lack of prioritisation, or a viewpoint that such design methods are "invisible labour" [335, par. 2] and thus not necessary to document as part of the wider methodology of the system. Rivard, Arnold, and Tilton in particular discuss the concept of "invisible labour" and the lack of value placed upon the process of creating the DRE:

> As many Digital Humanities (DH) scholars and workers know, this situation of having a product with little to no knowledge of the labor behind its construction is rather common. We create DH projects for particular audiences, but the decisions and labor behind their construction remain hidden as they are not the end goal of the projects. [335, par. 2].

However, this work would argue that elevating the methodology behind the creation of the DRE and foregrounding the design and methodology of the construction will increase the perception of the work involved and the academic rigour required to generate these resources.

Regarding funding and resources, Warwick, Nyhan, Terras, et al. state outright that the "cost of [conducting user studies]...is considerable" [332, p. 6]. However, leveraging solid methodologies around user analysis and implementing an iterative approach can actually lower the cost of development. Conducting surveys and focus groups is a low-cost endeavour, requiring only a few hours of the researcher's time to design and conduct, and while the analysis of such data can be more time intensive (especially concerning transcription and analysis of focus group content), much of this can be offloaded to less expensive resources or even incorporated into pedagogical settings and used as learning opportunities for students studying the field. If an iterative approach to development is involved (as will be discussed later), then these user studies can be given a narrow focus in order to understand a particular aspect of the system (thus reducing the time spent on the research activity even further). As these various studies are conducted, they can be constantly fed back into the development process, allowing for a refinement of the requirements without gate-keeping the actual construction of the resource.

The importance of a research-grounded approach cannot be overstated. This work alone highlights some of the challenges inherent in system design even when serious consideration is given to the overall design, thus underscoring the need for an iterative approach to system design (as discussed further on). Moreover, the notion of such a methodological, research-grounded approach to software design and implementation is certainly not new to the Digital Humanities. Ruecker, Radzikowska, and Sinclair note the importance of such a grounded approach, breaking the implementation into three phases: design, prototype, and production. During each phase, they engage with a community through user studies to evaluate their ideas [221]. The following section builds on Ruecker, Radzikowska, and Sinclair's work, updating it with more modern, iterative approaches to software design, while also contextualising the important role a research-based approach plays within the ecosystem of grant funding and proposal writing.

6.2.1 Conducting Grounded System Design

While it may be understood within the field of Digital Humanities that system design is an important piece of the larger puzzle, rarely is a proper, grounded methodological design undertaken, as previously noted. This can lead to a number of issues (such as *scope creep*, which refers to the addition of functionality outside the original scope of the project; ineffective designs; and budget overruns); however, as the DRE is often designed within an academic setting, it is equally important from the standpoint of academic rigour to adhere to principles of system and software design borrowed from the field of Computer Science and to document the methods selected and adhered to, justifying these choices based on collected data and the literature (just as all *traditional* scholarly outputs are wont to do). As Digital Humanities breaks new ground in challenging the traditional *publish or perish* model of research and strives to find ways of raising the inherent academic value of the DRE (juxtaposed against the value of the traditional publication), it must embrace the research-based principles of the disparate fields outside the Humanities from which it borrows.

This need is further justified when considering the value of a research-based approach on the learning and engagement process. The design phase of the *Alcalá Record Books* highlighted the importance of a need for discovery beyond standard search and browse mechanisms inherent in most DREs. This research led to the development of the interactive visualisation-based search. However, it was the case study that revealed the differences in modes of reasoning that played a vital role in how the users engaged with various aspects of the system, articulating the inductive versus deductive reasoner's individualised needs. In a larger-scoped project, this data could have then been fed back into the overall design, iterated upon and re-tested to flesh out an optimal tool for research (which is, after all, an extension of the larger learning process and involves many of the same pedagogical principles such as scaffolding, enculturation, contextualisation, and engagement).

Rather than re-hashing the various software design principles, which would involve a wide range of topics from requirements gathering to software architecture (which itself breaks down into multiple sub-domains) to testing and implementation [284], [336], [337], this work instead seeks to lay out the steps necessary for a research-based approach to usability and system design with an eye towards expanding this process to a larger, iterative approach (as will be discussed in the next section).

Step 1: Initial Discovery

While many creators of a DRE may assume knowledge of how a user will interact with a system, such assumptions are often dangerous, even when the primary audience of a system is the community in which the creator resides (i.e., a historian creating a DRE for other historians). Furthermore, if the project is in any way geared

towards public consumption or has a public engagement facet, it is imperative to understand the needs and desires of the larger public audience. Therefore it is worth the creator's time to perform initial discovery regarding the wants and needs of the larger community base. Below are some suggestions for how to conduct this type of research:

- Focus Groups - the most common type of discovery method, focus groups generally consist of 6-8 participants (ideally) and run no more than 90 minutes. Multiple focus groups should be conducted so as to contain a large sample size representative of the communities that will leverage the system. Each group should contain a diverse set of voices, although it is acceptable to run specialised focus groups in an effort to isolate a particular voice. For example, a focus group that consists of an even spread of historians and non-historians can provide for stimulating discussion around the divergent needs of the two groups; however, there is also value in running one focus group that consists solely of historians and another that consists solely of non-historians in order to isolate the distinctive needs of each of these groups. Additionally, a creator could choose to conduct three focus groups which incorporate all three types: a historian-specific group, a non-historian-specific group, and a mixed group of both historians and non-historians. However, regardless of the makeup, the creator should always ensure independence of observations—that is, every participant must be unique to their group; no one participant should belong to more than one focus group.

  In terms of methodology, the focus group should be designed around interactive tasks and should avoid long periods of pure discussion. This is to ensure engagement by all participants so that all voices are heard, not just the loudest. Many of the tasks detailed to Chapter 4 are solid, proven techniques for understanding the needs and desires of a group, but any activities selected *must* be geared towards answering particular questions the creator has regarding how the user community might leverage the system or what needs they may possess. The focus group should be recorded, and once completed, it should be transcribed and analysed using qualitative analysis techniques such as those detailed in Chapter 4.

- Surveys - these can be useful when trying to ascertain more generalised ideas about the target audience(s). Surveys should be short (take no more than 5-10 minutes) and should attempt to answer very specific questions (e.g. "On a scale of 1-10, how important is feature x to you?", etc.). Surveys should also be widely dispersed so as to gain as large a sample size as possible. Because they tend to be more generalised than focus groups, a larger sample size is necessary in order to obtain statistically significant results.

- Prototyping - this method generally involves the creation of some type of prototype of a particular aspect of a system or feature. It may involve paper prototypes, interactive wireframes (such as those created with Balsamiq Mockups or MockFlow), or an actual coded system that is in the early stages of development. Like both surveys and focus groups, the questions need to be focused so as to ensure the creator gathers the necessary data to further enhance the overall design. Typically, this approach involves one-on-one interviews in which participants are asked to perform a particular task or set of tasks using the prototype, after which feedback is gauged, often using a combination of various Lickert scales and free-form feedback. These interviews should be recorded so as to accurately capture the experience of the user for further analysis. Generally, prototyping draws from a small user base as it is a time-intensive process (each interview ideally takes about 30 minutes and the sample size should be a minimum of 10). However, like both of the above, it is important to incorporate a diverse set of voices across a range of backgrounds in order to ensure the creator is capturing all of the necessary data points. While the prototype itself can (and often does) follow on from initial research, such as that described in the previous two points, it can also stand on its own, allowing the creator to test her own original assumptions regarding the needs of the community and use the feedback gained from interviews of the prototype to iterate on the design. Thus prototyping becomes part of both the initial discovery and the design approach detailed further on.

Once the data from the above is compiled, it should be analysed, and user personas should be generated which describe the background of the various user types and their goals and motivations. How this data is analysed will likely depend on the methods chosen for discovery. Focus groups are often an effective means of gaining an in-depth understanding of a segment of the user community and, due to their discussion-based nature, lend themselves to a more qualitative approach to analysis. Surveys, however, offer a broader sample of the population and, if properly constructed, can provide quantitative data that can be used to support design decisions. Ultimately, it is the recommendation of this work to incorporate as many of these methods as possible in order to gain the widest sample of opinions regarding the needs of the user community. It should also be noted that this approach need not be linear in nature. In fact, as will be justified further on, the discovery (and nearly every facet of the design and implementation of the DRE) can and should be iterative in nature.

What remains the ultimate imperative, however, is obtaining the necessary data to craft personas which then drive the requirements analysis. While some may argue that personas are used *in lieu* of user studies [332], true personas are in fact developed from user studies and are based on the demographics, motivations, and needs

of the users that were uncovered during the studies.[13] There are numerous ways of crafting user personas and there exist a myriad number of websites, acting as persona generators, dedicated to assisting with the crafting of a user persona. However, the primary goal in all persona creation is to craft a character (or characters) that will drive the narrative of the system design. As the design progresses, these personas act as touchstones for the team, and any components designed and implemented must be done so within the context of a given persona or personas. In this way, the design is sure to reflect the needs and desires of the user base, as opposed to the needs and desires of the creator.

Step 2: Design

Ultimately, the compilation of data into a user persona as defined above drives the creator towards a narrative design approach. By centring the user(s) as the central character(s) in a story, the design is crafted around the stated goals and motivations of these characters. Such a design approach lends itself towards more immersive experiences, as it tends to build upon the expectations and emotional needs of the user [338], [339], an approach known as *narrative design*, or design through storytelling [340]. Storytelling is one of the most powerful devices available and is inherent across cultures. As Molnar and Kostkova states: "One of the most effective ways of conveying information and learning is through storytelling and narratives" [338, p. 200]. Thus the following discussion of design approach is situated within the context of narrative, as it not only serves as an effective tool but also incorporates the inherent principles of both Goal-Directed Design and Emotional Design.

In undertaking a narrative design approach, the design phase involves both crafting the larger story of the system and the individual stories of the various personas within the software ecosystem. Generally, the approach should begin with small, individual stories (also referred to as *use cases*) focusing on actual tasks or goals the user is attempting to accomplish with the system. Every use case must be tied not only to a persona defined in Step 1, but also to a goal or underlying motivation detailed in the persona. Additionally, these use cases should detail any pre- or post-conditions necessary for the completion of the use case.[14] Finally, each use case should define the actions that the user may take to accomplish the post-condition; these are often detailed in individual steps but can sometimes be more generalised. As a result, small individual stories are crafted and pieced together to construct a

[13] Warwick, Nyhan, Terras, et al. seems to conflate all personas (which can be role-based, goal-based, engagement-based, or fictional) with strictly fictional personas, which are created by the design team based on their own understanding of the user base. This work feels it is important to draw this distinction, as it advocates for *non-fictional* personas, whose creation is grounded in solid research.

[14] A pre-condition would be defined as a condition that the user must meet before being able to execute the use case, such as logging into the system with a username and password, or enrolment in a particular group, etc. A post-condition is typically the end result of the use case, for example "A visualisation is produced" or "A transcription is returned", etc.

larger picture of the overall design. In this way, the narrative design approach is actually reflective of the inductive reasoning approach and should feel more familiar to the typical humanities scholar, who is likely trained in this mode of reasoning.

Once the design documentation is completed, best practice would dictate the creation of wireframes or prototypes. These can then feed back into Step 1 to allow for additional user feedback prior to committing resources to actual development. The benefit here is that wireframes (and the aforementioned design documentation) can all be created without the use of programmers, and can thus mitigate some of the cost of the DRE construction. Once again, this process can and should be iterative, allowing the designs to morph and change in keeping with feedback from the broader community.

In addition to the smaller scale design inherent in narrative design (which focuses on user-centric design model), a larger overview of the system design from a technical perspective must also be undertaken. This is less to justify the design decisions in the User Interface (UI), and more to detail the reasoning behind the technical decisions such as database choice/configuration, design patterns, component architecture, etc. A system design in this scenario generally involves the creation of high-level documentation that defines the various necessary features (or *components*, as they will be referred as going forward) and how the general system will be designed. Choices will be made around the type of database server (RDBMS[15], NoSQL[16], etc.). While the relational model can offer much in terms of data analysis, a NoSQL implementation may better fit the model for most DREs, as this would allow the actual encoded document to be stored and processed (depending upon the database implementation). For example, eXistDB is an XML-based NoSQL database, and if the DRE is encoded using an XML-based schema such as TEI (Text Encoding Initiative), the implementation would be simplified by allowing for queries against the TEI directly. However, each of these systems also has its drawbacks. For example, eXistDB suffers from performance issues when leveraging a Lucene index and the KWIC module. On a small corpus (such as the *Alcalá Record Books*), this may not be a problem, but on a larger corpus, such a performance hit may prove detrimental. Hybrid approaches can also be implemented (such as was the case with *Alcalá Record Books*).[17] It is interesting to note here that this design document then mimics the deductive reasoning process: it begins with a higher level overview of the system, and then breaks down the system into smaller components with details justifying the decisions made. As this type of document is generally geared towards a more technical audience, it follows that it would structure itself in a deductive-friendly fashion.

[15] Relational Database Management System. Examples include MySQL, SQLServer, MariaDB, PostgreSQL, OracleDB, etc.

[16] Refers to a wide range of databases that do not rely upon the relational schema inherent in SQL-based systems. Common examples include MongoDB, eXistDB, IBM Domino, Apache Giraph, Virtuoso, etc.

[17] See Chapter 4 for more information about the design of the *Alcalá Record Books* and the use of multiple types of NoSQL databases.

Thus, both of these types of documents (a narrative design document, focused on a user-centred design approach, and a technical design document, detailing the system architecture and technical design decisions) blends both the Humanities (via the inductive-based narrative design) and Computer Science (via the deductive-based technical design). Ensuring the existence of this documentation and the research undertakings involved in their creation, the DRE establishes itself as a research-grounded output, crafted under the same methodologies inherent to its disparate fields and demonstrating a truly interdisciplinary approach.

6.2.2 Justifying an Iterative Approach and the Need for Continuous Feedback

While the above section justifies the need for a grounded approach, such an approach can be—and when used in Digital Humanities, is often—subsumed under the Waterfall paradigm of software research.[18] This traditional approach is even seen to some extent in Ruecker, Radzikowska, and Sinclair's model for interface design for the Digital Humanities [221]. However, as Computer Science has evolved, it has recognised the importance of an iterative approach to development, leveraging Agile methodologies such as Scrum or Kanban. The primary reason for such a shift—and the reason it is most applicable to the DRE—is that it would allow for case studies (such as those described in Chapter 4) to be conducted in conjunction with development. Aspects of the system could be created and then tested within a larger user community, and feedback could be solicited in the form of surveys. This data could be then analysed (using some of the statistical techniques described in Chapter 5) and changes to the development plan or feature sets could be easily integrated. This would allow for the creation of a more holistic system and would also demonstrate to funding bodies a level of public engagement uncommon in many of these types of projects. A system that has a one to two year development cycle from a funding perspective could run between 2 and 4 case studies during this time period to allow for constant improvements and iterations. Such an approach would prove the relevance of the techniques being utilised while also crafting a more emotionally satisfying system for the end-user, which in turn would lead to deeper levels of immersion and engagement (as discussed in Chapter 3 and borne up by the data presented in Chapter 5).

Continuous feedback would also allow researchers to begin to craft new paradigms with regard to search and browse techniques beyond those of the keyword search or even the visualisation-based search detailed within this work. Based on the research conducted, it would seem that both a more refined keyword search (perhaps with the addition of a faceted search or more granular advanced search options) in addition to a visualisation-based search are necessary to engage multiple types of users,

[18] The Waterfall software design approach stipulates the creation, implementation, and release of software in a singular, linear fashion. Each stage is a pre-requisite for the next and user feedback is not gathered until after the release of the final product [284], [336], [337].

and feedback from both types of search mechanisms indicates a need by the user for further functionality. This suggests that new paradigms are merely awaiting realisation, and the DRE can serve as the vehicle through which these paradigms can be designed and tested. Allowing the DRE to serve as a living laboratory for modes of exploration and learning within a constructed, education-based environment further positions its value not just within Digital Humanities but to academia as a whole.

6.3 Limitations

As with every project, there were limitations that occurred and are mentioned here with an eye towards understanding the constraints of the work. The greatest limitation was in the gathering of case study participants. The original plan was to conduct the case study in a classroom setting, thus granting the opportunity to situate the work within a real-world learning environment and using standard modes of assessment to determine any impact on learning. Unfortunately, it proved difficult to locate a classroom setting within the university with a big enough student population to provide statistically significant results. As such, it was decided to refocus the aims of the study onto engagement and leverage participants within a workshop and seminar setting. The advantage of obtaining data in a real-world setting still remained; however, the ability to measure learning itself became much more nebulous as a result of this shift. While effort was made to identify moments of learning based on the written feedback, further study is needed to ascertain the true impact visualisations can have on the learning process.

A second limitation was the size of the case study in terms of population. With a low population ($n = 35$), the case study was constrained to non-parametric tests. Several data points arose that show a lack of statistical significance (such as gender as it correlates to search preference or the comparison of number of boundary objects created across search preference groups); however these points were often close to the cut-off point for statistical significance ($p = .05$). A larger study may show stronger, more statistically relevant correlations among these various variables.

6.4 Looking Towards the Future

While this work has sought to elucidate some of the ways in which the use of data visualisations can have an impact on learning, engagement, and the DRE as a whole, it is by no means exhaustive or complete. In fact, it serves to lay the groundwork for future work to expand on the findings herein and explore new avenues. This work serves not as an end point, but rather as a beginning—a shifting of the field's understanding of the role data visualisations play and their importance in the learning and research processes. By laying the initial groundwork, this work hopes to establish a foundation upon which future research can be built.

6.4.1 New Areas of Exploration

When considering the findings put forth by this thesis, multiple avenues present themselves for exploration. Perhaps first and foremost would be to further solidify the link between inductive and deductive reasoning and how the user's preferred mode of reasoning affects their interaction with an interactive visualisation-based search. A further case study could be conducted that specifically tests from the outset whether the user leans more towards a deductive reasoning approach or an inductive reasoning approach. This would likely rely upon a collaboration with psychology in order to determine the best set of metrics for determining not only how to test for these two modes of reasoning, but also how to scale them (as it is unlikely that the results of such tests would lead to a purely dichotomous result). Leveraging this data in conjunction with further data collection similar to that detailed in Chapter 5 would allow for confirmation of the findings and further support the need for providing tools that accommodate divergent modes of reasoning.

A second avenue would be to expand some of the work conducted here into types of visualisations more commonly seen in the humanities. The *Alcalá Record Books* was somewhat unique in its construction and utilisation of data visualisations, as most of the data is of a financial nature and thus more tightly structured, lending itself to more *traditional* implementations of a data visualisation such as bar charts, pie charts, and line graphs. Conversely, the type of data (and the relatively small size of the corpus) did not lend itself as well to more common visualisations seen in Digital Humanities, such as those discussed in Chapter 2: topic modelling, vector space analysis, sentiment analysis, etc. This type of text mining tends to produce visualisations that not only need additional contextualisation—a challenge in and of itself—but also present technical challenges when attempting to tie nodes and edges to specific points within the data (as the statistical model upon which these types of visualisations are often built is obfuscates the underlying raw data). Thus, there is a potential Computer Science application here to develop frameworks and/or libraries that would allow for more interactive data visualisations, and for this type of tie to the underlying raw data to be easily exposed to facilitate a more interactive search via the visualisation itself.

Yet a third avenue would be a further exploration of the mechanism(s) for visual search, which was alluded to in the discussion section of Chapter 5. While the visualisation search in the *Alcalá Record Books* certainly facilitated both learning and engagement for deductive reasoners, both the data presented in Chapter 5 and the analysis presented earlier in this chapter suggest a need for more beyond the annotation mechanism and the basic visualised search. As such, an extension of the current visualisation-based search could be developed and tested that would seek to incorporate additional tools to facilitate the learning process for the deductive reasoner. Such a tool would likely allow for the following:

- the customisation of results - this would likely involve allowing the user to

remove specific cases from the visualisation, which may involve being able to see the detail records and selectively remove discriminate values. Such functionality is actually supported by the qualitative analysis, in which numerous participants requested a similar process.

- a log of every customisation - this log would act as a history of the deductive reasoning process, allowing the user to trace her reasoning through her respective customisation(s).

- additional reporting tools - the aforementioned log should also include the ability to visualise its contents in order to further facilitate the deductive mode of logical regression. If the visualisation of the log proves to be cumbersome, at minimum a function should exist that allows the user to export her search process, providing reproducible steps to defend her argument. In essence, this export should act as a model that could potentially be reloaded into the system and produce the same result.

- a regression component - working in conjunction with the log, this feature would provide the user with the ability to trace back to a particular point in her search process and create a branch of the model that allows her to explore additional paths through the data. Such a tool would facilitate a type of branching narrative wherein each branch produces its own unique model, and each of these models could then be integrated and compared.

This tool would seek to address some of the difficulties encountered by deductive reasoners with the inductive annotation process by shifting the process into a visualisation-based approach with the same learning outcomes. Whereas the annotation model follows the inductive process of building from the small to the general, this tool would leverage the deductive process by further facilitating and documenting the movement from general to specific. While still theoretical, this model would serve as a useful subject for evaluation to determine if deductive reasoners can be helped to scaffold information in a Humanities context using a deductive reasoning approach while relying upon tools which facilitate the deductive, rather than inductive, approach to learning.

6.4.2 Alternative Paths

It would be remiss to not also consider the alternative paths this work chose not to take. One such path is seen in the second avenue for further research listed above: incorporating more *traditional* humanities-based visualisations such as topic models, network graphs, etc. As previously explained, these types of visualisations were specifically avoided for the purposes of this work for two reasons:

- These methods are better suited towards larger corpra (or at least corpra with a higher amount of textual content). The data contained in *Alcalá* is small (only

slightly over 300 pages) and is much lighter in textual content, as most of the data is financial transactions—each consisting of only a short description followed by an amount. Attempting to use tools such as topic modelling, vector space analysis, sentiment analysis, etc. would simply be impractical on such a small corpus.

- Many of the aforementioned types of visualisations rely upon statistical models that are difficult to contextualise and accurately replicate. As part of this work was to focus on the need for reproducible search results that could be archived and accurately recreated, it was decided—in light of the previous point—to focus instead on data visualisations that highlight a temporal or financial pattern.

As this work set out to investigate what is a rather new area of enquiry for the Digital Humanities, the idea was to keep the visualisations as simple as possible in order to eliminate additional variables of complexity. That is not to say that more complex visualisations could not also benefit from the ideas presented within this work—quite the opposite—but for the purposes of this initial proof of concept, a more simplified version of the data visualisation was deemed to be the best course of action.

Alternatively, a different type of corpus could have been selected that was higher in textual content and complexity (such as the *1641 Depositions* [273] or the *Letters of 1916* [192]). However, the content of many of these larger corpra are often under copyright restrictions and would have added additional complexity and work to the construction of the environment. As the *Alcalá Record Books* already possessed all of the data in a transcribed format and was owned by the university, it presented the ideal opportunity to circumvent aspects of the DRE creation that would have relied on expertise beyond the means of a single researcher, while also solving issues of copyright restriction as the archive itself is owned by the Russell Library at Maynooth University.

6.5 Final Thoughts

It is well established that no two individuals are alike and thus their preferences related to learning and reasoning are also unique. Even within a dedicated field of enquiry (such as History, Biology, or Psychology), each scholar will approach the learning process differently depending upon their individual preferences and idiosyncrasies.[19] Traditionally within the academy, however, scholars have been able to rely upon a basic understanding of how other scholars within their domain reach conclusions, conduct research, or scaffold the learning process for future scholars based upon established norms within the field and underlying methodologies that inform the academic rigour of their work.

[19] Such ideas form part of the justification for Laurillard's *Conversational Framework*.

The Digital Humanities, however, is somewhat more unusual. While true that it has traditionally leaned more heavily towards scholars from a Humanities background (specifically Literature and History), as the field has grown, it has begun to attract a more diverse set of scholarship from fields outside the remit of the traditional modes of enquiry inherent to most Humanities work. As such, it becomes increasingly more difficult to draw upon a methodological framework from which to not only evaluate future scholarship but also train scholars in its work. The field itself is at a crossroads: it must recognise the differences in both learning and reasoning that exist between Humanities- and Non-Humanities-trained scholars and begin to construct tools that support divergent modes of thought and enquiry. Furthermore, it must establish solid methodologies on which to build and evaluate future scholarship that seek to incorporate this more interdisciplinary approach.

The use of a visualisation-based search is but one method to be further investigated and improved upon, but one which could also offer a solid methodological grounding from which the field can continue to grow. Visualisations assist with the learning and scaffolding process for both the inductive and deductive reasoner; however, *how* these users leverage a visualisation-based search (and other requisite tools for constructing and scaffolding knowledge) will differ based on each user's preferred mode of reasoning. Thus, it is essential to build tools that can be tested and deployed for use across the entire domain of Digital Humanities in order to facilitate this process. In addition, as visualisation-based searches universally improve overall engagement with the system, failing to incorporate such a robust search mechanism is to the DRE's ultimate detriment. The work here has clearly demonstrated that when used in conjunction with the traditional keyword search approach, the reader is ultimately benefited regardless of their enculturated experience. Thus, like distant reading acting as an *augmentation* to close reading, so too is the visualisation-based search an *augmentation* to the keyword search

Standard search and browse mechanisms have constrained the field since its inception. They have acted as a frame within which Digital Humanities constructs its modes of enquiry. Drucker notes that migrating to the digital environment should free scholars from many of the frames that have previously bound the field:

> For scholarly work, the ultimate focus of my inquiry, the relation among frames is integral to the relations of what are traditionally considered text and paratext. In a digital environment, those relations are loosened from their condition of fixity and can be reorganized and rearranged according to shifting hierarchies of authority and priority. A footnote to one text becomes the link to a text which becomes the primary text in the next window or frame, and so forth. [6, p. 157]

While her discussion centres on breaking the frames of rigidity of the page, it can be further expanded to the world of search and browse, which is simply another type of frame that has served the field well but is limited in its application. By expanding the frame of the traditional search and browse and incorporating new modes of

search through the use of interactive data visualisations, the Digital Humanities can construct environments that not only titillate, but also engage a wider audience and offer learning opportunities to scholars and students alike.

Digital Humanities finds itself on the verge of a paradigm shift. No longer is it strictly the domain of Humanist scholars who have learned to tinker with technology. Increasingly, Non-Humanists are joining the field, bringing with them a wealth of knowledge and methodologies that can only serve to strengthen the interdisciplinary nature of the field. Furthermore, by offering tools that assist Non-Humanists with research, the field also exposes the breadth and depth of Humanities research to a larger audience, thus bolstering the importance of Humanities research and allowing it to expand beyond the constraints of a singular field. Data visualisations offer perhaps what is the first step on this journey by allowing deductively-trained researchers to engage more effectively and authentically with Humanities data. These visualisations also offer new ways of teaching the Humanities, and their potential to contribute to learning and pedagogy uniquely positions both educational technologists and Digital Humanists to explore new methods of engagement that could serve to shift the paradigm of digital pedagogy, crafting a new way forward.

Appendix A

Focus Groups

A.1 Focus Groups: Code System

Following is the code system used for the creation of all coded segments for this focus group analysis.

TABLE A.1: Focus Groups: Coded Segments

Code System	Memo
Book Metaphor Disapprove Approve	
6 Part Story Negative Positive	Used to track responses to the 6 part story method
Emotive Desire	This code is used to track the notion of wanting some kind of emotional reaction to a resource.
Visualisation	Anything related to visualisations is coded here as a catch all
Safety	Used for anything related to discussions of safety and security
Problems Lack of Tangible Material Findability	 Identifies mentions of the problem of materiality. Some people want a more material or tangible experience than the digital can provide. Problems related to being able to locate information in a resource

Continued on next page

Table A.1 – continued from previous page

Code System		Memo
	Readability	Relates to problems with readability (visualisations or otherwise). This also encompasses issues with design (in terms of intuitiveness)
	Lack of Context	Relates to a lack of context—typically this involves visualisations which lack any kind of contextualisation; however this could also be related to help files, directions, etc. for the resource itself (non-visualisation related)
	Transparency	This code relates to anything regarding the transparency of the data itself: methodology, contents, source, etc.
	Technical Limitations	Problems related to technical limitations such as lack of high speed connectivity, storage of large data formats, cross platform interoperability, etc.
	Accessibility	This is related specifically to comments regarding accessibility issues such as visual impairments, hearing impairments, colour-blindness, lack of literacy, etc.
	Pay walls	Problems related to having to pay for information and the cost-prohibitive nature of such.
	Cognitive Overload	The "rabbit-hole" scenario
Reasons for Use		
	Confirmation	Specific to those situations where a user leverages a DHP to confirm information they already (think they) know. Sort of a secondary source.
	Engagement	This may be engagement with the public or with a particular community (knowledge dissemination or feedback solicitation, etc.)

Continued on next page

Table A.1 – continued from previous page

Code System		Memo
Collaboration		This is specifically related to leveraging a resource to promote collaboration. This may be through crowd sourcing (similar to letters of 1916 contribute functionality) or allowing other communities to somehow be involved in the creation and curation of the data. NOTE: This differs from engagement (which typically includes user forums) where you are attempting to use the resource to engagement with the community either for feedback or knowledge dissemination
Preservation		Using a resource as a preservation mechanism
Access		Gaining access to materials which may not be readily available in the analogue (or to use the resource due to its convenience for access)
Discovery		Leveraging the resource in order to discover new information
	Analysis	Used for anything that facilitates an analysis of the information (i.e. comparative analysis)
	Learning	Using the resource to learn about a new subject. This is slightly different from Discover which is related solely to finding new information but not necessarily using it as a resource to promote active learning.
	Pattern Recognition	Using the resource recognise patterns in the data. Strongly correlated to visualisations.
Emotive Reactions		This code and its sub codes are used to track specific emotional reactions to either a particular resource or to an aspect of interaction. This is different from a desire for an emotive interaction.

Continued on next page

Table A.1 – continued from previous page

Code System	Memo
Neutral Negative Positive	

A.2 Focus Group A

A.2.1 Focus Group A: Transcription

Following is the transcription for Focus Group A.

[00:00:02.15] IANC67EDF278: Ok we're going to go ahead and get started. First of all I want to thank all of you for agreeing to attend this focus group for me. This is part of my dissertation, part of my research to support my dissertation. Just to give you a little bit of background, my background is in interaction design and computer science. I am specifically looking at what I have dubbed "Digital Humanities Projects" or "DHPs". These are things such as digital scholarly editions, thematic research collections, digital archives, basically any of these kinds of electronic projects that we create within Digital Humanities that are very text heavy, that involve a lot of reading and interaction with multimedia. So they are going to be primarily history based and literary, literature based. So just a couple of things. Just to remind everyone of the group rules that you signed. They are pretty straight forward. Be respectful of everyone's opinion, don't interrupt anyone, let people make sure they finish their thoughts. The most important thing to remember with a group like this, the whole point of it is to generate ideas and think really big and outside the box. So there is no wrong answer no matter how off the wall or how crazy you think it might be. Don't be afraid to give voice to it. Everything has value because I never know what one person might say that might spark something else or even give me an idea. So the whole point of this is just to generate ideas and generate content. So do not be afraid to say anything. This is a safe space.

[00:01:41.24] IANC67EDF278: So at the end of the meeting, I'll be collecting any kind of paperwork that we generate outside of the consent forms. So we'll be doing some drawings, different kind of visual things. Don't be afraid if you aren't an artist. I'm by no stretch of the imagination an artist. I can't draw to save my life. So don't be intimidated by that. I will be collecting them so that I can scan them and analyse them as part of the study. Again, just kind of going back, there is no right or wrong. Just express yourself. That's kind of the whole point of this. And I will be referring to all of these as DHPs just so everyone knows what a DHP is when I refer to it.

It's kind of a term I made up, since there's really no one term to grasp all of these different things.

[00:02:36.13] IANC67EDF278: Your information is protected. We're not dealing with anything sensitive but from an ethical perspective, it is best for me to ensure everyone's anonymity. So if anything that I can out of this, if I publish any quotes or anything like that, no one will be, will have their identity exposed. Everyone will be de-identified in my analysis. If you have any concerns about any of that, please do feel free to contact me. If you decide after the fact that you don't want any of your quotations or anything being used, by all means let me know and I'll make sure none of those get included in my analysis. And I can't pull out all of your information. The only thing I can't pull out is your actual image from the video recording that I'm taking. However, that will not actually be published. It will simply go into my archive as part of my research and the data that I have to give over at the end of my study.

[00:03:30.11] IANC67EDF278: I'm hoping you all got a chance to go through some of the links that I sent you. We're not going to be looking heavily at specific sites. These were to give you an idea of the kinds of things we're going to be exploring today. So whenever I refer to some of the things I want you to think about or generate discussion on, we're going to be referring to these types of websites. Things like the Letters of 1916, who I think most of you here are familiar with: the Chymstry of Isaac Newton, the Burchardt Source project, the Woodman Diary, which was also created here at Maynooth. And then even going into some of the more esoteric data visualisation stuff that we see in Mapping the Republic of Letters or in Kindred Britain.

(ACTIVITY) ICEBREAKER

[00:04:11.21] IANC67EDF278: So the first thing we are going to do—we're going to do something a little bit different. Because of the whole promise of anonymity, I'm going to have everyone go around and say their name. If you want to tell people your names, you are more than welcome but I'm not going to make everybody. However, we are going to do a little bit of an icebreaker. Take a piece of paper. I want you to write down a word or phrase that when you think of these kinds of DHPs, what is something that you want to see or be able to do with them that maybe you can't do, or something that you like, or even something that you dislike. Just one word or phrase that comes to mind when you think about DHPs. Go ahead and write it down on that piece of paper and then put it facedown in the middle of the table.

[00:05:11.07] PAG7889FCFB7: What you like or don't like? Just one thing?

[00:05:13.00] IANC67EDF278: Just one thing. Anything that comes to mind.

[00:06:19.24] IANC67EDF278: Has everybody had a chance to write something down?

[00:06:33.26] IANC67EDF278: So I'm going to shuffle them up. Hopefully you don't get your own. Ok. Now, everybody go ahead and flip them over, and we are

going to go around the room and you're going to read out what your card says. So who wants to start.

[00:07:16.04] PAGC07A5FEDA: More visual stimulus to create engagement by a wider population.

[00:07:21.08] IANC67EDF278: Ok

[00:07:22.29] PAGB260EA6ED: Limitations, "question mark"

[00:07:26.11] IANC67EDF278: Interesting! Ok

[00:07:29.02] PAGA54E9A3C8: I would like more DHPs to look nice as well as having informative and primitive...something?....they can do.

[00:07:45.07] IANC67EDF278: Ok. Does someone, whoever wrote that, want to clarify?

[00:07:51.17] PANF10C348AA: Yeah, my writing's terrible. <LAUGHTER> I would like more DHPs to look nice as well as being informative and to prioritise UI.

[00:08:03.26] IANC67EDF278: Ah, ok. So prioritising the UI.

[00:08:08.10] PAG7889FCFB7: Access

[00:08:11.14] IANC67EDF278: That could go either way, right? That could be a good thing or a bad thing...

[00:08:15.25] PANF10C348AA: Pull data of different kinds, instantly.

[00:08:19.29] PAN4A69B0FD3: Limitations for people who have poor visibility or who are blind.

[00:08:26.10] IANC67EDF278: That's a good point. So thinking about, a lot of times when we talk about access, we talk about access to material but we also don't think about accessibility issues. So that's a really good point.

[00:08:40.29] IANC67EDF278: Ok so kind of dovetailing off that then, we're going to go into a little bit of discussion for a few minutes where I just kinda want to let you guys free form discuss amongst yourselves and with me. So I guess starting "Why would you choose to use something like this, a DHP?" Whether it be Letters [of 1916] or Mapping the Republic or the Woodman Diary. What about them is it that interests you that would draw you to look something like this up or use it for some reason.

[00:09:13.29] PANF10C348AA: My assumption just from using these projects myself is that they're <AUDIO UNCLEAR>.

[00:09:27.21] IANC67EDF278: Ok. Find new things?

[00:09:30.24] PAGA54E9A3C8: I use the Letters [of 1916] because I think that especially these more type of political <AUDIO UNCLEAR> people were writing their feelings and emotions in real time. And through this we have access to their thought process in real time a 100 years ago.

[00:09:54.22] IANC67EDF278: So almost like a window into a real life experience that we can't really get any more. That's fantastic.

[00:10:01.15] PAGB260EA6ED: I would be going on the same lines as this lady [PAGA54E9A3C8] because it's a real time snapshot. I come from a material, cultural history so I'm kind of picking things that are not captured in kind of traditional

histories. So like you might look at in the Letters [of 1916] in particular I like the fact that I can put one word in, maybe bread, which is really random from a history point of view, but kind of looking at everyday items that people are kind of discussing. You know instead of necessarily political history but like a kind of social history behind the actual political history.

[00:10:44.15] IANC67EDF278: So how their day to day lives remained affected by these really monumental moments.

[00:10:49.21] PAGB260EA6ED: Yeah so like what you would consider to be mundane things but they say so much about [an] historical period of history.

[00:10:59.15] IANC67EDF278: Anyone else?

[00:11:02.28] PAGC07A5FEDA: Yeah, I'm a casual user of these kinds of things myself and I would probably go into them just to get interest in something. Or alternatively I could be interested already. Like I've used a resource dedicated to Tolstoy's War and Peace that has multiple layers: time lines, relationships. And you know after you've read the book or are in the process, you want to dive more into what's going on there. I guess that sort of resonates with what you all said. And with the Irish letters [Letters of 1916], I have no idea because I'm not coming from Ireland but then I go in and I can get to more than a stack of letters; I get more of a picture. And that kind of interests me, like a movie or someone telling me about it.

[00:11:47.18] IANC67EDF278: Ok. So a different way of viewing history.

[00:11:51.05] PAGC07A5FEDA: Yeah or literature

[00:11:52.22] IANC67EDF278: Or literature. In a more interactive kind of way.

[00:11:55.11] IANC67EDF278: Anyone want to add to that any more?

[00:11:59.18] PAN4A69B0FD3: These projects are essentially very practical too. From an institutional point of view, putting it into an online resource, albeit it's expensive, it is easier to send people to a website than have them clogging up your reading room or your archive or what have you, which is much harder to add a new room than it is to build a website and stick the resource up there. But there's a counter to that too because if you shove stuff [like that] online with no context to it, all you are really doing is ticking a box. And the worry is that, from a historical point of view, the archives become more and more digitised and there's less and less context, you end up shifting through loads and loads of stuff trying to find it without the tools to do it.

[00:12:39.11] IANC67EDF278: Ok. Have any of you ever tried to use visualisations—not all of these [DHP examples] have visualisations built into them. It's actually something the Letters [of 1916] hasn't incorporated yet. But things like Mapping the Republic of Letters, which is kind of a prime example of this kind of thing. <MUTTERING> I always forget where the link is....

[00:13:07.24] PAGB260EA6ED: This is the reason I couldn't access the link on that. So I haven't seen how that uses the visualisation.

[00:13:21.14] IANC67EDF278: I always get lost here. Ok that's right. <Looking at Mapping the Republic of Letters>. So let's say we come in and look at Voltaire. We

can see kind of the different data visualisations of how his correspondence network, these types of things. Now these are just screenshots. The actual visualisations you can drill into. Now have any of you used anything like this, from other websites or I should say other DHPs?

[00:13:50.28] PANF10C348AA: With the 1641 depositions I was using during last years, and there's a network visualisation function with the, oh what was that consortium called, I think it begins with an "e", yeah they had like an optimist thing where you could see network visualisations. It didn't look as nice as [Mapping the Republic of Letters] cause I think I've seen that whole visualisation play and it's absolutely spectacular, but I'm not getting a huge amount of information from it.

[00:14:18.28] IANC67EDF278: Do you find the visualisations helpful? Or are they a hindrance?

[00:14:23.05] PANF10C348AA: You have to really know what is your looking for already. Going into it, trying to find an argument...not so much, from my limited experience.

[00:14:31.19] IANC67EDF278: So you find it difficult to read?

[00:14:32.27] PANF10C348AA: Yes. Yeah, I mean so much of it depends on the design.

[00:14:37.07] IANC67EDF278: Right right. Anybody else?

[00:14:40.21] PAG7889FCFB7: Yeah the transatlantic life trade, that database, there's some very good data visualisations of that but when you speed it up you see the enormous impact it had. You're just seeing ships. You're seeing people. But you're seeing data as a...everyone was accounted for as one unit of cargo. Important people. So when you see that amount of cargo moving, you know, around the triangle, but when speeded up, you don't see it until you see it over time, 20, 30, 40 years, impact. I think that one was quite good. And the Texas Slave Project did a lot of visualisations and stuff. I think they really tried to...they definitely tried to show something. But they went through all the ads and then they started to make visualisations of the ads of runaway slaves. It's not until you see some of their visuals that you realise that they impact they had, the historical impact even and on the population itself. How it moved about. Sometimes when you read something, it doesn't hit your brain until you see it in transit.

[00:16:03.02] IANC67EDF278: So in those two examples that you gave, would you say it really brought the data to life for you in a way that you couldn't get from just reading the content or the website?

[00:16:13.26] PAG7889FCFB7: You never get through the transatlantic site—we've done that. You'd never get through the amount of queries that you could do in their database. It's phenomenal, and don't forget it's all ship logs. Everything is cargo, is unit, where it starts and where it finishes is not necessarily the <AUDIO UNCLEAR> <LAUGHTER>, you know there's a lot of politics you know, with the high seas and so on. But I think when you see the data moving as an action over a period of time,

like spatial history, you know and across a geographical space, it's definitely, it's mind-blowing.

[00:16:56.00] IANC67EDF278: So it definitely gives you a way to look at it that you couldn't.

[00:16:59.09] PAG7889FCFB7: Huge impact on how you understand that type of data.

[00:17:04.23] IANC67EDF278: Hmm ok. That's very interesting. Thank you. I haven't actually look at those...

[00:17:09.07] PAGB260EA6ED: Sorry, I just couldn't hear. What was the name of the transatlantic site you were talking about? The transatlantic site?

[00:17:16.24] PAG7889FCFB7: The Slave Trade Database

[00:17:19.03] PAGB260EA6ED: The Slave Trade

(ACTIVITY) CARD SORTING

[00:17:23.07] IANC67EDF278: Ok. Alright the next thing we are going to do then, we're going to do a little bit of a new exercise. So I'm going to give you some post-it notes. And I want you to list all of the different things you would like to be able to do with these kinds of projects, the DHPs. Anything at all, as mundane or as crazy as you can think of. I want you to create, I want you to write just one word or phrase on each post-it note.

[00:18:02.02] PAG7889FCFB7: IANC67EDF278, is this just with the visual projects or with the digital humanities projects...

[00:18:06.07] IANC67EDF278: About anything. The Digital Humanities projects as well. So one word or phrase per post-it note but as many post-it notes as you want for the next 5 minutes or so.

[00:18:22.13] PAGA54E9A3C8: How you would like to use it....

[00:18:28.13] PAGB260EA6ED: Do you want to know which of those...

[00:18:31.18] IANC67EDF278: No just any...cause it's general. So any DHP at all, anything you would like to do. Whether it be beaming the information directly into your brain or stepping into a virtual simulation of something or being able to search and browse. It doesn't matter how mundane or crazy you want to get.

<5 MINUTE BREAK WHILE POST-IT NOTES ARE CRAFTED>

[00:23:03.07] IANC67EDF278: Ok go ahead and finish writing your final thought. And gather up all of your notes and bring them over to this front table. Just lay them out on the table.

[00:23:44.23] IANC67EDF278: So what we are going to do now as a group, I want you guys to read everyone's responses and as a group decide if we were going to implement these and you didn't have to think about technical limitations, which ones would be the most important and then which ones would be the least important. So we will divide the table down the middle and all the ones you think are most important should fall over here and all the ones you think are less important should fall over here. So this is the side we want for things that we think absolutely need to happen. And if everything ends over here, then fine.

[00:24:22.13] PAG7889FCFB7: So priority 1, priority 2.

[00:24:23.28] IANC67EDF278: Priority 1, yeah yeah. There you go. That's a good way to put it. Thank you.

[00:24:28.11] PAGB260EA6ED: So we need to agree in unison. <LAUGHTER>

[00:24:33.08] PAN4A69B0FD3: Let's hope we all want similar things.

[00:24:37.13] IANC67EDF278: So go ahead. You've got about 5 minutes to work on this.

<GROUP DISCUSSION> <AUDIO UNCLEAR>

[00:25:15.12] PAGB260EA6ED: A feel for the original...

<AUDIO UNCLEAR>

[00:25:19.28] PAGB260EA6ED: Yeah it's a feel for the original text, the archive materials, the source materials. How, I have no idea.

[00:25:28.06] PANF10C348AA: So that would be like digitised version...<AUDIO UNCLEAR>

[00:25:30.27] PAGB260EA6ED: It's just like the uh, the Woodman Diaries, like the layout of the calendar is just like any kind of standard calendar you'd see. Where I'd kind of like to see flowery pages or something that gives you a feel for the actual source material as well. Instead of just like a page of the text. I'm kind of expecting something I'd see...<AUDIO UNCLEAR>....the diary entries. I'm just a really kind of textual, kind of visual person. So I guess when I go to the library to see and feel the source material which obviously with digital is difficult. I don't know how we would do that. It's kind of difficult. Especially for kind of children or people who aren't used to working with source material like that, we get like history...

[00:26:28.13] PANF10C348AA: Yeah that's pretty important.

[00:26:32.20] PAGB260EA6ED: Yeah, builds a picture of an environment for the mundane aspects of the literature.

[00:26:40.10] PAN4A69B0FD3: Is that a visualisation?

[00:26:45.29] PAGB260EA6ED: <LAUGHTER> I'm so visual it's going to be difficult. I'm just interested in the history, I suppose that's what I describe this for. I want to be able to group themes together from simple search. So just like being able to search for more words so I can see the whole theme of one word of that being used to define. So I can say before like putting in "bread" in or "gold" or whatever from the [Chymstry of Isaac Newton] and getting like a list of ... the [Chymstry of Isaac Newton] is better than the Letters of 1916 because when do a search function in the Letters of 1916 you get all the letters. You don't get that word in context of the sentence. And the context of the sentence is probably more interesting to people. Rather than just getting a random title, you don't know if you're going to be interested in that. Whereas the word in a sentence...

[00:27:41.09] PAGC07A5FEDA: Sure

[00:27:41.24] PAGB260EA6ED: It gives you context to look further.

[00:27:43.07] PAGC07A5FEDA: Yeah yeah.

[00:27:45.00] PAGB260EA6ED: <AUDIO UNCLEAR> The archive systematically layer by layer, starting with a paired down UI as per Letters of 1916 so it's not like overwhelming people with like all this written information. But something simple that you can kind of get more engaged in the further you go into it.

[00:28:06.13] PAGC07A5FEDA: I think that's actually a great point. That's something I was thinking of as well because you know with <AUDIO UNCLEAR> you want somebody to guide you through it. And I rally think that it should have it because this isn't like an <AUDIO UNCLEAR> archive where you see all this stuff and you don't know where to start. And since this is, well it's also for researchers but also for people who just stumble upon it. You have to take them by the hand and at least make some path to...for them to explore and after that they can go on their own.

[00:28:41.24] PAGB260EA6ED: I mean researchers are used to digging through things. But to actually capture somebody's imagination you need to start gently I think. At least that's just what I was thinking.

[00:28:52.11] IANC67EDF278: Ok three more minutes.

[00:28:52.22] PAGB260EA6ED: Oh gosh. <LAUGHTER> Sorry! Don't start with me.

[00:28:57.24] PANF10C348AA: There's a lot of overlap in terms of what people have. I had something very similar. If possible to have some kind of expert testimony in terms of creating linear experience through the archive....

[00:29:09.16] PAGB260EA6ED: So that's this.

[00:29:11.01] PANF10C348AA: Yeah pretty much.

[00:29:11.02] PAGB260EA6ED: So we'll stick those two together.

[00:29:13.02] PANF10C348AA: It feels less um....

[00:29:14.22] PAGC07A5FEDA: We can probably stick this one as well. Get suggestions from the UI on what I can explore next. So...

[00:29:21.12] PAGB260EA6ED: Is there anybody else?

[00:29:23.02] PANF10C348AA: See other people's highlights and bookmarks related. I think that's in the same category and I think that's useful too.

[00:29:29.26] PAGB260EA6ED: So it's kind of search functions. But it's more kind of like...

[00:29:35.02] PANF10C348AA: Like the way kindles have the "most highlighted" options on things.

[00:29:36.27] PAGB260EA6ED: Yeah

[00:29:37.17] PAN4A69B0FD3: Mine's about more context which is kind of similar to what I'm saying

[00:29:41.21] PANF10C348AA: Less flashy graphics.

[00:29:43.06] PAN4A69B0FD3: Yeah

[00:29:46.23] PAGA54E9A3C8: Yeah we can probably add these two together as well. These are mine. One's like learning and teaching so they're kind of the same thing really!

[00:29:56.03] PAGB260EA6ED: Is there anything about collection policies?

[00:30:06.15] PANF10C348AA: Can I correct the amount? I'm going to put one of mine over in the less important column.

[00:30:10.26] PAG7889FCFB7: I think this is the relative to not just building on other projects. It's politics and everything. The why, the how, the when. Why that stuff is in that collection is very important as to why other stuff is not.

[00:30:34.10] PAGC07A5FEDA: Yes yes.

[00:30:36.05] PAG7889FCFB7: And then the second part of that is then how is DH mentoring more collaborative, more open access. People don't share their methodologies in creating the resource, which I would have an issue with. Secrecy around that type of thing. If you're going to call yourself a Digital Humanities project, you better be able to back it up with the policies that their claiming.

[00:31:00.10] PAN4A69B0FD3: Reproducibility is in there [as well], isn't it?

[00:31:02.11] PAG7889FCFB7: I need more. I need more from the resource. Not just from the materials of the resource but from entirety of the resource. I need more information. I need more stuff. That for me is my big problem with DHPs at the moment. And the second thing that I have a problem with is being able to download metadata in bulk. If their giving a metadata record for each single thing, why can we not just get a .zip folder for everything that we want?

[00:31:31.24] PANF10C348AA: So a complete bibliography...

[00:31:32.27] PAG7889FCFB7: Creates enormous amounts of work for people who want to use visualisation tools with XML records. So I'd love to see that. That would save a lot of time.

[00:31:42.07] PAN4A69B0FD3: I think that's very important. Let's put that over [in the important column].

[00:31:45.19] PAG7889FCFB7: Give me your stuff.

[00:31:47.28] IANC67EDF278: Ok how you guys coming? Do you have everything mostly....

[00:31:51.16] PAN4A69B0FD3: I had one about ocular. Everything is visual but none of them are audio based or for people who don't use your vision and that's a real problem. There are some audio archives but even they are problematic because they use so much text visualisation <AUDIO UNCLEAR>. But that's hard to do with the technology we have. We are both limited by the hardware, software, and the fact that everybody who codes uses their eyes. There's not a lot of blind coders. There are a few but there are very, very few.

[00:32:22.29] PAGB260EA6ED: Yeah I have a hearing problem so I'm coming from the visual side. You're coming from the hearing side...

[00:32:28.25] PAG7889FCFB7: You have a great point there. The government is introducing a bill on copyright. And they've amended the copyright to include I think the Marrakech Treaty which is now all fundamentally based on access. And their going to be talking about sources, somebody who is blind needs to access a

Digital Humanities project who has been funded by tax payer's money. They should have the option of audio.

[00:32:54.23] PAN4A69B0FD3: That's been in law any ways since '05. They just don't listen.

[00:32:57.14] PAG7889FCFB7: But it will. I think that's moving into legislation at the moment.

[00:33:03.12] PAN4A69B0FD3: I certainly hope they'll do it.

[00:33:06.03] PAG7889FCFB7: Yeah yeah, it's top order.

[00:33:08.28] PAGB260EA6ED: Is there any more access ones?

[00:33:11.07] PANF10C348AA: Create your mobile <AUDIO UNCLEAR> DHPs certainly fall under that.

[00:33:15.16] PAN4A69B0FD3: That's more about showing them on your phone. You could build a mobile one but it's like "why can't it scale down" and things like that?

[00:33:23.26] PAGB260EA6ED: Alright ok, so that's kind of like an aside from this one?

[00:33:27.29] PAN4A69B0FD3: Kind of connected, yeah.

[00:33:29.09] PAGC07A5FEDA: I think search and find kind goes here because it doesn't make much sense to not search and find the resource.

[00:33:35.29] PAN4A69B0FD3: Actually this one kind of goes into your idea of more information, the idea of clustered projects. So if you can link them back together at the backend, so we can get all the metadata you are talking about. So if this is standard, why can't we connect this to the other standard projects? Can we link all these at the backend? Can we actually look at how...

[00:33:53.19] PAG7889FCFB7: Have we the right to? Have we, you know, copyright? We aren't looking at anything that hasn't been patented, etc. etc. And if you're going to call it, for me, a digital humanities project and you're going to stand by collaborative, open access, open source and all of this. If you're going to stand by it, you have to give us guts. You have to give us the guts in order to prove the guts. So yeah I think that would be great to get into the backend of something.

[00:34:21.20] IANC67EDF278: Alright so let's stop it here. We could do this all day! You guys have generated a lot of really interesting stuff! So kind of...

[00:34:31.04] PAGB260EA6ED: We're not sorted yet at all...

[00:34:33.05] IANC67EDF278: Are you doing like a spread?

[00:34:34.15] PAN4A69B0FD3: We're sort of leaving things over there and more than getting rid of stuff.

[00:34:38.27] PANF10C348AA: Maybe the stuff that more people came up with that could be more democratic...

[00:34:43.18] PAGB260EA6ED: These are just little groups

[00:34:45.02] IANC67EDF278: So are these all different things?

[00:34:47.01] PAGB260EA6ED: These are similar things. That's a similar thing with an offshoot. And this is....<GROUP DISCUSSION> oh no, but I'm not doing it in priority. I'm just doing groups.

[00:35:01.24] PAN4A69B0FD3: Mother, child, and cousin.

[00:35:04.20] IANC67EDF278: Ok I just want to be able to read them.

[00:35:09.05] PAGB260EA6ED: Search and find. Oh no that's switching between different modes without having to search separately.

[00:35:18.00] PAG7889FCFB7: I don't think that's as important as some of the others. [I think we just] really need to just go really really important and then worry about the rest.

[00:35:28.15] PAGB260EA6ED: Further reading. I think that ones important. But is that like....

[00:35:34.08] PAN4A69B0FD3: I don't know what that means. I didn't write it.

<PANF10C348AA RAISES HAND>

[00:35:36.12] PAN4A69B0FD3: Oh you did?

[00:35:37.18] PANF10C348AA: Uh yeah well, it would like at the end of an article there would be some book suggestions for further reading and analogue resources...

[00:35:43.13] PAGB260EA6ED: Is this related to see other people's highlights and bookmarks, kind of?

[00:35:47.26] PANF10C348AA: Sure

[00:35:49.06] PAN4A69B0FD3: They're cousins.

[00:35:53.24] PAGB260EA6ED: Because by seeing other people's bookmarks as well you kind of....

[00:35:56.25] PANF10C348AA: They are engaging...

[00:36:00.18] PAG7889FCFB7: It's like when somebody makes a comment you have the option to go and see where they got that kind of comment and stuff as well?

[00:36:09.00] IANC67EDF278: Alright so let's move on to the second part of this. *<points to OAG53BAC308C>* She's just going to take a picture of this. So now we're going to take all of these and we're going to put them up on this little grid over here. We're going to look at what are the ideas that are kind of your run-of-the-mill, normal ideas versus ideas that are kind of original and you know people aren't really thinking about these kinds of things. Maybe they aren't really that common. And then we are going to put them on a grid of pretty easy to do and this is kind of impossible to do. And we're going to put them into separate columns. Make sense?

[00:36:51.21] PAG7889FCFB7: No. I don't get that?

[00:36:54.03] IANC67EDF278: Ok so any of the ideas that you guys have written down that are things that are kind of common place, things that are already implemented, commonly being used.

[00:37:06.20] PAG7889FCFB7: Then those would be implemented? They'd be normal?

[00:37:10.06] PAG7889FCFB7: Maybe, maybe not? I mean search and find is a more common one. Some of the accessibility ones were a little more, I don't want to say original, but different? They are not common things that we produce. So you know accessibility ones might go more this way, whereas something like search and find might go more this way. And then along the Y axis, will be what is something that is easy or we think is easy to do versus something that is like really impossible. So something like you know beam myself into the computer. That would be impossible.

[00:37:53.04] PAGB260EA6ED: Or me feeling a piece of manuscript through the computer is impossible! <LAUGHTER>

[00:37:59.09] IANC67EDF278: At least right now! So I just want you guys to put them into the quadrants and if you feel they are more original and you want to push them more this way on the axis, that's fine. But visually position them on this chart for me.

[00:38:13.19] PAG7889FCFB7: As a group?

[00:38:15.00] IANC67EDF278: Yeah as a group. So try not to spend too much time....<GROUP DISCUSSION>

(ACTIVITY) HOW, NOW, WOW

<GROUP DISCUSSION> <AUDIO UNCLEAR>

[00:38:28.29] PAGB260EA6ED: I don't see why that's a problem at all. This is the metadata thing with the .zip files that you were talking about.

[00:38:34.27] PAN4A69B0FD3: That is complicated though.

[00:38:36.28] PAGB260EA6ED: Is it complicated?

<GROUP DISCUSSION>

[00:38:39.26] PAGB260EA6ED: It's impossible at the moment.

[00:38:41.24] PAG7889FCFB7: No it's not really possible. Gotta ask and go around. You can't just, I mean you can't just take .zip....

[00:38:49.07] PANF10C348AA: If you were going to the National Library and use a picture you were going to share an archive between 5 different institutions, they wouldn't want to....

<GROUP DISCUSSION>

[00:38:58.27] PAGB260EA6ED: So put this over here then somewhere?

[00:39:01.15] PAN4A69B0FD3: Which is the impossible part?

[00:39:04.07] IANC67EDF278: Impossible is up here. Easy is down here.

[00:39:08.25] PAGB260EA6ED: Is it an original idea though? Cause I'd say every researcher just wants to be able to view....

<GROUP DISCUSSION>

[00:39:15.08] IANC67EDF278: Yeah I'd say it probably goes....

<GROUP DISCUSSION>

[00:39:16.18] PAG7889FCFB7: A good metadata is not, good metadata is impossible to find....good resources with good metadata. Then the other thing here...a

separate issue is the fact that you want chunks of metadata, whether it's good or bad, we want lumps of it. We want to be able to just take it...

[00:39:35.19] PAGB260EA6ED: So at the moment it's impossible....

[00:39:38.10] PAG7889FCFB7: Well good metadata IS possible and it is available and it is now....

[00:39:43.29] PAGB260EA6ED: So further up here then

<AUDIO UNCLEAR>

[00:39:52.13] PAGB260EA6ED: So good metadata ....

[00:39:52.13] PAG7889FCFB7: Is possible...

[00:39:55.13] PAGB260EA6ED: This is possible. This is impossible.

<AUDIO UNCLEAR>

[00:39:59.21] PAG7889FCFB7: I think metadata has really really improved in the past 5, 10 years.

[00:40:04.27] IANC67EDF278: You said if this one is possible....

[00:40:06.02] PAG7889FCFB7: This one is possible

<GROUP DISCUSSION>

[00:40:09.00] PAN4A69B0FD3: This is an easy one. More information on collection development, policies <AUDIO UNCLEAR>. It's just a matter of getting people to do it.

[00:40:17.29] PAG7889FCFB7: Not necessarily. It's access to the material...

[00:40:21.14] PAN4A69B0FD3: No. The policy.

<GROUP DISCUSSION>

[00:40:25.29] PAN4A69B0FD3: But this is about informational policy. This is like releasing pdfs of this is how we did it. This isn't them telling us. This is about them releasing information. This is about how you did it.

[00:40:36.13] PAG7889FCFB7: Yeah but you can't find that information.

[00:40:39.09] PAN4A69B0FD3: But they obviously have developed these...so if the pdf exists they just have to release it.

[00:40:45.25] PAGA54E9A3C8: Can that be under "how" then?

[00:40:46.16] PAG7889FCFB7: They are both software engineering plan.

<GROUP DISCUSSION>

[00:40:51.14] PAGC07A5FEDA: That's probably, why does it go in there then?

[00:40:54.05] PAN4A69B0FD3: We know it exists.

[00:40:56.06] PAGC07A5FEDA: You're saying it's very doable so it should be in that corner over there, no?

[00:41:00.17] PAGB260EA6ED: Easy to do is down at the bottom and impossible is at the top. And then the normal idea and original idea .... it's a normal idea so yeah, way over there.

[00:41:11.21] PANF10C348AA: I'm pretty sure I have 2 other contributions...

[00:41:12.05] PAGB260EA6ED: Speaking of method statements and everything that's in an archive so...

<AUDIO UNCLEAR>

[00:41:18.28] PAGB260EA6ED: How easy is audio visual to do? Your audio/visual thingy with the downloadable podcasts...

[00:41:24.20] PAN4A69B0FD3: That's not me.

[00:41:26.03] PAGB260EA6ED: Oh that's yours. Sorry.

[00:41:27.12] PANF10C348AA: I would say it's easy if we have the perfect funding lines. I don't know if we should take that into account.

[00:41:34.29] IANC67EDF278: Let's consider money is no object.

[00:41:40.12] PAN4A69B0FD3: I love this funding stream!

<LAUGHTER>

[00:41:45.29] IANC67EDF278: It's my dream world!

[00:41:48.05] PAGB260EA6ED: It's easy...easy, normal idea.

[00:41:50.08] PAGC07A5FEDA: I have a bunch here about when you're going through different layers and immersed in the environment. So do you think...

[00:41:57.07] PAGB260EA6ED: Well money is no object...

[00:42:01.17] PAGC07A5FEDA: Then everything is there...<LAUGHTER>

[00:42:03.09] PAGC07A5FEDA: So I think it goes kinda here-ish. So it's not really done that often because it requires so kind of curation, a person that would design that...I don't think it's very difficult to do.

[00:42:17.29] PAN4A69B0FD3: This is the one that easily switch between modes without having to search separately. What does "modes" mean? I don't know. What does "modes" mean in this context?

[00:42:26.20] PAGC07A5FEDA: I wrote this. You know sometimes you can go to a site and you can search by text or you can explore the graph or something. But what I want to do is like a single thing to go between those things. Because my mind does not go to like "oh let's do it by text or let's do it by picture". It wants to go where it wants to go.

[00:42:46.26] PAGB260EA6ED: Yeah I would want that as well. <LAUGHTER>

[00:42:49.16] PAGC07A5FEDA: Yeah it's just about creating a connection. You know, we have a data point we have represented in many different ways and I'm looking at it. And I want to be able to go those different ways.

[00:43:00.09] PAN4A69B0FD3: So yeah then, that's doable. It is doable but you have to design it that way. It's not technologically impossible. I suppose that's what we are saying about doable or not, isn't it? The technology is there or not.

[00:43:14.29] PAGC07A5FEDA: I think so, yeah.

[00:43:17.26] IANC67EDF278: 5 more minutes

[00:43:18.23] PAN4A69B0FD3: Now it's about will power. <LAUGHTER> That's how you define that. Money is no object but will power is not included in money is it?

[00:43:27.11] IANC67EDF278: Will power is not included in money.

<LAUGHTER>

[00:43:30.13] PAN4A69B0FD3: So where do you put this on the left / right? Do we say this is doable down here? What do you think?

<GROUP AGREES>

[00:43:42.00] PAN4A69B0FD3: So only one impossible.

[00:43:48.11] PANF10C348AA: Bookmarks. I assume this means for yourself. <AUDIO UNCLEAR> So I'd say that's pretty straight forward.

[00:43:55.06] PAN4A69B0FD3: This one is impossible.

<GROUP DISCUSSION>

[00:44:02.14] PAGC07A5FEDA: Why is it impossible? Why?

[00:44:03.27] PAN4A69B0FD3: Because they can't. The way they do it, they'd have to build the site would require a lot of resources to do a complete audio lines. Because your computer would have to be built in such a way to communicate <AUDIO UNCLEAR>

[00:44:17.16] PAGB260EA6ED: In other words the technology isn't...

[00:44:19.13] PAN4A69B0FD3: You can turn on a computer by talking to it...roughly. But you can't like...Siri is good but it's not that good. So even though I can dictate an entire essay. But then telling Siri to send that in an email <AUDIO UNCLEAR>. Because the technology is not yet seamless this will always have technological problems, until we get there. I don't even know if we're going to get there because everybody speaks slightly different. And on any given day you're going to sound slightly different to previous voices trained with you.

[00:44:50.23] PAGB260EA6ED: So that sounds like a technology issue.

[00:44:52.14] PAN4A69B0FD3: That's a technology problem.

[00:44:56.11] PAGC07A5FEDA: So can we put it in "How"? Because it's like, we definitely want it to happen.

<GROUP AGREES> <GROUP DISCUSSION>

[00:45:46.25] PANF10C348AA: We're saying that clustered projects are impossible.

<GROUP DISCUSSION>

[00:46:27.17] PANF10C348AA: Ok we're saying that visualisations are distracting and hard to use. And they should be able to have standard <AUDIO UNCLEAR>.

[00:46:37.24] PAGB260EA6ED: I suppose once you have an archive with dates anyways. It's functionality is going to date...

[00:46:43.00] PANF10C348AA: Yeah I guess I'm talking about design. Something as good as...I mean it is done. I think we can put it in a pretty narrow range.

[00:46:52.19] PAG7889FCFB7: I've never come across a DH project that provides an audio assistance. No I haven't. Apart from an oral history project where we have a transcription and a podcast or something that, you'd know about this <*LOOKING AT PAGA54E9A3C8*>. You have the audio and you can listen to the tape or you can read the transcription. But the problem then would be that we haven't gotten advanced enough to query the audio. We'd have to query the transcription. But surely when you hear it and when you're use to always using that sense and not your sight, that you would be able to pick up words and should be able to speak

<AUDIO UNCLEAR> and say for your <AUDIO UNCLEAR> 100 interviews of that time that were oral.

[00:47:48.12] PAGC07A5FEDA: That's very doable actually. It's really not that difficult. I mean you just have a layer of you know as you said transcription or something like that that you search from it and then you go straight to that place where the audio is....

[00:48:03.20] PAN4A69B0FD3: <AUDIO UNCLEAR> It depends on what language you are in. And what happens if you are from a country where the way you pronounce a word is not the way a typical [language speaker] would pronounce it...

<GROUP DISCUSSION>

[00:48:19.26] PAN4A69B0FD3: So if you're in South Africa...I just looked at a project recently which was the topic modelling of parliamentarian <AUDIO UNCLEAR> and I was wondering how they did that. Because it's a very complicated language to even hear. And the written version of it is very different than what it sounds like. So how would you even train a machine to learn that. So you're going to have a continual technological problem. Yes the kind of classical European languages will be somewhat easier but you get into different variants. You go into the Balkan regions and they have a very different languages.

[00:48:52.15] IANC67EDF278: Ok guys. Let's go ahead and wrap up the last...I think you have 3 left.

[00:48:58.14] PAGC07A5FEDA: I'm just going to go ahead and put this in the now because

[00:49:00.18] IANC67EDF278: We're running out of time.

[00:49:03.04] PAGB260EA6ED: Building a strong picture of the environment from the mundane aspects of the literature.

[00:49:07.13] PAN4A69B0FD3: That's possible.

[00:49:08.05] PAGB260EA6ED: That is possible yeah.

[00:49:12.20] PAGB260EA6ED: Get a feel for the original text and archive materials. That's depending on how deep you want to go with that.

<LAUGHTER> <GROUP DISCUSSION>

[00:49:23.21] PANF10C348AA: This is a general improvement one. Add to existing abilities and improving on the current <AUDIO UNCLEAR> of DHPs.

<LAUGHTER>

[00:49:37.29] PAGB260EA6ED: That's just the <AUDIO UNCLEAR>.

[00:49:41.05] PAG7889FCFB7: <AUDIO UNCLEAR> DHP sites. That's already been...it's already happened.

[00:49:51.19] PANF10C348AA: I'm speaking from my own experience of having DHP projects in between two institutions that in theory are quite close to one another. Academic politics are such that is just doesn't happen.

[00:50:02.19] PAN4A69B0FD3: It's the human element.

[00:50:04.12] PANF10C348AA: I'm talking about the human elements. yeah I suppose I'd like to see cultural institutions collaborate when it comes to Digital Humanities. <AUDIO UNCLEAR> It's just myself. Sorry, I'm not breaking the rules or anything....

(ACTIVITY) 6 PART STORY METHOD

[00:50:22.26] IANC67EDF278: Alright thank you guys. This is nice. You've got a lot of stuff. This is fantastic! Alright so we're going to move on to the next bit. Take a seat. You each have a white piece of paper in front of you. I'm going to ask that you take that white piece of paper and fold it into thirds. Like this. Then take that and fold it in half. And when you open it up, you've got 6 boxes. 3 on the top and 3 on the bottom. Ok? So the next thing we're going to do is called the 6 part story method. We're going to craft a story about using DHPs. And in each box, I'm going to give you specific instructions, you're going to draw something that is going to relate to the instructions I give you. But keep in mind as you're drawing things that you're connecting everything together in a narrative, into some kind of story. So the very first box here in the upper-left hand corner, I want you to draw a picture of who—this one <POINTS TO BOX>—who the kind of the main character of your story is. Is it a historian? Is it a librarian? Is it a student? Is it an educator? Is it a researcher?

[00:52:19.02] PAGA54E9A3C8: Are we writing the words down?

[00:52:20.08] IANC67EDF278: Nope you're going to draw a picture.

[00:52:21.10] PAG7889FCFB7: And is it the user?

[00:52:22.27] IANC67EDF278: Yeah it's basically the user. So who is the main character / who is the user of your DHP in your story. And feel free to use any of the pens. Any colours you want. Whatever you want to do. In fact, I'll even grab some markers in case people want to get <AUDIO UNCLEAR>.

[00:53:08.10] IANC67EDF278: And just so you know, we're going to go across the boxes, this way <POINTS HORIZONTALLY>. So it's going to be 1, 2, 3, 4, 5, 6 <POINTS TO INDIVIDUAL BOXES> as we draw. But just draw in the first box for now.

[00:53:27.25] PAG7889FCFB7: This is an imaginary DHP?

[00:53:29.21] IANC67EDF278: I'm sorry

[00:53:32.03] PAG7889FCFB7: Am I imagining a DHP? I have a specific thing in my head.

[00:53:37.11] IANC67EDF278: That's fine! Just draw a picture of your user or your main character. I just ask that you make it semi-realistic. So don't draw a dog.

[00:55:11.07] IANC67EDF278: Ok in the second box, the middle box on the top. We're going to draw what the user is trying to do. What our main character in our story, they are using the DHP to do what? What is the task or tasks that they are trying to accomplish. You can draw pictures of books. Or pictures of little Bunsen burners. Beakers.

[00:57:21.00] IANC67EDF278: Ok in our third box, on the left hand side, no right hand side. I'm backwards. We're going to draw what are things that get in the way. What are the things that stop the user from accomplishing what they are trying to do. What are the obstacles that they might face? What are the things that get in their way. The things that hinder them. Whatever that might be. Maybe it's accessibility issues. Maybe it's the only have a phone and it doesn't work on a phone. Or maybe it's, they have trouble reading it. Whatever the situation is, whatever your hinderance might be. What is something that gets in their way.

[00:58:06.07] PAG7889FCFB7: I'm sorry IANC67EDF278 my hearing's not great.

[00:58:09.00] IANC67EDF278: No that's alright. So in this box draw what is something that gets in the way of the user accomplishing the task in their story.

[00:58:17.07] PAG7889FCFB7: In the imaginary resource.

[00:58:18.25] IANC67EDF278: In your imaginary resource, yes. Something that gets in the way.

[00:59:23.18] PAG7889FCFB7: Draw something.

[00:59:25.21] IANC67EDF278: Yes. Draw something. Always draw.

[01:00:13.11] IANC67EDF278: Ok. In the bottom left, space number 4, we're going to draw something that HELPS the user. So here we drew something that got in their way. Now we're going to draw something that helps them. Helps them to accomplish their task. Something that HELPS the user. So here is something that hindered them. Now we're going to draw something that helps them, in your imaginary resource. Down in the bottom box.

[01:02:26.02] IANC67EDF278: Alright. In the bottom middle, box number 5, we're going to draw what is the turning point of the story. What is the thing that happens that makes everything just fall into place. It's like your climax moment. So you've had some things that have helped. You've had some things that have hurt them, some things that have helped them. But what is the thing that was just like their "Eureka!" moment. Something happened or they found something or there was some new technology that was introduced. Whatever it might be that just got them over the hump and got them to accomplish their goal.

[01:04:21.12] IANC67EDF278: Ok in the last box. What is the final result. What is your user, what is the final outcome for your user? What is the end result of your story? What is the final bit? How does the user end up? How does the main character feel about things? Where are they at? What's going on? How does your story end?

[01:06:05.27] IANC67EDF278: When you're finished just put your pen down so I know everyone is done. Great! Alright, who wants to share their story? Anyone?

[01:06:24.05] PAGC07A5FEDA: I could. So my user is...she looks kind of academic but not necessarily a teacher or a student. Just somebody who is curious I guess. And her goal is to get excited about something, learn about something. And she wants to use the resource to pretty much get inspiration about...she doesn't really know what exactly. And what stands in her way is all sorts of old peripherary

that I drew. Basically, something between the ideas that she's trying to get and her old brain because she has to use the keyboard, she has to use the mouse, the interface that might not be very good. But what helps her in the way is again something guiding her through the interface. Some kind of path through the data that other users made or designers made. Something that goes into some direction. But there's always things that allow to detour. Like there's something interesting here or people are discussing this. And the turning point for my story is virtual reality really. I know of, as a user I want to believe in some magic that's going to happen. And the end of the story is that she did get excited about something and she went and told somebody else. And now they also want to explore and perhaps share it.

[01:07:57.29] IANC67EDF278: Ok thank you. Who else wants to share?

[01:08:06.00] PAG7889FCFB7: Yeah mine. My user is an archaeologist. And there's not a lot of funding in Ireland for any of that at the moment. And for archaeological projects that aren't necessarily going to show something. So my resource is aerial photographs of Ireland where the archaeologist can go in and zoom in on specific areas where he wants to see if there are any activity underneath that aren't visible on the plane side. So the problem with the resource is that the internet connection keeps crashing. And what would be nice would be a good internet connection so it doesn't keep crashing. And then the archaeologist finds this big ring thing that looks unusually out of place and then he digs up and finds loads of stuff and brings it to the museum. Is that what you're looking for?

[01:09:02.06] IANC67EDF278: It works! Thank you! Anybody else want to share?

[01:09:08.20] PANF10C348AA: Mine is a casual reader of fiction and wants to use an archive, a literary archive of some kind to relevant to modern literature and of course with copyright and DRM and new legislation that's not possible. That's the obstacle. And working a job its hard to upscale the kind of technical know how he might be able to. So getting a grant and learning to collaborate with techie people is the turning point. And the result is a fully digitised and searchable archive.

[01:09:47.12] IANC67EDF278: Perfect. Thank you.

[01:09:53.21] PAGB260EA6ED: I'll go. So I've got the young person and they could have just something as simple as dyslexia. They are looking for some kind of creative, educational, inspiration. But the volume of text is really off putting to their self confidence as well as interacting with an archive. So I've got more visual and auditory kind of click buttons to kind of, all kinds of links to kind of spark their imagination. So I kind of started with that whole kind of paired down interface which leads them into looking at something, listening to something. Then maybe having the confidence to search for something, bringing something up which leads them to kind of more learning and kind of kick starts the whole kind of educational, creative kind of experience. And then so my final thing is them sharing both the information they have gained but the experience and confidence they've gained from that experience to a wide group of people whoever that is.

[01:11:04.02] IANC67EDF278: Ok. Great! PAN4A69B0FD3, do you want to share or would you rather not?

[01:11:10.26] PAN4A69B0FD3: I don't mind. It's not that complicated. Mine starts with Academic A. And Academic A finds a resource in an archive that is very get out. So I actually a thing in mind, I won't go into it. But this is my thing from earlier where we have loads of information and we have to get it out to the public but how do we do it? So they want to build a Digital Humanities project but their problems are administration, politics, access. So then the solution to that is THE LAW: digitisation and Digital Arts & Humanities. So we have this new technology. We've passed a bunch of laws about access and so on. And now with the type of technologies we can now digitise a lot of records in a much shorter and less expensive time. Then of course the solution is money, open-source, Web 2.0, that's the technological aspect of it. And the end of it as very workable and valuable resource and our hope is that Academic A gets tenure as a result of this and can continue on his merry way.

[01:12:13.00] IANC67EDF278: Ok thank you. Would you like to share?

[01:12:16.10] PAGA54E9A3C8: Mine is very simple. It's really about someone trying to do family history. And really what they want to do is like their own like diaries and personal letters and things like that that they've had in their families for a long time. So the problem is how to do it. Their probably not really skilled about how to actually access the right information as to how to get it into like a website for example in order to share with everyone. So the solution is to get the right help. And obviously there's someone there that's skilled. And basically the finished product is they get it all put together and are able to share it.

[01:13:11.05] IANC67EDF278: Wonderful! Alright, thank you guys! I appreciate it! If you could just write your initials on the back so I know whose is whose.

(ACTIVITY) THE ANTI-PROBLEM

[01:13:41.07] IANC67EDF278: Ok moving on. The next thing we're going to do is a little exercise that I call the anti-problem. So I'm going to give you guys a really outrageous, ridiculous statement that is clearly not something that we would ever want to do. And I want you to, actually let me give everybody pieces of paper if you don't have any. You can just take a piece of paper from the notebook and just kind of divide it down the middle, fold it lengthwise so you can have two columns. And in the left hand column for the next five minutes or so, I want you to write possible ways of supporting the problem I'm going to give you. So different, not solutions to the problem but different tasks that you would do to make sure that the task I'm going to give you can happen. And they can be as ridiculous as you want. So to kind of help explain this a little bit better: the solution that I'm going to give you, the problem that I'm going to give you is how do we make it impossible to read in DHPs?

[01:14:55.04] PAG7889FCFB7: Possible?

[01:14:54.25] IANC67EDF278: How do you make it IM-possible to read in DHPs?

[01:14:58.18] PAN4A69B0FD3: To read?

[01:14:59.08] IANC67EDF278: To read. To actually like reading the content and be able to understand it.

[01:15:04.03] PAN4A69B0FD3: How do we make it possible?

[01:15:05.14] IANC67EDF278: No how do we make it IM-possible.

[01:15:08.21] PAN4A69B0FD3: Oh ok. In the left hand column?

[01:15:09.23] IANC67EDF278: In the left hand column I want you to give me ideas about how we would make it just absolutely, bloody impossible to read the content that we put in DHPs.

[01:15:25.26] PAG7889FCFB7: To read content? Impossible in the left column.

[01:15:32.12] IANC67EDF278: Yup. Absolutely impossible in the left hand column, list out different ways that we could, different things we could do....

[01:15:38.19] PAG7889FCFB7: I mean, even cloak something out?

[01:15:41.14] IANC67EDF278: Yeah! In fact as OAG53BAC308C mentioned to me: "Black backgrounds with black text" as an impossible to read.Things like that.

[01:15:56.02] PAG7889FCFB7: So impossible to read something in a DHP?

[01:15:59.10] IANC67EDF278: Yes.

[01:16:09.01] IANC67EDF278: Give you guys about 5 minutes to do that.

[01:17:29.13] PAGB260EA6ED: What do you call...do you know the buttons that you click like Facebook, search and all those kind of things. What do you call them?

[01:17:38.19] PANF10C348AA: Social media links?

[01:17:44.02] PAGB260EA6ED: Not necessarily social media but like recognisable things so like you click on, OH! "Icons".

[01:18:57.09] IANC67EDF278: Ok. Now on the right hand side. I want you to go through and tell me the exact opposite of each task that came up with. So if you, if one of your tasks was "black background with black text", then the solution to that would be a "white background with black text" or a "black background with white text". So give me solutions to each of the various problems you've proposed.

[01:22:26.24] IANC67EDF278: Ok. Alright. Who wants to share a couple of their ideas? Give me kind of your original solution and then the rebuttal to it. You don't have to give me them all but just some of your top 10 picks or your top 5 picks.

[01:22:50.25] PAGA54E9A3C8: My first one was to write in the language that's least understood. Like for example, Sanskrit. People wouldn't know Sanskrit so that would be difficult for most people. And so obviously the solution is <AUDIO UNCLEAR>. And then <AUDIO UNCLEAR>. Or make the print so small that it wouldn't be readable and you couldn't change the font size. Or use like 20 different languages in the one sentence.

[01:23:29.24] IANC67EDF278: Wow that's really cruel! <LAUGHTER>

[01:23:31.26] PAGB260EA6ED: And and, write backwards, instead of frontwards.

[01:23:38.06] IANC67EDF278: Those are all really good!

[01:23:44.06] PAGC07A5FEDA: My worst nightmare is sources that publish scant pages. And well scant pages are good if you want to look at the original with all the

goodies like customisable fonts and nice, plain background as opposed to trying to figure out what's written on a picture.

[01:24:09.16] IANC67EDF278: Ok so when you say trying to figure out what's written on a picture, do you mean like people when they try to visually represent like notes scribbled in the margin and that kind of thing? Like where they try to reproduce it exactly the same as the image but in a plain text format?

[01:24:24.20] PAGC07A5FEDA: No I actually just meant releasing it as an original image.

[01:24:30.23] IANC67EDF278: And that's all?

[01:24:31.24] PAGC07A5FEDA: Yeah but it's also what you said. Reproducing it might also be annoying. I kind of want to separate content from the way it was presented. I know sometimes you want exactly the opposite. You want to see how it was. But in that case you can probably go and see the attached image or something like that.

[01:24:49.26] IANC67EDF278: Ok. Both of those are really great. Some other ones?

[01:24:54.11] PANF10C348AA: Make the font comic sans. <LAUGHTER> I'd say "Don't do that".

[01:25:01.04] IANC67EDF278: Ok.

[01:25:02.17] PAN4A69B0FD3: Mine was just don't add any text. So you're only looking at the artefacts. And the only text that comes up is that horrible latin. You know that it's not even any sentences. It's the default that kicks in.

[01:25:13.26] PANF10C348AA: The wing-dings?

[01:25:15.16] PAN4A69B0FD3: Yeah you just start looking at that and there's loads of images.

[01:25:19.26] IANC67EDF278: That would definitely make it unreadable.

[01:25:21.29] PAN4A69B0FD3: The other one was to make it a whole bunch of binary so it's all 1's and 0's. <LAUGHTER> So everything was a series of 1's and 0's. So you're just confused.

[01:25:31.19] IANC67EDF278: Ok. Anybody else want to share any?

[01:25:37.06] PAG7889FCFB7: I have <AUDIO UNCLEAR> no CSS sheet attached.

[01:25:46.11] IANC67EDF278: Ok so just no styling at all. It's just plain XML.

[01:25:49.21] PAG7889FCFB7: Everything's all mixed up. And Link rot.

[01:25:54.00] IANC67EDF278: What was that?

[01:25:55.11] PAG7889FCFB7: Link rot.

[01:25:55.19] IANC67EDF278: Link rot?

[01:25:57.29] PAG7889FCFB7: Broken link.

[01:25:59.17] IANC67EDF278: OH! Broken links.

[01:25:59.17] PAG7889FCFB7: So it's not going to show anything at all.

[01:26:04.24] IANC67EDF278: Yeah that would be very problematic, especially for people who are doing kind of non-linear reading approaches.

[01:26:10.20] PAG7889FCFB7: Especially if it's text that is hyperlinked and depending on the scholarly edition and the link goes nowhere. And so....

[01:26:20.19] PAN4A69B0FD3: Wouldn't it be better if the link kept bringing you back to the same page you were at so you click and reload and you click and reload and you're just going out of your mind.

[01:26:29.09] PAG7889FCFB7: It's like they're everywhere! Broken links, I'm telling you! They're in every resource. They're hidden all over the place.

[01:26:37.12] IANC67EDF278: Yeah they are! It's so annoying!

[01:26:39.01] PAG7889FCFB7: I tell you!

[01:26:40.22] IANC67EDF278: Alright. Any other ideas that anyone wants to share?

[01:26:44.02] PAGB260EA6ED: One of mine was unrecognisable icons for functions and links. Or just not being able to find what you want to do, like a search button. That's really difficult.

[01:26:57.21] IANC67EDF278: Sadly something that happens more often than it should.

[01:27:03.27] PAGC07A5FEDA: I actually just thought of one thing, I don't know if we brought it up or not. I would be very annoyed if the resource tried to say imitate the pages of the book. I mean sometimes you go to a resource and you have like an image of the book and you can click and it kind of, you know, folds the pages. I don't want to see that. Like if I'm going to a digital resource, I want it to scroll indefinitely while I'm reading. So like it doesn't have to tie so, be so tied to the original form because like if we are doing a digital resource we can use all these formless things that would make more sense in our context as opposed to trying to "Here's the book and it's on your screen but here's the pages that are turning."

[01:27:47.22] IANC67EDF278: So I'm curious, why does that...first of all let me preface this by saying I completely agree with you. But why specifically does that bother you? Like what is it about things that represent...because this kind of the thrust of my research. Is that trying to represent physical objects in digital space exactly as they are represented in their analogue counterparts is not helpful from a reading standpoint. So can you talk to me a little bit more about why that bothers you. What is it about it that really gets at you?

[01:28:19.02] PAGC07A5FEDA: Yeah because what I'm thinking about right now is that when I go to a site like this, I'm saying reading, right? And the way I'm reading right now is that I open the sites in medium and I scroll. And I know what it's going to look like. Like I prefer my reading to look the same more or less. Like I use to have a kindle so I really love that all of the books, like here's the font that I choose, here's how they look. And I can just kind of separate myself from...like it kind of dissolves, you know? So I'm just reading and I feel a more direct connection between the text and how I'm receiving it. And I don't want anything to get in the way of that. Like, I'm thinking maybe there should be a couple of different modes. Because sometimes you really want to go into text and you don't want anything to

stand between you. Sometime you want to get the feel for it, for what it was like and you don't want to go to the archive to get the actual book or anything. So maybe we could somehow switch or make it a continuum between how much of the medium you want in between. But in general, I just think that the screen is not a book and it shouldn't pretend to be one.

<NERVOUS LAUGHTER>

[01:29:40.15] OAG53BAC308C: So i have just one. Maybe it's a generational thing? I know I'm not really supposed to be talking, sorry, but maybe it's generational thing because I find that if I have to scroll up it actually distances me from the text. I think that maybe I'm more used to growing up holding a physical book so doing that seems more natural to me and that seems, that motion takes me out of what I'm reading and into my motion. This one is just part of reading.

[01:30:05.14] IANC67EDF278: So we've got two different opinions. Two thoughts of—ok a third. Let's get some information...

[01:30:12.08] PAGB260EA6ED: So this has really clarified what I said before in my mind about the whole visual thing. So if I want to get the feeling of a manuscript or something like that I want to be able to look at a visual of it but with the function of either turning a book page or however you would use the physical object but I don't want to use the transcription like that. I want to either scroll or like the kindle and set my own font and everything else and just have it like, if I'm absorbing the information, I just want to do it in my own format and then scroll or a page turn is alright but I want that separate from the actual visual kind of engagement with the actual source material.

[01:30:55.17] IANC67EDF278: So the hybrid approach, kind of what [PAGC07A5FEDA] suggested earlier of, if I'm depending on the kind of reading mode that you're in or what your goal is you want to be able to switch.

[01:31:09.19] PAGB260EA6ED: Yeah yeah

[01:31:11.07] PANF10C348AA: I think thats good. I mean I kind of agree with you as well. I'm going to do that annoying thing where I cite some anonymous study that I read, science says but maybe people have had that experience from flipping through books with footnotes versus endnotes. Your brain apparently treats thresholds like ok you've put that down, that page is gone. We don't need to worry about that anymore. And this is something that when you go into other rooms, what you've gone in for is kind of evaporated because your brain has left that behind. Apparently. So I kind of think of it in the same way. In a book, you're holding a physical object and it's never gone. But when you're turning the page in that terribly twee, stylised way across the screen, it's kind of treating it like a different thing. Now you have your own experience and it's subjective. Things are all different to tastes. It just seems very first generation iPad to me, that page animation.

[01:32:02.27] PAN4A69B0FD3: Can I ask a question about what do we mean by the book. Are we talking about a scan of a book from say 17th century which is a

scanned representation so it's a pdf or something and you're looking at the physical scanned copy? Or are we talking about a digital version of it? Are we talking about where we've transcribed it into some sort of publication or a mobile or something like that. Is that what we are talking about? Which one? Because it's kind of confusing cause I think you want both, right?

[01:32:33.19] PAGB260EA6ED: No I don't want the two things to be the same. I want to be able to switch between the two...

[01:32:38.08] PAN4A69B0FD3: A scan...

[01:32:38.19] PAGB260EA6ED: A scan and a transcription, yeah yeah.

[01:32:42.15] PAN4A69B0FD3: But also there's couple different ways of transcribing. So there's the one that you're not talking about, the e-pub, where you put it into your kindle or your iPad and it's <AUDIO UNCLEAR> to possibly move, right? Then there's also just PDF which you can just bring it up as a two page or a single page you and just go up and down. So....just trying to clarify what exactly we are talking about here. It's just a little confusing. Because I wasn't sure .... I think you all had different ideas in your head.

[01:33:09.24] IANC67EDF278: And they all exist in DHPs though, don't they? Those different versions exist so I think it's important to highlight you know how we use them IS different. And whether we are talking about a scan of a manuscript, that we might experience one way. Or a transcription that is dynamic and hyperlinked and it has annotations and all this fun stuff. Or we're talking about a transcription that is simply a PDF version of the scanned manuscript that they've just saved a transcription of in a Word document. So you're kind of ....

[01:33:48.00] PAG7889FCFB7: The other one of a 3D model of a book that you can move around to see the depth or something, which a lot of projects now where there's something has been found in kind of a virtual reality of an old site, of remaking, of things that have been found. So you can walk in and pick up a book in an artificial or virtual environment and twist the book and flip it upside down. But not necessarily open the book, do you know what I mean? You don't necessarily have to open it...is that what people want from the analogue of something? I'd like to see a book so I know how thick it is before picking it up to read....<LAUGHTER>. Oh it's too big, I'll leave that one over there.

[01:34:45.07] IANC67EDF278: You want to know what you're getting yourself into...

[01:34:49.26] PAG7889FCFB7: Do I want to see that or if I'm looking at a transcription and you're scrolling down to see how many pages is it...oh it's never going to end. Do you know what I mean? Or for copy and paste. You know, I think we do need several modes of adaptation...

[01:35:06.09] PAGB260EA6ED: Whether it's a pdf or a scan of the original document, I want to see how it was written. So just like to actually being able to see the physical book without being able to see inside it would be really horrible for me. <LAUGHTER>.

[01:35:22.27] PAN4A69B0FD3: The problem as well is...so for some...I just remember this when I was doing history at Trinity they were very clear. They said "You cannot use an ebook for citation. Because it does not have a page number. Because depending on the font you set, the page number will change."

[01:35:41.01] PAGB260EA6ED: I've had this problem, cite things. I've had this ancient book that I couldn't actually borrow anywhere. I had to buy it. And I had to buy the Kindle version of it. And I said to my professor, "Well I have to be able to cite this because I can't actually, physically get ahold of it anywhere and I need to cite it in this essay." So she said "Yeah that's the way we are going and you're going to have to come up with something to deal with that."

[01:36:05.09] IANC67EDF278: Yeah I noticed there are some who are starting with the Kindle you can reference the location which is kind of a dynamically calculated numerical value, but it's tricky because you have to know all their settings to kind of go along with it but there are rumblings, in the SOME of the citation formats of trying to say "How can we incorporate locations in ebooks."

[01:36:38.10] PAGB260EA6ED: And um just going back to that, in the [Letters] of 1916, so when I was looking at it in development...so I wanted to have this visual thing and I was like "If money is no object, I want to be able to click on a word and it tell me what that word is". So I see what they've done that you can click on it and get a visual zoom into it. But actually you've, once you've clicked into it, you've no idea where you are on the page. So I found that really confusing. Because you know you're getting an expanded view of what the page looks like and like a couple of lines of writing. But if you can't actually see what that is in comparison to the transcription, it's really confusing.

[01:37:14.28] IANC67EDF278: Gotcha. Wow, that's something really interesting to know.

(ACTIVITY) LOVE LETTER / BREAK UP LETTER

[01:37:20.27] IANC67EDF278: Ok alright, so we've got about 10 minutes left and I'm going to do one final activity. It's kind of a simple one, a fun one. Again, where's that notebook at? If everybody could take a piece of paper. We're going to write...you can pick any DHP that you've used, whichever one you like, any of the ones I've mentioned, any of the ones that you've used in the past. And you have the option, you can either write it a love letter extolling it's virtues and how much you love it and why love it and what's so great about it. Or you can write it a break up letter. "Dear Letters of 1916, I need to break up with you because..." and then tell everyone. So those are your options. Love letter? Or break up letter. Any DHP you like. And you have about 7 minutes.

[01:38:12.01] PAN4A69B0FD3: A Dear John to an archive....

[01:38:13.16] IANC67EDF278: Yeah, Dear John....

[01:38:16.15] PAG7889FCFB7: On the back of that sheet that we've been writing on, yeah? Or do we need new one?

[01:38:20.08] IANC67EDF278: Use a new sheet...new sheet.

[01:38:21.18] PAN4A69B0FD3: Does it have to be <AUDIO UNCLEAR>...

[01:38:23.00] IANC67EDF278: It can be anything.

[01:38:24.05] PAN4A69B0FD3: Oh perfect.

[01:38:25.24] PAGB260EA6ED: What's the name of the Isaac Newton one?

[01:38:28.25] IANC67EDF278: The Chymistry of Isaac Newton. C-H-Y-M-I-S-T-R-Y.

[01:38:34.18] PAG7889FCFB7: What about...so one of these Digital Humanities Projects or something that you've used?

[01:38:39.17] IANC67EDF278: Yeah

[01:38:41.00] PAG7889FCFB7: One of these?

[01:38:42.04] IANC67EDF278: These are ones that I've used for various different reasons or various essays that I've written or obviously worked on since I was on the Letters of 1916 team. But you don't have to use any of these. You can use other ones. You mentioned the slave project. If you want to use that one, that's fine.

[01:38:57.29] PAG7889FCFB7: What's Chimera? I'm, I don't know...there's too many of them...

<GROUP DISCUSSION>

[01:39:09.00] PAGB260EA6ED: So you want us to say what it is we love about the project or why we don't want to use it anymore if it's a break up?

[01:39:23.14] IANC67EDF278: Yeah, so you can either talk about loving it or if you have one that you really, really can't stand, then you can write a break up letter. And tell me why...explain why you are breaking up with it. So what is it you don't like about it. Or if you love it, what is it you DO love about it.

[01:39:40.17] PAGB260EA6ED: What if you've got two that you can love / hate? <LAUGHTER>

[01:39:45.01] IANC67EDF278: I actually had somebody...I did this with somebody yesterday and they actually ended up writing a love letter but then said "But if you want to keep dating me"...<LAUGHTER> "...you better do x, y, and z." So that's another valid approach!

[01:40:00.01] PAG7889FCFB7: Brilliant! What about then...we're talking about Digital Humanities Projects. And I'm working in Digital Archives and they tend to like to be called Digital Archives, even though Digital Humanities people might call them "Digital Humanities Projects."

[01:40:13.14] IANC67EDF278: Yes I know.

[01:40:16.06] PAG7889FCFB7: Digital librarians have that same issue as well so...

[01:40:18.17] IANC67EDF278: Well the point of my research is, you know...there's a lot of overlap, yeah...text...anything text based.

[01:42:33.26] PAN4A69B0FD3: Are you going to have to read this, IANC67EDF278? You're going to have problems with this chicken scrawl.

[01:42:39.12] IANC67EDF278: <LAUGHTER> If I have trouble, I'll get back to you and have you help me transcribe...

[01:42:42.06] PAN4A69B0FD3: Type it up....

[01:42:42.26] IANC67EDF278: Yeah....

[01:42:51.29] PAN4A69B0FD3: You know actually, I can just tell from this, you're going to have a hard time reading...

[01:42:55.09] IANC67EDF278: Well you can always read yours and then I'll have the audio recording and can transcribe it from there. That will also work.

[01:43:04.13] PAN4A69B0FD3: I'll do it on my laptop. That'll be easier. I've never broken up with a website before. Feels great!

[01:44:00.10] PANF10C348AA: IANC67EDF278, I don't suppose I might be able to finish the letter later and give it all my emotional energy. I just got a call and I think I need to take it. Is that ok?

<PANF10C348AA LEAVES THE ROOM>

[01:44:07.24] IANC67EDF278: Ok yeah.

[01:44:45.03] IANC67EDF278: Ok, somebody want to...ok we can't hear everybody's. But who wants to share? Anybody? You don't have to?

<LAUGHTER>

[01:44:57.12] PAGB260EA6ED: Why is everyone looking at me? <LAUGHTER>. I know i've got verbal diarrhoea but do I always have to be the one that talks!

[01:45:04.04] IANC67EDF278: Sorry...no. You're just in my line of sight. Sorry!

[01:45:07.03] PAGB260EA6ED: It's ok. I've got use to it. <LAUGHTER>

[01:45:15.21] PAG7889FCFB7: Break up over the phone! It's not worth it! <LAUGHTER>

[01:45:18.07] PAGB260EA6ED: Dear Chymistry, You really have inspired my quest to turn lead into gold. I'm amazed at how quickly my search for both elements leads me to contextual links within the archive, helping me to sort through the other chemical processing I do not need. Alchemy is magic so I would like to see more visual inspiration spurring me forward in my quest.

[01:45:38.25] IANC67EDF278: Fantastic! If you could just initial your letters, so I know whose is whose. Anybody else want to share?

[01:45:56.25] PAGC07A5FEDA: I guess I could. Dear White Pages, I want to thank you for being such a great companion. You run natively on my phone and I access you anywhere. You're trying hard to keep me entertained even though War and Peace is so hard to read, I find pleasure in discovering the underlying things. Your relationship graph really helps me to keep track of who is talking to whom and when because I can never remember anyway. I also love the War piece marking on each chapter because I hate the War pieces anyway and skip them. I know it took so many people to bring you to life but maybe we'll have a new book coming out soon.

[01:46:29.07] IANC67EDF278: Ok does anybody have a break up letter? We've got a couple love letters.

[01:46:33.21] PAGA54E9A3C8: I do. Dear 1641 Depositions, After a 5 year commitment to you, I've decided to call an end to our relationship. During our time together, I've found that you've been inflexible moving between the 4 or 5 sources, Dublin originals, and all the others, I find your language to be contradictory. And

also the website was not updated as promised. Therefore, our dating period has come to an end.

[01:46:57.27] IANC67EDF278: Excellent! Ok. Alright thank you guys so much for participating in this. I really appreciate your help. I got some great information. I can't wait to kind of dive into it, and divulge...figure everything out. I will, if you give me two seconds, I have gift cards, or $10, €10 vouchers to Starbucks for everybody that's here today as kind of a "Thank you!" So let me grab those and hand them out and we'll be good to go.

A.3 Focus Group B

A.3.1 Focus Group B: Transcription

Following is the transcription for Focus Group B.

[00:00:03.01] IBJ7A006F734: Alright. Welcome everybody and thank you for taking the time to participate in this focus group for my PhD dissertation. So once again, just to go over a little bit about what we will be doing and kind of the goals. I am specifically exploring what I've dubbed "Digital Humanities Projects" as a temporary placeholder. These may include Digital Scholarly Editions. They may include Digital Archives, Thematic Research Collections. I'm focused on anything that's text-based in the Digital Humanities Space. So mostly literary and historical content. I am interested in exploring how we use these and finding new ways of presenting us with kind of getting through the data in ways that support different ways of reading on the internet because we read differently in digital environments.

[00:00:58.29] IBJ7A006F734: So the point of today's focus group is to gather information about how you guys use these types of things, what works for you, what doesn't. What's missing. So that I can better understand how to move forward with my own design recommendations. Just remember this is a completely safe space so that means no idea is too ridiculous or too out there. The point is to generate ideas that ARE kind of out there or crazy that you know might even seem impossible to do. Because it's from those that I can kind of figure out different approaches that we might be able to take. So don't be afraid to speak your mind. Don't be afraid to speak out.

[00:01:46.00] IBJ7A006F734: Just to remind everyone of the rules. You want to be respectful of everyone. Don't talk over anyone. Don't threaten anyone which I don't think will happen. Don't you know call people names. <LAUGHTER> I'm looking at a couple of you. <LAUGHTER>

[00:02:07.02] IBJ7A006F734: So we'll be creating a few different things throughout the course. There will be some drawing and some writing and stuff. Don't be afraid about you know if you don't think your drawings are good. It's fine. I draw stick figures so there's no judgement there. But I will collect everything at the end

because I'll use that as part of my analysis. And just to clarify, any time I use the term "DHP", I'm referring to any of these types of resources. So DHP meaning "Digital Humanities Project". So things like the Letters [of 1916], the Chymstry of Isaac Newton, the Woodman Diary, Mapping the Republic [of Letters], Kindred Britain, Burchardt Source. There's tons and tons of these. So I've just kind of created this umbrella term to refer to everything for now because I'm not focused on just Digital Scholarly Editions or just Thematic Research Collections. I want to keep a little broad right now.

(ACTIVITY) ICE BREAKER

[00:03:03.12] IBJ7A006F734: I'm not going to have everybody introduce themselves because I did promise anonymity, although most of you know each other. But instead we are going to do a little bit of an icebreaker to start. So take a piece of paper. And on that piece of paper, write down one word or phrase that you think of when you think of a DHP. Whatever that word or phrase might be. Whatever first comes into your mind whenever you think of these types of scholarly exercises. And then once you've written it down, go ahead and put it face down in the middle.

[00:04:55.10] IBJ7A006F734: Ok. So I'm going to mix these up and hopefully you won't get your own. Now we're going to go around the room and I want each of you to read out what you have on your card. We'll start here.

[00:05:36.22] PBR9D1AD41E7: Documentary / diplomatic

[00:05:41.20] PBR1B0A881F1: Unfamiliar

[00:05:44.03] PBRF60C8C611: Research

[00:05:45.05] PBR99A4D800B: Letters

[00:05:47.05] PBJ6BAA5509A: Incomplete

[00:05:49.24] PBJ560B377FD: Hopeful

[00:05:51.07] PBJCA47575EA: Scholarly

[00:05:53.26] IBJ7A006F734: Ok those are all good. Some interesting ones: unfamiliar, incomplete, some of the ones I would have expected like "scholarly".

(ACTIVITY) GROUP DISCUSSION

[00:06:04.11] IBJ7A006F734: So let's kind of use this to spring board a little bit to have a short discussion about why you use these, if you do. If you haven't used them before, hopefully you got a chance to look through the ones I sent out ahead of time. But if you don't use them, why might you use them? Why might you be interested in them? Who wants to start off?

[00:06:26.13] PBJ6BAA5509A: As in projects?

[00:06:29.10] IBJ7A006F734: Yes as in these types of projects.

[00:06:32.09] PBR99A4D800B: Well I would often use these sorts of things just to see what's out there. Just to see whether there is anything on the computer that would short circuit what I'm having to go into town to do. Because up until even the last 10 / 15 years you had to go into town to dig through repositories. So I try this first to see if there's anything workable.

[00:06:57.25] IBJ7A006F734: And is there for personal research?

[00:07:00.05] PBR99A4D800B: No for college. No for projects for college. For my own work in college.

[00:07:11.19] IBJ7A006F734: Ok wonderful! Anyone else?

[00:07:14.10] PBR1B0A881F1: I use it for research, for college essays. I used stuff like the textual sources to like search through to find patterns if I wanted to try to confirm or check something.

[00:07:26.15] IBJ7A006F734: Ok

[00:07:29.24] PBJCA47575EA: I use particular ones like a repository of texts so I don't have to have a giant library.

[00:07:37.14] IBJ7A006F734: Ah ok. I imagine, especially in your field, that's probably really beneficial.

[00:07:45.28] PBJ6BAA5509A: I both use it for research purposes in terms of competency research and whatever it is that is in that resource but also from an interface perspective as well. That is one part of my work is about humanities interfaces. And I suppose it is that the content is always mediated through these channels, whether it's a search box; whether it's faceted search and so on and so forth. And I think there's a bit of training that is needed. I have not found yet something that is truly intuitive. But I know what can be done so I am fully efficient and I find that that is where <*AUDIO UNCLEAR*>. That is what probably takes me the longest. So it's a commitment. When I look at a resource, it's a commitment. Because I need two or three days to understand how the resource works and then I can actually use it.

[00:08:47.28] IBJ7A006F734: Anyone else want to add anything?

[00:08:50.01] PBR9D1AD41E7: I think just for pure information purposes. If you're just interested in learning more about a very specific subject, or a very specific person, if that happens to be your hobby. Some of these projects are, give you the opportunity to go very granular. And you know if I'm looking at the Letters [of 1916] and I have a very specific letter from a very specific person...maybe you're related to that person which interests you personally.

[00:09:27.23] PBJ560B377FD: Similar to what your saying, these projects, you might go looking for one specific thing and you find out more information than you intended to find out. And there you go. Your scope knowlege has increased and you've just spent a few hours looking through something....

[00:09:48.07] PBR99A4D800B: Sometimes you can be very surprised...hopeful

[00:09:58.23] PBJCA47575EA: Sometimes you can be very frustrated...

[00:10:00.02] PBR99A4D800B: That's why I said "hopeful"

[00:10:01.10] PBJ560B377FD: It's interesting as well sometimes that the frustration you have when you can't find something new leads to another digital repository. And you find out "Oh there isn't enough information in this particular one" but another example of something similar is online or can pick up the loose threads at that point. So there does seem to be a point at which the ends unravel but then they begin again in another repository.

[00:10:27.27] PBR9D1AD41E7: I was thinking of the same metaphor; you just keep pulling at the loose threads until you find something else and that leads you to the other guy.

[00:10:38.08] PBR99A4D800B: I think it teaches you to be open to the possibilities when you find these loose threads. You then think "oh I never thought about that; let's try that angle." And I think that's very useful. If the site is well done, it leads you on to other things and I think it can be very illuminating. But also my friend has this wonderful phrase: "down rabbit holes". And you can end up following something because it's really interesting and loose the thread of what it is you're doing. But I do think it's great...I think they're very useful.

[00:11:14.12] PBJ560B377FD: One thing that sort of curtails that experience though is the issue of access. Similar to what you were saying about going into town to visit these places can sometimes stimy the research that you're doing. At one point if you are pulling at loose threads and coming up with this very interesting thing, then you come against a pay wall. Or come up against something that says "Oh this content's available for €5.99" well that can be blocked. Even your idea is lost and you're just like "Nope. No more." Or you go to try to find another example.

[00:11:44.13] IBJ7A006F734: So would you say that the pay walls are definitely kind of gatekeepers in a sense, in that they keep you from...you're less likely to pay. You're just like "No way" say...

[00:11:59.04] PBR1B0A881F1: Yeah their off putting. Like Q. Most of it's like "You have to pay to get a print" or "Come in" and it's like it's not available. But then it's kind of like "Why do you have it digitised as a record if I can't see it? Why can you let me find it and then just be like 'No'".

[00:12:15.28] PBJ560B377FD: That's very much true. I understand why pay walls exist or why they need to but at the same time it is very much a gate, something that shuts off this resource. It tantalises you when you're there like a preview or something like that but you can't actually access the material unless you fork out money for it. Which again, I understand why but it is awkward.

[00:12:34.15] PBR9D1AD41E7: Particularly for researchers who are often not the most well-funded group of people. I was just going to add to PBJ560B377FD's idea of access goes hand-in-hand with accessibility. Access is one thing but it's also making sure it's accessible to the greatest number of users, many of whom, or some of whom, may not be able to read or see or hear in the same way that other people might be able to. So if it's just content in one form, you're cutting off potential other users if they're not able to access that because of their own unique...

[00:13:29.02] PBJ560B377FD: Limitations, yeah

[00:13:31.21] PBR9D1AD41E7: Limitations or abilities.

[00:13:33.20] PBJ560B377FD: It's interesting then as well, is there a conversation to be had about the level of fidelity or quality...like if they are digital copies, if they are photographed to the highest possible quality, say TIFF images versus JPEGS. The issue of access...it's one thing to have it open or available but if it's such large

data files, you're then locking it to people who might not have a high-speed internet connection or might be on dial up or something like that. So again, you're limiting....

[00:14:01.07] PBJCA47575EA: High speed is pretty common these days, though?

[00:14:05.15] PBJ560B377FD: No

[00:14:06.02] PBR9D1AD41E7: Well it depends.

<GROUP DISCUSSION>

[00:14:21.05] IBJ7A006F734: Ok um, just very briefly then...the other thing I want to look at is...how many of you have explored things like Mapping the Republic of Letters or Kindred Britain that leverages these types of data visualisations. And if you have used them, <MUTTERING>, do you find them useful? Are they difficult to read? Do they make it easier to explore? What are your thoughts, specifically on these types of data visualisations?

[00:14:55.06] PBJCA47575EA: I've very briefly looked, very unimpressed...especially at the last one.

[00:15:02.02] IBJ7A006F734: With the Kindred Britain?

[00:15:03.01] PBJCA47575EA: Yeah

[00:15:03.23] IBJ7A006F734: Why?

[00:15:06.29] PBJCA47575EA: Because it seemed like these projects that I've come across...they claim to have certain resources...but you get to a page filled with data with no link to the actual thing you're looking for. Or links that don't go anywhere.

[00:15:35.21] PBR99A4D800B: I'm using this kind of thing in my doctorate and I find you have absolutely have to have an explanation as to what that means. Because looking at that I just think, wow that's interesting and obviousvly there's a huge body of work behind that but you have to look at that body of work in order to appreciate what you've got there. You can't just take that map and say oh yeah that's it! I've got it!

[00:15:58.08] IBJ7A006F734: So there not really readable....?

[00:15:59.17] PBR99A4D800B: I don't think they are like that. They're kind of a process that you visualise what's on the page but you need the whole....

[00:16:07.17] PBR1B0A881F1: They're like a prompt. They like tell you to go look into it. They like kinda present an idea that you then might go and explore but you might not actually be able to extract anything from them, themselves. Like I wouldn't really get anything too much from that. But it would encourage me to look at the sources for that more closely, for like geo coordinates.

[00:16:28.04] PBR99A4D800B: It's very likely that the blue is everything mentioned in this person's letters. And so you...and the lighter blue is a heavy line so therefore he's done this trip many times and discussed it much or there's a connection...so you don't know quite what all these lines really really mean and you would absolutely have to read the letters.

[00:16:56.00] PBJ6BAA5509A: I suppose these kinds of visualisations <AUDIO UNCLEAR> present a perspective because all these kinds of visualisations have certain biases that you have analysed and <AUDIO UNCLEAR> certain models

and you have presented to us. Effectively for me this presents a way you can use the data stored in whatever repository this person has. And as you said [*points to PBR9D1AD41E7*], if there is no link from this perspective onto the data itself, then it is entirely worthless.

[00:17:34.15] PBRF60C8C611: I think it's partly the same. I find ... I come at this from a slightly different perspective, which is quite unpopular in the DH kind of course that I'm doing that is kind of scholarship for scholars. So I a lot of the time might go into these visualisations or these websites and find myself thinking "That's a fantastic resource. I don't know how to use it to advance my work. How does this benefit my research." So...and then I think images without context don't have meaning. They're just an image.

[00:18:14.20] PBJCA47575EA: Actually, it fits in with what you were saying earlier <AUDIO UNCLEAR> just give me the context, I don't know...I don't want to try and work out your research from a single page <AUDIO UNCLEAR>. Where's the help file? Just because it's online doesn't mean it's accessible in any way.

[00:18:38.10] PBR1B0A881F1: I was just going to say what PBJ6BAA5509A was saying with perspective...

[00:18:39.14] PBJ6BAA5509A: Anonymity

[00:18:41.19] PBR1B0A881F1: Sorry?

[00:18:42.01] PBJ6BAA5509A: Anonymity <*LAUGHTER*>

[00:18:43.24] PBR1B0A881F1: Sorry. As he was saying with perspective, if I don't know...like it says there that only a certain amount of them have location information but I would kind of want to know how many do and what; how many <*AUDIO UNCLEAR*> is that the large majority do or they don't? Because otherwise that could only be 5% of correspondence and then that doesn't actually really tell you that much because it's such a small amount. So we need much more definitive...not methodology because that would be too long to accompany a visualisation but kind of a better context of how that research worked.

[00:19:25.07] IBJ7A006F734: Ok. Great! Thank you. That was all very helpful. Alright, so now we're going to move on and I'm going to collect your cards.

(ACTIVITY) CARD SORTING

[00:20:13.00] IBJ7A006F734: Now...I'm going to use some post-it notes. Alright. And for the next um, 5 minutes: list all the things that you would like to be able to do with these type of DHP style resources. Whether it be something that as simple as: I want to be able to search. I want to be able to browse. I want to analyse word frequency. I want to look at multiple collections. You mentioned the issue of accessibility so I want audio transcriptions or whatever the different thing may be on each piece of paper, so for each idea use a different sticky note. So um, but for every little thing that you want to do just write down one word or phrase to describe it and put it on sticky note and then put that sticky note aside. Generate as many sticky notes as you want in the next 5 minutes.

[00:21:31.04] PBJ6BAA5509A: One word?

[00:21:31.14] IBJ7A006F734: One word or phrase to describe the idea. One idea per note.

[00:24:37.28] IBJ7A006F734: Two more minutes

[00:26:28.17] IBJ7A006F734: Ok go ahead and finish writing whatever you are working on. Alright now I want you to bring all your notes with you and come over here. As a group I want you to bring your ideas and then put them into these three categories. So how you as a group view them. Are they very important and are things you absolutely want to see. Are they somewhat important. Or we don't really care or really know how we feel about them. Just kind of group them into these 3 different categories. Alright? Just kind of work together as a group to kind of reach a consensus. No fighting! <*LAUGHTER*>

[00:27:22.06] PBJ6BAA5509A: Open Access. Very important.

[00:27:33.20] PBR99A4D800B: That one's crooked up there. Try to put it neatly. Surely we are all obsessive compulsives if we are doing our doctorates.

[00:27:50.08] PBJ560B377FD: Anyone else have something they feel is very important.

[00:27:52.13] PBR99A4D800B: Create a library or resource for future access.

[00:27:56.15] IBJ7A006F734: That's a kind of um like an archive. Preservation.

[00:28:04.24] PBR99A4D800B: That's somewhat important I think. Not always.

[00:28:11.04] PBR9D1AD41E7: I think it depends on the content.

[00:28:12.13] PBR99A4D800B: Yes exactly.

[00:28:19.13] PBRF60C8C611: Next um, specifically I also have create and contribute. They weren't specific I suppose to a library or a resource but I suppose.

[00:28:26.14] PBJCA47575EA: Something that anybody can join in.

[00:28:27.16] PBRF60C8C611: Um no. That's not the <AUDIO UNCLEAR> <LAUGHTER> but to be able to create something or contribute to something. So not just to be a user but a contributor.

[00:28:43.19] PBJ560B377FD: I think that's important. I had user forums, so a place for user's to engage in discussion.

[00:28:50.27] PBRF60C8C611: I had that too.

[00:28:54.27] PBJCA47575EA: That one yes. I think that's very important.

[00:28:54.28] PBJ560B377FD: Very important or....

[00:28:54.27] PBJCA47575EA: I think it's very important

[00:28:55.26] PBRF60C8C611: I think it's important.

[00:28:58.17] PBR99A4D800B: You disagree?

<*GROUP DISCUSSION*>

[00:29:03.19] PBJ6BAA5509A: I would say that's somewhat important.

[00:29:07.12] PBR9D1AD41E7: I'd say it varies depending on the content.

[00:29:10.26] PBJ560B377FD: Yeah it depends on what the material is.

[00:29:15.14] PBJCA47575EA: I guess for me there's just no way to get feedback or ask questions, so a user forum would be really important to me.

[00:29:21.23] PBR9D1AD41E7: I have <AUDIO UNCLEAR>

[00:29:23.26] PBJ560B377FD: I also have user feedback functionality. Some sort of web page where you could leave us information about the project.

[00:29:37.16] PBR99A4D800B: How about just a general search?

[00:29:38.08] PBJ560B377FD: Aye that's very important.

[00:29:41.09] PBJCA47575EA: A search that works.

[00:29:43.00] PBR1B0A881F1: If we have it multiple times do we just put one up?

[00:29:44.23] IBJ7A006F734: If you have them multiple times you just kind of cluster them around each other.

[00:29:55.04] PBJ560B377FD: I did an example for just "text searchable".

[00:29:59.05] PBR99A4D800B: Yeah that's grande.

[00:30:03.10] PBJ560B377FD: Would that be...I suppose somewhat important.

[00:30:07.13] PBR1B0A881F1: I kind of have this feeding into word search is like associated word. So I search for "British" but then I have to search "British" and then "Britain" and then "English" but it brings up similar type of contexts.

[00:30:20.20] PBJ560B377FD: Yeah that's a good idea. But I think it depends on the specific...

[00:30:24.15] PBR1B0A881F1: It does, yeah. But yeah it does depend on the specific resource.

[00:30:27.25] PBJ560B377FD: Mine sort of centres about just text annotation. Which is that if that term comes up you could hover over it and see specific information.

<GROUP DISCUSSION>

[00:30:44.18] PBJ6BAA5509A: How we use the resource.

[00:30:46.03] PBJCA47575EA: A lot of places have no explanation as to how find your way through.

[00:30:52.06] PBR99A4D800B: Ah ok navigate through.

[00:30:52.18] PBJ6BAA5509A: And maybe a tutorial...

[00:30:53.14] PBJ560B377FD: That'd be very good.

[00:30:54.15] PBR1B0A881F1: That's fantastic!

[00:31:01.26] PBJ6BAA5509A: I had 3 other things that were related which are: a clear statement of use, "How do I use this data?"; a clear note on data authenticity, "What is the source of this data? How authentic is it? How clearly...how was it made?";

[00:31:15.29] PBJCA47575EA: By clear can we say "lack of jargon".

[00:31:19.02] PBJ6BAA5509A: Yeah.

[00:31:18.18] PBR99A4D800B: That's very important. Clear *<AUDIO UNCLEAR>* to authenticity. I think that should be at the very top. You want to use...rubbish. So you want a clear statement of authenticity.

[00:31:33.14] PBJ6BAA5509A: And how I use that...what is the legal copyright...yada yada yada.

[00:31:38.25] PBR99A4D800B: Yes that's absolutely essential.

[00:31:42.05] PBJ6BAA5509A: And finally is it useful to have access to data that is downloadable? Or are you ok with...

[00:31:49.09] PBR9D1AD41E7: I had thought of that as access

<GROUP DISCUSSION>

[00:31:58.06] PBJ560B377FD: I think it needs to go with that.

[00:32:02.21] PBJ6BAA5509A: But is that something that...I'd say that's somewhat important. That's something you can do on the website itself. You don't need to download the data.

[00:32:09.24] PBRF60C8C611: Yeah not downloadable but to have access. I'd like to have access to it...

[00:32:19.00] PBR99A4D800B: You mean codes as in...you don't mean the....

[00:32:23.07] PBRF60C8C611: The script and the textual....

<GROUP DISCUSSION>

[00:32:40.18] PBR9D1AD41E7: Well we can't seem to agree so maybe we should put it in the "Don't Know"...

[00:32:44.17] PBR99A4D800B: Data access. We're all talking about that so I think that's very important. You want to be able to download it, surely, when you get it.

[00:32:53.24] PBR1B0A881F1: I think it depends on the angle. We're all approaching it from ...we're data people so....we want the data. But to a general user I don't know how important that is.

<GROUP DISCUSSION>

[00:33:08.26] PBR99A4D800B: I think easy data access is absolutely important.

[00:33:12.27] PBJ6BAA5509A: I had 3 other points.

[00:33:15.29] PBJCA47575EA: Actually download is somewhat....

[00:33:18.27] PBJ6BAA5509A: It's the same problem here around this area. The data itself. Three other ones. One is...my predominant problem is I don't know what the resource contains. It's a bunch of text. People say I have all these things but actually I don't know. So I want...this is like "Don't care...somewhat important" level. Visualise resource. What you have I want that visualised. So whether that is "This amount of what we have 70,000 words, 40,000 plays, 20 novels." But I want it as a visual narrative. So tell me what that resource overall-archingly contains. Perhaps even as a mode of access. I will..."Don't care"?

[00:34:10.24] PBJ560B377FD: I don't think it's "Don't care"...it's somewhat important. It's important to someone. It's important to you.

[00:34:17.24] PBR1B0A881F1: It's not what you'd dedicate your time to.

[00:34:19.16] PBJ6BAA5509A: It's important to me but I'm not sure as a group it's important.

<GROUP DISCUSSION>

[00:34:30.07] PBJ560B377FD: My one I think no one would care about is for textual projects. A comparative window view so you don't have to exit or go back. You have the two of them displayed together.

[00:34:37.06] PBR1B0A881F1: Yeah!

[00:34:41.24] PBR99A4D800B: Sometimes I use these things to solve a mystery. I really depend on myself or my own analysis but sometimes I need to go...<*AUDIO UNCLEAR*> Maybe you guys don't care about solving mysteries. <*LAUGHTER*>

[00:35:04.16] PBR1B0A881F1: Well we have...like annotations and context. So that's important.

[00:35:09.07] PBJ560B377FD: I'd go with that. That additional context so if it's a print out...

<*GROUP DISCUSSION*>

[00:35:47.26] IBJ7A006F734: Two more minutes

[00:35:49.02] PBJCA47575EA: No restrictions on what we can access?

<*GROUP DISCUSSION*>

[00:35:57.26] PBRF60C8C611: So anyone have...feel safe and respected. As in when I'm using these websites, that's what I want. It's not actually about the technical...it's just the user experience.

[00:36:09.03] PBJ560B377FD: It's the interaction with the team or something....

[00:36:10.28] PBRF60C8C611: Yeah

[00:36:12.15] PBJCA47575EA: Do you mean like them not collecting data on your use, or...?

[00:36:14.27] PBRF60C8C611: Um....yeah.

[00:36:18.05] PBJ560B377FD: User anonymity?

[00:36:19.21] PBRF60C8C611: But even I suppose what your talking about there's kind of a lot of about forums there as well. When you're contributing to a forum that it's a safe environment to do so.

[00:36:33.07] IBJ7A006F734: So no harassment...

[00:36:33.07] PBR9D1AD41E7: Community guidelines...

[00:36:34.29] PBRF60C8C611: Yeah. Maybe it's not ....again it depends on if you want to use these...

[00:36:44.10] PBR99A4D800B: But are these....are these actually a part of that open access though...that sounds much more intimate. And contacting real people than these imply

[00:36:55.09] PBR1B0A881F1: That has more to do with the resource....

[00:36:57.00] PBR99A4D800B: Yeah....these are resources. You're sounding like you're getting into something far deeper than what....

[00:37:01.01] PBRF60C8C611: But resources...not these ones specifically but other ones do have conversations and...I'd just like to see...I think my experience with scholarly editions in particular is that there's just like "there's the information" so I'd like to see more of a conversation around the information but then within that conversation I'd like to feel...<*GROUP DISCUSSION*> yeah yeah.

[00:37:25.19] PBR9D1AD41E7: So maybe it's about interactivity. A bit more interactivity.

[00:37:29.26] PBRF60C8C611: Yeah but monitored.

[00:37:32.23] PBR9D1AD41E7: Monitored interactivity.

[00:37:35.26] PBJ560B377FD: Sounds similar to like transcription based <AUDIO UNCLEAR>...that your interaction, your contribution is regarded well...

[00:37:41.19] PBRF60C8C611: Well even with the Letters [of 1916] because people aren't transcribing and then getting an email back going "No this was rubbish", which is not happening.

[00:37:51.00] PBR1B0A881F1: Oh I do that all the time. *<LAUGHTER>*

[00:37:52.27] IBJ7A006F734: Alright so we need to wrap...so any other ones you have...

[00:37:56.16] PBJCA47575EA: Well the actual content, not just keywords. Not just a set of ... *<AUDIO UNCLEAR>*

<GROUP DISCUSSION>

[00:38:05.04] PBJ6BAA5509A: I think that goes with the data authenticity thing. Plain old data from an authenticity point of view. I just want the damn data. I don't...access for data right over there.

[00:38:13.13] PBR9D1AD41E7: I had sort and filter data separate from search because I think once you get results, you should then be able to...

[00:38:21.09] PBJ560B377FD: Filter it out...

[00:38:22.13] PBR9D1AD41E7: Filter down.

[00:38:24.06] PBJ6BAA5509A: I had serendipity. Actually find ... just to be amazed by something! <LAUGHTER>

[00:38:30.05] PBR99A4D800B: That's really important important. I think that sometimes has been the most wonderful thing in some of my research...complete serendipity and people have been shocked, I've found. So it's really important that you are open to that.

[00:38:45.01] PBRF60C8C611: I have textual and linguistic analysis which I think is just specific to me so I can just stick that over here....

[00:38:50.12] PBJ6BAA5509A: No no no...you care about it...

[00:38:55.26] PBR99A4D800B: No I think that's important because if you can't look at your text cause that's how we do it...don't we all do it...that's a key thing, isn't it? Really...

[00:39:09.16] PBRF60C8C611: Yeah I suppose...yeah yeah. Sorry.

[00:39:16.20] PBJ6BAA5509A: I suppose that also feeds into PBRF60C8C611, if you over-visualise resources....

<GROUP DISCUSSION>

[00:39:22.27] PBJ6BAA5509A: ...you push it out on one side. This is perhaps more to filter out one text itself or ...but still provide some kind of yeah...related structure.

<GROUP DISCUSSION>

[00:39:36.10] IBJ7A006F734: You guys I'm the one that has to make sure everyone's anonymous. If you know each other and want to say each other's names and nobody cares, it's fine.

<GROUP DISCUSSION>

[00:39:46.27] IBJ7A006F734: Anything else anyone wants to put up there before we...

[00:39:49.07] PBR1B0A881F1: Mine are just minor like visualisations and having thematic collections so I kinda know what I'm getting into...so it's organised.

[00:39:58.15] PBJ6BAA5509A: Yeah I think it's the same thing we've been talking about it.

[00:40:01.03] PBJCA47575EA: It's kind of related to search as well.

<GROUP DISCUSSION>

[00:40:11.17] PBR99A4D800B: I have more but I don't think it's quite what you are looking for. It's what I use these things for.

[00:40:15.29] IBJ7A006F734: I'm certainly interested....

[00:40:17.07] PBR99A4D800B: I use them to illustrate my ideas...confirm my own ideas. If I'm floundering, I can find somebody else more important than me who has said things.

<GROUP DISCUSSION>

[00:40:34.23] IBJ7A006F734: Alright. Brilliant! Thank you! Ok take a seat, briefly.

(ACTIVITY) HOW NOW WOW

[00:40:46.00] IBJ7A006F734: Alright so the um, the next exercise that we are going to do, and we're going to have to kind of go through this one quickly because we are running out of time. I want you to take all the ones that you deem in your very important category and then we are going to move them to this sheet over here. This is called my "How, Now, Wow" grid. So, we're going to look at, the kind of axis here is along the Y axis...the higher up on this chart it goes the more "impossible" it is to do. So if it's just like...if one of you had the idea of "beaming yourself into the computer"...that would be impossible. And then lower is it's very easy to do. Then along the x axis, we're looking at these are things that are you know kind of normal, run-of-the-mill ideas, things that are already kind of happening or that are pretty common place versus this is something that's not really being done. This something that's really new or different. So kind of take all your stuff that's in this very important category and see where it fits over here for me.

[00:42:07.09] PBRF60C8C611: Just this...the very important category.

<GROUP DISCUSSION>

[00:42:14.00] PBJ6BAA5509A: It's done and it should be done.

[00:42:18.00] PBR1B0A881F1: Do you want me to take all the ones that are together...?

[00:42:19.26] IBJ7A006F734: Yeah keep them all together...

[00:42:26.20] PBR1B0A881F1: So now and...

[00:42:30.09] PBJ6BAA5509A: ...is it easy to do.

<GROUP DISCUSSION>

[00:42:38.04] PBJ6BAA5509A: Are we talking technically or infrastructurally?

[00:42:41.03] IBJ7A006F734: However you want to clarify it...

[00:42:42.25] PBJ560B377FD: Ok if it's infrastructure, it's near the top.

[00:42:47.01] PBR1B0A881F1: Ok so what? You want me to go up to impossible?

[00:42:47.27] PBJ560B377FD: No no, it's not impossible but its...

[00:42:50.10] PBJ6BAA5509A: Yeah it's in the middle. It's not technically difficult but infrastructurally might not be easy.

[00:42:56.15] PBR9D1AD41E7: I would say search is down. Everyone does it, it's already being used...

[00:43:11.26] PBR1B0A881F1: What about getting to the level of having associated words. I don't come across that often in a search.

[00:43:19.25] PBR9D1AD41E7: We could maybe put those a little...

[00:43:21.24] PBJ6BAA5509A: Maybe more towards an original idea...you don't see that very often.

[00:43:24.15] IBJ7A006F734: Yeah if one of the items in the cluster goes little bit differently, you can move it off the cluster...

[00:43:36.15] PBR1B0A881F1: So this is normal and easy. And this is slightly more over? Not that difficult but not really <AUDIO UNCLEAR>. Ok. What's next? Help files / FAQS

<GROUP DISCUSSION>

[00:43:50.29] PBJ560B377FD: Yeah that should be right at the bottom.

[00:43:55.00] PBRF60C8C611: It's technically easy but they're not actually that common.

[00:44:00.02] PBJ6BAA5509A: They're not common...but it's a normal idea...

<GROUP DISCUSSION>

[00:44:08.02] PBJCA47575EA: I feel like they're never done because the people who are designing these databases are designing them for people who understand databases.

[00:44:16.00] PBJ560B377FD: Yeah yeah or they're too familiar.

[00:44:17.22] PBJCA47575EA: Rather than end users who might not have any type of knowledge.

[00:44:22.11] PBR99A4D800B: The word is "self-evident". If the people who are making it feel it is self-evident, they don't explain it. You have to explain it.

[00:44:32.21] PBR1B0A881F1: Ok. What's next. Easy data access, access the code, download the data...

[00:44:34.29] PBJ6BAA5509A: Yeah they are all on the access one.

[00:44:37.16] PBJ560B377FD: Yeah

[00:44:46.29] PBJCA47575EA: So you were talking about having to get access to code. I came across someone who was claiming this was all accessible, downloadable but you just have to install a virtual machine and these 5 different programs...

<LAUGHTER>

[00:44:58.22] PBRF60C8C611: So how accessible is that...

[00:45:05.14] PBJ560B377FD: Uh clear statement of use.

[00:45:07.09] PBR1B0A881F1: Clear language. Clear statement of use, clear language. That's very easy to do.

[00:45:14.08] PBJ560B377FD: Yeah well it should be.

[00:45:17.00] PBJCA47575EA: Actually I don't know if it is. I mean it can be very difficult to explain concepts that we are working on in easy...

[00:45:24.00] PBR1B0A881F1: Yeah. I won't make it very easy.

[00:45:26.12] PBRF60C8C611: Particularly when you are in that area...people tend to forget what they didn't know when they started out.

[00:45:34.11] PBR1B0A881F1: So it's not exactly easy...Somewhere in the middle. And normal but maybe not always used. Ok. Clear statement of use. Clear note on data authenticity. Content. Find a scan of the original.

[00:45:51.23] PBJ6BAA5509A: Again, infrastructurally difficult but also might not be terribly easy to implement I think. Because this comes down to legal issues, 99% of the time. People who are working on these projects, such as yours truly, won't know the legal. And they vary from country to country. So that's a different problem entirely.

[00:46:17.09] PBR1B0A881F1: Ok so...

[00:46:19.04] PBJCA47575EA: If you're pulling yourself ... a repository of a certain type of data isn't there and you're just pointing to it, it's not really <AUDIO UNCLEAR>

[00:46:27.25] PBR1B0A881F1: So it's not easy. And then normal idea, original idea.

[00:46:31.10] PBJ6BAA5509A: Ok let's put it this way that we won't see this very often.

[00:46:35.17] PBR9D1AD41E7: So maybe somewhere in the middle.

[00:46:37.24] PBJ6BAA5509A: Yeah somewhere in the middle. Might not be terribly easy to implement but...

[00:46:41.26] PBR1B0A881F1: Ok. Effective, intuitive design, working links, see more on one screen.

[00:46:48.00] PBJ560B377FD: Working links should be...

[00:46:50.16] PBR1B0A881F1: Less links to click

[00:46:53.24] PBJ560B377FD: Again design is ... its both current and...yeah it's subjective. One person's good design could be another person's nightmare.

[00:47:03.27] PBR9D1AD41E7: I feel like it should be a normal idea but it's not. So maybe put it somewhere in the middle?

[00:47:10.11] PBJ560B377FD: Yeah. And a little more common. Because designers are commonplace now.

[00:47:14.03] PBJCA47575EA: I think that kind of what you are saying it's self-evident <AUDIO UNCLEAR> yeah here's the data! I know all this...I understand it...

[00:47:24.15] PBR1B0A881F1: Ok. It's done.

[00:47:26.15] IBJ7A006F734: Brilliant! Thank you guys. Have a seat.

(ACTIVITY) 6 PART STORY METHOD

[00:47:51.26] IBJ7A006F734: Alright, our next exercise is what I call the 6 Part Story Method. So you each need to get a piece of paper. And you'll all need a pen or pens or markers or whatever you want to use. This is a drawing exercise. So feel free to use different colours or anything else. Alright so...we're going to start. Take a piece of paper. Hold it horizontally and we're going to fold it into thirds.

[00:48:33.24] PBR99A4D800B: Is this a trick question?

[00:48:34.28] IBJ7A006F734: It is not a trick question. You'll have 3 columns if you want to unfold it. Alright and then fold it in half. So that when you are done, you have 6 boxes. So open it up. Alright so, in the first box, in the upper left hand corner. I want you to...so what we are going to do, is we are going to construct a story. We're going to construct a narrative about how the particular person, who will be a person of your choice, would use a digital humanities project, or one of these types of resources. So, in the very first box, you're going to draw me a picture of who your main character is. Are they a student? Are they a researcher? Are they grandma researching her long lost sister. Is it..you know who is it? Draw me a picture of your main character of your story is going to be. It should be in the upper, right-hand box. Or left-hand box. Sorry.

[00:51:15.00] IBJ7A006F734: Oh and please be realistic with your stories. So no um supernatural animals or anything...

[00:51:34.28] IBJ7A006F734: No magical giraffes.

[00:51:42.02] PBJCA47575EA: You are limiting people's imaginations! <LAUGHTER>

[00:52:02.29] IBJ7A006F734: Ok. In our second box, where you're going to draw me picture of what the main character is trying to do with this resource. What is the task they've set out to do. What is the thrust of why they are involved in the story. What is the purpose. What is it they are trying to accomplish.

[00:54:18.25] IBJ7A006F734: Ok, in the third box. Draw me a picture of something that gets in their way. Something that hinders them from accomplishing what they are trying to accomplish. Something that blocks them. Something that stops them, related to the digital humanities, the resource that they are using. Always tie it back to the resource that they are using.

[00:54:41.07] PBJCA47575EA: Do we have to be positive?

[00:54:43.11] IBJ7A006F734: It doesn't have to be positive. As long as it relates to the resource that they are using. Something that gets in their way.

[00:56:23.05] IBJ7A006F734: Ok in the bottom box, over here, underneath your main character. Draw me a picture of something that...something within the digital humanities resource that is helping them. Note that this should not be a solution to the problem, the hindrance that they get. It should be something else that is helping them accomplish their task. So it doesn't necessarily negate this. But it's something that is specifically contained within the DHP.

[00:58:51.00] IBJ7A006F734: Alright, in the bottom middle box. Every story, has it's climatic moment. It's "all is lost" or you know the big thing that happens towards the end of the story. What is that thing that happens while they are using this resource, what is it, you know did they stumble upon a specific piece of information or did they look at visualisation and go "wow! everything clicks". What is that one thing that happens that is the turning point in their story as they are using this resource. Draw me a picture. And again, tie it back to something that happens within the Digital Humanities Project.

[01:01:11.14] IBJ7A006F734: Ok. And in your very last box. What's the end of your story. Draw me a picture. And when you're finished just put down your pen or marker or whatever you're using, so I know you're done.

[01:03:27.12] IBJ7A006F734: Ok. Alright so who wants to share their story first.

[01:03:36.13] PBJ6BAA5509A: Ok. I'm going to stand up first. My character is called Richard. You don't know Richard. He likes...ok he is this hypothetical character I imagined with shaggy hair and beard and wears these long coats and lovely sweaters.

[01:03:55.29] PBR99A4D800B: And glasses. And wears glasses?

[01:03:58.11] PBJ6BAA5509A: Um yes. Sometimes when he reads he wears glasses. Yes.

[01:04:00.22] PBR1B0A881F1: Wonderful character development.

[01:04:01.19] PBJ6BAA5509A: Yes. He likes cigarettes. He likes the python programming language. The TEI. Not in any specific order. So he is a PhD candidate. And he's using these, in the second panel, he's using these resources to write his dissertation to get a better understanding of DSEs, like someone we know. But then the problem happens. Where he suddenly realises that he can't get the files that it has on the site that he's using in the format that he needs it in, which is Text Encoding Initiative (TEI) files. And he is heart broken because his PhD is not going to go anywhere unless he finds the TEI format. But Richard at the end he's fairly happy because he knows where the data comes from so at least he knows that's ok. "You know what, the data is fine. I know where it comes from. I know it's complete." So he's satisfied. Then he suddenly realises or that the project makes these files available in a different format. But it's still a standard format. And Richard is delighted because he can then convert it to TEI and then here, Richard gets his PhD. He takes a boat back to his home town, his homeland in the U.K. and he and Lucy who is another fictional character, get married. And the end!

<APPLAUSE>

[01:05:40.16] PBR99A4D800B: I think you need to run that...like always.

<GROUP DISCUSSION>

[01:05:54.25] IBJ7A006F734: Who would like to go next?

[01:05:57.29] PBJ560B377FD: Ok. So mine starts off with the fictional character, "General Public". So General Public one day, smoking his pipe, cries "Core Blimey! I would like to find, access documents related to the first World War." And decides

to do this by visiting a big genealogical site. And he visits big geological site in the panel there. And he says "Blast!" because there's a pay wall and he can only access those records for €9.99. But then in the fourth panel he comes across a solution and strokes his moustache and sees there's a free trial function. And "I can abuse the free trial function", which allows me to access the resource. Then in the 5th panel he remembers reading in the local newsletter, that there is in yonder university, a smaller project which is doing digitising <AUDIO UNCLEAR> resources which he needs. So he decides to support said project in the final panel. And everyone is happy because he can give them very good constructive criticism and feedback over what he needs to do and in turn, they provide him with a resource that he requires to do his research. Thanks.

<APPLAUSE>

[01:07:13.28] PBJCA47575EA: Hmm. My fine art skills I have...this is a frustrated student, Jebediah. He's got a pink tie. And his problem is that he's working on text which other people have transcribed from digital documents. And the problem is he wants to see the original manuscript. And the internet is completely useless because his computer is broke. The obstacle is this stupid web designer, the person who was designing this resource who thinks they know what they are doing but clearly don't. So the solution is to use the search box to find where is the library. Climax. Books good! Computer bad! And finally the ending poor Jebediah is having a nightmare, saying "Make it stop!" And the devil has invaded his dreams saying "your research will never be complete!".

[01:08:18.20] PBRF60C8C611: Oh I've had that nightmare!

[01:08:21.25] PBR9D1AD41E7: That took a turn I did not expect.

[01:08:32.17] PBR9D1AD41E7: I like colours. So my character is super student who likes books and computers...sometimes. And her problem is that she has to face the viva committee. And she's using these resources to finish her dissertation, when she gets the dreaded 404 error because whoever designed it, didn't know what they were doing. But thankfully a smarter person also enabled a download button, so she can just download the data and do whatever she wants with it. And then she has her "Eureka!" moment staring at an endless page of spreadsheets and data that looks very much like dots on a page. And she very happily finishes her dissertation and gets her PhD!

<APPLAUSE>

[01:09:33.12] PBR9D1AD41E7: And she even has the little hat.

[01:09:40.06] PBR1B0A881F1: My character doesn't have a name but like PBR9D1AD41E7 she also has a computer and books. So she's trying to look through the diaries of a certain figure because she's using them as a case study. And she's trying to find particular concepts to link together so this is about Irish female power. And the obstacle she comes across is that it is restricted. Or for no reason has a system error and will not find any results. And there's no help because the resource is pretty useless and they just uploaded texts. And then it can go either of two ways. She can try a new

word and have that work and get results. That's nice. Or you can't find anything so you throw the computer in the fire. Or out the window. It's open to interpretation. And as a result it could also go two ways. If you throw the computer in the fire, you can get a very boring, one-dimensional piece of work. Or you can get a lovely piece on Irish female identity in colonial India that presents many ideas and works out very well!

<APPLAUSE>

[01:10:49.26] PBRF60C8C611: So um that's my...I gave up on the *<AUDIO UNCLEAR>*...my character is a researcher / postgrad / postdoc or academic. That's their degree on the wall there. And their ultimate goal is to I suppose advance their career, so through publication on maybe an open access journal, their CV is there. Their thesis or monograph or oh just a normal research paper....so then I drew this sort of maze but a maze is really hard to draw so...but its basically the top is like it says "Public Access Only" so that members of the public can just bypass this because a lot of these projects are being made with the public in mind. And then "Academic Access" is for the...is the maze but I decided to give up on the maze. The helping thing is an office with an open door. So Digital Humanities Scholars, academics in general, really want people to come engage. My experience over the last few years is that they are just...they are more than happy to help. This researcher is a medieval Irish researcher for no particular reason. So what happens is they come across resources online like "Irish Script on Screen", "CELT" and "EDEL" and those resources will help them with their traditional research. So that's why you got the pen and paper there. And then end of the story is that this particular researcher ends up thinking that Digital Humanities is quite similar to a set of tools.

[01:12:39.12] PBJCA47575EA: Digital Humanities is a set of tools?

[01:12:43.26] PBRF60C8C611: Yeah! I'm not saying that is...I don't think it is.

[01:12:48.15] PBJCA47575EA: Just a bunch of tools?

[01:12:51.09] PBRF60C8C611: It's the ultimate...this is a narrative! *<LAUGHTER>*

[01:13:00.07] PBR1B0A881F1: Safe space! I didn't call PBJ560B377FD names. *<LAUGHTER>*

[01:13:06.16] PBRF60C8C611: So it's great than I can see a manuscript on screen. That's great because then I don't have to travel to that library just to see the document any more. So I can see how that helps me with my traditional-based research. But I can't see how these tools might advance other types of learning or research. Sorry.

[01:13:29.17] PBJCA47575EA: So not like how IBJ7A006F734 is a tool. *LAUGHTER>*

[01:13:34.03] PBR9D1AD41E7: You're not suppose...

[01:13:36.24] PBR1B0A881F1: You are in rare form! *<LAUGHTER>*

[01:13:39.21] PBR99A4D800B: He uses it as an adjective instead of a noun.

[01:13:42.05] PBR1B0A881F1: I feel like we should move on to the next one...

<APPLAUSE>

[01:13:48.25] PBR99A4D800B: An individual who is mature and really wants to do something besides housework. So decides to go to college and is given a whole series of assignments and also has ideas of her own. So she could be researching for a specific assignment but the problems are, is that you have access or things that are downloaded or digitised. You have no idea what's in the book. You just don't know where to begin. So you start by looking at the websites that are offered in the topics that you have. You get links which are frustrating. You get advanced...um searches which often don't result in anything. Very frustrating. But finally, after a great deal of effort, you get a result. And then you print everything. And print everything. And print everything. And you can make a book out of it. And you get your PhD hopefully.

<APPLAUSE>

[01:14:58.13] IBJ7A006F734: If everybody could just put their initials on the back of each paper and then I'll come around and collect those. And then you'll both tear out a piece of paper each.

[01:15:33.12] PBR99A4D800B: Are you working on *<AUDIO UNCLEAR>*?

[01:15:36.28] PBR1B0A881F1: I did a *<AUDIO UNCLEAR>* on Irish Female Identity and Power Structure

[01:15:40.29] PBR99A4D800B: Did you do...did you see recently on ... sorry. That was besides the point. I just found a great journal article about Lady Curzan (1898-1905).

[01:15:58.14] PBR1B0A881F1: I was looking at Lady Dunford.

[01:16:00.00] PBR99A4D800B: Yes. Oh really good. Very nice. I haven't looked at her. She's too late for me. I'm looking mid-nineteenth century. So...

(ACTIVITY) ANTI-PROBLEM

[01:16:07.26] IBJ7A006F734: Alright so the next project we are going to do...this is called the Anti-Problem. So, everybody should hopefully have a piece of paper at this point or going to in two seconds. Take that piece of paper and fold it in half, right down the middle. Sorry, length-wise, so you create columns. Now what we are going to do is I am actually going to propose a topic to you that is absolutely ridiculous. And you're going to come up kind of the answer to that problem. The reason we call this the "Anti-Problem" is because the problem I'm going to propose to you is obviously something we would never ever do. So in this situation your problem is, how do we make it impossible to read in these types of resources. So, for example, we could make it black background with black text. Obviously we can't do that. So that would be an idea you would jot down on the left-hand side of the paper. So for the next 5 minutes or so, I want you to go through and along the left-hand side, write down various different ways we could make it impossible to read in a DHP.

[01:17:37.16] PBJCA47575EA: Are you assuming that there is actually information there?

[01:17:38.19] OBJA85D3875B: <*LAUGHTER*> Delete!

[01:17:40.25] IBJ7A006F734: I don't know. That would certainly make it impossible to read!

[01:21:46.11] IBJ7A006F734: Ok you have 30 more seconds.

[01:22:35.28] IBJ7A006F734: Now on the right hand side of the paper, I want you to basically write the anti-thesis to the answer...to whatever solution you have on the left-hand side. So what is kind of the solution to your solution. So if you wrote on the left-hand side of the paper "Black text on black background" then you should have a ... to make it, you're now making it readable so "Black text on white background" or "White text on a black background".

[01:25:44.10] IBJ7A006F734: About 30 more seconds.

[01:26:31.26] IBJ7A006F734: Ok. Alright so who wants to share with me some of their ideas for how you make it impossible to read in these kinds of things.

[01:26:43.00] PBJ6BAA5509A: I just found this really hard. Because you said impossible, you didn't say <*AUDIO UNCLEAR*>.

[01:26:51.17] IBJ7A006F734: No I didn't say <*AUDIO UNCLEAR*>

[01:26:53.19] PBJ6BAA5509A: Font size too small and there is no way to increase the font size. Literally can't read it. Make it bigger. <AUDIO UNCLEAR> Use a font colour, poor contrast between the background and text. Please use contrast. Simple solution to that. One I have for myself is, the text is behind a pop-up that won't close. They forgot the close button. Don't have popups. Just not a good thing. <AUDIO UNCLEAR> Provide markup language. And the others were practical, but text as an image that is too large to open. Keeps me from signing out. I keep seeing a grey screen. Could be a PDF or a very very very 500MB image that doesn't load over a browser. Don't use images as a placement for text and use smaller file sizes. That was my five. That was difficult.

[01:27:59.00] PBJ560B377FD: I have a lot of what you said. Image is too low resolution to read. If it's say a digitised something like that...For something else that's possible if there's, if it's broken links, you can't access it, it's impossible. I said if it's a transcription based thing, these next three run together. But Unicode errors, transcription errors, where you can't actually read the document. I also said if it's a case of visualisation or something like that, well no key for visualisation, too much information, or if there's no logical starting point or narrative for the user to follow. I guess the solution for that is try to add a good key, that's informative and tells you where to start.

[01:28:47.17] PBR9D1AD41E7: I had stupid fancy script fonts. So...

[01:28:53.17] PBR1B0A881F1: I'd say wingdings...

[01:28:54.29] PBJ560B377FD: Wingdings

[01:28:58.15] PBR9D1AD41E7: Wingdings or webdings and my solution to that is just keep it simple, sans serif fonts. No need to get fancy. I also have exorbitant pay wall. So as opposed to just like pay wall that you don't want to pay, a pay wall—I've come across a couple of reports that I need. And they are like $500 please and

I'm like "No". Not giving you $500. So the solution to that would be obviously open access or at the very minimal, a reasonable paywall. And this, I don't think this is at all helpful for you, but I said "government restriction on websites" if you like in like North Korea or Russia or ... and my solution to that is to live in a democratic country.

[01:29:49.08] PBJCA47575EA: Oh wow. Ok.

[01:29:52.01] PBR1B0A881F1: I had a lot of the same but like formatting but bad formatting so if you don't have any spaces between the words. Which would make it pretty impossible to read. So good formatting then is important. And then if it's kind of an unusable platform, if it's not supported by most computers or by a certain kind of...like if you need to download a lot to use it and so to have it on a stable, commonly used platform.

[01:30:20.21] PBRF60C8C611: Just one other one because...um, don't let people know that it exists. They can't read it if they don't know it's there.

[01:30:32.08] IBJ7A006F734: What's your solution to that?

[01:30:33.29] PBRF60C8C611: Make your project visible through social media platforms, twitter, facebook, and then it says outreach as well. Let people know you are there.

[01:30:46.15] PBR99A4D800B: I just really had one. Sometimes if you, when you scroll down it freezes and it can't...so if you are very...and you are clicking around and it freezes. So you need to make sure to use it quickly to program it...I don't know anything about programming but you must be able to allow it move quickly through it's performances.

[01:31:12.23] PBR9D1AD41E7: Responsive

[01:31:13.02] PBR99A4D800B: Responses. Absolutely.

[01:31:15.07] PBJCA47575EA: I suppose the font thing, you could make the text <AUDIO UNCLEAR> you could have a button that changes the text colour like the way some games have options for colour-blind people.

[01:31:27.07] PBJ560B377FD: That's a good idea.

[01:31:31.06] PBJCA47575EA: You could use, if you want to make it impossible to read, just like pictures. Like ancient hieroglyphics or something. But add visual aids beyond just words to help you navigate. I saw one website which used this algorithm to mix up the letters in the words to kind of simulate how a dyslexic person might view it. So if there was some way to help dyslexic people navigate easier, more usable. dyslexic is a hard word to spell, just so you know.

[01:32:12.18] PBR9D1AD41E7 & PBR1B0A881F1: Irony

[01:32:16.20] PBJCA47575EA: Introduce a script that the window closes every time <*AUDIO UNCLEAR*> so the solution. proper bug testing. Doesn't happen very often. Kick computer out of window. Don't use computer <*AUDIO UNCLEAR*>. Remove all the files from the server. <*AUDIO UNCLEAR*>

[01:32:44.10] IBJ7A006F734: Ok so again, please initial all of those so I can collect them. And then go ahead and take another piece of paper from the notebook. We are going to do one final exercise.

[01:33:02.04] PBJ560B377FD: *<AUDIO UNCLEAR>* image and the text is too small and you go to zoom in but actually *<AUDIO UNCLEAR>*

[01:33:10.11] PBR99A4D800B: Or you go to zoom in and it turns into pixels

<GROUP DISCUSSION>

(ACTIVITY) LOVE LETTER / BREAK UP LETTER

[01:33:39.24] IBJ7A006F734: Ok so, for our last exercise you are going to, you have the option to write either a love letter or a break up letter to any one of these types of websites that you choose. So one that you are familiar with or that you've used in the past.

<LAUGHTER>

<GROUP DISCUSSION>

[01:34:05.28] PBJ560B377FD: "It's not you, it's me."

[01:34:09.19] PBR1B0A881F1: "It's not me, it's you..."

[01:34:10.13] PBJ560B377FD: "I've changed..."

[01:34:10.29] IBJ7A006F734: Guys it does need to be a DHP so make sure you're not just writing *<AUDIO UNCLEAR>* or something. But you have the option to write either a love letter or a break up letter. And we will share them afterwards so please don't be do...be as mean as you like but just don't...modify your language appropriately.

[01:34:32.12] PBJCA47575EA: Can I write about Kindred Britain?

[01:34:33.29] IBJ7A006F734: Yes. It doesn't have to be one of these. *<AUDIO UNCLEAR>*

<GROUP DISCUSSION>

[01:41:45.19] IBJ7A006F734: Ok go ahead and finish up your thought.

[01:41:50.13] PBJ6BAA5509A: That was too emotional. *<LAUGHTER>*

[01:42:19.14] IBJ7A006F734: Who would like to share theirs first?

[01:42:24.02] PBJ6BAA5509A: I'll go first like every time. Dear Letters of 1916, I'm sorry to have to convey this in a letter. I couldn't bear to do this face-to-face (I literally can't). I will fondly recall all of the times I have fixed your typos, and corrected your tags, the times I spent staring at you—wondering who made you and how—so complex, almost impossible to understand. All that will remain with me. I hope when you think back on the time we spent together, you will smile. Perhaps sometime in the future we will meet again in a different country, on a different server. For now, I must go. Yours, PBJ6BAA5509A.

[01:43:09.14] PBR99A4D800B: Ok. Dear British Parliamentary Papers, Although I went into this relationship with great hopes, you constructed impossible barriers to our relationship. First, you asked me questions I couldn't possibly know until I had gotten to know you better. Then when I accidentally discovered one of your little secrets you then changed the title from what it actually was to one of your own

coded interpretations. When I asked you in every possible combination of words I could think of you hid the information from me. I had to resort to lieing to trick you into revealing your sources. Then to cap things off, you provided me with flawless information I didn't want or need. I am stuck with you BPP but I'm not not happy. Forget the party I was bringing you to and if I meet you don't expect a welcome hug!

<APPLAUSE>

[01:44:05.28] PBRF60C8C611: Mines much <AUDIO UNCLEAR> than that. So it was written to the Saint Patrick's Confessio Hyperstack Project, so Dear Saint Patrick's Confessio, Hyperstack Project, I am writing to say I love you because I think you are a good example of the type of editions that medievalists can endeavour to create. You attempted to redefine the role of the text and the editor in a digital scholarly edition. Admittedly, you're not perfect. You can be a little hard to navigate and I'm still sour about not getting that internship but I love you all the same. With love, PBRF60C8C611.

[01:44:38.25] PBR1B0A881F1: Dear Letters of 1916. I'm sorry to say that our time has come. I've dedicated three years of my life to you and you've offered virtually nothing in return. We both know this to be true. The transcription desk is shoddy, your workflow stunted, and you can't seem to commit like I need you to. No matter how much time and energy I invest, we are still in the same place. I've tried over the years to expand your horizons but your persistence in scaring people away has undermined my efforts. It's like you're not even trying. So after three, long, arduous, abstract-filled years, I'm saying enough. Goodbye Letters and just know that I'll take comfort in the fact that it wasn't me, it was definitely you.

<APPLAUSE>

[01:45:30.29] PBRF60C8C611: I really think these should be sent.

[01:45:38.28] IBJ7A006F734: Don't worry, everything will be completely de-identified when I transcribe it so...

<LAUGHTER>

[01:45:44.03] OBJA85D3875B: We have the NDA as well...

[01:45:45.26] PBR1B0A881F1: So I don't feel like it'd be that hard to connect but...

[01:45:50.06] PBR9D1AD41E7: So Dear Jane Austen Fiction Manuscripts, I want to love you, I really do. I should be excited to visit you, but I'm not. You're static, unchanging, and limited. I mean, a girl can only zoom in on the title page of Persuasion so many times, you know? I want more. I NEED more. Your search function is laughable—yes that's harsh, but you're a digital resource. If you want to be resourceful and useful, you need to help people find things. How about some more features? Why not allow for comparisons between documents? What about content in other formats? Hell, even a transcription of the texts and marginalia would be nice. Don't say "Oh well, just read the book." That's not the point! If you're going to claim the mantel of a digital humanities project, you need to actually BE a digital humanities project). You can do better. PBR9D1AD41E7.

<APPLAUSE>

[01:46:46.15] PBJCA47575EA: Ok I have very limited experience with this type of thing. Dear Mapping map map, blah blah blah <*NOTE: Mapping the Republic of Letters*>. I was going to say I'm sorry but I'm not. You're a waste of my time. I try to encourage your company but I couldn't. You're as shallow as this piece of paper. Your links don't go anywhere useful, you hide your data from the ones who want to see it. It's like you don't want anyone to love you. And then you send me these stupid pictures with blue and yellow lines. What the hell, explain! And not just with one word answers, pointless keywords, show me an explanation! Make sense! That you come with a manual. Never mind, it's too late. I'm never talking to you again. I refuse to click your links. It never got me anywhere anyways. Your parents should be ashamed of themselves for bringing you up like this. Full on hate, me.

<*APPLAUSE*>

[01:47:37.09] PBJ560B377FD: Dearest Letters of 1916, this hurts me more than it could ever hurt you, because you are an unfeeling piece of software. I'm breaking up with you. It's over. Our summer dalliance was nothing more than fleeting dream. I know that you've tried to change for me but with every change, I fear that you hate me more and more. For the sake of the past, and for the happy memories we've spent together with Spender and Martin, I'm ending it. I want to remember the good in you, not the trouble shooting, not the sleepless nights full of errors, and pointless arguments over naming conventions. But the real you; the little site that could! Goodbye my dear friend.

[01:48:27.05] OBJA85D3875B: Catharsis therapy session!

[01:48:31.26] IBJ7A006F734: Those were fantastic. Thank you guys! And on that note, we are finished. So thank you very much. If you didn't sign your letters, just initial them for me so that I know whose is whose.

[01:48:43.16] PBR1B0A881F1: Oh it's signed. <*LAUGHTER*>

[01:48:47.05] IBJ7A006F734: Does anybody have any kind of comments that they want to make about anything we did today or...before we go? Anything they feel they need to get off their chest?

[01:48:59.01] PBJCA47575EA: The exercise with the finding solutions to the puzzle thing...that was good. I like the way you visualise things.

[01:49:08.25] IBJ7A006F734: Thank you

[01:49:12.10] PBJ6BAA5509A: I really enjoyed that

<*GROUP DISCUSSION*>

[01:49:17.09] IBJ7A006F734: Alright so I have €10 vouchers to Starbucks for all of you for participating so thank you. Thank you guys so much! Thank you for coming. I greatly appreciate it!

Appendix B

Focus Group Design

Appendix B details the design of the focus groups conducted during the initial discovery phase of the research. This appendix details the documents (such as the participant application, information sheet, focus group rules, consent form, and ethical approval) in addition to the distribution email that was sent out to gather participants.

B.1 Distribution Email

The following text was used to distribute the participant application for the focus group:

> Shane McGarry, a PhD candidate in Digital Humanities at Maynooth University, is seeking participants for an upcoming research study. The study (which comprises both a focus group and individual 1 on 1 interviews) focuses on how users interact with Digital Humanities Resources such as The Letters of 1916, The Woodman Diary, and Vincent Van Gogh The Letters. Shane is seeking participants for a focus group (to take place in late November / early December) which will explore ideas for how users interact with these websites and what changes or improvements they would wish to see for future types of editions. Additionally, Shane is also seeking participants for 1 on 1 interviews (to take place early 2017) to test out prototypes of new editions that will built based on the focus group feedback. All participants will receive a €10 voucher to Starbucks for their participation.
>
> If you are interested in participating or wish to learn more about the study, please go to http://www.shanemcgarry.com/call-applicants-focus-group/. From here, you can fill out the form to sign up as a participant or download the information sheet which has more detail regarding the study.

B.2 Participant Application

Research Study Participant Application

If you are interested in participating in a research study to help facilitate my work on my dissertation, please fill out the fields below. All information provided will be confidential.

* Required

1. **First Name** *

2. **Surname** *

3. **Email Address** *

4. **Phone Number (optional)**

5. **Age** *
   Please note: All participants must be at least 18 years of age or older.
   Mark only one oval.
   - ◯ 18 - 25
   - ◯ 26 - 40
   - ◯ 41 - 64
   - ◯ 65 and Above
   - ◯ Prefer Not to Say

6. **Gender** *
   Mark only one oval.
   - ◯ Male
   - ◯ Female
   - ◯ Other
   - ◯ Prefer Not to Say

7. **Highest Level of Education** *
Mark only one oval.

- ◯ Secondary School
- ◯ Bachelors Degree
- ◯ Masters Degree
- ◯ PhD
- ◯ Prefer Not to Say

8. **Nationality** *
Mark only one oval.

- ◯ Irish
- ◯ English
- ◯ Polish
- ◯ American
- ◯ South Asian
- ◯ Russian
- ◯ Other - European
- ◯ Other - Non European
- ◯ Prefer Not to Say

Experience with Digital Scholarly Editions

Please provide information regarding your exposure to Digital Scholarly Editions (such as the Letters of 1916, the Woodman Diary, or Vincent Van Gogh: The Letters).

9. **Have you ever used a Digital Scholarly Edition?** *
Mark only one oval.

- ◯ Yes
- ◯ No

10. **Please rate your overall experience with DSEs you have used** *
If you have never used a DSE, please indicate a "3" (which is Not Applicable / No Opinion).
Mark only one oval.

	1	2	3	4	5	
Extremely Negative	◯	◯	◯	◯	◯	Extremely Positive

Availability

Please provide information regarding your availability

11. **Are you willing to travel to Maynooth University?**
Mark only one oval.

- ◯ Yes
- ◯ No

12. **Please check which aspect of the study you are interested in participating** *
Please check all that apply
Check all that apply.

- [ ] Focus Group
- [ ] 1 on 1 Usability Interviews

13. **Please check dates you are available (focus group only)**
Please check all dates which you are available. Note: This is applicable only for the focus group participation. Participation in 1 on 1 interviews will be conducted around your schedule).
Check all that apply.

- [ ] Wednesday, 14 September
- [ ] Friday, 16 September
- [ ] Friday, 30 September
- [ ] Wednesday, 7 October
- [ ] Wednesday, 19 October
- [ ] Friday, 21 October

14. **Do you agree to be added to distribution groups?** *
In order to make communications easier, I will be creating distribution groups to reach out to participants for both the focus group(s) and the 1 on 1 interviews. Please indicate if you give your permission to be added to these groups. Please Note: you will only receive notifications regarding focus group(s) or interviews which you agree to participate in. You will not receive any SPAM or unsolicited email.
Mark only one oval.

- ◯ Yes
- ◯ No

B.3 Information Sheet

Information Sheet

Project Title:	Escaping the Book Metaphor: Exploring Interaction Paradigms in Digital Scholarly Editions
Researcher:	Shane A. McGarry, PhD Candidate, Digital Humanities
Department:	An Foras Feasa
Email Contact:	shane.mcgarry.2015@mumail.ie

About this Project

This project is part of the Researcher's PhD thesis, which seeks to shift the discourse around the design of and interaction within a specific area of digital scholarship: that of the Digital Scholarly Edition (DSE). While the DSE has been a staple in the Humanities community for the last 20 years, it developed out of a need to digitise its analogue counterpart: the scholarly edition. As such, the DSE has remained married to its analogue metaphors, mimicking a print object in a digital realm. While the scholarly edition in its physical form has served academia well, digital environments provide us with new horizons to explore—horizons unbound by physical constraints. As technology evolves and becomes more ubiquitous, younger generations are reading differently, embracing concepts such as hyper reading, which is defined as the ability to read and consume vast amounts of information into a "unified frame, to associate different ideas together" (Hayles, 2012; Hui, 2013). These different methods of reading are not reliant on traditional close reading of texts, and the ways in which we read and consume information are changing, thus the boundary of the book metaphor in a digital environment has become obsolete. By exploring new, dynamic ways of exploring text-based content digitally, my work will provide a paradigm shift in how we interact with text within the DSE, through the use of more interactive environments, such as Augmented Reality (AR), Mixed Reality (MR), and the use of data visualisations.

The purpose of this research project is two-fold. First, through the use of focus groups, I intend to work with both academics who rely upon the Digital Scholarly Edition for research as well as the general public who read through them for either personal interest or individual research needs. Every participant will be given an opportunity to discuss how he or she utilises these types of editions and voice likes and dislikes. Focus groups will then work collaboratively to brainstorm various ideas that would allow them to obtain new information or view information in different ways.

The second phase of the project will then involve individual one on one interviews. Users will have the opportunity to use a prototype which will highlight new interactions to support different modes of reading and research within the Digital Scholarly Edition. Users will provide feedback regarding their thoughts on how intuitive they find the interface and what aspects of the new system they like or dislike. Changes to the prototypes will then be made based upon

user feedback and users will be provided with the opportunity to conduct additional interviews to gauge their reactions and opinions to any changes.

Requirements of Participants

Participants of the study will be given the opportunity to participate in a focus group. Additionally, users will also be given the opportunity to participate in one-on-one usability interviews. All participants will have the option to participate in just the focus group or just the usability interviews or participants may opt to participate in both. All participants will be required to sign a consent form (regardless of which activity a participant selects). The requirements for the individual activities are listed below.

Requirements for Participation in Focus Groups

The following will be required of participants in focus groups:

- All participants consent to be video-taped (see consent form for further information)
- All participants will be required to actively participate in a group discussion setting of approximately seven (7) to ten (10) people.
- All participants will agree to adhere to a code of conduct and will treat all other focus group participants with dignity and respect.
- All participants will agree to meet for a single session of a minimum of two (2) hours and a maximum of four (4) hours.
- All participants will agree to attend a session in person at Maynooth University Campus in the Iontas Building. If in-person attendance is not possible, participants will make arrangements with the Researcher to participate virtually via Skype.

Requirements for Participation in One-on-One Interviews

The following will be required of participants in one-on-one interviews:

- All participants agree to utilising a computer with screen capture software installed and running.
- All participants consent to be audio-recorded (see consent form for further information)
- All participants agree to give open and honest feedback throughout the interview.
- All participants agree to meet for a duration of approximately 30 minutes (and no more than one (1) hour).
- All participants agree to meet on-campus at Maynooth University (no virtual meetings will be possible for one-on-one interviews due to the need for screen capture).
- All participants will agree to meet for at least one (1) interview. When available, participants may be asked back to participate in an additional two (2) interviews (for a maximum of three (3) total one-on-one interviews). Participation in these subsequent two (2) interviews is optional but desirable.

Project Data

All data for this project will be used to develop prototypes which will form the basis of the analysis of the Researcher's PhD thesis. While certain demographic data will be used to highlight differences in gender or national identity as they relate to culture, such data will only be used in its aggregate sense and will not be used to personally identify individual participants. Each participant will be assigned an identifier in order to promote anonymity and personally identifiable information will not be disclosed within the thesis itself.

Participants of the focus group are aware, however, that they will not be completely anonymous as they will be: a) working in a group setting with other participants and b) will be video-taped. No contact information will ever be shared, and names will not be mentioned in the final write up of the dissertation. Transcripts of any audio will leverage anonymous identifiers in the place of names wherever names might occur.

All data will be analysed and then stored on a secure, encrypted drive. Per University policy, this data will be maintained in a secure archive for a maximum of 7 years at which point, it will be destroyed.

Confidentiality

Every effort will be made to keep each participants information confidential. However, it must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.

B.4 Consent Form

Consent Form

Project Title:	Escaping the Book Metaphor: Exploring Interaction Paradigms in Digital Scholarly Editions
Researcher:	Shane A. McGarry, PhD Candidate, Digital Humanities
Department:	An Foras Feasa
Email Contact:	shane.mcgarry.2015@mumail.ie
Supervisor:	Prof. Susan Schreibman, Directory, An Foras Feasa
Email Contact:	susan.schreibman@nuim.ie

All participants in this research project must sign this consent form in order to participate. As outlined in the Information Sheet, every effort will be made to keep information private and confidential.

If at any time, up to the point of publication, any participant decides to withdraw his or her consent for this project, he or she may do so by contacting the researcher. If the undersigned was a participant in the one-on-one interviews, all audio recordings and screen grabs will be discarded and any transcripts will be deleted. If the undersigned participates in the focus group and later opts to withdraw consent, he or she understands that while any transcripts containing the participants words will be discarded, the video itself cannot be edited to remove the participant's likeness.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at research.ethics@nuim.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

I, _____, understand and consent to all information disclosed in the Information Sheet. I acknowledge I have read the Information Sheet and understand the nature of this research project. I consent to be contacted by the above stated researcher or supervisor regarding my participation in this project. I also understand that every effort will be made to de-identify my personal information and all contact information will be held privately and confidentiality in a secure location by the researcher and An Foras Feasa and will not be shared for any reason. Furthermore, I acknowledge I have read and have received a copy of this consent sheet for my own personal records.

- ☐ By checking here, I agree to participate in the focus group as outlined in the Information Sheet. I consent to be video recorded as part of the exercise and understand that while my personal information will be be published in the final results, due to the nature of the group setting and the video recording, my anonymity cannot be completely guaranteed. Furthermore, I agree to the rules of the set forth for focus groups as outlined in the attached "Focus Group Rules" document and understand that if I violate the rules, I will be asked to leave the group; however, I understand that any participation prior to my being asked to leave will be utilised in accordance with the research.

- ☐ By checking here, I agree to participate in one-on-one interviews as outlined in the Information Sheet. I consent to be audio recorded as part of the exercise and understand that I will be utilising a computer with screen capture software installed and running during the exercise.

Print Name: _____

Signed: _____

Date: _____

B.5 Focus Group Rules

Focus Group Rules

Project Title: Escaping the Book Metaphor: Exploring Interaction Paradigms in Digital Scholarly Editions
Researcher: Shane A. McGarry, PhD Candidate, Digital Humanities
Department: An Foras Feasa
Email Contact: shane.mcgarry.2015@mumail.ie

Supervisor: Prof. Susan Schreibman, Directory, An Foras Feasa
Email Contact: susan.schreibman@nuim.ie

The purpose of this focus group is to provide an open forum for participants to discuss how they utilise Digital Scholarly Editions and what aspects of these types of systems require improvement. Due to the nature of a group setting, it is natural for participants to become excited or passionate regarding a particular aspect of the system. However, it is imperative that participants conduct themselves at all times with the utmost professionalism and respect for their fellow participants. As such the following rules have been devised to ensure fairness and equality within the group setting.

1. All participants recognise and understand that the Researcher is the moderator and arbitrator of the group. It is the Researcher's responsibility to ensure the group remains on task and disputes which arise in the course of discussion are ultimately settled by the Researcher and at his discretion.
2. All participants agree to speak to other participants in a calm and respectful manner. The use of derogatory language will not be tolerated and participants understand that doing so will result in immediate expulsion from the group.
3. The use of physical violence or the threat of such is completely unacceptable and will result in the immediate expulsion from the group. In the case of physical violence or threats, the Gardai will be contacted and information regarding the incident will be passed on to the Gardai in compliance with local laws.
4. All discussions will be held in an orderly manner. If a participant is speaking, every effort will be made by other participants to not interrupt or keep other participants from speaking their own thoughts.
5. Participants acknowledge that this focus group is an entirely free-form exercise. There is no such thing as a "bad idea". All ideas, regardless of feasibility, are welcome.

I, _____, understand and acknowledge the above rules. I acknowledge that I have read them and agree to comply with them. I further acknowledge that I have been given a copy of said rules and agree to comply with any decisions made by the Researcher regarding the course or direction of the conversation. I also agree to

adhere to any decisions made by the Researcher for the purposes of this project regarding my continued participation.

Print Name: _____

Signed: _____

Date: _____

B.6 Ethical Approval

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

20 May 2016

Shane A. McGarry
An Foras Feasa
Maynooth University

RE: Application for Ethical Approval for a project entitled: Escaping the Book Metaphor: Examining User Interface Metaphors in Digital Scholarly Editions

Dear Shane,

The Ethics Committee evaluated the above project and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30 September 2017.

Kind Regards,

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

C.c. Professor Susan Schreibman, An Foras Feasa

Reference Number
SRESC-2016-035

Appendix C

Design Documents

This appendix details the documentation used to define the overall design of the *Alcalá Record Books*. The appendix begins by providing user persona sheets for each of the three personas. The initial wireframes used for the system design and concept face are then included. Finally, the main portions of the backend system code along with some of the front end code, which detail:

1. how the transactions were loaded into the system for the training and classification aspects detailed in Chapter 4
2. the Python module which handles the interactions with the eXistDB datastore
3. the Python module which handles the interactions with the MongoDB datastore
4. the Abstract Factory component used by the front end system (written in Typescript / Angular 6.0) which generates the chart requested by the user

C.1 User Personas

C.1.1 The Researcher: Laura

Laura

age: 30-45
residence: Dublin 07
education: PhD, History
occupation: Postdoctoral Researcher
marital status: Single

"This archive is too massive and the search isn't providing me with the right hits."

Laura is post-doctoral scholar at a university in Dublin, Ireland. She is a historian interested in learning more about early Irish diaspora to continental Europe and how the diaspora not only effected Ireland but also the various cultures across continental Europe.

As a digital native, Laura is very comfortable with technology. She enjoys using digital resources and prefers them over physical ones which she finds difficult to search through. She is well-read within her field and has a strong grasp of keywords she needs to use when searching these types of systems. However, she often finds herself frustrated by a lack results from her searches in the various DREs she leverages. She feels there are items missing from her results that could have a significant impact on her work.

Comfort With Technology

INTERNET

SOFTWARE

MOBILE APPS

SOCIAL NETWORK

Criteria For Success:

Laura needs a method to better facilitate her exploration of a manuscript. The system also needs to provide her with a level of confidence that she is truly able to see all of the items relevant to her search.

Needs

- A different mode of search beyond a keyword search
- Confidence in her result set

Wants

- A system that is visually attractive and fun to use
- The ability to download her result set

Values

- Ease of use
- Contextualisation of data
- Reproducible results

Fears

- Missing important documents due to obscure wording
- Being unable to publish her findings as the result set cannot be reproduced

C.1.2 The Learner: Jenny

Jenny

age: 18-24
residence: Maynooth, Co. Kildare
education: Bachelors
occupation: Student
marital status: Single

"I think learning online is the next step for the future but all my history classes are about looking at books in the library."

Jenny is a student at Maynooth University working on her Bachelors degree in History. She has always loved history but is much more comfortable with digital tools than pen and paper.

Jenny appreciates the necessity of learning how to conduct historical research "the old-fashioned way", but, she wants to bring History into the modern age. As a digital native, she believes that using digital tools can provide her with deeper insights than relying solely on pen and paper. As such, she has enrolled in a Digital History course where she hopes to learn how to conduct historical research using online tools.

Comfort With Technology

INTERNET

SOFTWARE

MOBILE APPS

SOCIAL NETWORK

Criteria For Success:

What's needed in order to make him or her feel successful.

Needs

- These should be required
- This might be Spanish localized content
- Support from higher ups

Wants

- Wants are something that is not required but delights
- This might be the holy grail feature
- It delights him or her with its simplicity or completeness

Values

- Qualities that he or she values
- High level planning is an example

Fears

- Fears are the things that keep him or her up at night
- This might be something like failure
- But might also have to do with real fear

C.1.3 The General User: Matthew

Matthew

age: 25-35
residence: Dublin 01
education: Masters
occupation: Computer Programmer
marital status: Married

"I need to find information about one of my ancestors who attended Alcalá. But I'm not sure where to start!"

Matthew is a computer programmer with a small tech-startup in Dublin. As a digital native who holds a Masters in Computer Science, Matthew is very comfortable with technology. He has had a sudden interest in his family history and recently discovered one of his ancestors was educated at the Royal Irish College in Alcalá. He wants to see if he can find more information but not he's not really sure where to begin. As he is not a historian, he doesn't have access to the archives and needs to rely on information he can find on the internet. While he has been able to find numerous mentions of the Royal Irish College in Alcalá and he knows the name of his ancestor, he is overwhelmed by the sheer volume of data.

Comfort With Technology

INTERNET

SOFTWARE

MOBILE APPS

SOCIAL NETWORK

Criteria For Success:

To successfully locate information about his ancestor that can provide a general time-frame for when he was in Alcalá and when he took his viaticum to return home to Ireland.

Needs

- An interface for locating an individual person
- The ability to capture information from within the archive so he can save it as part of the family history

Wants

- An attractive system that is also fast and stable. As a computer programmer, he tends to be quite picky about what the systems he uses.

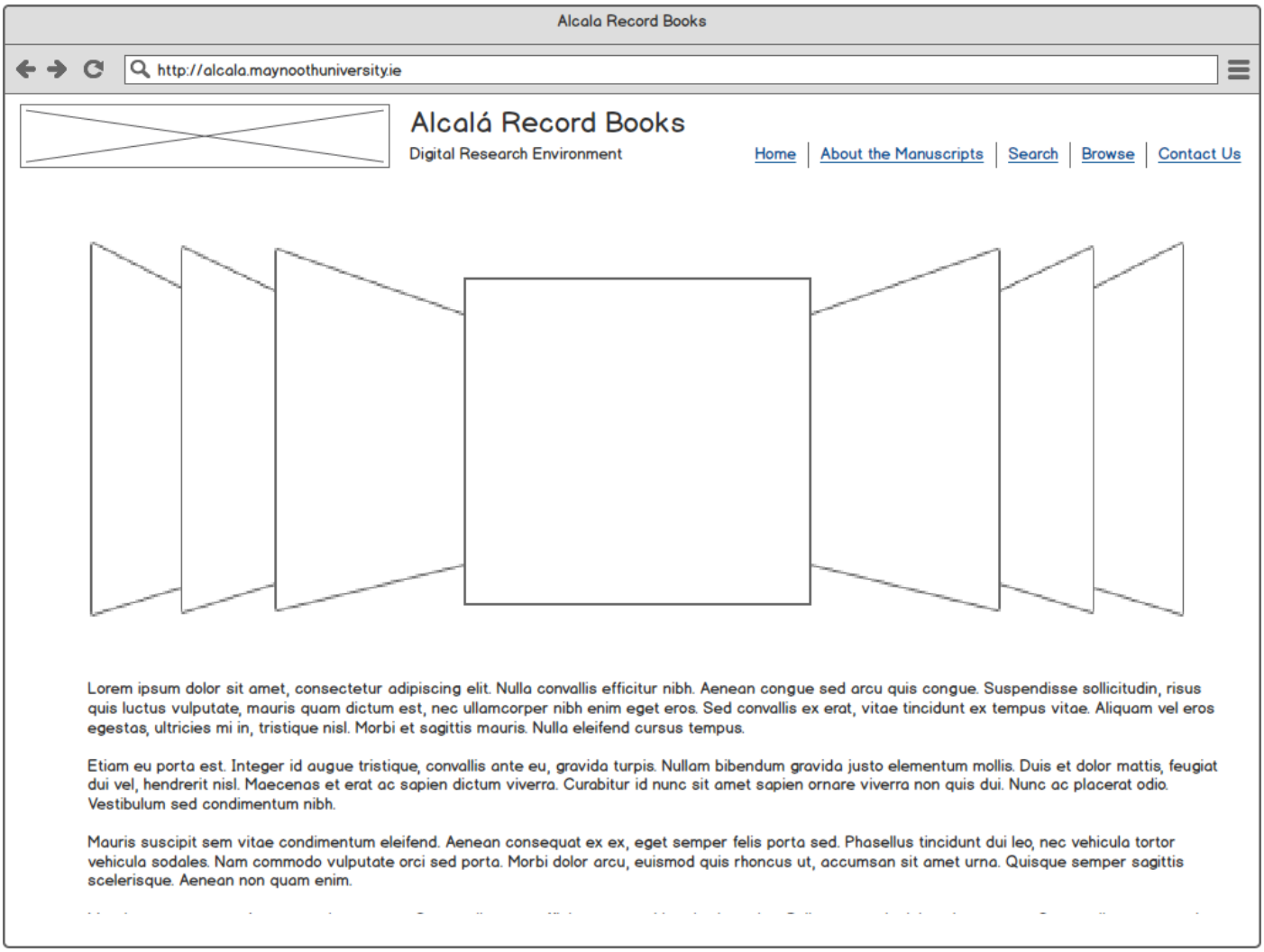
Values

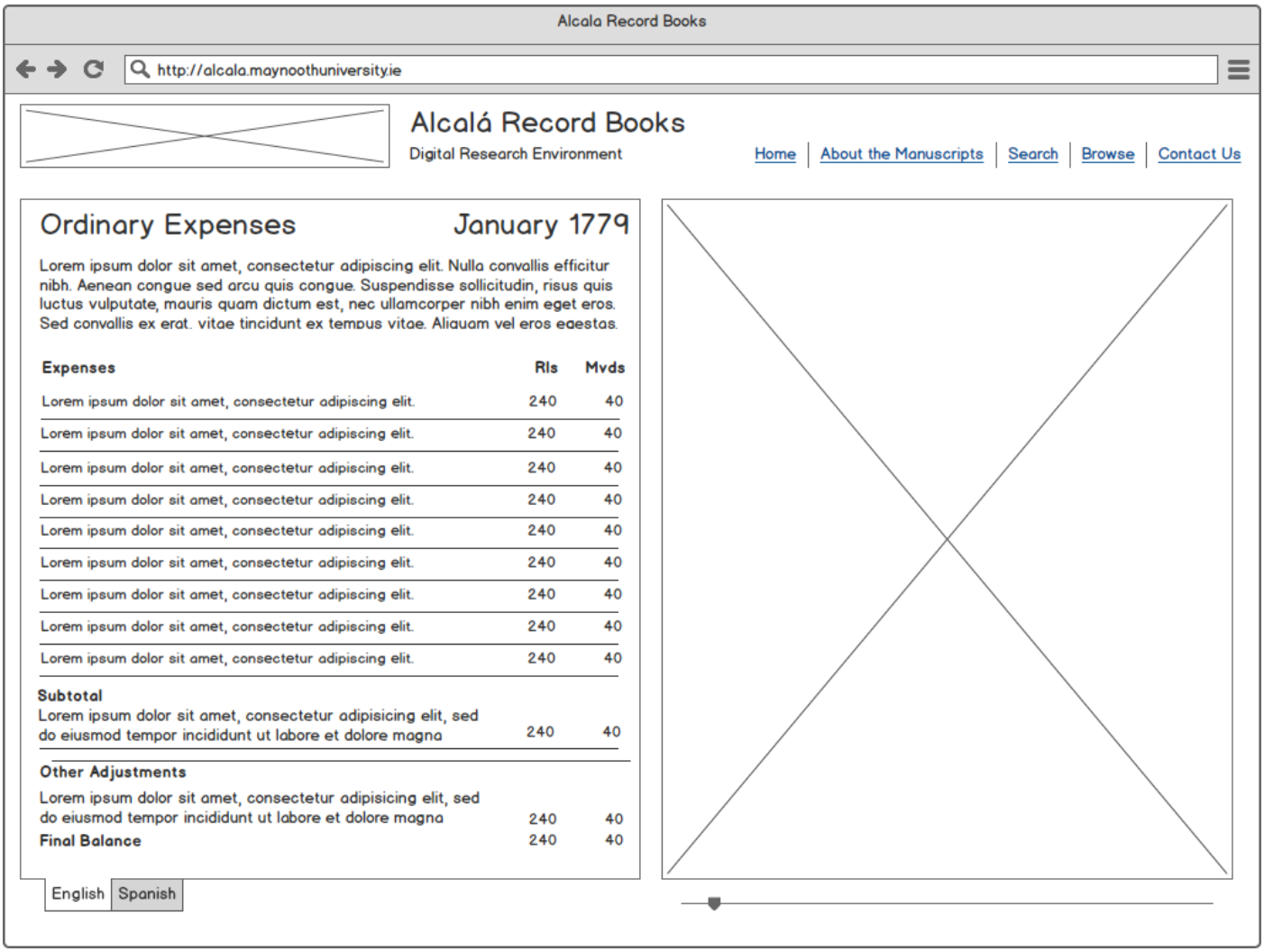
- Efficiency
- Succinct results

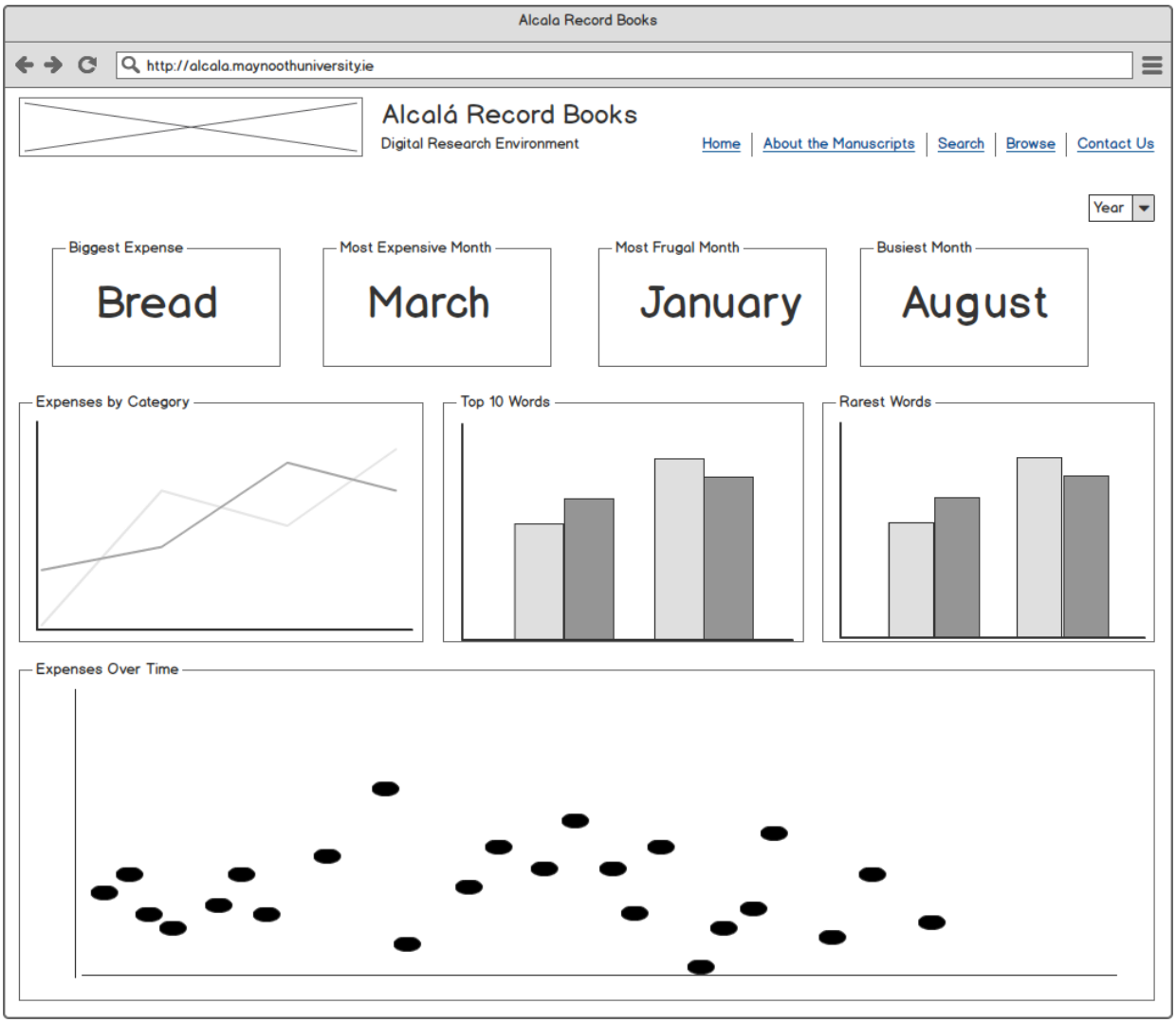
Fears

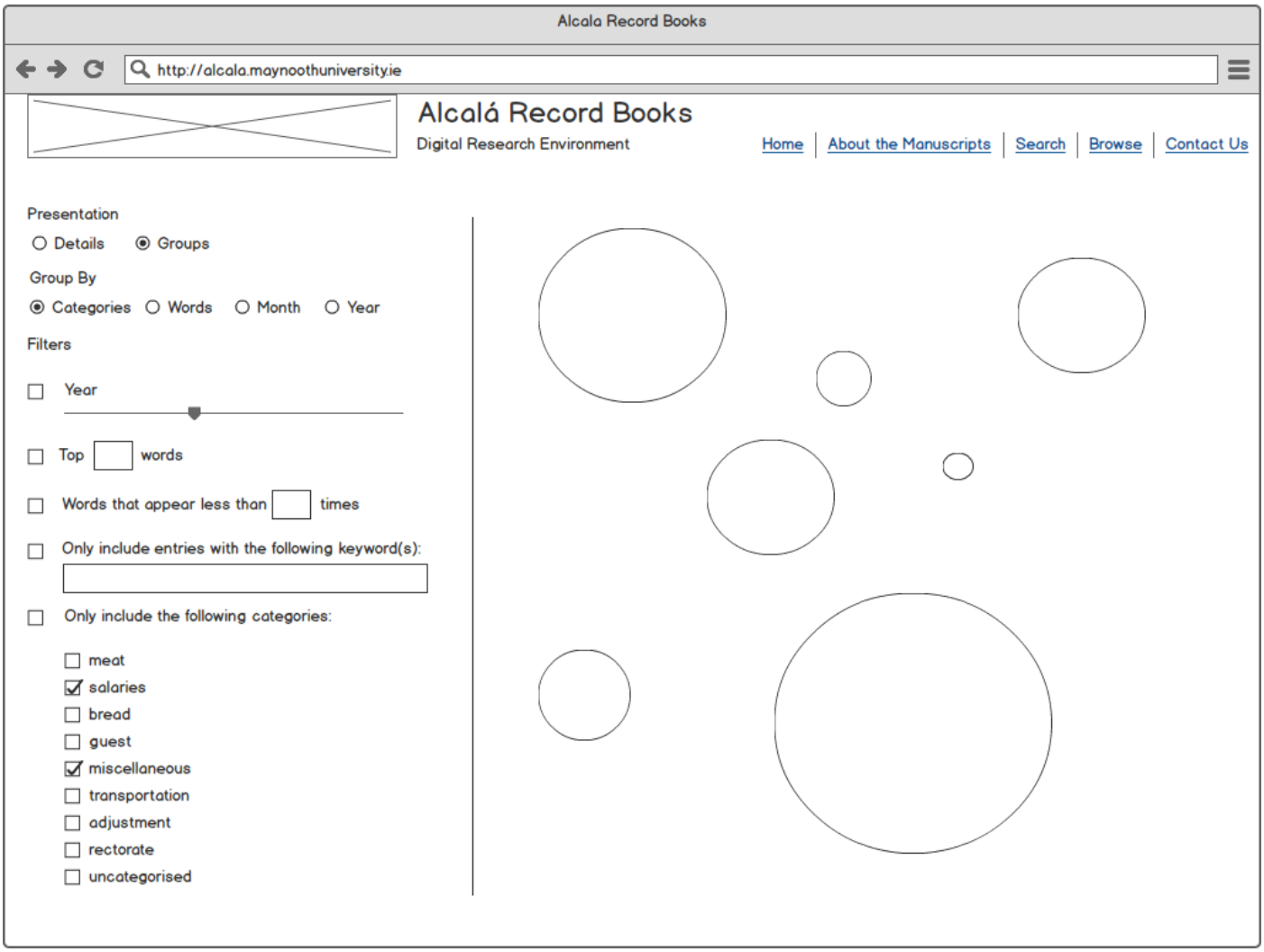
- Being given too much information
- Not being able to locate the right information

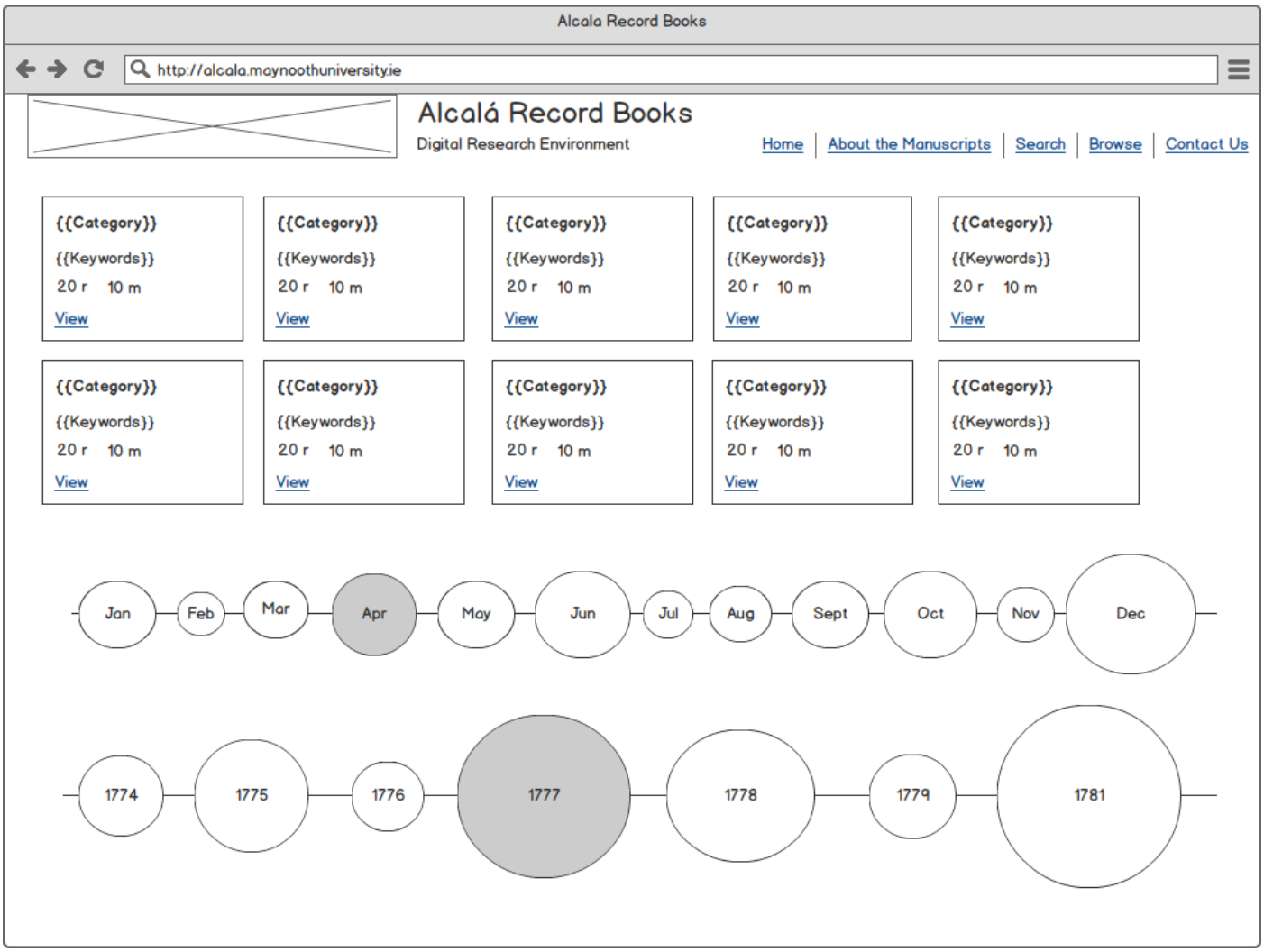
C.2 Wireframes

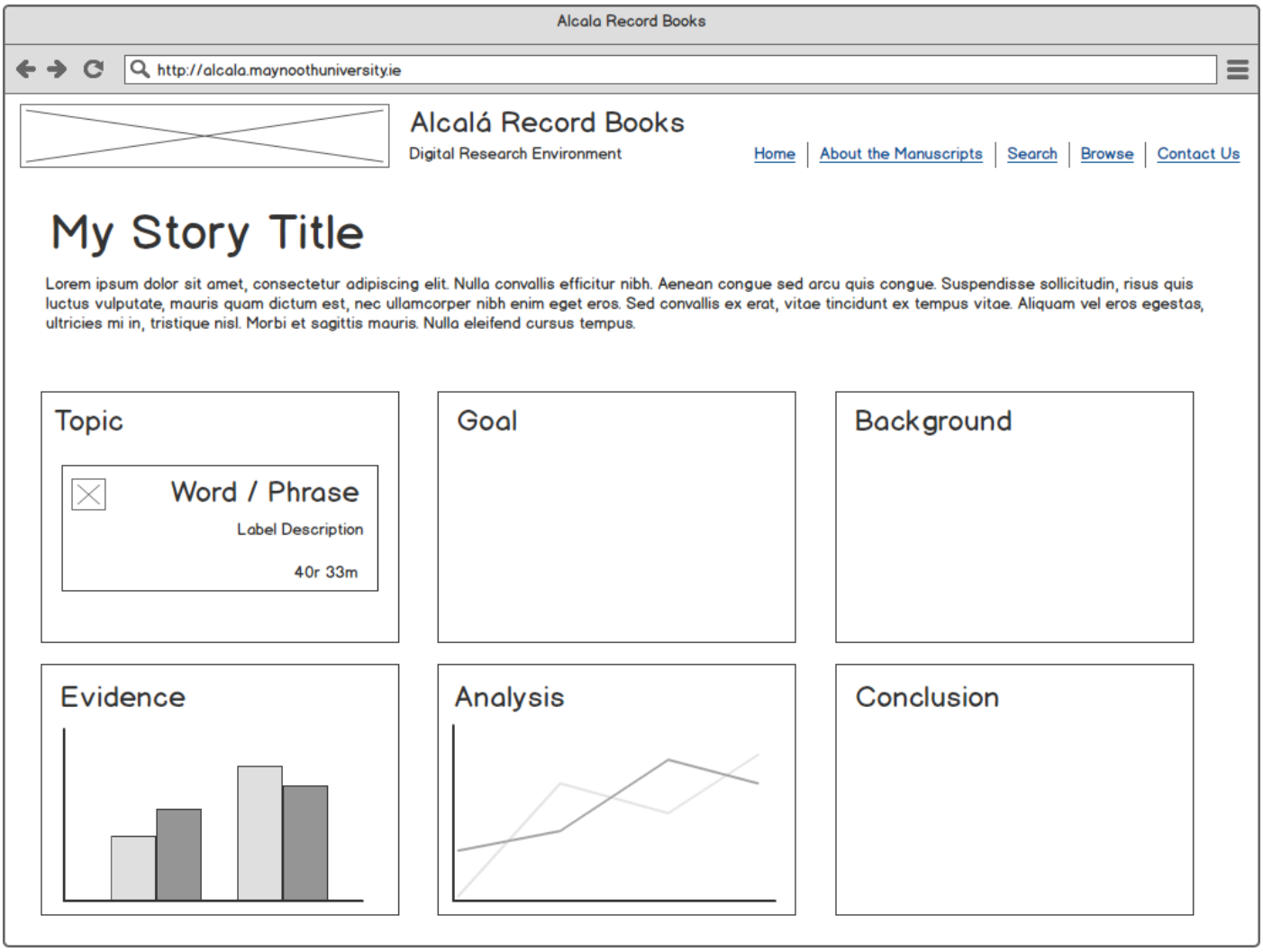


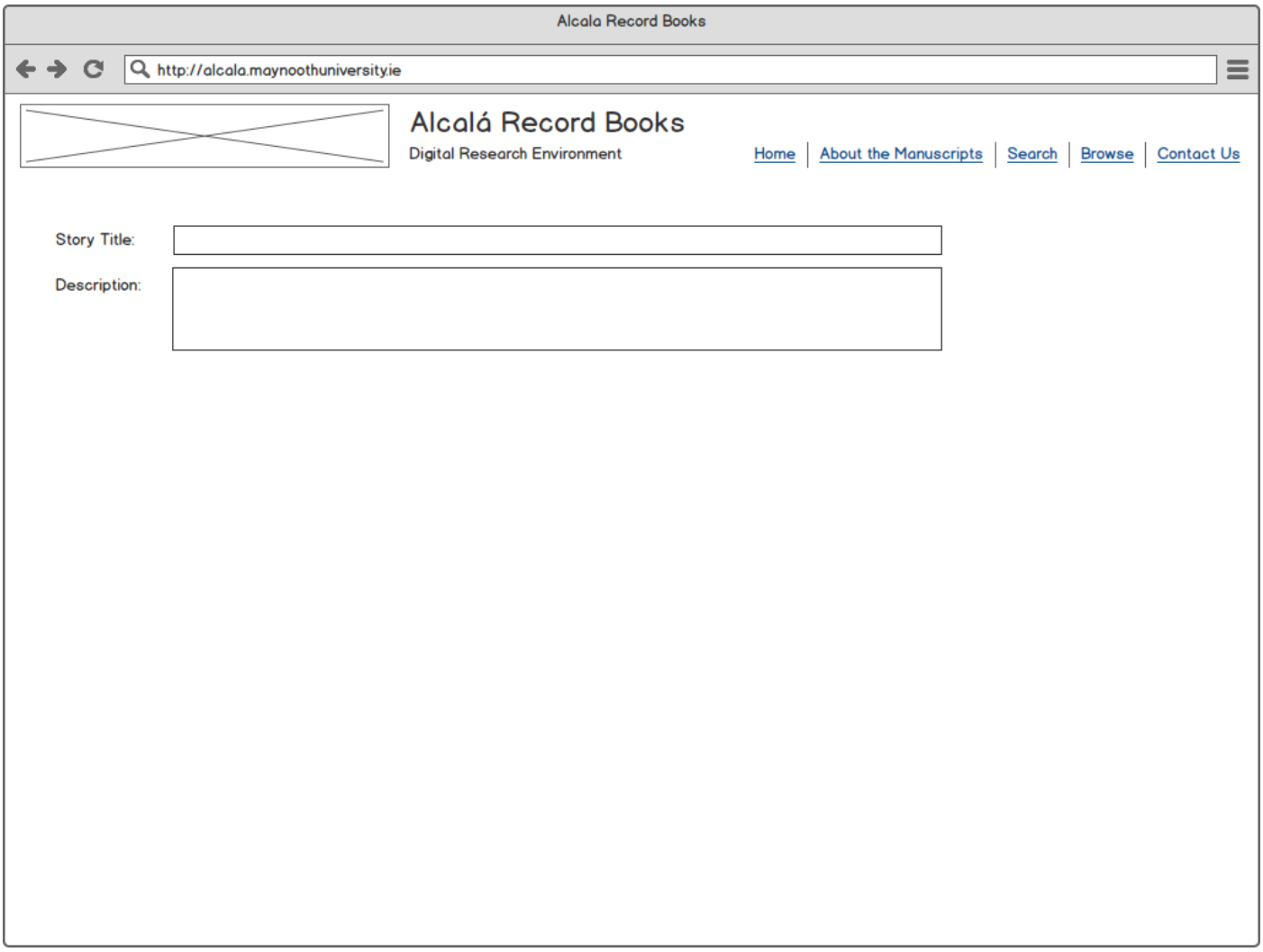

Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
Ordinary Expenses
January 1779
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla convallis efficitur nibh. Aenean congue sed arcu quis congue. Suspendisse sollicitudin, risus quis luctus vulputate, mauris quam dictum est, nec ullamcorper nibh enim eget eros. Sed convallis ex erat, vitae tincidunt ex tempus vitae. Aliquam vel eros egestas.
Expenses
Rls
Mvds
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Lorem ipsum dolor sit amet, consectetur adipiscing elit. 240 40
Subtotal
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna 240 40
Other Adjustments
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna 240 40
Final Balance 240 40
English
Spanish


Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
Year
Biggest Expense
Bread
Most Expensive Month
March
Most Frugal Month
January
Busiest Month
August
Expenses by Category
Top 10 Words
Rarest Words
Expenses Over Time


Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
Presentation
Details
Groups
Group By
Categories
Words
Month
Year
Filters
Year
Top
words
Words that appear less than
times
Only include entries with the following keyword(s):
Only include the following categories:
meat
salaries
bread
guest
miscellaneous
transportation
adjustment
rectorate
uncategorised


Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
{{Category}}
{{Keywords}}
20 r 10 m
View
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sept
Oct
Nov
Dec
1774
1775
1776
1777
1778
1779
1781


Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
My Story Title
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla convallis efficitur nibh. Aenean congue sed arcu quis congue. Suspendisse sollicitudin, risus quis luctus vulputate, mauris quam dictum est, nec ullamcorper nibh enim eget eros. Sed convallis ex erat, vitae tincidunt ex tempus vitae. Aliquam vel eros egestas, ultricies mi in, tristique nisl. Morbi et sagittis mauris. Nulla eleifend cursus tempus.
Topic
Word / Phrase
Label Description
40r 33m
Goal
Background
Evidence
Analysis
Conclusion


Alcala Record Books
http://alcala.maynoothuniversity.ie
Alcalá Record Books
Digital Research Environment
Home
About the Manuscripts
Search
Browse
Contact Us
Story Title:
Description:

C.3 Software Code

C.3.1 Data Loading and Classification

```
# Step 1: Load the transactions from eXist-db using a custom build ExistData class (which is wrapper for the pyexistdb module)
# and the AnalysisItem data object

# AnalysisItem class definition
class AnalysisItem(JsonSerializable):
    def __init__(self, words, _id=None, categories=None, year=None, month=None, reales=None, maravedises=None, pageid=None, *args, **kwargs):
        self.words = words
        self.categories = categories
        self.pageid = pageid
        self.year = year
        self.month = month
        self.reales = reales
        self.maravedises = maravedises
        self._id = _id
        self.monthName = self.get_month_name()

        import uuid
        if self._id is None:
            self._id = uuid.uuid4().hex

    def get_month_name(self):
        import calendar
        return calendar.month_name[self.month]

# Utilities class for static methods
class Utilities:
    @staticmethod
    def build_word_list(analysisItems):
        results = list()
        for t in analysisItems:
            results.extend(t.words)
        return results

    #
    ----------------------------------------------------------------------------------------------------

    # Code Adapter From: RStudio-pubs, The Analytics Store
    # URL: https://rstudio-pubs-static.s3.amazonaws.com/79360_850b2a69980c4488b1db95987a24867a.html
    #           http://www.theanalyticsstore.com/
    # Remarks: This code was adapted from the above website as well as a course on NLP offered
    #                    by the Analytics Store and modified to work for this application
```

```
40      #
        ----------------------------------------------------------------------------------------

41      @staticmethod
42      def preprocess(text, to_lowercase=True, remove_num=True, tokenize=
        True, remove_stopwords=True, normalise_method="lem"):
43          if to_lowercase:
44              text = text.lower()
45
46          if tokenize:
47              text = Utilities.get_tokenizer().tokenize(text)
48
49          if remove_num:
50              text = [w for w in text if not w.isdigit()]
51
52          if remove_stopwords:
53              text = [w for w in text if not w in Utilities.get_stopwords
        ('en')]
54
55          if normalise_method == "lem":
56              text = [Utilities.get_lemmatizer().lemmatize(w) for w in
        text]
57          elif normalise_method == "stem":
58              text = [Utilities.get_stemmer().stem(w) for w in text]
59          else:
60              raise ValueError("You have provided an invalid value for
        the normalise_method argument. Current valid values are 'stem' and
        'lem'.")
61
62
63          return text
64
65      @staticmethod
66      def get_lemmatizer():
67          lemmatizer = WordNetLemmatizer()
68          return lemmatizer
69
70      @staticmethod
71      def get_tokenizer():
72          tokenizer = RegexpTokenizer(r'\w+')
73          return tokenizer
74
75      # This method loads a set of additional stop words and appends it
        to the standard stop word list.
76      @staticmethod
77      def get_stopwords(language='en'):
78          exclusion_file = Path(config.NLP_CONFIG['data_dir'] + "
        excludedWords.txt")
79          excluded_words = list()
80          with exclusion_file.open(mode='r') as ef:
81              for line in ef:
82                  word = line.replace('\n', '').replace('\r', '')
```

```
                if word is not None:
                    excluded_words.append(word)
        stopwords = get_stop_words(language)
        stopwords.extend(excluded_words)
        return stopwords
    @staticmethod
    def get_stemmer():
        stemmer = PorterStemmer()
        return stemmer


class DataImporter:
from mongoDB import MongoData # Custom written wrapper classes for handling data transactions using pymongo
from existDB import ExistData # Custom written wrapper classes for handling data transactions using pyexistdb
#Code to load data
db = ExistData()
all_pages = db.get_all_pages()
exist_data = list()
for p in all_pages:
    for m in p.months:
        for ex in m.expenses:
            for e in ex.entries:
                if e.description is not None:
                    exist_data.append(AnalysisItem(Utilities.preprocess(e.description.english),
                                                   pageid=p.id, year=p.year, month=m.month,
                                                   reales=e.amount.reales, maravedises=e.amount.maravedises))
            if ex.adjustment is not None and ex.adjustment.description is not None:
                exist_data.append(AnalysisItem(Utilities.preprocess(ex.adjustment.description.english),
                                               pageid=p.id, year=p.year, month=m.month,
                                               reales=ex.adjustment.amount.reales, maravedises=ex.adjustment.amount.maravedises))
        if m.otherAdjustments is not None and m.otherAdjustments.description is not None:
            exist_data.append(AnalysisItem(Utilities.preprocess(m.otherAdjustments.description.english),
                                           pageid=p.id, year=p.year, month=m.month,
                                           reales=m.otherAdjustments.amount.reales, maravedises=m.otherAdjustments.amount.maravedises))


# Step 2: Insert the transactions into mongodb using the pymongo module
mdb = MongoData()
documents = []
```

```
122 for e in exist_data:
123     mdb.insert_one_transaction(e)
124     documents.append(e)
125 
126 #Step 3: Create 3 sets of training data, each containing 100
    transactions
127 # First training set consists of the first 100 objects in the full
    dataset
128 training_first100 = documents[:100]
129 # Second training set consists of 100 random items from the full
    dataset
130 import random
131 training_random100 = random.sample(documents, 100)
132 # Third training set contains a random sample of 100 items pulled from
    a subset of the full dataset (the top 10 most frequently used words
     in the corpus)
133 from nltk import FreqDist
134 
135 # First, determine the most frequently used words and sort by the
    frequency distribution (desc)
136 fdist = FreqDist([w.lower() for w in Utilities.build_word_list(
    documents)])
137 sorted_fd = sorted(fdist.items(), key=lambda kv: kv[1], reverse=True)
138 filtered_data = list()
139 
140 # Now, grab the 10 most frequently used words
141 dict_freq = dict(sorted_fd[:10])
142 
143 # Next, create a filtered list of all the data where the 10 most
    frequently used words are mentioned
144 for word in dict_freq.keys():
145     temp_list = list(filter(lambda i:word in i.words, documents))
146     filtered_data.extend(temp_list)
147 
148 # Finally, filter out any duplicate records and create a random sample
149 unique_ids = set([x._id for x in filtered_data])
150 training_items = [x for x in documents if x._id in unique_ids]
151 training_topwords100 = random.sample(training_items, 100)
152 
153 # Write each training set into mongodb using the MongoData() wrapper
154 mdb.update_multiple_transactions(transaction_list=training_first100,
    use_training=True)
155 mdb.update_multiple_transactions(transaction_list=training_random100,
    use_training=True)
156 mdb.update_multiple_transactions(transaction_list=training_topwords100,
     use_training=True)
157 
158 #Step 4: After the training data has been categorised, use it to
    classify the rest of the manuscript
159 # We start with creating a document classififerr using the sklearn
    package
160 from sklearn.preprocessing import MultiLabelBinarizer
```

```
161 from sklearn.pipeline import Pipeline
162 from sklearn.feature_extraction.text import CountVectorizer
163 from sklearn.svm import LinearSVC
164 from sklearn import svm
165 from sklearn.feature_extraction.text import TfidfTransformer
166 from sklearn.multiclass import OneVsRestClassifier
167 import random
168
169
170 class DocumentClassifier:
171     def __init__(self):
172         self.label_bin = MultiLabelBinarizer()
173
174     def classify(self, data_train, data, test_mode=False):
175         train_labels = self.label_bin.fit_transform([i.categories for i
     in data_train])
176         train_data = [' '.join(i.words) for i in data_train]
177
178         if test_mode:
179             data_classify = [' '.join(i.words) for i in random.sample(
    data, 100)]
180         else:
181             data_classify = [' '.join(i.words) for i in data]
182
183         classifer = Pipeline([
184             ('vectorizer', CountVectorizer()),
185             ('tfidf', TfidfTransformer()),
186             ('clf', OneVsRestClassifier(LinearSVC()))
187         ])
188
189         classifer.fit(train_data, train_labels)
190         predicted = classifer.predict(data_classify)
191         predicted_labels = self.label_bin.inverse_transform(predicted)
192
193         if test_mode:
194             null_labels = [x for x in predicted_labels if not x]
195             return 1 - (len(null_labels)/100)
196         else:
197             return predicted_labels
198
199 # Combine the trraining data into a single dataset
200 training_data = training_first100
201 training_data.extend(training_random100)
202 training_data.extend(training_topwords100)
203
204 # Make sure all of the categories in the training data are lowercase
205 for td in training_data:
206     new_categories = []
207     for c in td.categories:
208         new_categories.append(c.lower())
209     td.categories = new_categories
210
```

```
# # Get a list of predicted labels
classifier = DocumentClassifier()

# Loop through and update the transactions with the predicted category
predicted_categories = classifier.classify(data_train=training_data,
    data=documents)
for t, c in zip(transactions, predicted_categories):
    t.categories = c

# Write everything to mongo
mdb.update_multiple_transactions(transaction_list=documents)
```

C.3.2 ExistData Module

```
from pyexistdb import db as edb
from pyexistdb.exceptions import ExistDBException
from models.alcalaPage import AlcalaPage
from models.alcalaEntry import AlcalaEntry, AlcalaAdjustment,
    AlcalaOtherAdjustments
from models.alcalaSignOff import AlcalaSignOff
from models.pageResult import PageResult, PageResultList
from models.alcalaBase import AlcalaBase
from lxml import etree
from lxml.etree import tostring
import config


class ExistData:
    """
    Class used to access eXist-db instance. As this project does not
    update eXist, all methods are for read-only
    operations.

    Requirements:
    * pyexistdb
    * eulxml
    * lxml
    * config.py
    * Custom Alcala datamodel classes (see import statement)
    """

    EXCEPTIONS_PAGEID_TOOMANYRESULTS = """Querying for page {0} caused
    more than 1 page to return. This should not
                                        happen. Please contact an
    administrator to inspect the data."""

    def __init__(self):
        """
        Initialises an instance of eXist using the pyexistdb module.
    Settings such as host, username, and password are
        drawn from config.py.
        """
```

```
34
35          self.db = edb.ExistDB(config.EXISTDB_CONFIG['host'], username=
    config.EXISTDB_CONFIG['username'], password=config.EXISTDB_CONFIG['
    password'])
36
37      def get_all_pages(self, pageIndex=1, limit=500):
38          """Return a list of all AlcalaPage objects in the eXist-db
    instance"""
39          xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    return $x'
40          qr = self.db.query(xquery, pageIndex, limit)
41          results = list()
42          for i in range(0, qr.hits):
43              page = AlcalaPage(etree.XML(tostring(qr.results[i])))
44              results.append(page)
45
46          return results
47
48      def get_all_page_ids(self):
49          """Return a list of all pageIDs in the eXist-db instance"""
50          xquery = 'for $x in distinct-values(doc("alcala/books/ledger.
    xml")//pages/page/pageID) order by $x return $x'
51          qr = self.db.query(xquery, 1, 500)
52          results = list()
53          for i in range(0, qr.hits):
54              temp_item = AlcalaBase(qr.results[i])
55              results.append(temp_item.get_element_value("."))
56          return results
57
58      def get_all_signoffs(self, pageIndex=1, limit=2000):
59          """
60          Return a list of all AlcalaSignOff objects in the eXist-db
    instance
61          NOTE: This is primarily used to load data into mongodb for nlp
    / analysis.
62          """
63
64          xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    /*/*/signOff return $x'
65          qr = self.db.query(xquery, pageIndex, limit)
66          results = list()
67          for i in range(0, qr.hits):
68              sign_off = AlcalaSignOff(etree.XML(tostring(qr.results[i]))
    )
69              results.append(sign_off)
70
71          return results
72
73      def get_all_entries(self, pageIndex=1, limit=2000):
74          """
75          Return a list of all AlcalaEntry objects in the eXist-db
    instance.
```

```
76          NOTE: This is primarily used to load data into mongodb for nlp
    / analysis.
77          """
78
79          xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    /*/*/expenditure/entry return $x'
80          qr = self.db.query(xquery, pageIndex, limit)
81          results = list()
82          for i in range(0, qr.hits):
83              entry = AlcalaEntry(etree.XML(tostring(qr.results[i])))
84              results.append(entry)
85
86          return results
87
88      def get_all_adjustments(self, pageIndex=1, limit=2000):
89          """
90          Return a list of all AlcalaAdjustment objects in the eXist-db
    instance.
91          NOTE: This is primarily used to load data into mongodb for nlp
    / analysis.
92          """
93
94          xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    /*/*/expenditure/adjustment return $x'
95          qr = self.db.query(xquery, pageIndex, limit)
96          results = list()
97          for i in range(0, qr.hits):
98              entry = AlcalaAdjustment(etree.XML(tostring(qr.results[i]))
    )
99              results.append(entry)
100
101         return results
102
103     def get_all_otherAdjustments(self, pageIndex=1, limit=2000):
104         """
105         Return a list of all AlcalaOtherAdjustment objects in the eXist
    -db instance.
106         NOTE: This is primarily used to load data into mongodb for nlp
    / analysis.
107         """
108
109         xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    /*/*/otherAdjustments return $x'
110         qr = self.db.query(xquery, pageIndex, limit)
111         results = list()
112         for i in range(0, qr.hits):
113             entry = AlcalaOtherAdjustments(etree.XML(tostring(qr.
    results[i])))
114             results.append(entry)
115
116         return results
117
```

```
118     def get_entry_count(self, pageIndex=1, limit=50):
119         """Retrieves a count of all AlcalaEntry objects in the eXist-db
         instance."""
120         xquery = 'let $x := doc("alcala/books/ledger.xml") return count
        ($x//pages/page/*/*/expenditure/entry )'
121         qr = self.db.query(xquery, pageIndex, limit)
122         return qr.results[0]
123 
124     def get_page(self, pageid, pageIndex=1, limit=50):
125         """Returns a specific AlcalaPage object (via the supplied
        pageid paramenter) of the eXist-db instance."""
126         xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
        where $x/pageID="%s" return $x' % pageid
127         try:
128             qr = self.db.query(xquery, pageIndex, limit)
129             if qr.hits > 1:
130                 raise ExistDBException(Exception(str.format(self.
        EXCEPTIONS_PAGEID_TOOMANYRESULTS, pageid)))
131 
132             page = AlcalaPage(etree.XML(tostring(qr.results[0])))
133             return page
134         except ExistDBException as dberr:
135             print(dberr.message())
136 
137     def get_pages_by_keyword(self, keyword, year=None, pageIndex=1,
        limit=50):
138         """Conducts a keyword search against the entire eXist-db
        instance and returns a list of PageResult objects"""
139         query_string = '$hit//textContent[ft:query(.,"%s*")]' % keyword
140         if year is not None:
141             query_string += ' where $hit/content[@yearID = %d]' % year
142 
143         xquery = """
144                 import module namespace kwic="http://exist-db.org/
        xquery/kwic";
145                 let $hits :=
146                     for $hit in doc("alcala/books/ledger.xml")//pages/
        page
147                     where %s
148                     order by ft:score($hit) descending
149                     return
150                         $hit
151                 let $total-hits := count($hits)
152                 for $hit in $hits
153                 return
154                     <result>
155                         {$hit}
156                         <matches>
157                             {kwic:summarize($hit, <config width="40"/>)
        }
158                         </matches>
159                     </result>
```

```
160                 """ % query_string
161         qr = self.db.query(xquery, start=pageIndex, how_many=limit + 1)
162 
163         result = PageResultList(total_hits=qr.hits, current_index=
    pageIndex, result_limit=limit)
164         for i in range(0, qr.count):
165             xml = etree.XML(tostring(qr.results[i - 1]))
166             page = AlcalaPage(xml)
167             match_xml = xml.find('.//matches')
168             result_page = PageResult(page, match_xml)
169             result.add_page(result_page)
170 
171         return result
172 
173     def get_by_year(self, year, pageIndex=1, limit=50):
174         """Returns a list of AlcalaPage objects in the eXist-db
    instance for a specified year."""
175         xquery = 'for $x in doc("alcala/books/ledger.xml")//pages/page
    where $x/content[@yearID="%s"] return $x' % year
176         qr = self.db.query(xquery, pageIndex, limit)
177         results = list()
178         for i in range(0, qr.count - 1):
179             page = AlcalaPage(etree.XML(tostring(qr.results[i])))
180             results.append(page)
181 
182         return results
183 
184     def get_by_year_and_month(self, year, month, pageIndex=1,  limit
    =50):
185         """Returns a list of AlcalaPage objects in the eXist-db
    instance for a specified year and month."""
186         xquery = str.format("""for $x in doc("alcala/ledger.xml")//
    pages/page
187                                where $x/content[@yearID="{0}"] and $x/
    content/month[@monthID="{1}"]
188                                return <page>{
189                                        $x/pageID
190                                        !$x/content[@yearID="(0)"]
191                                        !<content>{
192                                            @*, month[@monthID={1}]
193                                                !<month>{@*, node()}</
    month>
194                                        }</content>
195                                    }</page>""", year, month)
196         qr = self.db.query(xquery, pageIndex, limit)
197         results = list()
198         for i in range(0, qr.count - 1):
199             page = AlcalaPage(etree.XML(tostring(qr.results[i])))
200             results.append(page)
201 
202         return results
```

C.3.3 MongoData Module

```
from pymongo import MongoClient
from sshtunnel import SSHTunnelForwarder
import pymongo
from bson.objectid import ObjectId
from models.analysisItem import AnalysisItem, CategoryData,
    AnalysisUserItem, TimeSeriesData, KeyTimePivotData, \
     TimeSummary, DataPackage
from models.search import SearchLogEntry
from models.dashboard import CustomPosterInfo, CustomDashboardInfo,
    CustomChartInfo, CustomInfoBox, BoundaryObject
from models.pivotData import CategoryMonthPivotItem,
    CategoryYearPivotItem, MonthYearPivotItem, AllPointsPivotItem, \
     WordFreqPivotItem, CategoryPivotItem, YearPivotItem,
    WordFreqMonthPivotItem, WordFreqYearPivotItem
from models.users import SiteUser
import config
import math


class MongoData:
    """
    Handles all interactions with an instance of MongoDB. Unlike the
    ExistData class, MongoData allows for both read
    and write operations as it is the analysis aspect of the system (
    where eXist-db is a system of record).

    Requirements:
    * pymongo
    * config.py
    * math
    * various custom data models (see import statements).
    """

    def __init__(self,database=None):
        """Initialises an instance of mongodb based on parameters
    supplied in config.py"""
        # Check to see which database to use (if specified. If none
    specified default to the config)
        if database is None:
            use_database = config.MONGODB_CONFIG['database']
        else:
            use_database = database

        print(use_database)

        if 'username' in config.MONGODB_CONFIG:
            self.client = MongoClient(host=config.MONGODB_CONFIG['
    server'],
                                      port=config.MONGODB_CONFIG['port'
    ],
```

```
41                                      username=config.MONGODB_CONFIG['
    username'],
42                                      password=config.MONGODB_CONFIG['
    password'],
43                                      authSource=use_database,
44                                      authMechanism='SCRAM-SHA-1')
45          elif 'ssh' in config.MONGODB_CONFIG: # if ssh is specified then
     we need to connect over an ssh tunnel
46              self.server = SSHTunnelForwarder(config.MONGODB_CONFIG['
    server'],
47                                               ssh_username=config.
    MONGODB_CONFIG['ssh']['user'],
48                                               ssh_password=config.
    MONGODB_CONFIG['ssh']['password'],
49                                               remote_bind_address=(
    config.MONGODB_CONFIG['ssh']['localhost'],
50
    config.MONGODB_CONFIG['ssh']['localport']))
51              self.server.start()
52              self.client = MongoClient(config.MONGODB_CONFIG['ssh']['
    localhost'], self.server.local_bind_port)
53          else:  # if there is no username, we supply fewer parameters to
     the constructor
54              self.client = MongoClient(host=config.MONGODB_CONFIG['
    server'],
55                                        port=config.MONGODB_CONFIG['port'
    ],
56                                        authSource=use_database)
57          # This class only works with alcala data. Although the
    constructor could be expanded to work with other datasets
58          self.db = self.client[use_database]
59
60      def __del__(self):
61          if hasattr(self, 'server') and self.server is not None:
62              self.server.stop()
63
64      def get_search_log(self, userID, searchType=None):
65          """Gets a list of search logs for a specified user (and type if
     requested)"""
66          if searchType is not None:
67              json_doc = self.db.search_log.find({'userID': userID, 'type
    ': searchType}).sort('dateCreated', pymongo.DESCENDING)
68          else:
69              json_doc = self.db.search_log.find({'userID': userID}).sort
    ('dateCreated', pymongo.DESCENDING)
70
71          results = []
72          for j in json_doc:
73              results.append(SearchLogEntry(**j))
74
75          if len(results) == 0:
76              results = None
```

```
77          return results
78
79      def log_search(self, userID, searchParams, searchType, totalHits=
    None):
80          """Logs the search and returns a valid search id"""
81          import datetime
82          if userID is not None:
83              search_obj = { 'userID': userID, 'type': searchType, '
    dateCreated': datetime.datetime.now(), 'params': searchParams.
    get_properties() }
84              if totalHits is not None:
85                  search_obj['totalHits'] = totalHits
86              search_id = self.db.search_log.insert_one(search_obj)
87          else:
88              search_id = None
89
90          return str(search_id.inserted_id)
91
92      def log_features(self, searchID, search_features):
93          """Logs the features of visualisation search for reproduction
    """
94          print('SearchID is %s' % searchID)
95          update_result = self.db.search_log.update({'_id': ObjectId(
    searchID)}, {
96              '$addToSet': {'features': {'$each': [search_features.
    get_properties()]}}
97          })
98          json_doc = self.db.search_log.find({'_id': ObjectId(searchID)})
99          return SearchLogEntry(**json_doc[0])
100
101     def get_custom_charts(self, userID=None):
102         """Gets a list of all custom charts (associated with userID if
    supplied)"""
103         results = []
104         if userID is not None:
105             json_doc = self.db.user_charts.find({'userID': userID})
106         else:
107             json_doc = self.db.user_charts.find()
108
109         for j in json_doc:
110             results.append(CustomChartInfo(**j))
111
112         if len(results) == 0:
113             results = None
114
115         return results
116
117     def get_custom_infoboxes(self, userID=None):
118         """Gets a list of all custom info boxes (associated with userID
     if supplied)"""
119         results = []
120         if userID is not None:
```

```
121             json_doc = self.db.user_infoboxes.find({'userID': userID})
122         else:
123             json_doc = self.db.user_infoboxes.find()
124 
125         for j in json_doc:
126             results.append(CustomInfoBox(**j))
127 
128         if len(results) == 0:
129             results = None
130 
131         return results
132 
133     def get_custom_posters(self, userID=None):
134         """Gets a list of all custom stories (associated with userID if
     supplied)"""
135         results = []
136         if userID is not None:
137             json_doc = self.db.user_posters.find({'userID': userID})
138         else:
139             json_doc = self.db.user_posters.find()
140 
141         for j in json_doc:
142             results.append(CustomPosterInfo(**j))
143 
144         if len(results) == 0:
145             results = None
146 
147         return results
148 
149     def get_boundary_objects(self, userID=None):
150         """Gets a list of all boundary objects (associated with userID
    if supplied"""
151         results = []
152         if userID is not None:
153             json_doc = self.db.boundary_objects.find({'userID': userID
    })
154         else:
155             json_doc = self.db.boundary_objects.find()
156 
157         for j in json_doc:
158             results.append(BoundaryObject(**j))
159 
160         if len(results) == 0:
161             results = None
162 
163         return results
164 
165     def get_custom_dashboard(self, userID):
166         """Gets the dashboard for the specified user"""
167         result = None
168         json_doc = self.db.user_dashboard.find({'userID': userID})
169         for j in json_doc:
```

```
170             result = CustomDashboardInfo(**j)
171
172         if result is None:
173             result = CustomDashboardInfo(userID=userID, charts=[],
    infoBoxes=[], stories=[], boundaryObjects=[])
174             result._id = self.insert_custom_dashboard(result)
175
176         return result
177
178     def insert_custom_infobox(self, infoBox):
179         """Inserts a saved, customised info box into the database"""
180         infoBox_id = self.db.user_infoboxes.insert_one(infoBox.
    get_properties())
181         return str(infoBox_id.inserted_id)
182
183     def update_custom_infobox(self, infoBox):
184         update_result = self.db.user_infoboxes.update({'_id': ObjectId(
    infoBox._id)}, {
185             '$set': {'type': infoBox.type, 'icon': infoBox.icon, 'label
    ': infoBox.label, 'colour': infoBox.colour}
186         })
187         query = self.db.user_infoboxes.find({'_id': ObjectId(infoBox.
    _id)})
188         return CustomInfoBox(**query[0])
189
190     def delete_custom_infobox(self, infobox_id):
191         delete_result = self.db.user_infoboxes.remove({'_id': ObjectId(
    infobox_id)}, {'justOne': True})
192         self.db.user_dashboard.update({}, {
193             '$pull': {'infoBoxes': {'$in': [infobox_id]}}
194         })
195         return True # This means we didn't raise an exception
196
197     def insert_custom_chart(self, chartObj):
198         """Inserts a saved, customised chart into the database"""
199         chart_id = self.db.user_charts.insert_one(chartObj.
    get_properties())
200         return str(chart_id.inserted_id)
201
202     def update_custom_chart(self, chartObj):
203         update_result = self.db.user_charts.update({'_id': ObjectId(
    chartObj._id)},{
204             '$set': {'title': chartObj.title, 'description': chartObj.
    description}
205         })
206         json_doc = self.db.user_charts.find({'_id': ObjectId(chartObj.
    _id)})
207         return CustomChartInfo(**json_doc[0])
208
209     def delete_custom_chart(self, chart_id):
210         delete_result = self.db.user_charts.remove({'_id': ObjectId(
    chart_id)}, {'justOne': True})
```

```
211         self.db.user_stories.update({}, {
212             '$pull': {'charts': {'$in': [chart_id]}}
213         })
214         self.db.user_dashboard.update({}, {
215             '$pull': {'charts': {'$in': [chart_id]}}
216         })
217         return True # this means we didn't raise an exception
218 
219     def get_custom_poster_by_id(self, id):
220         json_doc = self.db.user_posters.find({'_id': ObjectId(id)})
221         return CustomPosterInfo(**json_doc[0])
222 
223     def insert_custom_poster(self, posterObj):
224         """Inserts a customised user story into the database"""
225         poster_id = self.db.user_posters.insert_one(posterObj.
    get_properties())
226         return str(poster_id.inserted_id)
227 
228     def update_custom_poster(self, posterInfo):
229         """Updates a custom poster object"""
230         self.db.user_posters.update({'_id': ObjectId(posterInfo._id)},
    {
231             '$unset': {'sections': ''}
232         })
233 
234         self.db.user_posters.update({'_id': ObjectId(posterInfo._id)},
    {
235             '$set': {'title': posterInfo.title, 'description':
    posterInfo.description},
236             '$addToSet': {'sections': {'$each': posterInfo.sections}}
237         })
238 
239         json_doc = self.db.user_posters.find({'_id': ObjectId(
    posterInfo._id)})
240         return CustomPosterInfo(**json_doc[0])
241 
242     def delete_custom_poster(self, poster_id):
243         delete_result = self.db.user_posters.remove({'_id': ObjectId(
    poster_id)}, {'justOne': True})
244         return True # this means we didn't raise an exception
245 
246     def get_boundary_object_by_id(self, id):
247         json_doc = self.db.boundary_objects.find({'_id': ObjectId(id)})
248         return BoundaryObject(**json_doc[0])
249 
250     def insert_boundary_object(self, boInfo):
251         bo_id = self.db.boundary_objects.insert_one(boInfo.
    get_properties())
252         return str(bo_id.inserted_id)
253 
254     def update_boundary_object(self, boInfo):
```

```
255         self.db.boundary_objects.update({'_id': ObjectId(boInfo._id)},
    {
256             '$set': {
257                 'type': boInfo.type,
258                 'title': boInfo.title,
259                 'description': boInfo.description,
260                 'totalItems': boInfo.totalItems,
261                 'params': boInfo.params,
262                 'features': boInfo.features,
263                 'pageID': boInfo.pageID
264             }
265         })
266         json_doc = self.db.boundary_objects.find({'_id': ObjectId(
    boInfo._id)})
267         return BoundaryObject(**json_doc[0])
268 
269     def delete_boundary_object(self, bo_id):
270         delete_result = self.db.boundary_objects.remove({'_id':
    ObjectId(bo_id)}, {'justOne': True})
271         self.db.user_dashboard.update({}, {
272             '$pull': {'boundaryObjects': {'$in': [bo_id]}}
273         })
274         self.db.user_posters.update({}, {
275             '$pull': {'sections.$[].boundaryObjects': {'$in': [bo_id]}}
276         })
277         return True # This means we didn't throw an exception
278 
279     def insert_custom_dashboard(self, dashboardObj):
280         """Inserts a customised dashboard object into the database"""
281         dashboard_id = self.db.user_dashboard.insert_one(dashboardObj.
    get_properties())
282         return str(dashboard_id.inserted_id)
283 
284     def update_custom_dashboard(self, dashboardObj):
285         query = self.db.user_dashboard.find({'_id': ObjectId(
    dashboardObj._id)})
286         db_dashboard = CustomDashboardInfo(**query[0])
287         removed_charts = [x for x in db_dashboard.charts if x not in
    dashboardObj.charts]
288         removed_stories = [x for x in db_dashboard.stories if x not in
    dashboardObj.stories]
289         removed_infoBoxes = [x for x in db_dashboard.infoBoxes if x not
     in dashboardObj.infoBoxes]
290         removed_boundaryObjects = [x for x in db_dashboard.
    boundaryObjects if x not in dashboardObj.boundaryObjects]
291 
292         self.db.user_dashboard.update({'_id': ObjectId(dashboardObj._id
    )}, {
293             '$addToSet': {
294                 'charts': {'$each': dashboardObj.charts},
295                 'stories': {'$each': dashboardObj.stories},
296                 'infoBoxes': {'$each': dashboardObj.infoBoxes},
```

```
297                 'boundaryObjects': {'$each': dashboardObj.
    boundaryObjects}
298             }
299         })
300         self.db.user_dashboard.update({'_id': ObjectId(dashboardObj._id
    )}, {
301             '$pull': {
302                 'charts': {'$in': removed_charts},
303                 'stories': {'$in': removed_stories},
304                 'infoBoxes': {'$in': removed_infoBoxes},
305                 'boundaryObjects': {'$in': removed_boundaryObjects}
306             }
307         })
308         query = self.db.user_dashboard.find({'_id': ObjectId(
    dashboardObj._id)})
309         return CustomDashboardInfo(**query[0])
310 
311     def insert_one_transaction(self, transaction, use_training=False):
312         """Inserts a single transaction (AnalysisItem object) into
    either the transaction or training collection."""
313         if use_training:
314             transaction_id = self.db.transactions_training.insert_one(
    transaction.get_properties())
315         else:
316             transaction_id = self.db.transactions.insert_one(
    transaction.get_properties())
317         return transaction_id
318 
319     def insert_multiple_transactions(self, transaction_list,
    use_training=False):
320         """Inserts multiple transactions (AnalysisItem object) into
    either the transaction or training collection."""
321         json_docs = list(map(lambda t: t.get_properties(),
    transaction_list))
322         if use_training:
323             result = self.db.transactions_training.insert_many(
    json_docs)
324         else:
325             result = self.db.transactions.insert_many(json_docs)
326         return result.inserted_ids
327 
328     def update_multiple_transactions(self, transaction_list,
    use_training=False):
329         """Updates multiple transactions (AnalysisItem object) in
    either the transaction or training collection."""
330         if use_training:
331             bulk = self.db.transactions_training.
    initialize_ordered_bulk_op()
332         else:
333             bulk = self.db.transactions.initialize_ordered_bulk_op()
334 
335         for t in transaction_list:
```

```
336             bulk.find({'_id': t._id}).update({'$set': {'categories': t.
    categories}})
337 
338         bulk.execute()
339 
340     def update_training_data(self, id, categories):
341         """Updates the cateogries on training data in the curated
    training data collection."""
342         db_result = self.db.curated_training.update_one({'_id': id}, {'
    $set': {'categories': categories}}, upsert=False)
343         query = self.db.curated_training.find({'_id': id})
344         return AnalysisUserItem(**query[0])
345 
346     def get_word_summary(self, year=None):
347         """
348         Returns a list of WordFreqPivotItem objects that provides
    aggregates of amounts for each word in the system.
349         NOTE: Can be filtered by year.
350         """
351 
352         pipeline = []
353         pipeline.append({"$unwind": "$words"})
354         if year is not None:
355             pipeline.append({"$match": {"year": year}})
356         pipeline.append({"$group": {
357                                 "_id": {
358                                     "word": "$words"
359                                 },
360                                 "reales": {"$sum": "$reales"},
361                                 "maravedises": {"$sum": "$maravedises"
    },
362                                 "transaction_count": {"$sum": 1}
363                             }
364                         })
365 
366         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
367         results = list()
368         for j in json_list['cursor']['firstBatch']:
369             results.append(WordFreqPivotItem(word=j['_id']['word'],
    frequency=j['transaction_count'], **j))
370         if len(results) == 0:
371             results = None
372         return results
373 
374     def get_full_summary(self, filters=None):
375         """
376         Returns a list of AllPointsPivotItem objects that provides
    aggregates of amounts for each category in the system
377         broken down by year and month.
378         NOTE: Can be filtered by year.
379         """
```

```
380
381         pipeline = []
382         if filters is not None:
383             if filters.keywords is not None:
384                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
385             if filters.year is not None:
386                 pipeline.append({"$match": {"year": filters.year}})
387             if filters.filteredCategories is not None:
388                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
389
390         pipeline.append({"$unwind": "$categories"})
391         pipeline.append({"$group": {
392                                 "_id": {
393                                     "year": "$year",
394                                     "month": "$month",
395                                     "category": "$categories"
396                                 },
397                                 "reales": {"$sum": "$reales"},
398                                 "maravedises": {"$sum": "$maravedises"
    },
399                                 "transaction_count": {"$sum": 1}
400                             }
401                         })
402         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
403         results = list()
404         for j in json_list['cursor']['firstBatch']:
405             results.append(AllPointsPivotItem(category=j['_id']['
    category'], month=j['_id']['month'], year=j['_id']['year'], **j))
406         if len(results) == 0:
407             results = None
408         return results
409
410     def get_category_summary(self, filters=None):
411         """
412         Returns a list of CaetegoryPivotItem objects that provides
    aggregates of amounts for each category in the system
413         NOTE: Can be filtered by year.
414         """
415
416         pipeline = []
417         if filters is not None:
418             if filters.keywords is not None:
419                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
420             if filters.year is not None:
421                 pipeline.append({"$match": {"year": filters.year}})
422             if filters.filteredCategories is not None:
423                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
```

```
424
425         pipeline.append({"$unwind": "$categories"})
426         pipeline.append({"$group":  {
427                                     "_id": "$categories",
428                                     "reales": { "$sum": "$reales"},
429                                     "maravedises": { "$sum": "$maravedises"
    },
430                                     "transaction_count": {"$sum": 1}
431                                 }
432                         })
433
434         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
435         results = list()
436         for j in json_list['cursor']['firstBatch']:
437             results.append(CategoryPivotItem(category=j['_id'], **j))
438         if len(results) == 0:
439             results = None
440         return results
441
442     def get_time_series_data(self, keyName, listData, timeType):
443         """
444         :param keyName: the name of the key item which you are slicing
    time data for (typically category or word)
445         :param listData: the data from which you are deriving the time
    data
446         :param timeType: whether you are slicing by year (y) or month (
    m)
447         :return: A list of KeyTimePivotData objects that shows the
    total amount and transaction count broken down by the
448         requisite time series (year or month) for each key in the
    dataset.
449         """
450
451         results = []
452         curr_key = None
453         time_data = []
454
455         # If we are dealing with years, there are certain years which
    exist in the database. Otherwise if we are dealing
456         # with months then we want to get a list of valid months. This
    will be used to "fill in" missing times in the data
457         # for each key
458         if timeType == 'y':
459             validTimes = [1774, 1775, 1776, 1777, 1778, 1779, 1781]
460             timeAttr = 'year'
461         elif timeType == 'm':
462             validTimes = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
463             timeAttr = 'monthNum'
464         else:
465             raise Exception('Invalid timeType supplied')
466
```

```
467         prevTime = validTimes[0] - 1
468         if listData is not None:
469             for item in listData:
470                 if curr_key != item[keyName] and curr_key is not None:
471                     lastIndex = validTimes.index(prevTime)
472                     for i in range(lastIndex + 1, len(validTimes), 1):
473                         time_data.append(TimeSeriesData(timeValue=
    validTimes[i], timeType=timeType, totalAmount=0, transactionCount
    =0))
474                     results.append(KeyTimePivotData(key=curr_key,
    timeSeries=time_data))
475                     time_data = []
476                     curr_key = item[keyName]
477                     prevTime = validTimes[0] - 1
478                 elif curr_key is None:
479                     curr_key = item[keyName]
480 
481                 if prevTime not in validTimes:
482                     startLoop = 0
483                 else:
484                     startLoop = validTimes.index(prevTime) + 1
485 
486                 for i in range(startLoop, validTimes.index(item[
    timeAttr]), 1):
487                     time_data.append(TimeSeriesData(timeValue=
    validTimes[i], timeType=timeType, totalAmount=0, transactionCount
    =0))
488 
489                 time_data.append(TimeSeriesData(timeValue=item[timeAttr
    ], timeType=timeType, totalAmount=item.totalAmount,
    transactionCount=item.transactionCount))
490                 prevTime = item[timeAttr]
491 
492         # Need to add in the last of the data:
493         if prevTime == validTimes[0] - 1:
494             lastIndex = validTimes.index(validTimes[0])
495         else:
496             lastIndex = validTimes.index(prevTime)
497         for i in range(lastIndex + 1, len(validTimes), 1):
498             time_data.append(
499                 TimeSeriesData(timeValue=validTimes[i], timeType=
    timeType, totalAmount=0, transactionCount=0))
500         results.append(KeyTimePivotData(key=curr_key, timeSeries=
    time_data))
501         return results
502 
503     def get_word_time_data(self, searchParams=None):
504         """
505         :param year: Specify a year if you wish to see a monthly
    breakdown. Set to None if you wish to see a yearly breakdown
506         :return: A DataPackage object which contains overall summary
    information plus a breakdown of timeseries data for each
```

```
507         word.
508         """
509 
510         if searchParams is not None and searchParams.year is not None:
511             temp_data = self.get_word_by_month_summary(filters=
    searchParams)
512             if temp_data is not None:
513                 temp_data = sorted(temp_data, key=lambda x: (x.word, x.
    monthNum))
514             time_summary = self.get_month_summary(filters=searchParams)
515             timeType = 'm'
516             timeKey = 'month'
517         else:
518             temp_data = self.get_word_by_year_summary(filters=
    searchParams)
519             if temp_data is not None:
520                 temp_data = sorted(temp_data, key=lambda x: (x.word, x.
    year))
521             time_summary = self.get_year_summary(filters=searchParams)
522             timeType = 'y'
523             timeKey = 'year'
524 
525         results = self.get_time_series_data(keyName='word', timeType=
    timeType, listData=temp_data)
526         rawData = self.search_transactions(searchParams)
527         summary_info = self.get_total_spent(filters=searchParams)
528         if summary_info is not None:
529             grand_total = summary_info['reales'] + math.floor(
    summary_info['maravedises'] / 34) + (
530                           (summary_info['maravedises'] % 34) / 100)
531             total_reales = summary_info['reales']
532             total_maravedises = summary_info['maravedises']
533             total_transactions = summary_info['transaction_count']
534         else:
535             grand_total = None
536             total_reales = None
537             total_maravedises = None
538             total_transactions = None
539 
540         time_results = list()
541         if time_summary is not None:
542             for t in time_summary:
543                 time_results.append(TimeSummary(timeValue=t[timeKey],
    timeType=timeType, reales=t.reales,
544                                                 maravedises=t.
    maravedises, totalAmount=t.totalAmount,
545                                                 transactionCount=t.
    transactionCount))
546 
547         return DataPackage(reales=total_reales, maravedises=
    total_maravedises, grandTotal=grand_total,
```

```
548                         totalTransactions=total_transactions,
    timeSummary=time_results, data=results,
549                         rawData=rawData)
550
551     def get_category_time_data(self, searchParams=None):
552         """
553         :param year: Specify a year if you wish to see a monthly
    breakdown. Set to None if you wish to see a yearly breakdown
554         :return: A DataPackage object which contains overall summary
    information plus a breakdown of timeseries data for each
555         category.
556         """
557
558         if searchParams is not None and searchParams.year is not None:
559             temp_data = self.get_category_by_month_summary(filters=
    searchParams)
560             if temp_data is not None:
561                 temp_data = sorted(temp_data, key=lambda x: (x.category
    , x.monthNum))
562             time_summary = self.get_month_summary(filters=searchParams)
563             timeType = 'm'
564             timeKey = 'month'
565         else:
566             temp_data = self.get_category_by_year_summary(filters=
    searchParams)
567             if temp_data is not None:
568                 temp_data = sorted(temp_data, key=lambda x: (x.category
    , x.year))
569             time_summary = self.get_year_summary(filters=searchParams)
570             timeType = 'y'
571             timeKey = 'year'
572
573         results = self.get_time_series_data(keyName='category',
    timeType=timeType, listData=temp_data)
574         rawData = self.search_transactions(searchParams=searchParams)
575         summary_info = self.get_total_spent(filters=searchParams)
576         if summary_info is not None:
577             grand_total = summary_info['reales'] + math.floor(
    summary_info['maravedises'] / 34) + (
578                           (summary_info['maravedises'] % 34) / 100)
579             total_reales = summary_info['reales']
580             total_maravedises = summary_info['maravedises']
581             total_transactions = summary_info['transaction_count']
582         else:
583             grand_total = None
584             total_reales = None
585             total_maravedises = None
586             total_transactions = None
587
588         time_results = list()
589         if time_summary is not None:
590             for t in time_summary:
```

```
591                 time_results.append(TimeSummary(timeValue=t[timeKey],
    timeType=timeType, reales=t.reales,
592                                                 maravedises=t.
    maravedises, totalAmount=t.totalAmount,
593                                                 transactionCount=t.
    transactionCount))
594
595         return DataPackage(reales=total_reales, maravedises=
    total_maravedises, grandTotal=grand_total,
596                            totalTransactions=total_transactions,
    timeSummary=time_results, data=results,
597                            rawData=rawData)
598
599     def get_categories(self):
600         json_list = self.db.transactions.distinct('categories')
601         return json_list
602
603     def get_total_spent(self, filters=None):
604         pipeline = []
605         if filters is not None:
606             if filters.keywords is not None:
607                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
608             if filters.year is not None:
609                 pipeline.append({"$match": {"year": filters.year}})
610             if filters.filteredCategories is not None:
611                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
612
613         # pipeline.append({"$unwind": "$categories"})
614         pipeline.append({"$group": {
615             "_id": None,
616             "reales": {"$sum": "$reales"},
617             "maravedises": {"$sum": "$maravedises"},
618             "transaction_count": {"$sum": 1}
619         }})
620
621         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
622         if len(json_list['cursor']['firstBatch']):
623             return json_list['cursor']['firstBatch'][0]
624         else:
625             return None
626
627     def get_word_by_year_summary(self, filters=None):
628         pipeline = []
629         if filters is not None:
630             if filters.keywords is not None:
631                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
632             if filters.year is not None:
633                 pipeline.append({"$match": {"year": filters.year}})
```

```
634             if filters.filteredCategories is not None:
635                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
636 
637         pipeline.append({"$unwind": "$words"})
638         pipeline.append({"$group": {
639             "_id": {
640                 "year": "$year",
641                 "word": "$words"
642             },
643             "reales": {"$sum": "$reales"},
644             "maravedises": {"$sum": "$maravedises"},
645             "transaction_count": {"$sum": 1}
646         }})
647 
648         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
649         results = list()
650         for j in json_list['cursor']['firstBatch']:
651             results.append(WordFreqYearPivotItem(year=j['_id']['year'],
     word=j['_id']['word'], **j))
652         if len(results) == 0:
653             results = None
654         return results
655 
656     def get_word_by_month_summary(self, filters=None):
657         pipeline = []
658         if filters is not None:
659             if filters.keywords is not None:
660                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
661             if filters.year is not None:
662                 pipeline.append({"$match": {"year": filters.year}})
663             if filters.filteredCategories is not None:
664                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
665 
666         pipeline.append({"$unwind": "$words"})
667         pipeline.append({"$group": {
668             "_id": {
669                 "month": "$month",
670                 "word": "$words"
671             },
672             "reales": {"$sum": "$reales"},
673             "maravedises": {"$sum": "$maravedises"},
674             "transaction_count": {"$sum": 1}
675         }})
676 
677         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
678         results = list()
679         for j in json_list['cursor']['firstBatch']:
```

```
680             results.append(WordFreqMonthPivotItem(monthNum=j['_id']['
    month'], word=j['_id']['word'], **j))
681         if len(results) == 0:
682             results = None
683         return results
684 
685     def get_category_by_year_summary(self, filters=None):
686         pipeline = []
687         if filters is not None:
688             if filters.keywords is not None:
689                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
690             if filters.year is not None:
691                 pipeline.append({"$match": {"year": filters.year}})
692             if filters.filteredCategories is not None:
693                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
694 
695         pipeline.append({"$unwind": "$categories"})
696         pipeline.append({"$group": {
697                             "_id": {
698                                 "year": "$year",
699                                 "category": "$categories"
700                             },
701                             "reales": {"$sum": "$reales"},
702                             "maravedises": {"$sum": "$maravedises"},
703                             "transaction_count": {"$sum": 1}
704                         }
705                     })
706 
707         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
708         results = list()
709         for j in json_list['cursor']['firstBatch']:
710             results.append(CategoryYearPivotItem(year=j['_id']['year'],
     category=j['_id']['category'], **j))
711         if len(results) == 0:
712             results = None
713         return results
714 
715     def get_category_by_month_summary(self, filters=None):
716         """
717         Returns a list of CategoryMonthPivotItem objects that provides
    aggregates of amounts for each category in the system
718         broken down by month.
719         NOTE: Can be filtered by year.
720         """
721 
722         pipeline = []
723         if filters is not None:
724             if filters.keywords is not None:
```

```
725                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
726             if filters.year is not None:
727                 pipeline.append({"$match": {"year": filters.year}})
728             if filters.filteredCategories is not None:
729                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
730 
731         pipeline.append({"$unwind": "$categories"})
732         pipeline.append({"$group": {
733                             "_id": {
734                                 "category": "$categories",
735                                 "month": "$month"
736                             },
737                             "reales": {"$sum": "$reales"},
738                             "maravedises": {"$sum": "$maravedises"},
739                             "transaction_count": {"$sum": 1}
740                         }})
741 
742         json_list = self.db.command('aggregate', 'transactions',
    pipeline=pipeline, explain=False)
743 
744         results = list()
745         for j in json_list['cursor']['firstBatch']:
746             results.append(CategoryMonthPivotItem(month_name=self.
    get_month_name(j['_id']['month']), month=j['_id']['month'],
747                                                   category=j['_id']['
    category'], **j))
748         if len(results) == 0:
749             results = None
750         return results
751 
752     def get_month_summary(self, filters=None):
753         pipeline = []
754 
755         if filters is not None:
756             if filters.keywords is not None:
757                 pipeline.append({"$match":{"$text": {"$search": filters
    .keywords}}})
758             if filters.year is not None:
759                 pipeline.append({"$match": {"year": filters.year}})
760             if filters.filteredCategories is not None:
761                 pipeline.append({"$match": {"categories": {"$in":
    filters.filteredCategories}}})
762 
763         # pipeline.append({"$unwind": "$categories"})
764         pipeline.append({"$group": {
765                             "_id": {
766                                 "year": "$year",
767                                 "month": "$month"
768                             },
769                             "reales": {"$sum": "$reales"},
```

```
                                "maravedises": {"$sum": "$maravedises"},
                                "transaction_count": {"$sum": 1}
                            }
                        })

        json_list = self.db.command('aggregate', 'transactions', pipeline=pipeline, explain=False)
        results = list()
        for j in json_list['cursor']['firstBatch']:
            results.append(MonthYearPivotItem(year=j['_id']['year'], month=self.get_month_name(j['_id']['month']), **j))
        if len(results) == 0:
            results = None
        return results

    def get_year_summary(self, filters=None):
        pipeline = []
        if filters is not None:
            if filters.keywords is not None:
                pipeline.append({"$match":{"$text": {"$search": filters.keywords}}})
            if filters.year is not None:
                pipeline.append({"$match": {"year": filters.year}})
            if filters.filteredCategories is not None:
                pipeline.append({"$match": {"categories": {"$in": filters.filteredCategories}}})

        # pipeline.append({"$unwind": "$categories"})
        pipeline.append({"$group": {
                                "_id": {
                                    "year": "$year"
                                },
                                "reales": {"$sum": "$reales"},
                                "maravedises": {"$sum": "$maravedises"},
                                "transaction_count": {"$sum": 1}
                            }
                        })

        json_list = self.db.command('aggregate', 'transactions', pipeline=pipeline, explain=False)
        results = list()
        for j in json_list['cursor']['firstBatch']:
            results.append(YearPivotItem(year=j['_id']['year'], **j))
        if len(results) == 0:
            results = None
        return results

    def get_month_name(self, month):
        import calendar
        return calendar.month_name[month]
```

```
815
816     def insert_category_colours(self, category_colours):
817         json_docs = list(map(lambda t:t.get_properties(),
    category_colours))
818         result = self.db.category_info.insert_many(json_docs)
819         return result.inserted_ids
820
821     def get_category_colours(self):
822         json_list = self.db.category_info.find()
823         results = list()
824         for j in json_list:
825             results.append(CategoryData(**j))
826         if len(results) == 0:
827             results = None
828         return results
829
830     def insert_user(self, user):
831         db_user = self.get_user(user.username)
832         if db_user is None:
833             new_user = self.db.user_info.insert_one(user.get_properties
    ())
834             result = new_user.inserted_id
835         else:
836             result = db_user._id
837         return str(result)
838
839     def get_training_data_by_user(self, user_id):
840         json_doc = self.db.curated_training.find({'userId': user_id})
841         if json_doc is None:
842             return None
843         else:
844             results = list()
845             for j in json_doc:
846                 results.append(AnalysisUserItem(**j))
847             return results
848
849     def insert_multiple_training_for_curation(self, transaction_list):
850         json_docs = list(map(lambda t: t.get_properties(),
    transaction_list))
851         result = self.db.curated_training.insert_many(json_docs)
852         return result.inserted_ids
853
854     def get_all_users(self):
855         json_list = self.db.user_info.find()
856         results = list()
857         for j in json_list:
858             results.append(SiteUser(**j))
859
860         if len(results) == 0:
861             results = None
862
863         return results
```

```
864
865     def get_user(self, username):
866         json_doc = self.db.user_info.find_one({'username': username})
867         if json_doc is None:
868             return None
869         else:
870             return SiteUser(**json_doc)
871
872     def get_user_by_id(self, id):
873         json_doc = self.db.user_info.find_one({'_id': ObjectId(id)})
874         if json_doc is None:
875             return None
876         else:
877             return SiteUser(**json_doc)
878
879     def delete_user(self, user):
880         self.db.user_infoboxes.remove({'userID': user._id})
881         self.db.user_charts.remove({'userID': user._id})
882         self.db.user_stories.remove({'userID': user._id})
883         self.db.user_dashboard.remove({'userID': user._id})
884         self.db.search_log.remove({'userID': user._id})
885         result = self.db.user_info.remove({'_id': ObjectId(user._id)},
    {'justOne': True})
886         return result;
887
888     def update_user(self, user):
889         query = self.db.user_info.find({'_id': ObjectId(user._id)})
890         db_user = SiteUser(**query[0])
891         removed_roles = [x for x in db_user.roles if x not in user.
    roles]
892
893         self.db.user_info.update_one({'_id': ObjectId(user._id)}, {'
    $set': {
894             'allowLogging': user.allowLogging,
895             'emailAddress': user.emailAddress,
896             'firstname': user.firstname,
897             'password': user.password,
898             'username': user.username,
899             'surname': user.surname,
900             'loginToken': user.loginToken,
901             'lastLogin': user.lastLogin
902             },
903             '$addToSet': {'roles': {'$each': user.roles} }
904         })
905         self.db.user_info.update({'_id': ObjectId(user._id)}, {
906             '$pull': {'roles': {'$in': removed_roles}}
907         })
908         query = self.db.user_info.find({'_id': ObjectId(user._id)})
909         return SiteUser(**query[0])
910
911     def search_transactions(self, searchParams=None):
912         filters = []
```

```
913         if searchParams is not None:
914             if searchParams.year is not None:
915                 filters.append({'year': searchParams.year})
916             if searchParams.keywords is not None:
917                 filters.append({'$text': {'$search': searchParams.
    keywords}})
918             if searchParams.filteredCategories is not None:
919                 filters.append({'categories': {'$in': searchParams.
    filteredCategories}})
920 
921         if len(filters) > 0:
922             query = {'$and': filters}
923         else:
924             query = None
925 
926         json_list = self.db.transactions.find(query)
927         results = list()
928         for j in json_list:
929             results.append(AnalysisItem(**j))
930         if len(results) == 0:
931             results = None
932         return results
933 
934     def get_transactions(self, use_training=False, year=None):
935         if use_training:
936             collection_name = 'transactions_training'
937         else:
938             collection_name = 'transactions'
939 
940         if year is not None:
941             json_list = self.db[collection_name].find({'year': year})
942         else:
943             json_list = self.db[collection_name].find()
944 
945         results = list()
946         for j in json_list:
947             results.append(AnalysisItem(**j))
948         if len(results) == 0:
949             results = None
950         return results
951 
952     def ticks_since_epoch(self, year=None, month=None):
953         from datetime import datetime
954 
955         start_time = datetime.utcnow()
956         if year is not None:
957             start_time = datetime(year, 1, 1)
958         elif month is not None:
959             start_time = datetime(1781, month, 1)
960 
961         #ticks_per_ms = 10000
962         ms_per_second = 1000
```

```
963         #ticks_per_second = ticks_per_ms * ms_per_second
964         span = datetime(1970, 1, 1) - start_time
965         ticks = int(span.total_seconds() * ms_per_second)
966         return ticks
```

C.3.4 Abstract Factory Chart Component

```
1  import { DataSummaryPackage } from '../../../models/analysis-result';
2  import { EventDispatcher } from 'strongly-typed-events';
3
4  export interface IBaseChart {
5    allowableXFields: string[];
6    allowableYFields: string[];
7    allowableSizeFields: string[];
8    allowableGroupFields: string[];
9    selectedData: any[];
10
11   formatAxisData(axis: string, dataValue: any): any;
12   getAxisLabel(axis: string): string;
13   formatTimeTicks(date_value: number, date_type: string): any;
14   createOptions(): any;
15   formatData(chartData: DataSummaryPackage): any;
16 }
17
18 export abstract class BaseChart implements IBaseChart {
19   abstract allowableXFields: string[];
20   abstract allowableYFields: string[];
21   abstract allowableSizeFields: string[];
22   abstract allowableGroupFields: string[];
23   protected _onElementClick = new EventDispatcher<IBaseChart, any>();
24   selectedData: any[] = [];
25
26   abstract createOptions(): any;
27   abstract formatData(chartData: DataSummaryPackage): any;
28   get onElementClick() {
29     return this._onElementClick.asEvent();
30   }
31
32   protected constructor(public xField: string, public yField: string,
       public height: number, public width: number, public groupField?:
       string, public sizeField?: string ) {}
33
34   formatAxisData(axis: string, dataValue: any) {
35     switch (axis) {
36       case 'year':
37       case 'monthNum':
38         const timeFormat = axis === 'year' ? '%Y' :  '%b';
39         const ticks = dataValue > 0 ? this.formatTimeTicks(dataValue,
     axis === 'year' ? 'y' : 'm') : dataValue;
40         return d3.time.format(timeFormat)(new Date(ticks));
41       case 'totalAmount':
42         return d3.format('0f')(dataValue);
```

```
      case 'transactionCount':
        return d3.format('d')(dataValue);
      default:
        return dataValue;
    }
  }

  getAxisLabel(axis: string) {
    let result = axis;
    switch (axis) {
      case 'totalAmount': result = 'Total Spent'; break;
      case 'transactionCount': result = '# of Occurrences'; break;
      case 'year': result = 'Years'; break;
      case 'monthNum': result = 'Months'; break;
      case 'category': result = 'Categories'; break;
      case 'word': result = 'Words'; break;
    }

    return result;
  }

  formatTimeTicks(date_value: number, date_type: string): any {
    if (date_type === 'm') {
      return new Date(1781, date_value - 1, 1).getTime();
    } else if (date_type === 'y') {
      return new Date(date_value,  0, 1).getTime();
    } else {
      throw new Error(`${date_type} is not a valid date_type argument. Current valid options are 'm' and 'y'.`);
    }
  }
}

import {BaseChart} from '../base.chart';
import {DataSummaryPackage} from '../../../../models/analysis-result';

export class NVD3StackedAreaChart extends BaseChart {
  allowableXFields: string[] = ['time'];
  allowableYFields: string[] = ['totalAmount', 'transactionCount'];
  allowableSizeFields: string[] = [];
  allowableGroupFields: string[] = [];

  constructor(xField: string, yField: string, height: number, width: number) {
    super(xField, yField, height, width);
  }

  formatData(chartData: DataSummaryPackage): any {
    const results = [];
    const self = this;
    chartData.data.forEach(x => {
      const plot_values = [];
```

```
93      x.timeSeries.forEach(y => {
94        const timeDataValues = [];
95        timeDataValues.push(self.formatTimeTicks(y.timeValue, y.
    timeType));
96        switch (self.yField) {
97          case 'totalAmount':
98            timeDataValues.push(y.totalAmount);
99            break;
100         case 'transactionCount':
101           timeDataValues.push(y.transactionCount);
102           break;
103         default:
104           throw new Error(`The stacked area chart does not allow ${
    self.yField} for the yAxis.`);
105       }
106       plot_values.push(timeDataValues);
107     });
108     results.push({'key': x.key, 'values': plot_values});
109   });
110   return results;
111 }
112
113 createOptions(): any {
114   const self = this;
115   const options = {
116     chart: {
117       type: 'stackedAreaChart',
118       height: this.height,
119       width: this.width,
120       showLegend: false,
121       showControls: false,
122       margin: {
123         top: 20,
124         right: 20,
125         bottom: 40,
126         left: 60
127       },
128       x: function(d) { return d[0]; },
129       y: function(d) { return d[1]; },
130       useVoronoi: false,
131       clipEdge: true,
132       duration: 100,
133       useInteractiveGuideline: false,
134       callback: function(chart) {
135         chart.stacked.dispatch.on('elementClick', function(e) {
136           self._onElementClick.dispatch(self, e);
137         });
138       },
139       xAxis: {
140         axisLabel: super.getAxisLabel(this.xField),
141         showMaxMin: false,
142         tickFormat: function(x) {
```

```
            return self.formatAxisData(self.xField, x);
          }
        },
        yAxis: {
          axisLabel: super.getAxisLabel(this.yField),
          tickFormat: function(y) {
            self.formatAxisData(self.yField, y);
          },
          zoom: {
            enabled: true,
            scaleExtent: [1, 10],
            useFixedDomain: false,
            useNiceScale: false,
            horizontalOff: false,
            verticalOff: true,
            unzoomEventType: 'dblclick.zoom'
          }
        }
      }
    };
    return options;
  }
}
import {BaseChart} from '../base.chart';
import {DataSummaryPackage, TimeSeriesData} from '../../../../models/
    analysis-result';

export class NVD3ScatterChart extends BaseChart {
  allowableXFields: string[] = ['totalAmount', 'transactionCount'];
  allowableYFields: string[] = ['totalAmount', 'transactionCount'];
  allowableSizeFields: string[] = ['totalAmount', 'transactionCount'];
  allowableGroupFields: string[] = [];

  constructor(xField: string, yField: string, height: number, width:
    number, sizeField: string) {
    super(xField, yField, height, width, undefined, sizeField);
  }

  formatData(chartData: DataSummaryPackage): any {
    const results = [];
    const sizeTotals = this.calculateSizeTotals(chartData);
    chartData.data.forEach(x => {
      const plot_data = [];
      x.timeSeries.forEach(y => {
        plot_data.push({'x': this.calculateX(y), 'y': this.calculateY(y
    ), 'size': this.calculateSize(sizeTotals, x.key, y)});
      });
      results.push({'key': x.key, 'values': plot_data});
    });
    return results;
  }

```

```
192  createOptions(): any {
193    const self = this;
194    const options = {
195      chart: {
196        type: 'scatterChart',
197        height: this.height,
198        width: this.width,
199        scatter: {
200          onlyCircles: false
201        },
202        showDistX: true,
203        showDistY: true,
204      },
205      duration: 350,
206      xAxis: {
207        axisLabel: this.getAxisLabel(this.xField),
208        tickFormat: function(d) {
209          return self.formatAxisData(self.xField, d);
210        }
211      },
212      yAxis: {
213        axisLabel: this.getAxisLabel(this.yField),
214        tickFormat: function(d) {
215          return self.formatAxisData(self.yField, d);
216        },
217        axisLabelDistance: -5
218      },
219      zoom: {
220        enabled: true,
221        scaleExtent: [1, 10],
222        useFixedDomain: false,
223        useNiceScale: false,
224        horizontalOff: false,
225        verticalOff: false,
226        unzoomEventType: 'dblclick.zoom'
227      }
228    };
229    return options;
230  }
231
232  formatAxisData(axis: string, dataValue: any) {
233    switch (axis) {
234      case 'year':
235      case 'monthNum':
236        return dataValue;
237      case 'totalAmount':
238        return d3.format('0f')(dataValue);
239      case 'transactionCount':
240        return d3.format('d')(dataValue);
241      default:
242        return d3.format('')(dataValue);
243    }
```

```
244   }
245
246   private calculateSizeTotals(chartData: DataSummaryPackage): any[] {
247     const results = [];
248     chartData.data.forEach(x =>  {
249       let totalAmount = 0, transactionCount = 0;
250       x.timeSeries.forEach(y => {
251         totalAmount += y.totalAmount;
252         transactionCount += y.transactionCount;
253       });
254       results.push({'key': x.key, 'totalAmount': totalAmount, '
    transactionCount': transactionCount});
255     });
256     return results;
257   }
258
259   private calculateX(d: TimeSeriesData): number {
260     let x;
261     switch (this.xField) {
262       case 'year':
263       case 'monthNum':
264         x = d.timeValue;
265         break;
266       case 'totalAmount':
267         x = d.totalAmount;
268         break;
269       case 'transactionCount':
270         x = d.transactionCount;
271         break;
272       default:
273         throw new Error(`${this.xField} is not an allowable field for
    the x-axis`);
274     }
275     return x;
276   }
277
278   private calculateY(d: TimeSeriesData): number {
279     let y;
280     switch (this.yField) {
281       case 'year':
282       case 'monthNum':
283         y = this.formatTimeTicks(d.timeValue, d.timeType);
284         break;
285       case 'totalAmount':
286         y = d.totalAmount;
287         break;
288       case 'transactionCount':
289         y = d.transactionCount;
290         break;
291       default:
292         throw new Error(`${this.yField} is not an allowable field for
    the y-axis`);
```

```
293     }
294     return y;
295   }
296
297   private calculateSize(sizeTotals: any[], key: string,  d:
        TimeSeriesData): number {
298     let size;
299     const timeSeriesTotal = sizeTotals.find(x => x.key === key);
300     switch (this.sizeField) {
301       case 'totalAmount':
302         size = d.totalAmount / timeSeriesTotal.totalAmount;
303         break;
304       case 'transactionCount':
305         size = d.transactionCount / timeSeriesTotal.transactionCount;
306         break;
307       default:
308         throw new Error(`${this.sizeField} is not an allowable size
        field.`);
309     }
310     return size;
311   }
312 }
313 import { BaseChart } from '../base.chart';
314 import { DataSummaryPackage } from '../../../../models/analysis-result
      ';
315
316
317 export class NVD3PieChart extends BaseChart {
318   allowableXFields: string[] = ['time', 'key'];
319   allowableYFields: string[] = ['totalAmount', 'transactionCount'];
320   allowableSizeFields: string[] = [];
321   allowableGroupFields: string[] = [];
322
323   constructor(xField: string, yField: string, height: number, width:
        number) {
324     super(xField, yField, height, width);
325   }
326
327   formatData(chartData: DataSummaryPackage) {
328     const baseData = [];
329     switch (this.xField) {
330       case 'category':
331       case 'word':
332         chartData.data.forEach(x => {
333           let valueAmount = 0.0;
334           x.timeSeries.forEach(y => {
335             switch (this.yField) {
336               case 'totalAmount':
337                 valueAmount += y.totalAmount;
338                 break;
339               case 'transactionCount':
340                 valueAmount += y.transactionCount;
```

```
341                 break;
342               default:
343                 throw new Error(`${this.yField} is not an allowable
    value.`);
344             }
345           });
346           baseData.push({'key': x.key, 'value': valueAmount});
347         });
348         break;
349       case 'year':
350       case 'monthNum':
351         chartData.data.forEach(x => {
352           x.timeSeries.forEach( y => {
353             const timeElement = baseData.find(t => t.key === y.
    timeValue);
354             let valueAmount = 0;
355             switch (this.yField) {
356               case 'totalAmount':
357                 valueAmount = y.totalAmount;
358                 break;
359               case 'transactionCount':
360                 valueAmount = y.transactionCount;
361                 break;
362               default:
363                 throw new Error(`${this.yField} is not an allowable
    value`);
364             }
365             if (timeElement) {
366               timeElement.value += valueAmount;
367             } else {
368               baseData.push({'key': y.timeValue, 'value': valueAmount})
    ;
369             }
370           });
371         });
372         break;
373     }
374     return baseData;
375   }
376   createOptions(): any {
377     const self = this;
378     const options = {
379       chart: {
380         type: 'pieChart',
381         height: this.height,
382         width: this.width,
383         callback: function(chart) {
384           chart.pie.dispatch.on('elementClick', function(e) {
385             self._onElementClick.dispatch(self, e);
386           });
387         },
```

```
388         x: function(d) { return self.formatAxisData(self.xField, d.key)
    ; },
389         y: function(d) { return self.formatAxisData(self.yField, d.
    value); },
390         showLabels: false,
391         duration: 500,
392         labelThreshold: 0.01,
393         labelSunbeamLayout: true,
394         legend: {
395           margin: {
396             top: 5,
397             right: 35,
398             bottom: 5,
399             left: 0
400           }
401         }
402       }
403     };
404     return options;
405   }
406 }
407 import { BaseChart } from '../base.chart';
408 import {DataSummaryPackage, TimeSeriesData} from '../../../../models/
    analysis-result';
409
410 export class NVD3MultiBarChart extends BaseChart {
411   allowableXFields: string[] = ['time', 'key'];
412   allowableYFields: string[] = ['totalAmount', 'transactionCount'];
413   allowableSizeFields: string[] = [];
414   allowableGroupFields: string[] = ['time', 'key'];
415
416   constructor(xField: string, yField: string, height: number, width:
      number, groupField: string) {
417     super(xField, yField, height, width, groupField);
418 }
419
420   formatData(chartData: DataSummaryPackage): any {
421     switch (this.xField) {
422       case 'category':
423       case 'word':
424         return this.formatDataForKey(chartData);
425       case 'year':
426       case 'monthNum':
427         return this.formatDataForTime(chartData);
428       default:
429         throw new Error(`${this.xField} is not an allowable group
    option.`);
430     }
431   }
432
433   private formatDataForTime(chartData: DataSummaryPackage): any {
434     const results = [];
```

```
435     chartData.data.forEach(x => {
436       const timeData = [];
437       x.timeSeries.forEach(y => {
438         let xValue;
439         switch (this.xField) {
440           case 'year':
441           case 'monthNum':
442             xValue = this.formatTimeTicks(y.timeValue, y.timeType);
443             break;
444           default:
445             throw new Error(`${this.xField} is not an allowable field
    for the x axis`);
446         }
447         timeData.push({'x': xValue, 'y': this.getYDataPoint(y)});
448       });
449       results.push({'key': x.key, 'values': timeData});
450     });
451     return results;
452   }
453
454   private formatDataForKey(chartData: DataSummaryPackage): any {
455     const results = [];
456     chartData.data.forEach(x => {
457       x.timeSeries.forEach( y => {
458         const timeObj = results.find(z => z.key === y.timeValue);
459         if (!timeObj) {
460           results.push({'key': y.timeValue, 'values': [{'x': x.key, 'y
    ': this.getYDataPoint(y)}]});
461         } else {
462           const catObj = timeObj.values.find(c => c.x === x.key);
463           if (!catObj) {
464             timeObj.values.push({'x': x.key, 'y': this.getYDataPoint(y)
    });
465           } else {
466             catObj.y += this.getYDataPoint(y);
467           }
468         }
469       });
470     });
471     return results;
472   }
473
474   private getYDataPoint(data: TimeSeriesData): number {
475     let yValue;
476     switch (this.yField) {
477       case 'totalAmount':
478         yValue = data.totalAmount;
479         break;
480       case 'transactionCount':
481         yValue = data.transactionCount;
482         break;
483       default:
```

```
484         throw new Error(`${this.yField} is not an allowable field for
      the y axis`);
485       }
486       return yValue;
487     }
488
489     createOptions(): any {
490       const self = this;
491       const options = {
492         chart: {
493           type: 'multiBarChart',
494           height: this.height,
495           width: this.width,
496           showLegend: false,
497           margin: {
498             top: 50,
499             right: 50,
500             bottom: 50,
501             left: 60
502           },
503           callback: function(chart) {
504             chart.multibar.dispatch.on('elementClick', function(e) {
505               self.selectedData = e;
506               self._onElementClick.dispatch(self, e);
507             });
508           },
509           x: function(d) { return d.x; },
510           y: function(d) { return d.y; },
511           clipEdge: true,
512           duration: 500,
513           stacked: true,
514           xAxis: {
515             axisLabel: this.getAxisLabel(this.xField),
516             axisLabelDistance: 10,
517             tickFormat: function (d) {
518               return self.formatAxisData(self.xField, d);
519             },
520             showMaxMin: false,
521             staggerLabels: true
522           },
523           yAxis: {
524             axisLabel: this.getAxisLabel(this.yField),
525             tickFormat: function(d) {
526               return self.formatAxisData(self.yField, d);
527             }
528           }
529         }
530       };
531       return options;
532     }
533 }
534 import {BaseChart} from '../base.chart';
```

```
import {DataSummaryPackage} from '../../../../models/analysis-result';

export class NVD3LineChart extends BaseChart {
  allowableXFields: string[] = ['time'];
  allowableYFields: string[] = ['totalAmount', 'transactionCount'];
  allowableSizeFields: string[] = [];
  allowableGroupFields: string[] = [];

  constructor(xField: string, yField: string, height: number, width:
    number) {
    super(xField, yField, height, width);
  }

  formatData(chartData: DataSummaryPackage): any {
    const results = [];
    chartData.data.forEach(x => {
      const timeData = [];
      x.timeSeries.forEach(y => {
        let xValue, yValue;
        switch (this.xField) {
          case 'year':
          case 'monthNum':
            xValue = this.formatTimeTicks(y.timeValue, y.timeType);
            break;
          default:
            throw new Error(`${this.xField} is not an allowable x axis
    value`);
        }
        switch (this.yField) {
          case 'totalAmount':
            yValue = y.totalAmount;
            break;
          case 'transactionCount':
            yValue = y.transactionCount;
            break;
          default:
            throw new Error(`${this.yField} is not an allowable y axis
    value`);
        }
        timeData.push({'x': xValue, 'y': yValue});
      });
      results.push({'key': x.key, 'values': timeData});
    });
    return results;
  }

  createOptions(): any {
    const self = this;
    const options = {
      chart: {
        type: 'lineChart',
        height: this.height,
```

```
        width: this.width,
        showLegend: false,
        margin: {
          top: 20,
          right: 20,
          bottom: 40,
          left: 55
        },
        x: function(d) { return d.x; },
        y: function(d) { return d.y; },
        useInteractiveGuideline: false,
        callback: function(chart) {
          chart.lines.dispatch.on('elementClick', function(e) {
            self._onElementClick.dispatch(self, e);
          });
        },
        xAxis: {
          axisLabel: this.getAxisLabel(this.xField),
          tickFormat: function (d) {
            return self.formatAxisData(self.xField, d);
          }
        },
        yAxis: {
          axisLabel: this.getAxisLabel(this.yField),
          tickFormat: function(d) {
            return self.formatAxisData(self.yField, d);
          },
          axisLabelDistance: -10
        }
      }
    };
    return options;
  }
}
import {BaseChart} from '../base.chart';
import {DataSummaryPackage} from '../../../../models/analysis-result';

export class NVD3DiscreteBarChart extends BaseChart {
  allowableXFields: string[] = ['time', 'key'];
  allowableYFields: string[] = ['totalAmount', 'transactionCount'];
  allowableSizeFields: string[] = [];
  allowableGroupFields: string[] = [];

  constructor(xField: string, yField: string, height: number, width:
    number) {
    super(xField, yField, height, width);
  }

  formatData(chartData: DataSummaryPackage) {
    const baseData = [];
    switch (this.xField) {
      case 'category':
```

```
635       case 'word':
636         chartData.data.forEach(x => {
637           let valueAmount = 0.0;
638           x.timeSeries.forEach(y => {
639             switch (this.yField) {
640               case 'totalAmount':
641                 valueAmount += y.totalAmount;
642                 break;
643               case 'transactionCount':
644                 valueAmount += y.transactionCount;
645                 break;
646               default:
647                 throw new Error(`${this.yField} is not an allowable
    value.`);
648             }
649           });
650           baseData.push({'key': x.key, 'value': valueAmount});
651       });
652       break;
653     case 'year':
654     case 'monthNum':
655       chartData.data.forEach(x => {
656         x.timeSeries.forEach( y => {
657           const timeElement = baseData.find(t => t.key === y.
    timeValue);
658           let valueAmount = 0;
659           switch (this.yField) {
660             case 'totalAmount':
661               valueAmount = y.totalAmount;
662               break;
663             case 'transactionCount':
664               valueAmount = y.transactionCount;
665               break;
666             default:
667               throw new Error(`${this.yField} is not an allowable
    value`);
668           }
669           if (timeElement) {
670             timeElement.value += valueAmount;
671           } else {
672             baseData.push({'key': y.timeValue, 'value': valueAmount})
    ;
673           }
674         });
675       });
676       break;
677   }
678   const dataLabel = `${this.xField.charAt(0).toUpperCase() + this.
    xField.slice(1)} Breakdown`;
679   const results = [];
680   results.push({'key': dataLabel, 'values': baseData});
681   return results;
```

```
682   }
683
684   createOptions(): any {
685     const self = this;
686     const options = {
687       chart: {
688         type: 'discreteBarChart',
689         height: this.height,
690         width: this.width,
691         margin: {
692           top: 20,
693           right: 20,
694           bottom: 50,
695           left: 55
696         },
697         callback: function(chart) {
698           chart.discretebar.dispatch.on('elementClick', function(e) {
699             self._onElementClick.dispatch(self, e);
700           });
701         },
702         x: function(d) { return d.key; },
703         y: function(d) { return d.value; },
704         showValues: true,
705         valueFormat: function(d) {
706           return self.formatAxisData(self.yField, d);
707         },
708         duration: 500,
709         xAxis: {
710           axisLabel: this.getAxisLabel(this.xField),
711           tickFormat: function(d) {
712             return self.formatAxisData(self.xField, d);
713           }
714         },
715         yAxis: {
716           axisLabel: this.getAxisLabel(this.yField),
717           axisLabelDistance: -10,
718           tickFormat: function(d) {
719             return self.formatAxisData(self.yField, d);
720           }
721         }
722       }
723     };
724     return options;
725   }
726 }
727 import { LabelValue } from '../../models/visualisation.models';
728 import { NVD3StackedAreaChart } from './types/nvd3/stacked-area.chart';
729 import { NVD3DiscreteBarChart } from './types/nvd3/discrete-bar.chart';
730 import { NVD3PieChart } from './types/nvd3/pie.chart';
731 import { NVD3MultiBarChart } from './types/nvd3/multi-bar.chart';
732 import {NVD3LineChart} from './types/nvd3/line.chart';
733 import { NVD3ScatterChart } from './types/nvd3/scatter.chart';
```

```
734 import {CJSDiscreteBarChart} from './types/cjs/discrete-bar.chart';
735 import {HCJSDiscreteBar} from './types/hcjs/discrete-bar.chart';
736
737 export class ChartFactory {
738   static getAllowableCharts(): LabelValue[];
739   static getAllowableCharts(): LabelValue[] {
740     const results: LabelValue[] = [];
741     results.push(new LabelValue('Stacked Area Chart', 'stackedArea', {
      library: 'nvd3'}));
742     results.push(new LabelValue('Discrete Bar Chart', 'discreteBar', {
      library: 'nvd3'}));
743     results.push(new LabelValue('Pie Chart', 'pie', {library: 'nvd3'}))
      ;
744     results.push(new LabelValue('MultiBar Chart', 'multiBar', {library:
       'nvd3'}));
745     results.push(new LabelValue('Line Chart', 'line', {library: 'nvd3
      '}));
746     // results.push(new LabelValue('Discrete Bar Chart (Highchart)', '
      hcjsDiscreteBar', {library: 'hcjs'}));
747     // results.push(new LabelValue('Discrete Bar Chart (CJS)', '
      cjsDiscreteBar', {library: 'cjs'}));
748     // results.push(new LabelValue('Scatter Chart', 'scatter'));
749     return results.sort((a, b): number => {
750       if (a.value < b.value) { return -1; }
751       if (a.value > b.value) { return 1; }
752       return 0;
753     });
754   }
755
756   static createChart(type: Object);
757   static createChart(type: 'stackedArea'): NVD3StackedAreaChart;
758   static createChart(type: 'discreteBar'): NVD3DiscreteBarChart;
759   static createChart(type: 'pie'): NVD3PieChart;
760   static createChart(type: 'multiBar'): NVD3MultiBarChart;
761   static createChart(type: 'line'): NVD3LineChart;
762   static createChart(type: 'scatter'): NVD3ScatterChart;
763   static createChart(type: 'cjsDiscreteBar'): CJSDiscreteBarChart;
764   static createChart(type: 'hcjsDiscreteBar'): HCJSDiscreteBar;
765
766   static createChart(chartOptions): NVD3StackedAreaChart |
      NVD3DiscreteBarChart | NVD3PieChart | NVD3MultiBarChart |
      NVD3LineChart |
767                                     NVD3ScatterChart |
      CJSDiscreteBarChart | HCJSDiscreteBar {
768     switch (chartOptions.type) {
769       case 'stackedArea':
770         return new NVD3StackedAreaChart(chartOptions.xField,
      chartOptions.yField, chartOptions.height, chartOptions.width);
771       case 'discreteBar':
772         return new NVD3DiscreteBarChart(chartOptions.xField,
      chartOptions.yField, chartOptions.height, chartOptions.width);
773       case 'pie':
```

```
774         return new NVD3PieChart(chartOptions.xField, chartOptions.
    yField, chartOptions.height, chartOptions.width);
775       case 'multiBar':
776         return new NVD3MultiBarChart(chartOptions.xField, chartOptions.
    yField, chartOptions.height, chartOptions.width, 'key');
777       case 'line':
778         return new NVD3LineChart(chartOptions.xField, chartOptions.
    yField, chartOptions.height, chartOptions.width);
779       case 'scatter':
780         return new NVD3ScatterChart(chartOptions.xField, chartOptions.
    yField, chartOptions.height, chartOptions.width, chartOptions.
    sizeField);
781       case 'cjsDiscreteBar':
782         return new CJSDiscreteBarChart(chartOptions.xField,
    chartOptions.yField, chartOptions.height, chartOptions.width);
783       case 'hcjsDiscreteBar':
784         return new HCJSDiscreteBar(chartOptions.xField, chartOptions.
    yField, chartOptions.height, chartOptions.width);
785       default:
786         throw new Error(`${chartOptions.type} is not a valid chart
    option for the ${chartOptions.library} library`);
787     }
788   }
789 }
```

Appendix D

Case Study Design

The following pages detail the design of the case study articulated in Chapter 5. This appendix contains the pre- and post-test questionnaires, the information sheet provided to all participants, a sample consent form signed that was distributed to (and signed by) all participants, and the letter of ethical approval from Maynooth University's Research Ethics Department. As part of the analysis of the case study involved a qualitative analysis of some of the post-test questionnaire responses, the final sections of this appendix detail the code system used for the qualitative analysis as well as the transcripts and coded segments for the three pertinent aspects of the qualitative analysis of the post-test questionnaires: responses related to the keyword search, responses related to the visualisation search, and final thoughts on the system as a whole.

D.1 Pre-test Questionnaire

Alcalá Workshop Registration

Thank you for your interest in the Alcalá Case Study! Please take a moment to fill up this registration form. PLEASE NOTE: Registration is mandatory so we have numbers in advance. Same day registration will not be possible. Also, you must be at least 18 years of age to participate.

* Required

1. **Email address** *

Demographic Information

2. **First Name** *

3. **Last Name** *

4. **Email Address** *

5. **Gender** *
   Mark only one oval.
   - ◯ Male
   - ◯ Female
   - ◯ Prefer not to say
   - ◯ Other: _____

6. **Age** *
   Mark only one oval.
   - ◯ 18-21
   - ◯ 22-30
   - ◯ 31-40
   - ◯ 41-50
   - ◯ 50+
   - ◯ Prefer not to say

7. **What is your highest level of education?**
   Mark only one oval.

   - ◯ Bachelors
   - ◯ Masters
   - ◯ PhD
   - ◯ Prefer not to say
   - ◯ Other: _____

8. **What is your area of scholarship or expertise?**

   _____

Study Registration

As part of the Alcalá workshop, a case study will be conducted showcasing the new software tool built to around the Alcalá archive. This case study will take approximately 2 hours and will give you the opportunity to utilise the new software and provide feedback on its development and direction. If you agree to participate, it is imperative you are able to attend both hours of the study.

9. **Do you wish to participate in the case study portion of the workshop?**
   Clicking yes will lead you to a short survey that will be used to collect information about your technology usage. This survey should not take you more than 5 minutes to complete.
   Mark only one oval.

   - ◯ Yes *Skip to question 9.*
   - ◯ No *Skip to question 32.*

Technology Ownership

Please tell us a little about the types of technology you own.

10. **Do you own a laptop or personal computer?** *
    Mark only one oval.

    - ◯ Yes
    - ◯ No

11. **Do you own a smart phone, tablet, or other mobile device?** *
    Mark only one oval.

    - ◯ Yes
    - ◯ No

Technology Use

For the following questions, please indicate your answer on a scale of 1 to 10 with the following indicators:

- 1 = Never
- 2 = Once a month
- 3 = Several times a month
- 4 = Once a week
- 5 = Several times a week
- 6 = Once a day
- 7 = Several times a day
- 8 = Once an hour

- 9 = Several times an hour
- 10 = All the time

12. **Send and receive / read emails on a laptop or personal computer** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

13. **Send and receive / read emails on a mobile phone or tablet** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

14. **Send and receive text messages on a mobile phone or tablet** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

15. **Get directions or use GPS on a mobile phone** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

16. **Browse the internet on laptop / personal computer** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

17. **Browse the internet on a mobile phone or tablet** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

18. **Search the internet (on any device) for news** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

19. **Search for internet (on any device) for videos** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

20. **Search the internet (on any device) for images / photos** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

21. **Use apps for any purpose (other than games) on a mobile phone or tablet** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

22. **Utilise any form of social media** *
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	All the time

Technology Importance

For the following, please indicate on a scale of 1 to 5 whether or not you agree with the statement:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither agree nor disagree
- 4 = Agree
- 5 = Strongly Agree

23. **I think it is important to be able to find information on the internet** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

24. **I find it easy to locate information when I search the internet.** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

25. **I get anxious when I have to locate information on the internet** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

26. **I find it easy to adopt new technology or learn a new app** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

27. **I get anxious when I have to utilise technology I am unfamiliar with** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

28. **Technology makes my life easier** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

29. **I feel I can accomplish more when using technology** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

30. **I prefer to work on numerous projects, rather than focusing on one at a time** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

31. **When completing assignments, I prefer to switch between back and forth** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

32. **When I have to complete a task, I prefer to break it into steps and switch between the steps** *
Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

D.2 Post-test Questionnaire

Alcalá Account Books Post-test Assessment

Please fill out the following questionnaire as it relates to your overall experience with the particular search model you leveraged in today's experiment.

* Required

Demographic Information

1. **First Name** *

2. **Last Name** *

3. **User Login** *

4. **Email Address** *
   Please use the same email address you listed in your registration.

General Impressions

Please answer the following questions as it relates to your experience with the version of the search function you used today. Each of the following questions contains a pair of contrasting attributes, with a scale of 1-7 separating them. Please select the option that runs closest to your impression as it relates to that attribute.

When answering these questions, please answer spontaneously. Don't think too long about your decision as we are interested in your initial reactions to the site. Even if the attributes don't quite fit with your impression, try to select an answer for every item.

Remember, there is no right or wrong answer! Negative feedback is just as valuable as positive feedback.

5. **Annoying vs Enjoyable** *
   Mark only one oval.

	1	2	3	4	5	6	7	
annoying	⬭	⬭	⬭	⬭	⬭	⬭	⬭	enjoyable

6. **Not Understandable vs Understandable** *
Mark only one oval.

	1	2	3	4	5	6	7	
not understandable	◯	◯	◯	◯	◯	◯	◯	understandable

7. **Creative vs Dull** *
Mark only one oval.

	1	2	3	4	5	6	7	
creative	◯	◯	◯	◯	◯	◯	◯	dull

8. **Easy to Learn vs Difficult to Learn** *
Mark only one oval.

	1	2	3	4	5	6	7	
easy to learn	◯	◯	◯	◯	◯	◯	◯	difficult to learn

9. **Valuable vs Inferior** *
Mark only one oval.

	1	2	3	4	5	6	7	
valuable	◯	◯	◯	◯	◯	◯	◯	inferior

10. **Boring vs Exciting** *
Mark only one oval.

	1	2	3	4	5	6	7	
boring	◯	◯	◯	◯	◯	◯	◯	exciting

11. **Not Interesting vs Interesting** *
Mark only one oval.

	1	2	3	4	5	6	7	
not interesting	◯	◯	◯	◯	◯	◯	◯	interesting

12. **Unpredictable vs Predictable** *
Mark only one oval.

	1	2	3	4	5	6	7	
unpredictable	◯	◯	◯	◯	◯	◯	◯	predictable

13. **Fast vs Slow** *
Mark only one oval.

	1	2	3	4	5	6	7	
fast	◯	◯	◯	◯	◯	◯	◯	flow

14. **Inventive vs Conventional** *
Mark only one oval.

	1	2	3	4	5	6	7	
inventive	◯	◯	◯	◯	◯	◯	◯	conventional

15. **Obstructive vs Supportive** *
Mark only one oval.

	1	2	3	4	5	6	7	
obstructive	◯	◯	◯	◯	◯	◯	◯	supportive

16. **Good vs Bad** *
Mark only one oval.

	1	2	3	4	5	6	7	
good	◯	◯	◯	◯	◯	◯	◯	bad

17. **Complicated vs Easy** *
Mark only one oval.

	1	2	3	4	5	6	7	
complicated	◯	◯	◯	◯	◯	◯	◯	easy

18. **Unlikable vs Pleasing** *
Mark only one oval.

	1	2	3	4	5	6	7	
unlikable	◯	◯	◯	◯	◯	◯	◯	pleasing

19. **Usual vs Leading Edge** *
Mark only one oval.

	1	2	3	4	5	6	7	
usual	◯	◯	◯	◯	◯	◯	◯	leading edge

20. **Unpleasant vs Pleasant** *
Mark only one oval.

	1	2	3	4	5	6	7	
unpleasant	◯	◯	◯	◯	◯	◯	◯	pleasant

21. **Secure vs Not Secure** *
Mark only one oval.

	1	2	3	4	5	6	7	
secure	◯	◯	◯	◯	◯	◯	◯	not secure

22. **Motivating vs Demotivating** *
Mark only one oval.

	1	2	3	4	5	6	7	
motivating	◯	◯	◯	◯	◯	◯	◯	demotivating

23. **Meets Expectations vs Does Not Meet Expectations** *
Mark only one oval.

	1	2	3	4	5	6	7	
meets expectations	◯	◯	◯	◯	◯	◯	◯	doesn't meet expectations

24. **Inefficient vs Efficient** *
Mark only one oval.

	1	2	3	4	5	6	7	
inefficient	◯	◯	◯	◯	◯	◯	◯	efficient

25. **Clear vs Confusing** *
Mark only one oval.

	1	2	3	4	5	6	7	
clear	◯	◯	◯	◯	◯	◯	◯	confusing

26. **Impractical vs Practical** *
Mark only one oval.

	1	2	3	4	5	6	7	
impractical	◯	◯	◯	◯	◯	◯	◯	practical

27. **Organised vs Cluttered** *
Mark only one oval.

	1	2	3	4	5	6	7	
organised	◯	◯	◯	◯	◯	◯	◯	cluttered

28. **Attractive vs Unattractive** *
Mark only one oval.

	1	2	3	4	5	6	7	
attractive	◯	◯	◯	◯	◯	◯	◯	unattractive

29. **Friendly vs Unfriendly** *
Mark only one oval.

	1	2	3	4	5	6	7	
friendly	◯	◯	◯	◯	◯	◯	◯	unfriendly

30. **Conservative vs Innovative** *
Mark only one oval.

	1	2	3	4	5	6	7	
conservative	◯	◯	◯	◯	◯	◯	◯	innovative

Feedback

For each of the following questions, please provide feedback in your own words. Remember, this is all your opinion so there is no right or wrong answer! All opinions are valid and beneficial!

31. **Which version of search did you use today?** *
Mark only one oval.

- ◯ Keyword *Skip to question 32.*
- ◯ Visualisation *Skip to question 38.*

Keyword Search

Provide answers as it relates to the keyword search

32. **Describe your initial reaction to the Keyword Search.** *

33. **What did you like most about the keyword search?** *

34. **What did you like least about the keyword search you used today?** *

35. **Was there anything about the keyword search that made answering particular questions easy? If so, what?** *

36. **What was something interesting you learned using the keyword search?** *

37. **Have you also completed the visualisation search exercise?** *
Only answer yes if this is the second hour of the experiment and you have had an opportunity to do both the keyword and the visualisation exercises.
Mark only one oval.

- ◯ Yes *Skip to question 44.*
- ◯ No *Skip to question 49.*

Visualisation Search

Provide answers to the following questions as it relates to the visualisation search.

38. **Describe your initial reaction to the visualisation search.** *

39. **What did you like most about the visualisation search?** *

40. **What did you like least about the visualisation search?** *

41. **Thinking about the visualisations you constructed, was there anything interesting that immediately jumped out at you (such as connections between data points that might not have been obvious otherwise)? If so, describe it.** *

42. **What was something interesting you learned using the visualisation search?** *

43. **Have you also completed the keyword search exercise?** *
Only answer yes if this is the second hour of the experiment and you have had an opportunity to do both the keyword and the visualisation exercises.
Mark only one oval.

- ◯ Yes *Skip to question 44.*
- ◯ No *Skip to question 49.*

Comparison

Compare your experience between the keyword search and the visualisation search.

44. **Did you find the keyword search helpful in order to locate information?** *
Mark only one oval.

	1	2	3	4	5	
not helpful at all	◯	◯	◯	◯	◯	very helpful

45. **Did you find the visualisation search helpful in order to locate information?** *
Mark only one oval.

	1	2	3	4	5	
not helpful at all	◯	◯	◯	◯	◯	very helpful

46. **If you were doing research for a project / journal article / etc, would you prefer the keyword search or the visualisation search?** *
Mark only one oval.

- ◯ Keyword
- ◯ Visualisation

47. **Describe why you picked the above answer.** *

48. **Is there a scenario you can think of where one type of search mechanism is better than the other? If so, describe.** *

Final Thoughts

49. **What did you learn while using the system?** *

50. **Is there anything about the system you would change that would make using it a better experience for you?** *

D.3 Information Sheet - Workshop

Information Sheet

Project Title: Escaping the Book: Investigating Engagement with Humanities-based Visualisations to Promote Knowledge Acquisition in Digital Research Environments
Researcher: Shane A. McGarry, PhD Candidate, Computer Science
Department: Computer Science
Email Contact: shane.mcgarry.2015@mumail.ie

About this Project

This project is part of the Researcher's PhD thesis, which seeks to shift the discourse around the design of and interaction with search mechanisms inherent in Digital Research Environments (DRE). Many DREs are attempts to recreate analogue, historical manuscripts within a digital setting (such as the Alcalá Account Books which is the focus on this research project). As such, the DRE has remained married to many of its analogue metaphors, mimicking a print object in a digital realm and focusing the user experience on reading without providing alternative means of engagement with the text. Research has shown that attempting to read in digital environments creates unnecessary strain on cognitive processes and decreases comprehension. However, due to the nature of the digital object, alternative modes of engagement can be presented to the user to facilitate knowledge acquisition.

As technology evolves and becomes more ubiquitous, younger generations are reading differently. These different methods of reading are not reliant on traditional close reading of texts, and the ways in which we read and consume information are changing. Thus, the boundary of the book metaphor in a digital environment has become obsolete. By exploring new, dynamic ways of exploring text-based content digitally, my work seeks to validate alternative methods of textual engagement and support new research into modes of experience design.

The purpose of this research project is to ascertain how users can leverage different search mechanisms to construct new knowledge through a series of proscribed tasks. Participants will be asked to engage with these tasks through two means: through a keyword search (which requires the participant to enter keywords and comb through a list of results, searching for the specific data required to complete the aforementioned tasks), and through the use of data visualisations (which requires participants to analyse data visualisations, and encourages them to construct their own visualisations to explore the and engage with the data).

Requirements of Participants

Participants of the study will be given the opportunity to participate in a case study. All participants will be required to sign a consent form. The requirements for the case study are listed below.

Requirements for Participation in Case Study

The following will be required of participants in focus groups:

- All participants consent to have their activity on the Alcalá Account Books website stored and analysed (including but not limited to search keywords, result lists, log in / log out times, and data visualisations constructed).
- All participants will be required to complete a pre-test questionnaire to ascertain their level of engagement with technology.
- All participants will agree to complete a post-test questionnaire to ascertain their individual experience with the software.
- All participants will agree to attend two (2) one hour sessions: one which will be used to explore the keyword search task and the other which will be used to explore the visualisation search task.
- All participants will agree to attend a session in person at Maynooth University Campus in a computer lab (Eolas Hardware Lab: Room 011). No virtual attendance will be possible.

Project Data

All data for this project will be used to develop prototypes which will form the basis of the analysis of the Researcher's PhD thesis. While certain demographic data will be used to highlight differences in gender, age, or educational background, such data will only be used in its aggregate sense and will not be used to personally identify individual participants. Each participant will be assigned an identifier in order to promote anonymity and personally identifiable information will not be disclosed within the thesis itself.

All data will be analysed and then stored on a secure, encrypted drive. Per University policy, this data will be maintained in a secure archive for a minimum of 10 years in accordance with Maynooth University's Research Integrity Policy at which point, it will be destroyed.

Confidentiality

Every effort will be made to keep each participants information confidential. However, it must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.

D.4 Consent Form - Workshop

Consent Form

Project Title:	Escaping the Book: Investigating Engagement with Humanities-based Visualisations to Promote Knowledge Acquisition in Digital Research Environments
Researcher:	Shane A. McGarry, PhD Candidate, Computer Science
Department:	Computer Science
Email Contact:	shane.mcgarry.2015@mumail.ie
Supervisor:	Dr John Keating, Senior Lecturer, Computer Science
Email Contact:	john.keating@mu.ie

All participants in this research project must sign this consent form in order to participate. As outlined in the Information Sheet, every effort will be made to keep information private and confidential. However, it must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances, the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.

If at any time, up to the point of anonymisation, any participant decides to withdraw his or her consent for this project, he or she may do so by contacting the researcher. As such, any participant who opts to withdraw will have all data collected removed from the system of record and discarded.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at research.ethics@nuim.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

I, _____, understand and consent to all information disclosed in the Information Sheet and consent to participation in the case study. I acknowledge I have read the Information Sheet and understand the nature of this research project. I consent to be contacted by the above stated researcher or supervisor regarding my participation in this project and that all data will be anonymised at the end of the Fall 2018 term. I also understand that prior to anonymisation, all contact information will be held privately and confidentiality in a secure location by the researcher and will not be shared for any reason. Furthermore, I acknowledge I have read and have received a copy of this consent sheet for my own personal records.

☐ By checking here, I agree to participate in one-on-one interviews as outlined in the Information Sheet and consent to an audio recording of the interview.

Print Name: _____

Signed: _____

Date: _____

D.5 Ethical Approval - Workshop

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

20 May 2016

Shane A. McGarry
An Foras Feasa
Maynooth University

RE: Application for Ethical Approval for a project entitled: Escaping the Book Metaphor: Examining User Interface Metaphors in Digital Scholarly Editions

Dear Shane,

The Ethics Committee evaluated the above project and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30 September 2017.

Kind Regards,

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

C.c. Professor Susan Schreibman, An Foras Feasa

Reference Number
SRESC-2016-035

D.6 Information Sheet - Seminar

Information Sheet

Project Title:	Escaping the Book: Investigating Engagement with Humanities-based Visualisations to Promote Knowledge Acquisition in Digital Research Environments
Researcher:	Shane A. McGarry, PhD Candidate, Computer Science
Department:	Computer Science
Email Contact:	shane.mcgarry.2015@mumail.ie
Pre-test Info:	https://goo.gl/S2arDb

About this Project

This project is part of the Researcher's PhD thesis, the goal of which is to change how we think about searching in "Digital Research Environments" (DRE). Many DREs are attempts to recreate historical manuscripts on the computer (such as the Alcalá Account Books which is the focus on this research project). But as a result, the DRE tries to look and act like a regular book, focusing the user's attention on reading and ignoring the possibilities that computers can provide. Research has shown that attempting to read in digital environments actually creates a lot of difficulty for our brains, making reading digitally more difficult and, by extension, decreasing our ability to remember what we've read. Computers, however, can provide us with powerful tools, such as browsing via links on the page, and various formatting options, that would be impossible to implement in a printed book.

As technology evolves and becomes more commonplace, younger generations are reading differently. These different methods of reading are not the older, standard methods of reading (where you read from the top of the page to the bottom of the page and have to build on each page in your mind as you read it—a process known as "close reading"), and the ways in which we read and consume information are changing. As a result, attempting to mimic a printed book no longer makes much sense. By exploring new, dynamic ways of exploring these types of text-based manuscripts digitally, my work seeks to verify that there are other, more appropriate ways of engaging with these types of objects.

The purpose of this research project is to understand how users can use different ways of searching as part of the learning process by following a set of required tasks. Participants will be asked to complete these tasks in two ways: through a keyword search (which requires the participant to enter keywords and comb through a list of results, searching for the specific information that answers their questions), and through the use of data visualisations (which requires participants to analyse images, such as charts and graphs, and encourages them to construct their own visualisations to explore the manuscript).

Requirements of Participants

Participants of the study will be given the opportunity to participate in a case study. All participants will be required to sign a consent form. The requirements for the case study are listed below.

Requirements for Participation in Case Study

The following will be required of participants in focus groups:

- All participants consent to have their activity on the Alcalá Account Books website stored and analysed (including but not limited to search keywords, result lists, log in / log out times, and data visualisations constructed).
- All participants will be required to complete a pre-test questionnaire to ascertain their level of engagement with technology.
- All participants will agree to complete a post-test questionnaire to ascertain their individual experience with the software.
- All participants agree to submit a task worksheet that details the results of their efforts in obtaining the requisite answers to the aforementioned task(s).
- All participants will agree to attend two (2) tutorials: one which will be used to explore the keyword search task and the other which will be used to explore the visualisation search task.
- All participants will agree to attend a session in person at Maynooth University Campus in a computer lab (Eolas Hardware Lab: Room 011). No virtual attendance will be possible.

Project Data

All data for this project will be used to develop prototypes which will form the basis of the analysis of the Researcher's PhD thesis. While certain demographic data will be used to highlight differences in gender, age, or educational background, such data will only be used in its aggregate sense and will not be used to personally identify individual participants. Each participant will be assigned an identifier in order to promote anonymity and personally identifiable information will not be disclosed within the thesis itself.

All data will be analysed and then stored on a secure, encrypted drive. Per University policy, this data will be maintained in a secure archive for a minimum of 10 years in accordance with Maynooth University's Research Integrity Policy at which point, it will be destroyed.

Confidentiality

Every effort will be made to keep each participants information confidential. However, it must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.

D.7 Consent Form - Seminar

Consent Form

Project Title:	Escaping the Book: Investigating Engagement with Humanities-based Visualisations to Promote Knowledge Acquisition in Digital Research Environments
Researcher:	Shane A. McGarry, PhD Candidate, Computer Science
Department:	Computer Science
Email Contact:	shane.mcgarry.2015@mumail.ie
Supervisor:	Dr John Keating, Senior Lecturer, Computer Science
Email Contact:	john.keating@mu.ie

All participants in this research project must sign this consent form in order to participate. As outlined in the Information Sheet, every effort will be made to keep information private and confidential. However, it must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances, the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.

If at any time, up to the point of anonymisation, any participant decides to withdraw his or her consent for this project, he or she may do so by contacting the researcher. As such, any participant who opts to withdraw will have all data collected removed from the system of record and discarded.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at research.ethics@nuim.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

I, _____, understand and consent to all information disclosed in the Information Sheet and consent to participation in the case study. I acknowledge I have read the Information Sheet and understand the nature of this research project. I consent to be contacted by the above stated researcher or supervisor regarding my participation in this project and that all data will be anonymised at the end of the Fall 2018 term. I also understand that prior to anonymisation, all contact information will be held privately and confidentiality in a secure location by the researcher and will not be shared for any reason. Furthermore, I acknowledge I have read and have received a copy of this consent sheet for my own personal records.

☐ By checking here, I agree to participate in one-on-one interviews as outlined in the Information Sheet and consent to an audio recording of the interview.

Print Name: _____

Signed: _____

Date: _____

D.8 Ethical Approval - Seminar

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

20 May 2016

Shane A. McGarry
An Foras Feasa
Maynooth University

RE: Application for Ethical Approval for a project entitled: Escaping the Book Metaphor: Examining User Interface Metaphors in Digital Scholarly Editions

Dear Shane,

The Ethics Committee evaluated the above project and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30 September 2017.

Kind Regards,

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

C.c. Professor Susan Schreibman, An Foras Feasa

Reference Number
SRESC-2016-035

D.9 Case Study: Code System

Following is the code system used for the creation of all coded segments for this case study analysis.

TABLE D.1: Code System Used for Qualitative Analysis of Post-test Responses for the Case Study

Code System	
Keyword Preference	Used whenever a participant states a preference for the keyword search over the visualisation search. NOTE: This code is only used for explicit statements of preference.
Viz Preference	Used whenever a participant states a preference for the visualisation search over the keyword search. NOTE: This code is only used for explicit statements of preference.
Cognitive Overload	Issues related to cognitive overload or a feeling of being "overwhelmed"
Learning	Any data related to the acquisition of new knowledge
Positive	Used to code any positive feedback about the system.
Relevance	Any items related to relevance of the data as it relates to the search process. This differs from clarity, which focuses on items which are easily understood.
Clarity	Used to code positive comments regarding clarity of the data
Usability	Used to code positive usability comments
Negative	Used to code any negative feedback about the system.
Relevance	Any items related to relevance of the data as it relates to the search process. This differs from clarity, which focuses on items which are easily understood.
Clarity	Used to code issues regarding clarity of the data
Usability	Used to code negative usability comments about the system

D.10 Case Study Transcripts & Coded Segments

D.10.1 Keyword Transcripts

Participant: 5c793c999e02291bb0072d47

What was your initial reaction? The keyword search algorithm is very good, it gave all the results very quickly. One possible improvement according to me would be, give user/researcher option of sorting with respect to years

What did you like most? It was super fast

What did you like least? nothing, it was really good only sort functionality can be added

Was there anything that particularly appealed to you? besides quick, it was efficient as well, so i got the info i wanted within minutes.

What was something interesting you discovered? it is crucial to understand that search can make our job very easy provided, we must know what we are wanting to do in that job.

What did you learn while using the keyword search? we can easily understand about the trend of food like spices etc with the help of this website.

What would you change about the keyword search? yes as said, the back button should take me to results page rather than home page after the initial search is performed. secondly in dashboard kindly removes the cursor type-able symbol as it may confuse the user.

Participant: 5c793c189e02291ec95d87f0

What was your initial reaction? I didn't see what I expected

What did you like most? Nothing in particular

What did you like least? Whatever word I type, I felt like it displays the whole content in the database

Was there anything that particularly appealed to you? Nothing, Indeed it made it hard

What was something interesting you discovered? People survived without eating food for days and its common during those times

What did you learn while using the keyword search? I was amazed that people recorded what they ate everyday

What would you change about the keyword search?

1. Display relevant search results
2. Some of the records are not mentioned. For example: October 1779, there is expense record mentioned.
3. Advanced search is not working properly.(taking too long to load).

Participant: 5c793ec19e02291baf2ff332

What was your initial reaction? Easy to use, limited set of options.

What did you like most? The ability to specifically target the data you want from when you want.

What did you like least? When I searched for a particular keyword and I was brought to the page where I select the document that I want to view, when I am done viewing it there is no button to go back to the previous page and select a new document, instead I had to click the back button on the browser and then that would bring me back to start my search all over again.

Was there anything that particularly appealed to you? What was spent on what.

What was something interesting you discovered? The amount of money spent on wine was pretty crazy.

What did you learn while using the keyword search? I learnt about when certain foods were unavailable throughout the year and with more time I probably could have figured out why. **What would you change about the keyword search?** I would like if there was more instructional information especially in the visualization tool.

Participant: 5c793bdf9e02291bb384f4d7

What was your initial reaction? Enter input, find something different.

What did you like most? Bread or wine.

What did you like least? Army or politician. Some words do not have result.

Was there anything that particularly appealed to you? Bread or wine.

What was something interesting you discovered? How many matches according to the corresponding word.

What did you learn while using the keyword search? The words happened during the years.

What would you change about the keyword search? I think the best way to have good user experience is that using the Visualisation search first and keeping the keyword search as a candidate search. Because it is not always to meet the user's expect by Visualisation search. Such as some words are quite less used.

Participant: 5c793b7b9e02291bb250e588

What was your initial reaction? a list of records appeared with that contain the keyword I searched for

What did you like most? It showed me the list of all the the records with the one word I searched for. along with page tittle, number of times that keyword is repeated in a particular record and little text as well

What did you like least? the back button. whenever I click on a record from the searched list, when I click the back button, It does not take me to the searched list but take me back to initial page where I have to write the keyword again.

Was there anything that particularly appealed to you? I do not know.

What was something interesting you discovered? the page tittle shows different dates of a particular record. This was interesting as it showed me which record occurred on which date.

What did you learn while using the keyword search? with just giving a keyword system can display all the records with that keyword in it. I learned format of different receipts with different money amount of each item that a person has ordered. It increases my knowledge in this aspect.

What would you change about the keyword search? there are few things:

1. there should be an order button on the page where searched records are displayed so that user can order the results in ascending or descending order as currently records are not in order

2. when I right click on view page of any particular page and click on option 'open link in new tab', it asks me to login again but I am already logged in in the same browser.

3. When I save my search and go back to see them in my saved notes, the record shown in my saved notes is not in same order and it is little confusing that I am unable to know that whether system has searched my right record or not.

Participant: 5c793f8e9e02291bb1295dec

What was your initial reaction? Thought of Information Retrieval

What did you like most? One word giving loads of information from every month to year giving a very fine detail about it.

What did you like least? Easy to understand but wasn't easy to grasp.

Was there anything that particularly appealed to you? Was already familiar with it

What was something interesting you discovered? Easy access to information keeps the user engaged into what he or she is actually looking for.

What did you learn while using the keyword search? How to get the resources fast for what I'm looking at.

What would you change about the keyword search? Nope, it had all the information crisp and short to make it easily understandable.

Participant: 5c7940829e02291baf2ff334

What was your initial reaction? Very easy to use, even advanced search options

What did you like most? It was easy to find the exact piece of data you were looking for.

What did you like least? It exits to dashboard when you have viewed a particular budget entry in detail instead of returning to your search. For example when you look at May 1777 and then wish to look at June, you must search again.

Was there anything that particularly appealed to you? Yes, it provided comprehensive results with much less categorizing by the user, in comparison with the visual search

What was something interesting you discovered? I wished to see if particular entries had been duplicated for the month of November 1778. it was easy to observe hen looking at a detailed version of each entry that they were different.

What did you learn while using the keyword search? My first time making a academic poster.

What would you change about the keyword search? No I found the system extremely user friendly.

Participant: 5c7941a29e02291bb1295dee

What was your initial reaction? Standard keyword Search

What did you like most? did what it says

What did you like least? If you tried to open a search item to a new tab you were asked for Login details again, or if you clicked it it went to that page and then you lost your search. The previous searches wasnt too clear when you went back so did the search again

Was there anything that particularly appealed to you? not really

What was something interesting you discovered? NA

What did you learn while using the keyword search? it is a good system, made me look at why we do the things we do when using systems like this.

What would you change about the keyword search? Having the login persistent across new tabs when opened from a search, like with visualistaion search, or when you click view, it opens to a new tab rather than open on the current page and losing your search or where you were in the search results, especially when the search results returns multiple pages.

Participant: 5c7954449e02291bb250e590

What was your initial reaction? Easy to understand. Basic. Really liked the past searches log.

What did you like most? the past searches log.

What did you like least? Top _____ words and the Words appearing less than _____ times. These two features dint work as per expectation as the page kept on showing the loading symbol and dint landed upon result.

Was there anything that particularly appealed to you? Keyword search is a usual way of searching. It is easy to understand and learn.

What was something interesting you discovered? Total spent of the log of expenses of Alcala during a span of time upon different categories.

What did you learn while using the keyword search? total spent over number of years.

What would you change about the keyword search? Help or suggestions for user prevent invalid entries to search bar. Improve upon two features for Top _____ words and Words appearing less than _____ times.

Participant: 5c79553a9e02291bb384f4df

What was your initial reaction? A very traditional way of searching.

What did you like most? Very convenient advanced index and clear results display

What did you like least? If the amount of data is too large, it will cause the index to be slow.

Was there anything that particularly appealed to you? Yes, you can index a keyword directly, and you can quickly search for results without any restrictions. But often it is so difficult to get the results we want.

What was something interesting you discovered? In the upper part of the index bar, you can see the results of the past index, instead of looking at other indexing software, you can see the historical index in the index bar.

What did you learn while using the keyword search? Learned the typesetting format of keyword index and the novel concept of chart index

What would you change about the keyword search? Increase the index speed, if possible, the font size of the index result can be increased appropriately

Participant: 5c793c629e02291bb1295de8

What was your initial reaction? On entering the keyword, it quickly showed the results so I was happy with it.

What did you like most? Tiled view looked good to me.

What did you like least? There was no sorting mechanism available to sort the data based on year or the money spent.

Was there anything that particularly appealed to you? It was little difficult to answer a particular question because I had to open an individual page to see the balance or the money spent.

What was something interesting you discovered? I searched for category spices and had opened certain pages for the year 1781 and 1774 and found that the trend of money spent was decreasing.

What did you learn while using the keyword search? Pictures are usually more appealing than words. so visualization search was better than keyword search.

What would you change about the keyword search? Sorting should be implemented and in case of data corruption, error should be handled properly so that it doesn't keep on loading.

Participant: 5c7d378f9e022938b24bc147

What was your initial reaction? It helped that I got a preview of the extract that matched my keyword(s)

What did you like most? That it gave the year of the occurrence of each extract and gave me a preview into these things

What did you like least? I couldn't visualise the data very well or extrapolate much information from the accounts

Was there anything that particularly appealed to you? Not really. It was hard to draw conclusions from looking at text.

What was something interesting you discovered? Bread was in high demand for the college for a long time.

What did you learn while using the keyword search? That Irish eating habits resemble that of Irish people today which shows that even people who emigrated stayed true to their heritage

What would you change about the keyword search? No. I felt that it was very user-friendly and fun to use.

Participant: 5c7955da9e02291bb1295df0

What was your initial reaction? It was as expected with basic search providing both basic and advance options.

What did you like most? Its easy to use. Easy to relate with other search systems. Need less understanding.

What did you like least? There should be additional sorting options other than relevance. Will help to analyse data more efficiently.

Was there anything that particularly appealed to you? The in-text search and the relevance sorting was actually relevant most of the time to what i searched

What was something interesting you discovered? That even though the price of food did not vary over the years. There has sometimes been more expense than credit.

What did you learn while using the keyword search? Colleges were probably dependent on outer donations sometimes as the expense was more than normal income sometimes.

What would you change about the keyword search? Transition between search results and opening pages. Also the error handling of some pages where the page keeps on loading.

Participant: 5c793dc19e02291bb1295dea

What was your initial reaction? the search time is really fast. sometimes the screen gets hang may be due to server error

What did you like most? easily finds the keyword in the paper and tells how many times it appeared

What did you like least? every time i click back button i have to search again with the same keyword that i used earlier

Was there anything that particularly appealed to you? advance search

What was something interesting you discovered? advance search helps to dig more into the system and get backs to you with what you want

What did you learn while using the keyword search? how people survived in 17-18C. and what were the key factors

What would you change about the keyword search? some times after searching ?? appears as the paper name. and user has no idea what it is

Participant: 5c793d909e02291ec95d87f2

What was your initial reaction? it was good, all results fetched at once. I could visit results individually and know about the commodity.

What did you like most? I could visit results individually and know about the commodity in details corresponding to a particular date.

What did you like least? Search was good, there is no particular dislike.

Was there anything that particularly appealed to you? Yes, the keyword search made answering easier. I could trace each of the result in details.

What was something interesting you discovered? I learnt about old records of transaction of irish colleges.

What did you learn while using the keyword search? About the irish colleges and their expense on food items. Had used an online tool to search through written assembled record of history first time.

What would you change about the keyword search? nothing as such.

Participant: 5c793dfa9e02291bb0072d49

What was your initial reaction? The keyword search is easy to understand where we just type a keyword of what to search and results are generated which contain these keywords.

What did you like most? The overall interface is really easy to understand for the user while using this keyword search

What did you like least? When back button is clicked , the website navigates to first page where we need to again type the keyword we were searching.

Was there anything that particularly appealed to you? As I am not that familiar with the content of the website , I just went on through the keyword search feature and saw if I could get any historical info from this that seems interesting.

What was something interesting you discovered? I found some interesting facts about the amount spent on food between 1700-1800 and where most of the expenditure went.

What did you learn while using the keyword search? I found some interesting historical info and for first time,I came across the currency Reales and Maravedises.

What would you change about the keyword search? I did not went through the overall system but just one feature of it and I did not find much changes that are to be made in the keyword search except there might be change which can be made while pressing the back button after going through the search results, the website navigates back to first page and asks to enter the keyword again.

Participant: 5c0a59609e02299a92d16b4e

What was your initial reaction? I was initially pleased because it seemed to be giving me a more straightforward means of accessing the kind of specific or granular data I might have felt the visual search was lacking

What did you like most? Just as in the above, the capacity it offers to search across a wide range of documents easily, seeing the text snippets and where the data is coming from

What did you like least? It very quickly brought me onto documents that hadn't been marked up or at least the transcriptions were not displaying as they were in the previous exercise. Furthermore, when I clicked on document icons, they didn't open in a different tab, but changed the one I was currently on. Given how difficult it was to get back to the previous search without searching again this became frustrating. 'Open in a new tab' generally led to request timeout, or back to the login screen

Was there anything that particularly appealed to you? The basic functionality of a keyword search, such as 'city' or 'book' immediately gave an insight into how the school was relating in tangible terms to its immediate geographic context. Just optimising the function by resolving these issues outlined above would render it a far more successful feature

What was something interesting you discovered? The students were reading Thomas Acquinas, the school seems to have an arrangement with a local apothecary to get access to medicine

What did you learn while using the keyword search? The students marked the feast day of Saint Patrick and Christmas with cover-all expense, rather than detailed information about what it was they were purchasing, as is the usual arrangement

What would you change about the keyword search? The loading, display and interface issues outlined above

Participant: 5c0a63559e02299a9087d38e

What was your initial reaction? This is a straightforward, intuitive and immediately useful way of searching this record.

What did you like most? Its simplicity and appropriateness to a record of this nature which, while containing some quantitative information, is not one that I would really regard as containing enough quantitative data to make visualization by means of charts etc. particularly useful.

What did you like least? Keyword' seems slightly misleading here to me. In fact, it seems to be a complete word search of all the transcriptions in Spanish and their English translation. Would it not be useful to make this clear? Also, I can't help wondering if it would be useful to generate dropdown lists in alphabetical order (one for English, one for Spanish) of all 'keywords' (which is effectively all words) appearing in the text being searched. For me, a problem about keyword or any other kind of word searches is that it's much more useful if I have an idea of what words occur in the text. I don't want to waste time searching for words which do not occur, and which will give no results.

Was there anything that particularly appealed to you? Displaying the search term highlighted in bold in a snippet of the contextual text is most useful. In general the search interface (for ordinary searches and visual searches) for this record is very well designed - attractive, legible, intuitive.

What was something interesting you discovered? Well, no real new historical information. This is, broadly speaking, a type of record I am familiar with, and they tend to follow a certain pattern. Of course, if I had research interests in Irish colleges, it might be different.

What did you learn while using the keyword search? Not sure about learning! But I had the opportunity to try out a well designed search tool, for a record which, while hardly unique, is an interesting specimen of its genre. Thank you for all your work, and for organising this event, and giving us the opportunity to give feedback. :-)

What would you change about the keyword search? Nothing beyond a few comments already made above.

Participant: 5c0a5f259e02299a914f6574

What was your initial reaction? It doesn't look like other databases that I use for research. I wasn't able to have fun in data exploration at the beginning, to see how many things are in the database and the breadth. I also was unfamiliar with the college, and found that there wasn't enough information about the college for me to pick apart a picture of kind of what it was like. Are there any pictures? I expected there to be an About page that explained more than the small description on the home page. There was also little additional information when I did a google search.

What did you like most? I liked that there were little sassy descriptions of people who they have an economic relationship with.

What did you like least? The thing that bugged me the most was that when I hit the back button, it didn't go back to my search and I had to type it in again.

Was there anything that particularly appealed to you? No, other than how the words they used to describe the people they had an economic relationship with. But I wanted to know more.

What was something interesting you discovered? I learned what a drover is, and that they had a boy who swept the college hired out, rather than it being a chore of the priests and trainees.

What did you learn while using the keyword search? I learned that in the 17th-18th century Irish people in Spain were not treated as foreigners according to law and the king, that is kind of interesting. From here: https://www.reddit.com/r/ireland/comments/86oxbo/til_irish_emigrants_to_spain_during_the_1600s_and/

I also learned I don't know old currencies of Spain or what they mean.

What would you change about the keyword search? I would like confirmation that an object has been saved, so that I can be sure. Where has it gone?

It doesn't make sense to me that a search gets called an object, just like an item. Why is the save button at the bottom? Should be at the top.

I would like an all option in number of results

When I tried a search in quotes for a particular phrase "drover and slaughterer" it didn't work.

Once I've saved a search, I would like to go back to all the search results by clicking on it, or see saved searches in the search menu. Instead I have to go back and type in the same search, and sometimes the results are in a different order!!

I would like to be able to filter out specific line items, and to aggregate certain line items that we think are similar, such as cost of slaughterer that they pay each month.

I would like to know if I can do Boolean searchers, nested searches, etc. Searching for something like: boy who sweeps the college

The results show that one line item contains all these words, so it repeats these line items.

Participant: 5c0a62599e02299a8f45ce04

What was your initial reaction? Quite basic to us; would benefit from more faceted key word searching or more advanced searching e.g. 'search exact phase only'

What did you like most? Very clearly presented results page

What did you like least? Having to search through the results myself, without being able to facet or filter results. It would also be useful to click back to results page rather than having to start search from scratch or click through to previous searches.

Was there anything that particularly appealed to you? The results page having words that were searched for highlighted

What was something interesting you discovered? I didn't know anything about this topic of history so the collection/site is very interesting.

What did you learn while using the keyword search? I learned about the aspect of history and these particular records.

What would you change about the keyword search? Better search and filtering of results. Going back to search results easier. When clicking to next page of search results, could go to top of the page rather than having to scroll up.

Participant: 5c0a5d509e02299ab8d1cd42

What was your initial reaction? No reaction.

What did you like most? NA.

What did you like least? Related terms might be useful?

Was there anything that particularly appealed to you? No.

What was something interesting you discovered? NA.

What did you learn while using the keyword search? I value the ability to manipulate the results after obtaining them e.g. arranging them in an order that suits my research needs.

What would you change about the keyword search? Hover/click info boxes (with examples) might make the commands easier to use.

Participant: 5c0a591c9e02299a914f656f

What was your initial reaction? Straightforward, simple, familiar, easy to use

What did you like most? The keyword was bolded in the results list so I could clearly see the keyword in relation to the document.

What did you like least? It was a bit boring. I couldn't reorganise the search results by year or month, for example, and the results themselves didn't appear to be in any particular order (chronological, number of times keyword was mentioned, etc).

Was there anything that particularly appealed to you? The basic and advanced options were both simple and easy to use. There wasn't anything complicated about them.

What was something interesting you discovered? The keyword "wine" appears more times than "bread" or "meat" so clearly those guys liked to drink!

What did you learn while using the keyword search? Folks in the late 1700's really liked their wine! They also considered things like new bedsheets and fixing pans an "extraordinary" expense, which suggests a different understanding of the definition of "extraordinary" in the late 1700's versus today's contemporary world.

What would you change about the keyword search? The ability to sort search results based on the categories (year/month, for example) and perhaps some expanded advanced search options to be able to pick more than one year at a time.

Participant: 5c0a5fdc9e02299a92d16b52

What was your initial reaction? It operates like pretty much all Keyword searches. But I was surprised by my own reaction to it as a result of using the visualisation tool first. The results required a different skill set to see what I was looking for.

What did you like most? It did exactly what it is supposed to do

What did you like least? My main concerns were about functionality. If I clicked in on an image or particular entry I then lost my search and had to go back to the start again. This slowed the process down.

Was there anything that particularly appealed to you? I started to identify the data around the key pieces of information I was looking for. For example, I wanted to use the visualisation tool to get a sense of when more wine was purchased. Then, subconsciously, I was using the keyword search tool to find out more information around mentions of wine in the archive. That's when I began to see other patterns and identify other questions, such as what about abstinence days? How can I incorporate the cheaper rates into my previous visualisations.

What was something interesting you discovered? It reminded me that subconsciously I was using the two tools to address the same theme but my expectations of what answers I would get out of the results modified my research questions to meet the search/exploration tools.

What did you learn while using the keyword search? I learned more about the archive, about the Irish in Spain and about how we can use visualisations at a very basic level to help us explore historical data.

What would you change about the keyword search? Most of my comments would be about functionality. I would like to be able to use the tools to greater extent. Thank you for the opportunity.

Participant: 5c0a59b69e02299ab8d1cd40

What was your initial reaction? useful, expected in a digital resource

What did you like most? quick to return query

What did you like least? did not provide an option to return to search, rather one has to back click, and go to recent search, which is annoying

Was there anything that particularly appealed to you? useful, expected in a digital resource

What was something interesting you discovered? while there is a mention of a teapot, there is no mention of tea...

What did you learn while using the keyword search? n/a
What would you change about the keyword search? return to search option

Participant: 5c0a5aa79e02299a92d16b50
What was your initial reaction? Clean, simple user interface
What did you like most? That it was functional and worked very effectively
What did you like least? The Advanced Search was lacking and the stop words section did not work
Was there anything that particularly appealed to you? The keyword search throws up very general answers and was useful for the specific question.
What was something interesting you discovered? Yes, it appears that fruit & veg were not being purchases or claimed for as expenses
What did you learn while using the keyword search? I learned about a new archival source and I enjoyed the interface of the DRE
What would you change about the keyword search? A more nuanced Advanced Search that allows for monthly segregation of search terms and a more in-depth filter for disaggregating large search results.

Participant: 5c0a62329e02299a936b4940
What was your initial reaction? Easy to use, very familiar format.
What did you like most? The simplicity of the search and results.
What did you like least? It was unable to search for full phrases.
Was there anything that particularly appealed to you? The search was straightforward.
What was something interesting you discovered? The results were easier to understand than through the visual search.
What did you learn while using the keyword search? They ways data can be searched for and presented. The keyword search was simpler and the results more easily understood.
What would you change about the keyword search? Ability to search full phrases in keyword search.

Participant: 5c0a617f9e02299ab8d1cd44
What was your initial reaction? Negative, since at least on this browser, opening a search result in a new tab required a new login, and pressing back to return to search results meant having to redo the search.
What did you like most? The fact that the advanced tab was simple.
What did you like least? Functionality, presumably due to the browser, since I had to save searches and access those through the dashboard for detailed searches, rather than simply moving back to a set of search results.
Was there anything that particularly appealed to you? .
What was something interesting you discovered? That searching for a word in both English and Spanish, e.g. meat/carne, returned results for me in a different order.

What did you learn while using the keyword search? I learned that wage increases for the cook and laundrywoman were very rare indeed.

What would you change about the keyword search? Being able to rearrange the saved items on the dashboard, a click and drag sort of thing to move the individual 'modules' manually. Also a way to perhaps colour-code them according to types of search, or put them in sub-folders.

Participant: 5c0a8bf79e02299a92d16b5d

What was your initial reaction? I found it very easy to navigate the keyword search.

What did you like most? It was easy to see search results at a glance, along with the number of times the search result appeared in the text.

What did you like least? I found it difficult to combine search terms e.g. bacon and salt, as opposed to just bacon.

Was there anything that particularly appealed to you? The keyword search made it easy to contextualise search data e.g. to see how much money was spent on salt for curing bacon rather than how much was spent on cured bacon.

What was something interesting you discovered? That the price of salt increased from 1776 to 1779 by about 0.5 Rs.

What did you learn while using the keyword search? How to save data to the dashboard in order to compare results.

What would you change about the keyword search? No

Participant: 5c0a5a289e02299a936b493c

What was your initial reaction? straightforward google type search

What did you like most? its simplicity

What did you like least? inability to combine two given years and exclude others, losing last search when hitting the back button and having to redo it

Was there anything that particularly appealed to you? easy to locate information on a keyword or given year or over the entire timespan

What was something interesting you discovered? it retrieves a keyword regardless of whether capitals are used eg wine and Wine. This is good.

What did you learn while using the keyword search? creation of saved searches to create a dashboard

What would you change about the keyword search? not losing the search I've just done when hitting back arrow after examining a document. I don't want to save every single search or look in recent searches at the top every time necessarily.

Participant: 5c0a59f59e02299a9087d388

What was your initial reaction? Easy to start navigating

What did you like most? It behaved as I expected (returned organized search results)

What did you like least? Figuring out variations in spelling within the data

Was there anything that particularly appealed to you? It depends on the question!

What was something interesting you discovered? Variations in spelling (and possibly translations) exist within the data

What did you learn while using the keyword search? I learned about diet in Alcala

What would you change about the keyword search? A glossary of (old) spanish terms

Participant: 5c0a5ae19e02299a9ecf7d96

What was your initial reaction? as I expected, lots of results

What did you like most? completeness

What did you like least? overwhelming amount of data

Was there anything that particularly appealed to you? No particularly. I tended to start both searches with a 'keyword' as it formed my question. However, the graphic search produced the most attractive result, in that it bundled information into results

What was something interesting you discovered? Yes, that the expenditure variations over time are less important than particularly once off expenses that distort overall budgets

What did you learn while using the keyword search? That the best results come from a combination of searchword and visualisation

What would you change about the keyword search? I would make it easier to move between searches. Also, I think it would be good to indicate how the data was processed for the visualisation.

Participant: 5c0a61ff9e02299a9087d38c

What was your initial reaction? It worked adequately

What did you like most? That it functioned

What did you like least? It was limited to filtering by year, I could not filter for the cost of the search item year by year or seasonally

Was there anything that particularly appealed to you? Costs could be broken down into categories

What was something interesting you discovered? I liked that the search results could be viewed in English or Spanish

What did you learn while using the keyword search? That a search for 'spice' returned 88 results :-)

What would you change about the keyword search? It would be better if searches could be more flexible

Participant: 5c0a5ee69e02299a8f45ce02

What was your initial reaction? It was... very simple (not in the good way). Even the "Advanced Search" offered a paucity of options.

What did you like most? Because it immediately brought up a list of entries, I could easily tell if I'd hit the jackpot with a search term. Also, seeing the terms in context sometimes made it possible to discover other words to use as terms that I would not have previously expected.

What did you like least? Navigation was a pain. Navigating to the single page display would move away from the search page and require me to enter the search term again, and trying to open a page in a new tab always required a new login.

Was there anything that particularly appealed to you? It was easy to see if particular types of food were consumed, as well as how often they were used or if something was a special case.

What was something interesting you discovered? Spices! Lots and lots of spices were consumed (88 line items!). Also, chickpeas popped up on a handful of occasions, which I would not have expected.

What did you learn while using the keyword search? I stumbled across an item on March 1781 detailing what the expected daily rations were, which was very interesting. I also learned that Chickpeas were normally supplied by the Visit to last for a year, but clearly they occasionally ran out. I also appreciate how difficult it is to make effective search mechanisms... some elements of the system are unclear!

What would you change about the keyword search? I would definitely have persistent login across tabs, and maintain the ability to modify previous searches. More detail on the line-item level would be useful.

Participant: 5c0a6aa79e02299a9087d394
What was your initial reaction? Simple, easy **What did you like most?** Simple
What did you like least? Not sure of reasonable searches to make on this data
Was there anything that particularly appealed to you? No
What was something interesting you discovered? No
What did you learn while using the keyword search? Some Spanish
What would you change about the keyword search? Dictionary of terms, glossary of possible search items, export options for poster etc.

Participant: 5c0a5a749e02299a914f6571
What was your initial reaction? Seemed like a standard keyword search interface. Still don't know what the word count filters are for - trying to used them hung the browser.

What did you like most? Full text search - easier to find records of interest compared to more restrictive visualization search.

What did you like least? Results not date sorted. No return to search results (recent searches didn't always have the latest). Returning to basic search unexpectedly applied advanced search filters.

Was there anything that particularly appealed to you? Search made it faster to find specific information - therefore faster to answer questions.

What was something interesting you discovered? No

What did you learn while using the keyword search? Could be improved to make research easier.

What would you change about the keyword search? More information about options

D.10.2 Visualisation Transcripts

Participant: 5c793c999e02291bb0072d47

What was your initial reaction? Very impressive at the glance.

What did you like most? we can change our filters and understand information within minutes.

What did you like least? few points are not clear like in multibars.

Was there anything that particularly appealed to you? yes for example the bread consumption in 1777 made obvious that in the last quarter of year the bread consumption reduced significantly then other time of the year.

What was something interesting you discovered? visual information can be interpreted within few minutes provided it must be clearly represented.

What did you learn while using the visualisation search? The trend of bread-consumption for year 1777 was very easily understood and can be conveyed with the help of visual search functionality

What would you change about the visualisation search? yes would have made multi-bars much more clearer if possible

Participant: 5c793c189e02291ec95d87f0

What was your initial reaction? It's easy to understand the search results compared to keyword search

What did you like most? it displays relevant information in the charts

What did you like least? Everything is good

Was there anything that particularly appealed to you? I observed that sometimes the spices were not at all used during some months of a particular year and overall there is a decline in the usage of spices.

What was something interesting you discovered? The demand for a particular item(wine,spices etc) varies depending upon the month and the year.

What did you learn while using the visualisation search? Demand for a particular item like spices, wine, bread etc is varies a lot.

What would you change about the visualisation search? The keyword search should display appropriate search results.

Participant: 5c793ec19e02291baf2ff332

What was your initial reaction? My initial reaction was that there was a lot of information to take in on the screen, at times I found it difficult to know what I was looking at in regards to the charts and what the colours represented on the charts (specifically the bar charts where information was stacked on top of information).

What did you like most? I thought that the pie charts gave the best visual representation of the data.

What did you like least? I found that some of the charts where a bit clustered, especially the line charts and the stacked bar charts. Also when I did look at the data by the years I could see data in certain years, then when I viewed that year specifically no data would be shown.

Was there anything that particularly appealed to you? Yes, it made it all very clear in regards to what was going on diet wise, in regards to meat only mainly in december and january and this showed it very well.

What was something interesting you discovered? I learned that the meat was only available in january and december and this could have been because of the lack of salt in other months to keep it fresh.

What did you learn while using the visualisation search? Diets of people in the 1700's

What would you change about the visualisation search? More information on what I was looking at.

Participant: 5c793bdf9e02291bb384f4d7

What was your initial reaction? It looks just a data summary.

What did you like most? UI gives some help for user, for example, it already summary the category for user, user can depend on what he/she wants to search to choose the content to search without any typing.

What did you like least? Some button or icon is not clear with description, sometime I did not know what it is.

Was there anything that particularly appealed to you? For me, I am interested about the change of wine in the period.

What was something interesting you discovered? Check the change during the years.

What did you learn while using the visualisation search? What the different graph means. And figure out some interested thing of change from the data.

What would you change about the visualisation search? Some icon with more description.

Participant: 5c793b7b9e02291bb250e588

What was your initial reaction? it showed me a visualization of data based on my searched keyword and type of selected visualization mode like a pie chart or bar

What did you like most? how a same data can be shown with different modes of visualization and how the result can be downloaded as an image.

What did you like least? there is no detail or tittle with searched results that can show that this record is showing what type of searched keyword especially when I download an image, its quite confusing to know that this image is showing results about what data?

Was there anything that particularly appealed to you? I did not observe anything like that

What was something interesting you discovered? when I searched with the bread keyword, I liked how in different years, there is different information about this keyword that is then shown with interesting colors on the chart.

What did you learn while using the visualisation search? different colors of the graph showing different information about the keyword and filters user selected

What would you change about the visualisation search? 1) back button is not working correctly as if I go to dashboard or any other page after searching a particular keyword, when I click the back button, It does not take me to previous page but it takes me to initial page where I have to select all filters again.

Participant: 5c793f8e9e02291bb1295dec

What was your initial reaction? Gives a brief idea of what we are looking for.

What did you like most? More easy to understand when compared to key search

What did you like least? Nothing specific All of it was quite clear

Was there anything that particularly appealed to you? How there was a drastic change all the years and suddenly it was constant for quite some time and drastically dropped

What was something interesting you discovered? Easy to learn and understand when compared to key search

What did you learn while using the visualisation search? Visualization eases the way of understanding the concept more better.

What would you change about the visualisation search? Giving a visualization with the information so that the user will be more connected with the content.

Participant: 5c7940829e02291baf2ff334

What was your initial reaction? I found the visualization search more difficult that I anticipated as even though the data was clearly presented, more details on this were difficult to obtain.

What did you like most? Clear representation of the data, it was easy to see discrepancies among different years and budget categories.

What did you like least? When one clicks on a bar in the table representing a particular point of budget (i.e. meat), it provides more details. However if one, exited this page by pressing back, it returns to the homepage instead of the table of data. I found this annoying.

Was there anything that particularly appealed to you? Yes. I searched for comparisons between budget spent on meat and food respectively. I noted immediately there was two years 1776 and 1778, where the budget for meat actually exceeded the budget for other food items.

What was something interesting you discovered? Yes, details in 1776 revealed that a portion of the budget had also been spent on preservatives and spices, enabling them to preserve the extra meat

What did you learn while using the visualisation search? The system allowed me to visualize the budget needed to maintain the running of the refectory.

What would you change about the visualisation search? Being able to return to the table form details page

Participant: 5c7941a29e02291bb1295dee

What was your initial reaction? You get drawn to the visualization first and dont pay attention to the default search parameters, Which makes it a bit confusing. But once you get to grips with the search parameters it makes more sense

What did you like most? The data straight away at a glance.

What did you like least? Some of the charts were confusing with too much happening in them, no key to tell what was what

Was there anything that particularly appealed to you? no

What was something interesting you discovered? Possibly different wine suppliers owing from different amounts of wine at different prices, or possibly getting better wine for some reasson

What did you learn while using the visualisation search? The importance to highlighting where the user should focus when using a system like this

What would you change about the visualisation search? not to have a default search displayed when first open, as distracting and caused delay.

Participant: 5c7954449e02291bb250e590

What was your initial reaction? Visualizations are properly color contrasted with pleasant color scheme. Multiple options to switch between the graphs adds to the understanding of the user. Features like downloading the chart and saving the chart along with the filters is a good serves well to the users. Drill down facility present in the chart helps drill down to the descriptive/detailed log of the search.

What did you like most? Features like downloading the chart and saving the chart along with the filters is a good serves well to the users. Custom search options and group by facility. Drill down facility present in the chart helps drill down to the descriptive/detailed log of the search.

What did you like least? It becomes bit overwhelming when there are multiple tags attached with the entry. Like Meat and bread also include food with same total spent amount. There are no alert or error box as when entered wrong value to the custom search it takes long time to search a wrong or invalid entry.Also if legend pallet would have been available user could save some time hovering upon every partition in the visualization.

Was there anything that particularly appealed to you? Nothing as such.

What was something interesting you discovered? The spent record of the alcala across various categories from 1774-1782 (1780 excluded).

What did you learn while using the visualisation search? Details upon expenditure and value of the currency during that period.

What would you change about the visualisation search? Already mentioned in previous question for what I least liked about the visualization.

Participant: 5c79553a9e02291bb384f4df

What was your initial reaction? I think it's a very intuitive way to connect queries from charts and then.

What did you like most? Double-click on an attribute on the chart to jump directly to the study page where the value is located.

What did you like least? The data corresponding to each year is displayed according to the month, then the data for one year is divided into 12 copies. From the personal point of view, it is more hopeful that you can see the data and reason analysis of each year, instead of expecting a month and a month to go. There may also be summary data for each year, but I didn't notice it.

Was there anything that particularly appealed to you? In terms of data links, clicking on the data will automatically jump to the detailed data, but the detailed data list is a bit too much. Then the chart of the main page is not very accustomed to stacked as the default, customary histogram or line drawing is more intuitive.

What was something interesting you discovered? First of all, from the subject, from 1779 to 1781, during these years, the demand for red wine suddenly increased, and the data of various foods produced great abnormal values. This should be related to the historical background. From the perspective of IXD, UX feels good, with more connectivity than traditional stand-alone display and search. The UI design is simple and clear.

What did you learn while using the visualisation search? From the visual system, I learned more about the details of the interface design. For example, better jumps, a clear interface, and a quick search method.

What would you change about the visualisation search? If the detailed statistics behind the data can be more clearly reduced, it is better to make a list form. It would be better to click on the ones that can be viewed each year.

Participant: 5c793c629e02291bb1295de8

What was your initial reaction? At first, it was difficult to understand how the search is working and what the value in the graph are showing.

What did you like most? Different colors used in chart.

What did you like least? The graphs are little confusing as what does the value shown ion the chart means is not clear.

Was there anything that particularly appealed to you? I had searched the bread category and what I found is the money spent on bread decreased tremendously in 6 years from around 10,000 to about 2000.

What was something interesting you discovered? I had always thought that expenditure on food would be increasing across the years but it was just the opposite case.

What did you learn while using the visualisation search? I learned that bread utilization decreased to one fourth in 1781 as that of the bread consumed in 1774.

What would you change about the visualisation search? For the new user, it is difficult to understand the graph.The Y axis shows the money spent but in terms of what currency is not clear.

Participant: 5c7d378f9e022938b24bc147

What was your initial reaction? I really enjoyed the colour schemes and the different ways to display the information about accounts of a particular month, in order to investigate a theory/hypothesis.

What did you like most? Reading and acknowledging the information was much easier using the visualisation search rather than the keyword search because the different graphs helped for displaying different facts to the user. I also feel that screenshots of these are much more useful for presentations as people can simply view how certain trends started to appear on a year-to-year or month-to-month basis.

What did you like least? There was much support in the way of investigating the documents for hard evidence as to why certain trends appeared. Also the term "adjustments" is a little bit vague and so perhaps a key at the bottom of each graph could be displayed with an explanation to that corresponding feature/expense.

Was there anything that particularly appealed to you? There was a significant drop on the expenditure for wine in the year 1777 and this was very unexpected as the month of Christ's birth would have thought to have been the month of most celebrations, it being a primarily christian college.

What was something interesting you discovered? That trends over years tell us a lot about how people lived back then.

What did you learn while using the visualisation search? I learned that the highest expense for priests in this era was wine.

What would you change about the visualisation search? A key/legend at the bottom of each graph that matches the colour to a particular expense and then an explanation of the expense.

Participant: 5c7955da9e02291bb1295df0

What was your initial reaction? Visually page looked great but the chart displayed wasn't accurately conveying what its purpose was. Was it in terms of search term or money spent? Took few moments to realize its purpose. And the legends were not displayed for bar chart had to hover over each color to find out its purpose.

What did you like most? It had all the elements an advance search should have.But they need to be still brushed up and fine tuned. To make them more easy to understand.

What did you like least? Each search click opens a new tab. If you click on chart to open details views there is no visual back button. By mistake several time I used browser back button which took me to home page.

Was there anything that particularly appealed to you? Nothing specific.

What was something interesting you discovered? That there is a no meat day for Catholics which as a non catholic i did not know. Scholars and servants have similar rations for entire month.

What did you learn while using the visualisation search? That how food has been important part in connecting history.

What would you change about the visualisation search? Transition between the elements and page should be simpler.

Participant: 5c793dc19e02291bb1295dea

What was your initial reaction? I saw the chart at the beginning. I have to move my mouse on every color to get to know that category they belong

What did you like most? by looking at the search it helps me to easily find the paper

What did you like least? I took some time to figure out how the search is working. But then it was quite easy for me

Was there anything that particularly appealed to you? nope

What was something interesting you discovered? visualisation search helps to find the objects easily and quickly. And thus i was able to go through many different papers and read them. And come on a conclusion

What did you learn while using the visualisation search? The system helps to give a lot of information. it was easy experience but initially was little hard for me to understand

What would you change about the visualisation search? for a person like me who is not having any history background it was little bit difficult to relate between the things. If at the bottom of page you can give more information about which currency is used as you told in the class. Also what are the monthly expenses amount that they have initially before spending.

Participant: 5c793d909e02291ec95d87f2

What was your initial reaction? it was good, liked it better than the keyword search. Approach wise it was a better search

What did you like most? charts were very nice and gave insights of pattern of consumption without going into much details.

What did you like least? it was good, no dislike.

Was there anything that particularly appealed to you? nothing as such. But yes it was better presentation.

What was something interesting you discovered? pattern of food consumption in irish colleges

What did you learn while using the visualisation search? about irish colleges food consumption from alcala account book.

What would you change about the visualisation search? nothing.

Participant: 5c793dfa9e02291bb0072d49

What was your initial reaction? It is visually appealing and there are lot of functionalities provided, number of filters ,types of visualizations and so on.

What did you like most? The pie chart is really good to understand the visualized content more easily.

What did you like least? Nothing in particular, but the overall interface takes a bit of time to learn and get used to.

Was there anything that particularly appealed to you? Nothing in particular but for one or two years the data related to a few months was missing.

What was something interesting you discovered? I found some interesting facts using visualization search that bread usage was more in few years .

What did you learn while using the visualisation search? It is more about the history and I did not dig deep in to the findings but just was going through all of the interface and it is really good experience for me.

What would you change about the visualisation search? Nothing in particular

Participant: 5c0a59609e02299a92d16b4e

What was your initial reaction? Initially it was confusing, I wasn't sure what to click where in order to make it do what I wanted it to

What did you like most? The bar chart layout was very attractive, and once I had gotten the hang of the filters it became relatively easy to zoom in on what I wanted to and see the macro trends the data amounted to

What did you like least? When we filter it according to particular details often the individual documentation wouldn't show up, or the request would time out

Was there anything that particularly appealed to you? I looked into the ratio of amounts of wine to student to find out how wine consumption varied throughout the year. The patterns I found were calculated manually, if there was a way of doing this with the visualisation search I couldn't figure it out

What was something interesting you discovered? Wine expenditure either disappears in 1776 or enters into another category of expenditure

What did you learn while using the visualisation search? 1 arroba is equal to seventeen quarts

What would you change about the visualisation search? Loading times, primarily. More filtering options would be nice for a greater degree of specificity

Participant: 5c0a63559e02299a9087d38e

What was your initial reaction? I am not sure for myself how valuable it is to be able to visualize a relatively small record such as this.

What did you like most? Not sure.

What did you like least? This record is not predominantly a data series, and I'm not sure how valuable it is to be able to visual the occurrences of various key words occurring in the account book.

Was there anything that particularly appealed to you? See my last answer. Can chart visualizations of occurrences of key words in a relatively small record such as this really add very much value to the records themselves? Are the variations from year to year, e.g., terribly significant?

What was something interesting you discovered? Not sure.

What did you learn while using the visualisation search? I got to browse some of the original pages of this account book, with transcriptions of the original Spanish and English translation.

What would you change about the visualisation search? I think a few paragraphs about the size of the record (how many pages), a description of what lies behind the searches (images of pages, transcription, translation and key words). In particular, if searching key words, I think it's useful to know something about what I am searching. I learnt that quickly enough through browsing results, where I could see the key word files, but a few words on the home page might have been useful.

Participant: 5c0a5f259e02299a914f6574

What was your initial reaction? The first thing I noted is that it is missing tools that I use in my data analytics using R, excel, or other data tools. It is just year, I am unsure what group by means, I am unsure if my ticks have been listened to, I am unsure what top _____ words means, I am unsure what words appear less than _____ times means, and I am unsure where the categories come from, what line items are assigned to what categories, and how this is done.

What did you like most? I liked that I found that there was a boy who was tasked with carrying around 18000 reales.

What did you like least? It is inaccessible, it took control away from me to do what I want and had t work in the confines of the tool.

Was there anything that particularly appealed to you? NO

What was something interesting you discovered? NA

What did you learn while using the visualisation search? I learned that it can be very constricting working within someone elses database, and that just because something is creative commons does not mean it is open source.

What would you change about the visualisation search? Please let me download the data as a csv.

This isn't open source. What I would really like to do is to be able to export line items so that I can do data exploration in a way that I am comfortable with and tools that I use. Who owns this data? Is it the church? Is it Maynooth University or St. Patricks? It isn't even described on the site.

I think also that there is something I am missing in the goal oriented project. I don't come into the archive with a research question because the first thing I do is data exploration. As a data scientist we collect data with a research question, but rarely do we first go into a dataset already collect with a research question. This comes from what is available in the dataset and patterns that emerge.

Participant: 5c0a62599e02299a8f45ce04

What was your initial reaction? Found it easy to use and very interactive

What did you like most? Bring able to browse the data easily

What did you like least? Slightly confusing inputting the keywords/searching by category and words but this was a minor issue

Was there anything that particularly appealed to you? Being able to browse by category/keyword is really useful and having data displayed by year makes it very clear and useable.

What was something interesting you discovered? I learned how useful visualisation is when looking at humanities/archival data. I think it's very important to provide this type of information/functionality for researchers.

What did you learn while using the visualisation search? How useful visualisations can be in the understanding of data.

What would you change about the visualisation search? Better coordination between results pages

Participant: 5c0a5d509e02299ab8d1cd42

What was your initial reaction? Confusion

What did you like most? Ability to toggle between chart types easily

What did you like least? Scales were unclear

Was there anything that particularly appealed to you? Visualisation didn't reflect the data, so it wasn't possible to make connections.

What was something interesting you discovered? N/A.

What did you learn while using the visualisation search? That I should be a little more openminded about whether visualising first might offer new ways of exploring my research question (even if it doesn't prove the most effective way).

What would you change about the visualisation search? N/A

Participant: 5c0a591c9e02299a914f656f

What was your initial reaction? It was pretty! But it was also a bit confusing. While I'm familiar with different kinds of charts, it wasn't immediately clear what purpose each chart type had and which ones would be best for which kinds of questions.

What did you like most? The colours and the easy-to-view information. It was very clear right away what you were looking at.

What did you like least? It was a bit confusing and I didn't know if it was the search or the system. For instance, I filtered by category and selected 3 specific categories (bread, wine, meat) and yet the charts often included things like observance, adjustments and domestic as well. I wanted to isolate just bread, wine, and meat but couldn't figure out how to.

Was there anything that particularly appealed to you? Despite my attempts to isolate the bread, wine, and meant keywords, other keywords and categories kept showing up on the chart, which suggests that the three original keywords were likely

used in combination with other words and perhaps couldn't be isolated out of context.

What was something interesting you discovered? When I did the keyword search, I noticed wine appeared more often than bread and meat. But the visualisations showed that while wine was mentioned more, meat and bread accounted for a larger percentage of costs - so my perception of wine being the biggest expense from the keyword search was wrong when I realised the visualisations were showing actual total cost.

What did you learn while using the visualisation search? Folks like wine, but not as much as I originally thought. The overall costs per year varied greatly and I imagine it had a lot to do with who was present during each respective year. I also thought I'd like the visualisation search more than I did - which is not to say I didn't like it, but I'm surprised I found it more confusing than the keyword search. I still like the colours though.

What would you change about the visualisation search? I would reiterate adding the ability to sort the keyword search results by year or month, and perhaps a brief tutorial or explanation of how to best maximise the visualisation search would be helpful.

Participant: 5c0a5fdc9e02299a92d16b52

What was your initial reaction? As I was unfamiliar with the archive I was not sure what to search for. When I finally found something that had a result, the visualisation search provided a clear indication of the variation of spending on wine over the course of 1774-1779 & 1781.

What did you like most? It was reasonably intuitive to use although there are still some bugs (which is understandable).

What did you like least? I would have liked to have conducted more comparative searches. i.e. I would have liked to have a line graph that compares the spending over the course of twelfth months compared with the different years.

Was there anything that particularly appealed to you? What I noticed from the visualisations were the spikes in wine purchases in some years. Upon further investigation sometimes this was easy to identify the reasons why: new students; greater wine rations etc etc

What was something interesting you discovered? That wine purchases cannot be directly linked to specific times of years & feast days . I also wanted time to look in more detail at abstinence days but I needed more time on that.

What did you learn while using the visualisation search? I am unsure of what this question means. If you mean what did I learn as a result of using the system about the potential in using digital resources and tools, I learned, I suppose, the value of using visual data as a historian; but I question the reasons behind the choice of visualisations. I appreciate that this is a work in progress and I think that this work is very valuable but I would love to be able to crunch more data and to visualise this data in different formats.

What would you change about the visualisation search? In terms of interface I think that the system is reasonably intuitive. What I would like to be able to do is to interrogate the sources (via the visualisation tool) in greater detail and with greater sophistication. This may not be possible, of course. If anything, what this shows is that the great work done here whets the appetite for what could be achieved.

Participant: 5c0a59b69e02299ab8d1cd40

What was your initial reaction? innovative way to engage with a close /distant view of the data

What did you like most? different ways to view/cluster the data

What did you like least? took me a while to get the hang of it

Was there anything that particularly appealed to you? not sure

What was something interesting you discovered? I need time to adjust my brain to different experiences/forms of engagement with data

What did you learn while using the visualisation search? my brain needs time to adjust to different modes of engagement

What would you change about the visualisation search? not sure

Participant: 5c0a5aa79e02299a92d16b50

What was your initial reaction? Initially very positive to the visualisation search as it allowed for more nuanced answers to research questions.

What did you like most? The ability to use categories and create a variety of visualisations

What did you like least? The lack of nuanced answers to search queries and issues regarding the number of words filter

Was there anything that particularly appealed to you? It was interesting how the mulitbar charts overestimated visually the density of the search parameters. The ability to change charts from this to a more nuanced line chart by occurrences in the text allowed for a more balanced visualisation.

What was something interesting you discovered? Yes, the ability to examine holidays and wine purchases.

What did you learn while using the visualisation search? I learned that although the basic search function was useful for a general overview, the type of visual search implemented allows for a greater level of detail to be ascertained.

What would you change about the visualisation search? I would add a list of stop-words which would prevent extraneous data being added to the overall searches of the texts.

Participant: 5c0a62329e02299a936b4940

What was your initial reaction? Interesting to see graphs of results for various inputs.

What did you like most? Being able to change how the result data looked in different categories and graphs.

What did you like least? I didn't understand why the results differed when searched by keyword vs category, for the same thing (eg. bread).

Was there anything that particularly appealed to you? The difference between results when searched by keyword or category.

What was something interesting you discovered? How the same data can be presented in different ways.

What did you learn while using the visualisation search? How to change my search terms to find different results.

What would you change about the visualisation search? An explanation of what the results show.

Participant: 5c0a617f9e02299ab8d1cd44

What was your initial reaction? Confusion. Too much information without being able to filter out the irrelevant stuff.

What did you like most? Once I realised I could get relatively comprehensible results on the details tab, that was helpful, but I still could not narrow down the search as much as I wanted.

What did you like least? The visualisations. They were confusing and overwhelming.

Was there anything that particularly appealed to you? The connections between points were difficult to grasp, I could not immediately see the relationships between elements in the various visualisations. Some sort of key or guide per visualisation might have helped.

What was something interesting you discovered? I learned that I really don't like visualisations. They are far less clear to me than a list of text.

What did you learn while using the visualisation search? That I'm a traditional sort that likes lists.

What would you change about the visualisation search? More information on what the graphs meant and how they displayed information. The end user experience should not be so confusing.

Participant: 5c0a8bf79e02299a92d16b5d

What was your initial reaction? I found it a little bit confusing to begin with.

What did you like most? The visualisation search made it very easy to see trends in the data over a given time period.

What did you like least? If data did not exist for a particular year this impacted on the overall appearance of the chart.

Was there anything that particularly appealed to you? I was interested in the amount of salt purchased at various times of the year and particularly on Meat days as opposed to Abstinence days. The level of salt used was related to the amount of bacon purchased (because it was used for curing bacon). I had not anticipated this connection.

What was something interesting you discovered? That there was a relationship between the purchase of salt and the purchase of bacon.

What did you learn while using the visualisation search? That data can be extracted from archival sources for analytical purposes. Also, how important it is to have a research question before starting this process.

What would you change about the visualisation search? No

Participant: 5c0a5a289e02299a936b493c

What was your initial reaction? I liked it. Colourful and pleasing on the eye. I opted for the bar chart (not multipart)

What did you like most? The presentation of results in a nice colourful visual with option to try out different chart/display types

What did you like least? Bit confused with some of the options on the left eg category vs word, or what (if any difference) there was between choosing say wine as a keyword and again in the category, or just choosing wine as keyword and leaving category unticked. Don't think there was an option to just pick two given years from all the options there.

Was there anything that particularly appealed to you? Thought it was interesting that there was no spend on meat in Feb and Mar 1774 which wasn't as obvious when I was searching this using the keyword search only. The visual search made this really stand out.

What was something interesting you discovered? Saving the charts created to the dashboard

What did you learn while using the visualisation search? saving to dashboard and saving charts

What would you change about the visualisation search? more clarity on the search labels on left eg categories etc

Participant: 5c0a59f59e02299a9087d388

What was your initial reaction? I wondered how to start navigating

What did you like most? The graphics are really clear and clean

What did you like least? I would prefer to end (present retrieved data) with visualisations rather then start (search) with visualisations

Was there anything that particularly appealed to you? The different spending from month to month within any year was striking when viewed with a line chart.

What was something interesting you discovered? I need to be more focused in my initial queries

What did you learn while using the visualisation search? I learned about diets in Alcala

What would you change about the visualisation search? Perhaps a spanish-english dictionary

Participant: 5c0a5ae19e02299a9ecf7d96

What was your initial reaction? confused initially

What did you like most? the choice in data presentation options, the graphs were cool, especially the option to display monthly and then yearly values

What did you like least? I found it difficult to compare the different sets of graphic results and would have liked a function that allowed me to compare them

Was there anything that particularly appealed to you? Yes. i was interested in bread and was surprised by the decline in expenditure over the time period and also by variation in expenditure during specific years. There did not seem to be a regular pattern.

What was something interesting you discovered? That yearly and monthly visualizations give a more rounded overall view of consumption

What did you learn while using the visualisation search? this is an excellent way to access the data in a structured environment and yields new results

What would you change about the visualisation search? I would have like more information on the processing of the data from source to VRE

Participant: 5c0a61ff9e02299a9087d38c

What was your initial reaction? A little more flexiblity in searches than keyword search

What did you like most? Added flexibility of searches

What did you like least? E.g. results of search for total amount spent on wine was different depending on whether the search was conducted by keyword or category. Keyword seems to include mentions of wine in other categories which probably explains the different results but it engenders uncertainty or lack of confidence in the results

Was there anything that particularly appealed to you? I liked the fact that charts could be grouped or stacked

What was something interesting you discovered? To be aware that keyword searches will return results from other categories

What did you learn while using the visualisation search? To be aware that keyword and category searches can return different results.

What would you change about the visualisation search? More flexibility in searches

Participant: 5c0a5ee69e02299a8f45ce02

What was your initial reaction? My initial reaction was that I couldn't figure out how to use it. There were check boxes and fields that were greyed out and I couldn't get them to work as intended without trial and error.

What did you like most? It was a swift way to produce graphs and other data objects. I liked being able to switch between the graphs and the text references on the fly, and that (some) of the graphs could easily be swapped between modes.

What did you like least? Items and categories could not be removed from the graph. For example, every instance of "Meat" also created a category "Food" and

there was no way to exclude this second category from showing in addition and duplicating the result. I could not zoom in on the graph to compare smaller elements in detail, which was a problem because the "food" total was inflating the Y axis.

Was there anything that particularly appealed to you? I was interested to see that there were huge jumps in meat expenditure in 1775 and 1776; drilling down I could see that these expenses were for the bacon remainder. I wonder why they were only catalogued for those two years...?

What was something interesting you discovered? Expenses for food itself and expenses for labour were categorized together. I wonder if totals would be different otherwise? Also, It was interesting to see differences in fasting days... I didn't even realize fasting days were held with such regularity.

What did you learn while using the visualisation search? The specific expense for wine (as opposed to for other liquids) was interesting... was it just water, wine, meat, and bread that they drank? Also, I was interested to see that bookkeeping practices varied greatly from year to year.

What would you change about the visualisation search? I would like the ability to modify graphs after they are made, to exclude certain categories that are duplicating data, etc. I would also like the ability to reload saved graphs from the dashboard and modify their search parameters.

Participant: 5c0a6aa79e02299a9087d394

What was your initial reaction? Didn't know what kind of items you can search on

What did you like most? graphs

What did you like least? Not obvious what the data contained so what you could search on.

Was there anything that particularly appealed to you? No

What was something interesting you discovered? 18th century Spanish weights and measures

What did you learn while using the visualisation search? Nothing really apart from some minor facts about Alcala college life

What would you change about the visualisation search? No, can't think of anything.

Participant: 5c0a5a749e02299a914f6571

What was your initial reaction? Interesting approach but would prefer a more traditional keyword based entry point for which the results could then be expanded on using the visualisation

What did you like most? Easy graph generation, boundary object creation

What did you like least? Having to log in for every source viewed via the interface. Disconnect between chart labels and details i.e. clicking July bar opens June

source - model issue? Search interface non-obvious - hard to determine what categories vs. keywords is doing. Didn't figure out what word count options were doing.

Was there anything that particularly appealed to you? Easy to confirm expected answers e.g. no meat in Lent. Graphical presentation makes is easier to quickly identify trend changes.

What was something interesting you discovered? Not really - would need more time

What did you learn while using the visualisation search? Examining financial records faster with easy graphical interface

What would you change about the visualisation search? More information about function of different search parameters at start. Graphs would be easier to interpret if colours remained consistent for catagories.

D.10.3 Final Response Transcripts

Participant: 5c0a591c9e02299a914f656f

What was your preferred method? Keyword

Why did you choose this method? The keyword search brought me straight to the documents containing the keywords I was looking for. The visualisation search was useful for visualising the data and comparing categories and years, but it was less useful in terms of locating specific keywords in specific documents.

Is there a scenario where you would place importance of one type of search over another? The visualisations are better for comparison purposes, I think. It was more difficult to compare costs per year or costs across a specific year when using the keyword search. On the flip side, if you're looking for specific documents or mentions of a specific keyword, then the keyword search would be more useful. I think both types of search work well for individual needs and can complement each other.

Participant: 5c0a59609e02299a92d16b4e

What was your preferred method? Keyword

Why did you choose this method? The kind of information I generally look for will be textual, especially in the early stages of using a resource where I may not be sure what it is that I am actually looking for. Knowing that information is there is necessary before I start to quantify it or view it in more macroscopic terms as the visual search allows for.

Is there a scenario where you would place importance of one type of search over another? In an instance in which one were investigating how expenditure changes according to feast days, the visual search would be better

Participant: 5c0a59b69e02299ab8d1cd40

What was your preferred method? Keyword

Why did you choose this method? keyword is the choice of habit, but I am sure once I became accustomed to using visual search options, I would use both

Is there a scenario where you would place importance of one type of search over another? not sure

Participant: 5c0a59f59e02299a9087d388

What was your preferred method? Keyword

Why did you choose this method? It seems more targeted (more of a search), while visualisation seems more of a browsing activity.

Is there a scenario where you would place importance of one type of search over another? It may depend on personal preference but I think that familiarity with the data may be a factor (I would prefer keyword searching if I was more familiar with the data).

Participant: 5c0a5a289e02299a936b493c

What was your preferred method? Visualisation

Why did you choose this method? It was just easier to see a snapshots of the results for the type of research I was doing today

Is there a scenario where you would place importance of one type of search over another? The keyword search prompts you to look at the source document and its translation which from a historical or humanities point of view may be preferable. The visualisation search is useful to get a quick overview or snapshot of the information. I don't think the two would be mutually exclusive. There are instances where either one could be preferable and where a mix of the two is best.

Participant: 5c0a5a749e02299a914f6571

What was your preferred method? Keyword

Why did you choose this method? Keyword search gives me more opportunity to evaluate records myself

Is there a scenario where you would place importance of one type of search over another? Keyword search better for less organized records e.g. letters

Participant: 5c0a5aa79e02299a92d16b50

What was your preferred method? Visualisation

Why did you choose this method? Although both approaches are useful, the visualisation approach with its added filters allows for more nuanced questions to be posed to the dataset.

Is there a scenario where you would place importance of one type of search over another? In the context of very specific types of textual data, the visual approach would denature the original structure of the information, which would make the visuals not fit for purpose. An example of this would be highly specific military data.

Participant: 5c0a5ae19e02299a9ecf7d96

What was your preferred method? Visualisation

Why did you choose this method? for a journal article of project I'd like to present my results, the visualisation permits this more efficiently and quickly.

Is there a scenario where you would place importance of one type of search over another? the visualisation is best for presenting the sum of the information gathered, the searchword for granularity.

Participant: 5c0a5d509e02299ab8d1cd42

What was your preferred method? Keyword

Why did you choose this method? Aside from problems with data numbers, familiarity with keyword searching is a powerful influencer. I would naturally attempt to explore the data through the texts before visualising it and would often (somewhat narrowmindedly, I admit) consider visualisation a tool for explaining connections I had made, rather than generating new ones.

Is there a scenario where you would place importance of one type of search over another? N/A.

Participant: 5c0a5ee69e02299a8f45ce02

What was your preferred method? Visualisation

Why did you choose this method? The visualization gives a better high-level view (though it has huge issues), though if I were doing an article focusing on something in detail, it would not be sufficient.

Is there a scenario where you would place importance of one type of search over another? Visualisation is good for the high-level view of the data, the keyword search is ideal for detail.

Participant: 5c0a5f259e02299a914f6574

What was your preferred method? Keyword

Why did you choose this method? There is little difference. I had to choose.

Is there a scenario where you would place importance of one type of search over another? NA

Participant: 5c0a5fdc9e02299a92d16b52

What was your preferred method? Visualisation

Why did you choose this method? I think it is important that historians represent their data in a number of ways. We are light on visualisations (unless we are economic or art historians) and as a result I relished the opportunity to attack this data with the visualisation tool. This is not to say, however, that I think the keyword search is an inferior tool - rather a different and important tool.

Is there a scenario where you would place importance of one type of search over another? I think both are equally useful in different contexts.

Participant: 5c0a617f9e02299ab8d1cd44

What was your preferred method? Keyword

Why did you choose this method? The keyword search was clear, the visualisation search was overwhelming and lacking the same functionality to narrow down the search I wanted. I can see how it would be useful for a different goal, but visualisations are not part of my research goals in general.

Is there a scenario where you would place importance of one type of search over another? Depends entirely on the person searching and the purpose of the search, I would've thought.

Participant: 5c0a61ff9e02299a9087d38c

What was your preferred method? Visualisation

Why did you choose this method? I picked visualisation because it was a little more flexible in its search options

Is there a scenario where you would place importance of one type of search over another? Visualisation illustrates contrasts in results more readily

Participant: 5c0a62329e02299a936b4940

What was your preferred method? Keyword

Why did you choose this method? I would be able to understand and explain the results from the keyword search.

Is there a scenario where you would place importance of one type of search over another? When looking at broad data for general information, the visualisation search presents results in a really interesting way.

Participant: 5c0a62599e02299a8f45ce04

What was your preferred method? Visualisation

Why did you choose this method? Being able to browse easier in the visualisation search and having data displayed very clearly in graphs etc. really helped my understanding of the materials.

Is there a scenario where you would place importance of one type of search over another? Visualisation search would be better for analysing data over time, i.e. the years covered in the manuscripts.

Participant: 5c0a63559e02299a9087d38e

What was your preferred method? Keyword

Why did you choose this method? It's really to do with the nature of the record; see my answers above.

Is there a scenario where you would place importance of one type of search over another? Again, see my answers above.

Participant: 5c0a6aa79e02299a9087d394

What was your preferred method? Visualisation

Why did you choose this method? Visualisation really includes the keyword search anyway so you get extra ways to explore data

Is there a scenario where you would place importance of one type of search over another? Visualisation is based on quantification of data. Where this is important.

Participant: 5c0a8bf79e02299a92d16b5d

What was your preferred method? Keyword

Why did you choose this method? I found the results of the keyword search to be more reliable.

Is there a scenario where you would place importance of one type of search over another? The visualisation search enables you to view trends over time. However, I would find this data to be less reliable than the results from the keyword search.

Participant: 5c793b7b9e02291bb250e588

What was your preferred method? Keyword

Why did you choose this method? keyword because it gives more explanation with each record and this was more easy to understand for a lay person like me who did not have any previous knowledge about it. However I believe that Visualization technique can be more useful if there is just more detail available in tittle or sub tittle along with the graph to tell user more about the data which is being displayed

Is there a scenario where you would place importance of one type of search over another? yeah when I searched for bread keyword I could see better detailed result in keyword search as compared to the visualization one as I can read each of the record separately with a brief of each record.

Participant: 5c793bdf9e02291bb384f4d7

What was your preferred method? Visualisation

Why did you choose this method? I prefer to find some data which is organized.

Is there a scenario where you would place importance of one type of search over another? Some data might less and it must not appear on the Visualisation search.

Participant: 5c793c189e02291ec95d87f0

What was your preferred method? Visualisation

Why did you choose this method? Because it displays the appropriate search results

Is there a scenario where you would place importance of one type of search over another? I think Visualisation is better because it displays the content in the brief manner and anyone can understand it in less time

Participant: 5c793c629e02291bb1295de8

What was your preferred method? Visualisation

Why did you choose this method? Visualization search helped in visualizing the trends in much better way. It was easy to find whether the trend was increasing or decreasing and what was the exact value spent on a particular year.

Is there a scenario where you would place importance of one type of search over another? I guess in all types of scenarios visualization search was better than keyword search.

Participant: 5c793c999e02291bb0072d47

What was your preferred method? Visualisation

Why did you choose this method? visual search helps to understand information very quickly and can be more effective then reading

Is there a scenario where you would place importance of one type of search over another? yes any trend scenario can be better done with the help of visual search as it is more efficient as well as effective in results.

Participant: 5c793d909e02291ec95d87f2

What was your preferred method? Visualisation **Why did you choose this method?** it was more comprehensive.

Is there a scenario where you would place importance of one type of search over another? none.

Participant: 5c793dc19e02291bb1295dea

What was your preferred method? Visualisation

Why did you choose this method? at a time i can see different data. In search i might miss some of the type that are included in the system

Is there a scenario where you would place importance of one type of search over another? after reading a paper when you click back button the keyword search again asks for the keyword where as the visual search maintains the same state

Participant: 5c793dfa9e02291bb0072d49

What was your preferred method? Visualisation

Why did you choose this method? It is easy to understand and make notes using visual search and it also has many filters to apply .

Is there a scenario where you would place importance of one type of search over another? In terms of comparing ,visual search made it easy for me to find conclusions.

Participant: 5c793ec19e02291baf2ff332

What was your preferred method? Visualisation

Why did you choose this method? It is much easier to get an idea about what is going on with the visualization tool.

Is there a scenario where you would place importance of one type of search over another? When you are simply looking at money spent through a visualization chart it us much easier instead of just looking at the numbers.

Participant: 5c793f8e9e02291bb1295dec

What was your preferred method? Keyword

Why did you choose this method? The initial step for any research would be first through the keyword search followed by visualization.

Is there a scenario where you would place importance of one type of search over another? Key word search is the prime base to gather information about the data.

Participant: 5c7940829e02291baf2ff334

What was your preferred method? Keyword

Why did you choose this method? When looking for a particular point of information, it was returned by keyword search much more directly.

Is there a scenario where you would place importance of one type of search over another? If I were providing an overall breakdown of data, visualization is easier. However, when trying to identify a particular discrepancy, keyword search is much easier.

Participant: 5c7941a29e02291bb1295dee

What was your preferred method? Visualisation

Why did you choose this method? Felt more intuative having that much information displayed in a wall of results is hard to decipher, while having it displayed graphically is easier to use

Is there a scenario where you would place importance of one type of search over another? The visualisation search, as you can see what it returns with a time period involved so you get a better idea how important your item is straight away, rather than scrolling through search results and checking the dates

Participant: 5c7954449e02291bb250e590

What was your preferred method? Visualisation

Why did you choose this method? Users have the flexibility for custom search. Visualization help understand any piece of information in a quick and effective way. It conveys a lot more than a normal search result with a log. Visualization search helps see a holistic view.

Is there a scenario where you would place importance of one type of search over another? When we want to search total spent for Meat upon a large span of time, Visualization really helps.

Participant: 5c79553a9e02291bb384f4df

What was your preferred method? Visualisation

Why did you choose this method? My undergraduate major is mathematics, and I believe that anything with data will be more convincing. Therefore, according to the data to carry out the corresponding index can guarantee the right medicine, instead of what I want to search for.

Is there a scenario where you would place importance of one type of search over another? Keyword indexing is more advantageous than chart indexing when there is less information you know. Because of the keyword, we can list a large segment of data, we can find all the keywords we need and then index the chart.

Participant: 5c7955da9e02291bb1295df0

What was your preferred method? Visualisation

Why did you choose this method? Although visualization searches has some brush ups required it would be more convenient to analyze data using that rather than normal search. Once the user is accustomed with visualization UI

Is there a scenario where you would place importance of one type of search over another? Visualization helps sort data. Basic search can help more if re-arranging data options were available.

Participant: 5c7d378f9e022938b24bc147

What was your preferred method? Visualisation

Why did you choose this method? I feel as though certain unexpected statistics about the past can be better seen using pie charts and multibar charts.

Is there a scenario where you would place importance of one type of search over another? perhaps a regular expression-based search would be helpful insofaras for matching patterns in these documents

Appendix E

Figures

Appendix E is meant to provide larger scale versions of certain figures in the thesis that required additional sizing in order to adequately view the image. Not all figures in the thesis will be referenced here: only those figures that are difficult to see fine details at a smaller scale embedded in the body of the work. Each figure will reference it's original figure number in the caption.

E.1 Chapter 1

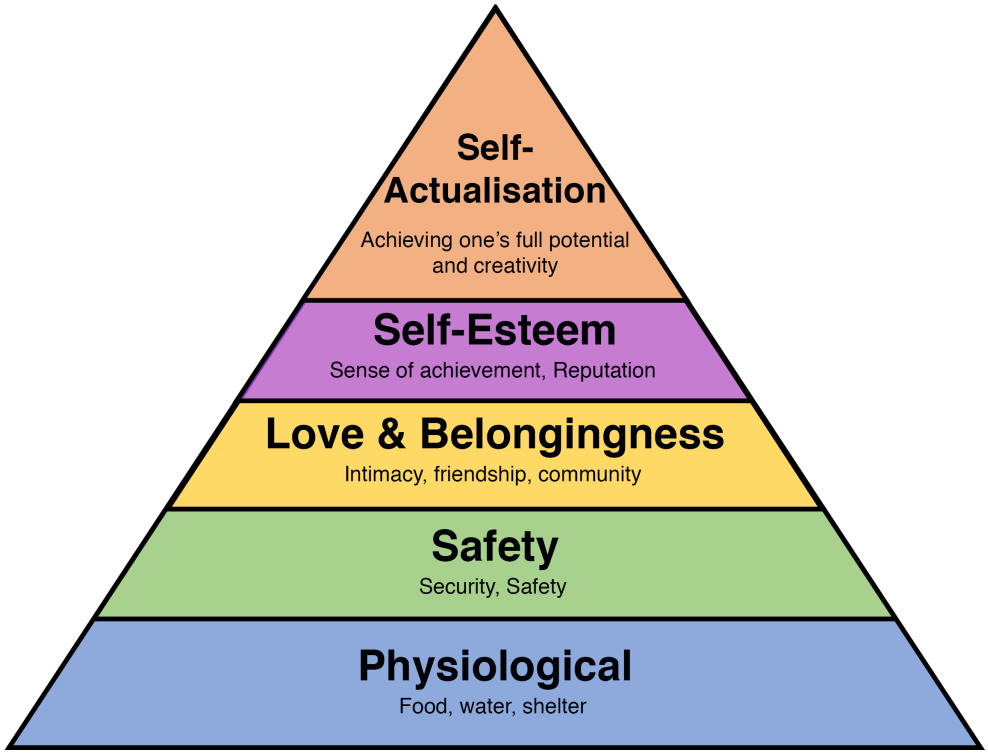


FIGURE E.1: Recreation of Maslow's Hierarchy of Needs as seen in Figure 1.9

E.2 Chapter 2

FIGURE E.2: Map of murders in London between 1674 and 1819 as visualised in *Locating London's Past* [182] as seen in Figure 2.2

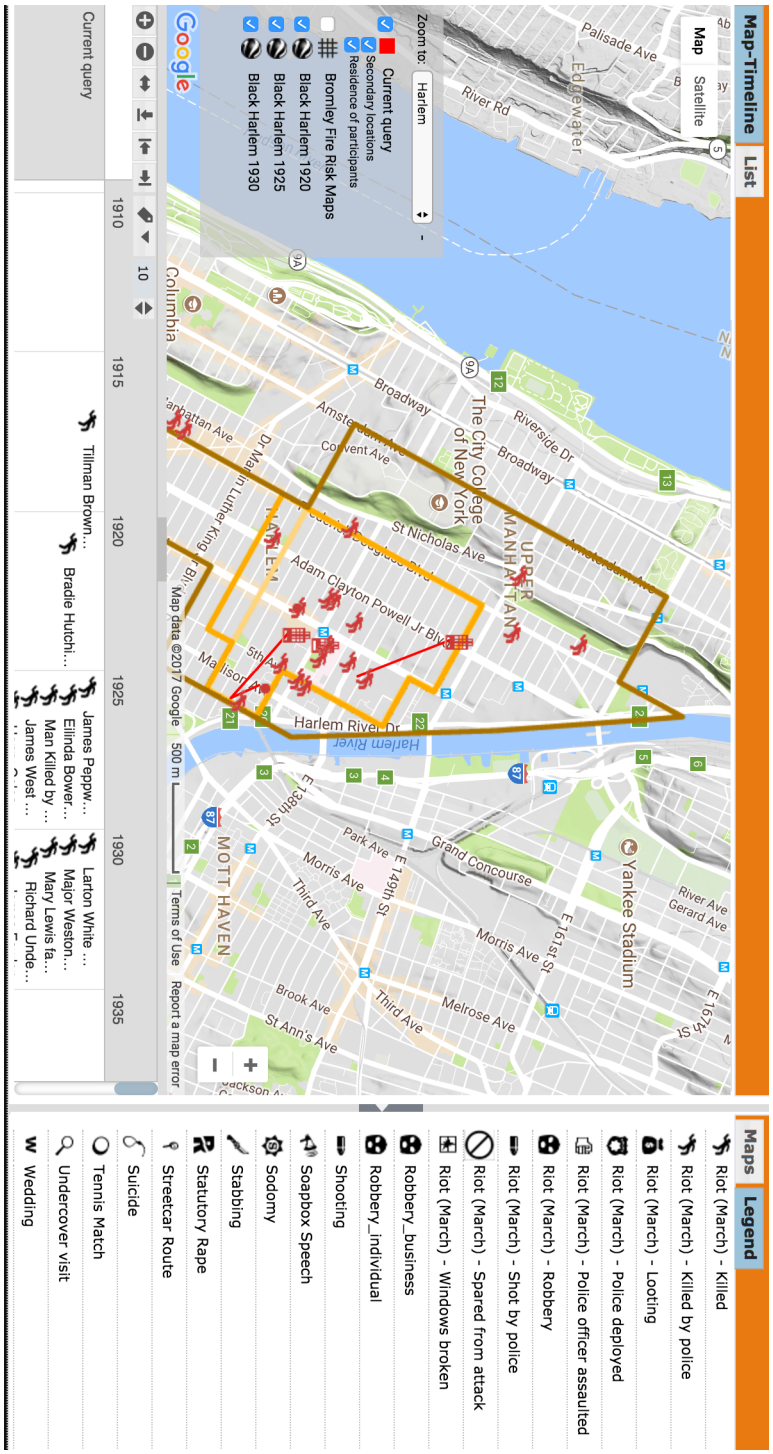


FIGURE E.3: Map of murders in Harlem between 1915 and 1930 as visualised in *Digital Harlem* [141] as seen in Figure E.3

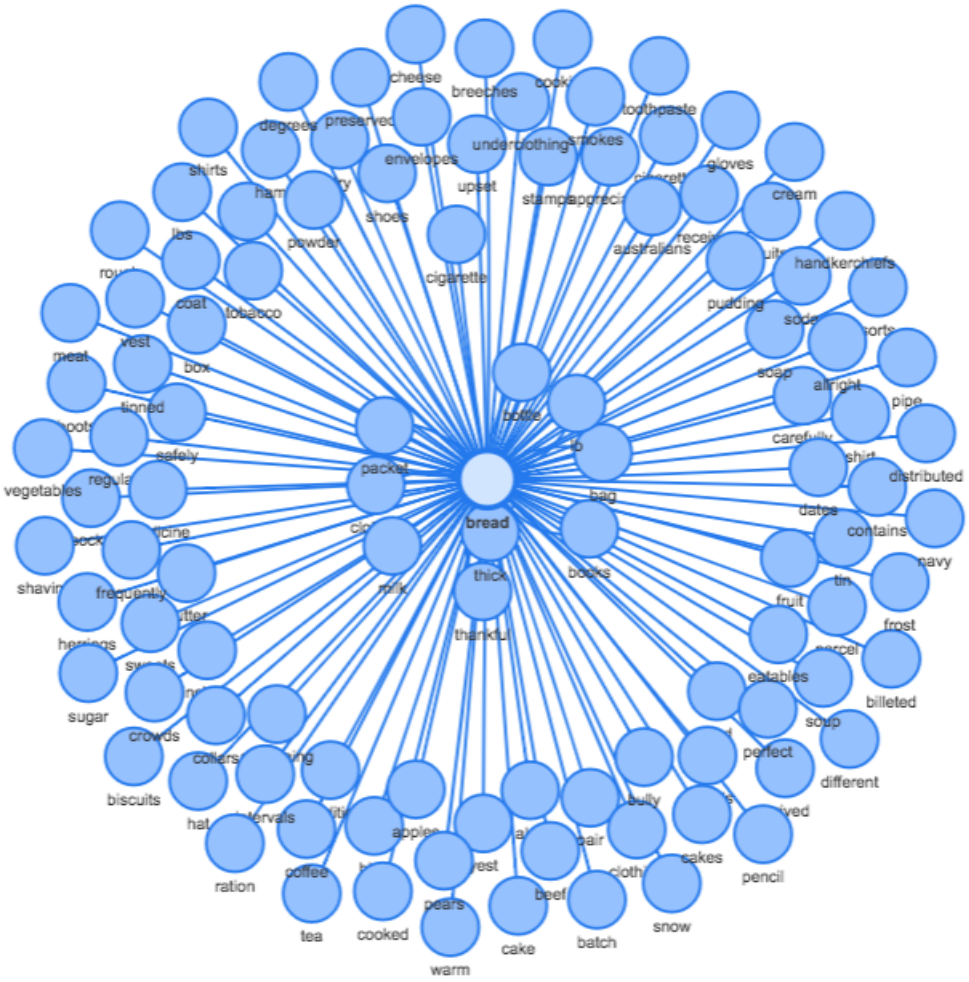


FIGURE E.4: Use of proximity in a data visualisation in the *Letters of 1916* which implies unintended relationships between keywords [192] as seen in Figure 2.7

FIGURE E.5: List of letters by category as seen on the original *Letters of 1916* website [192] as seen in Figure 2.10

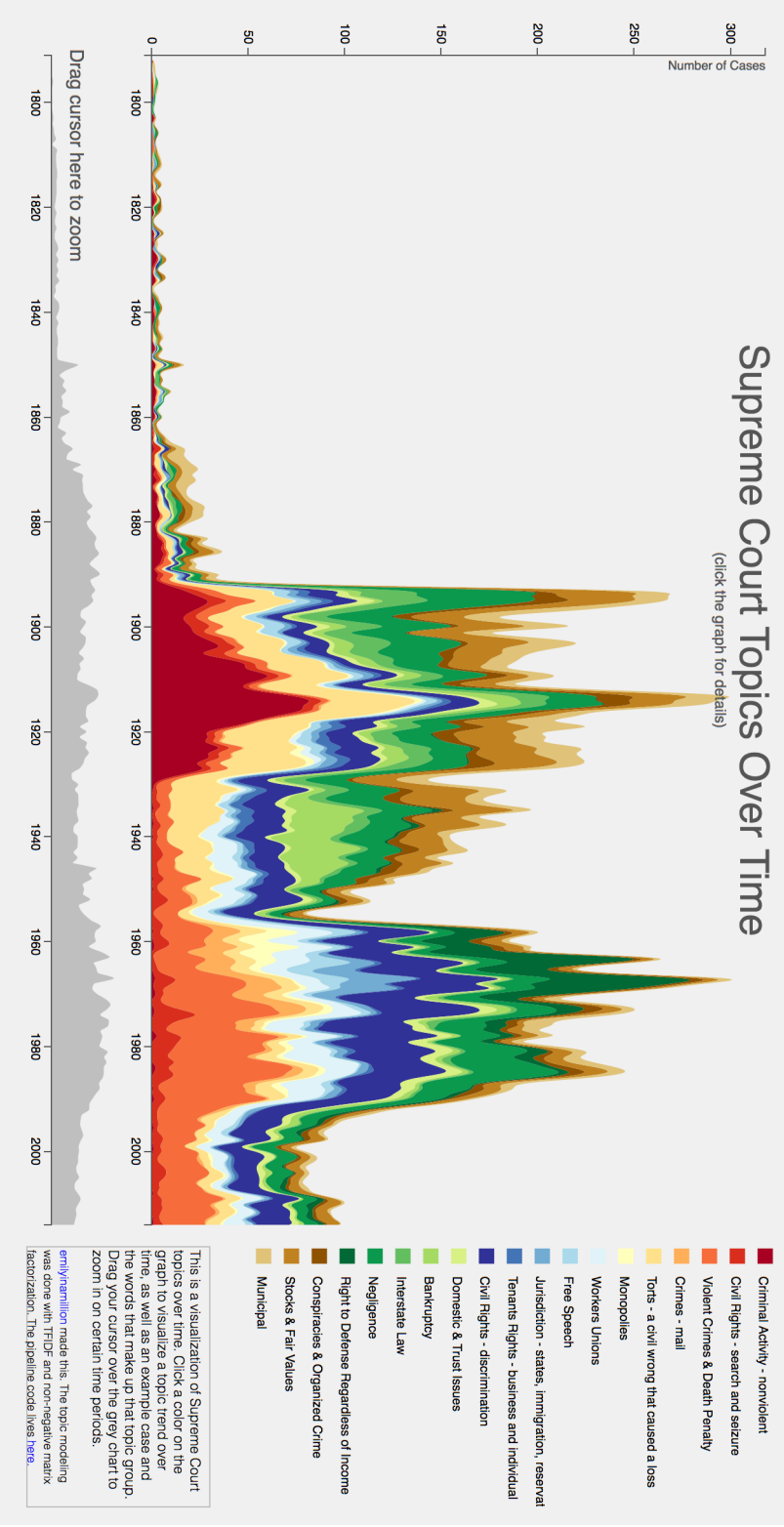


FIGURE E.6: Topic model leveraging Non-negative Matrix Factorization showing the distribution of topics in U.S. Supreme Court cases over time [204] as seen in Figure 2.19

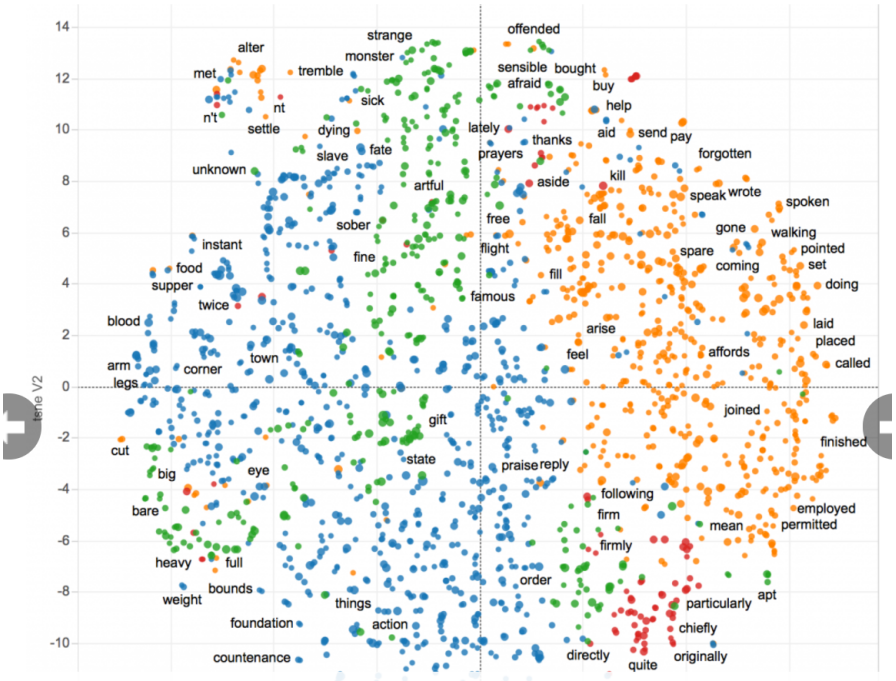


FIGURE E.7: Use of Word Vector Analysis to show syntatic relationships between words [206] as seen in Figure 2.20

E.3 Chapter 3

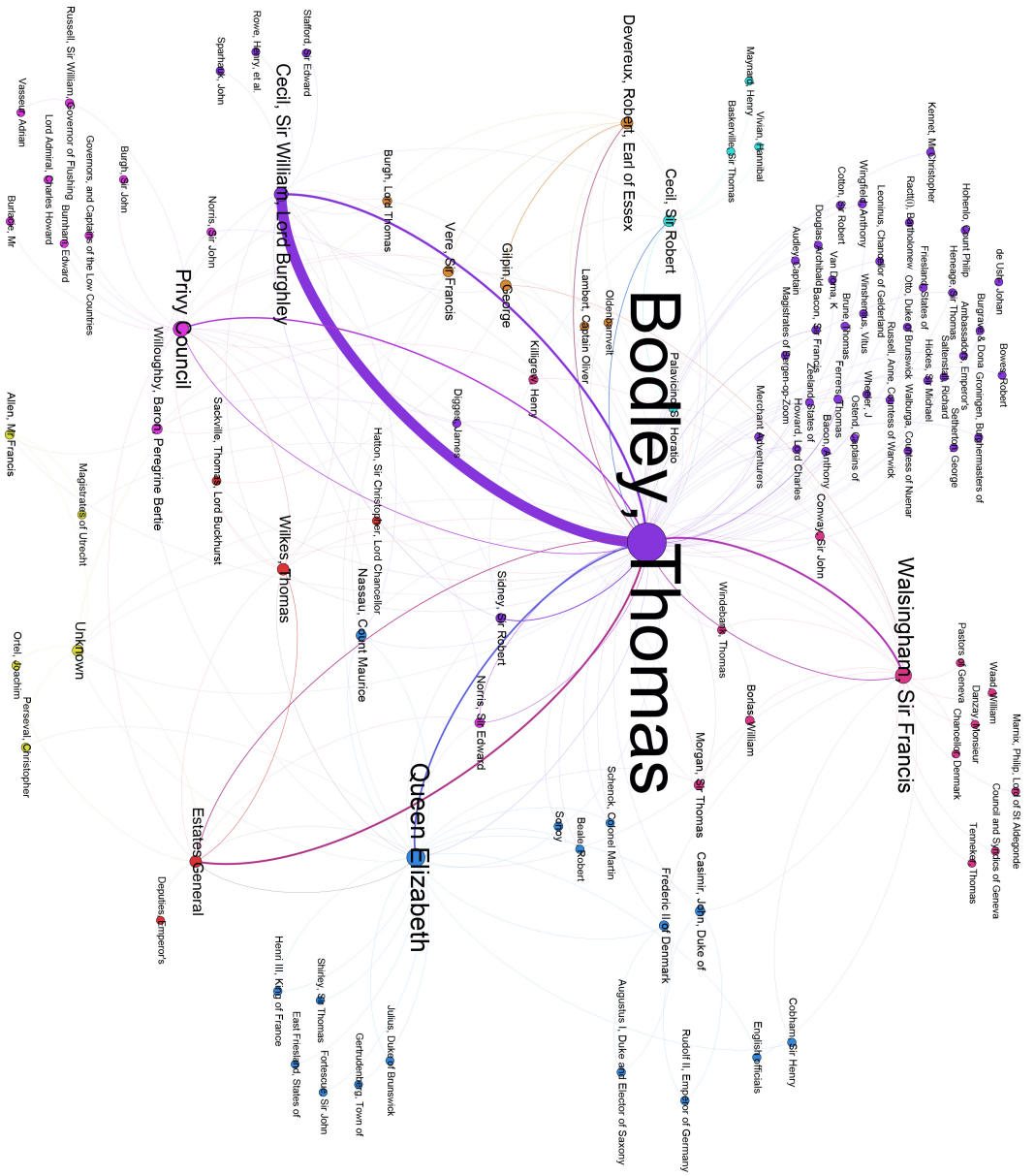


FIGURE E.8: Network graph of people mentioned in *The Diplomatic Correspondence of Thomas Bodley* [199] as seen in Figure 3.7

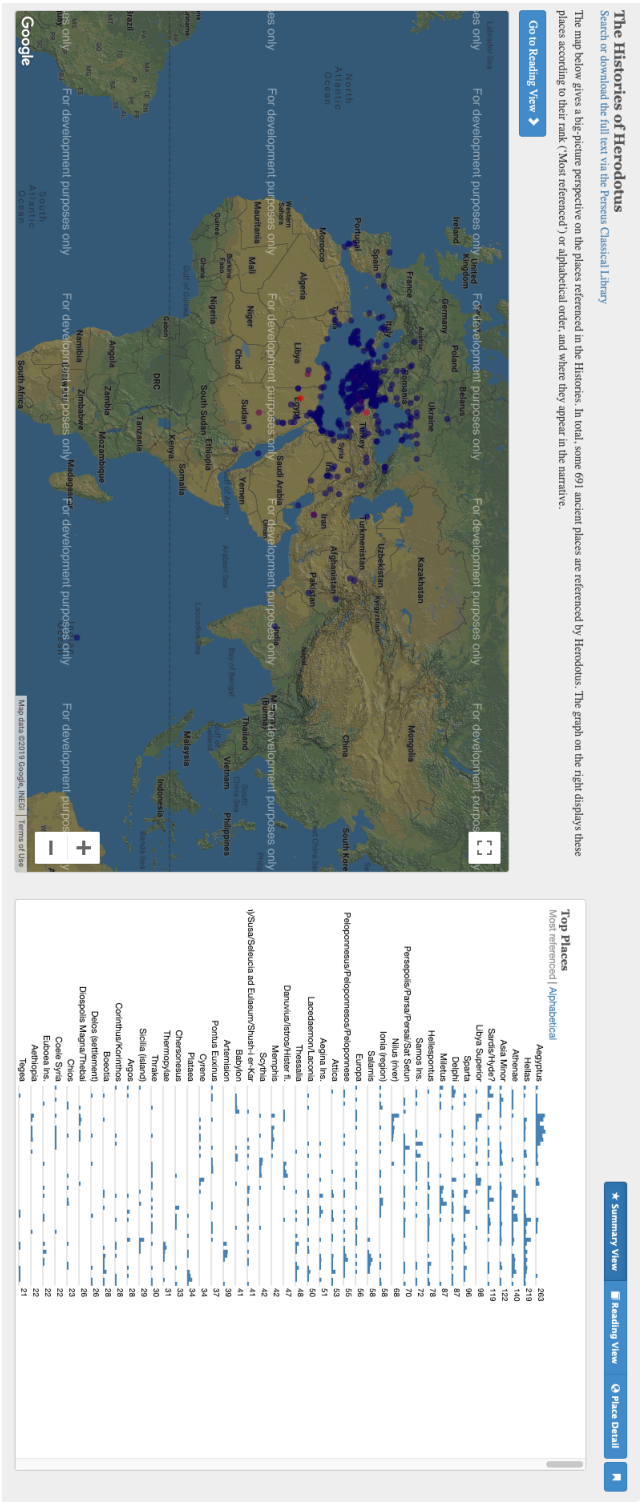


FIGURE E.9: Interactive visualisation of places in the *Hestia* project [278] as seen in Figure 3.8

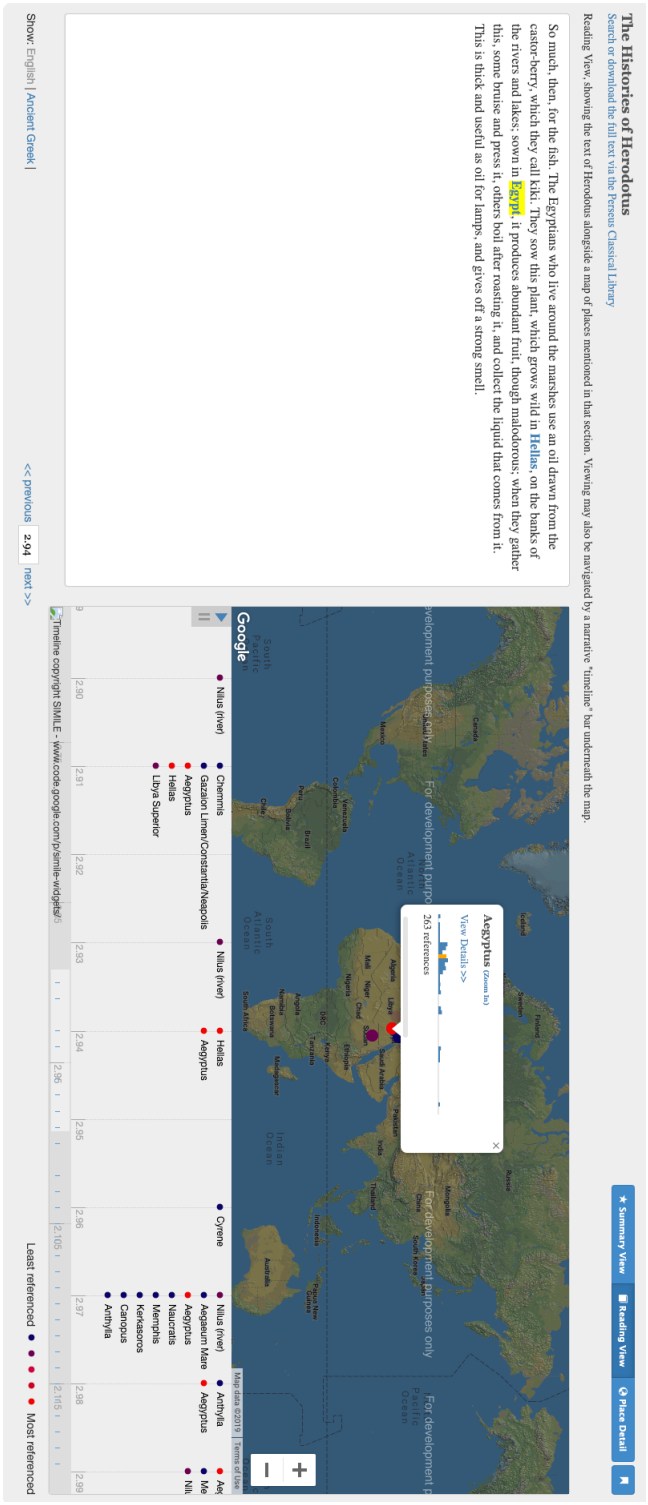


FIGURE E.10: Viewing the underlying data associated with a visualisation in the *Hestia* project [278] as seen in Figure 3.9

E.4 Chapter 4

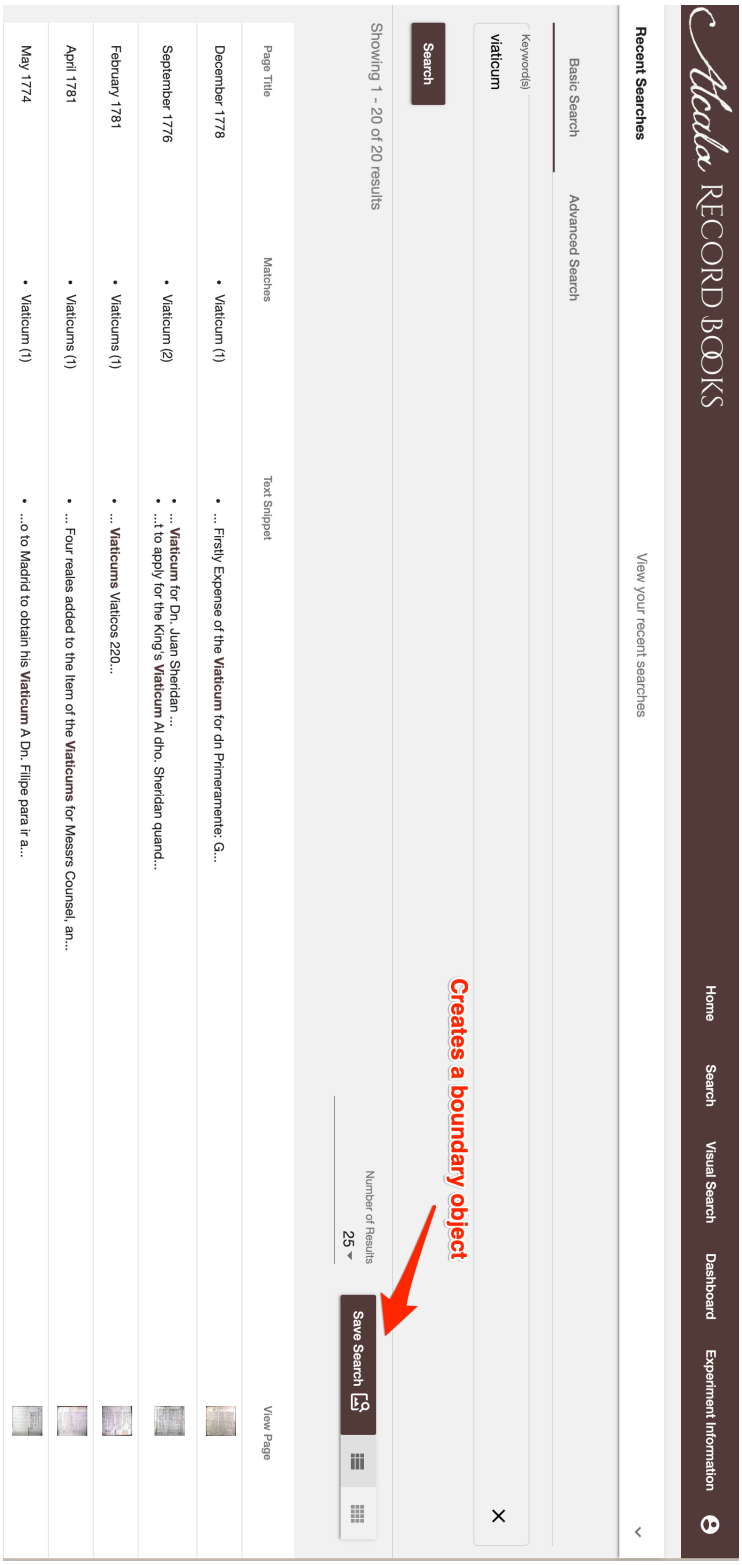


FIGURE E.11: Annotated screen grab highlighting the ability to create a boundary object from a keyword search as seen in Figure 4.6

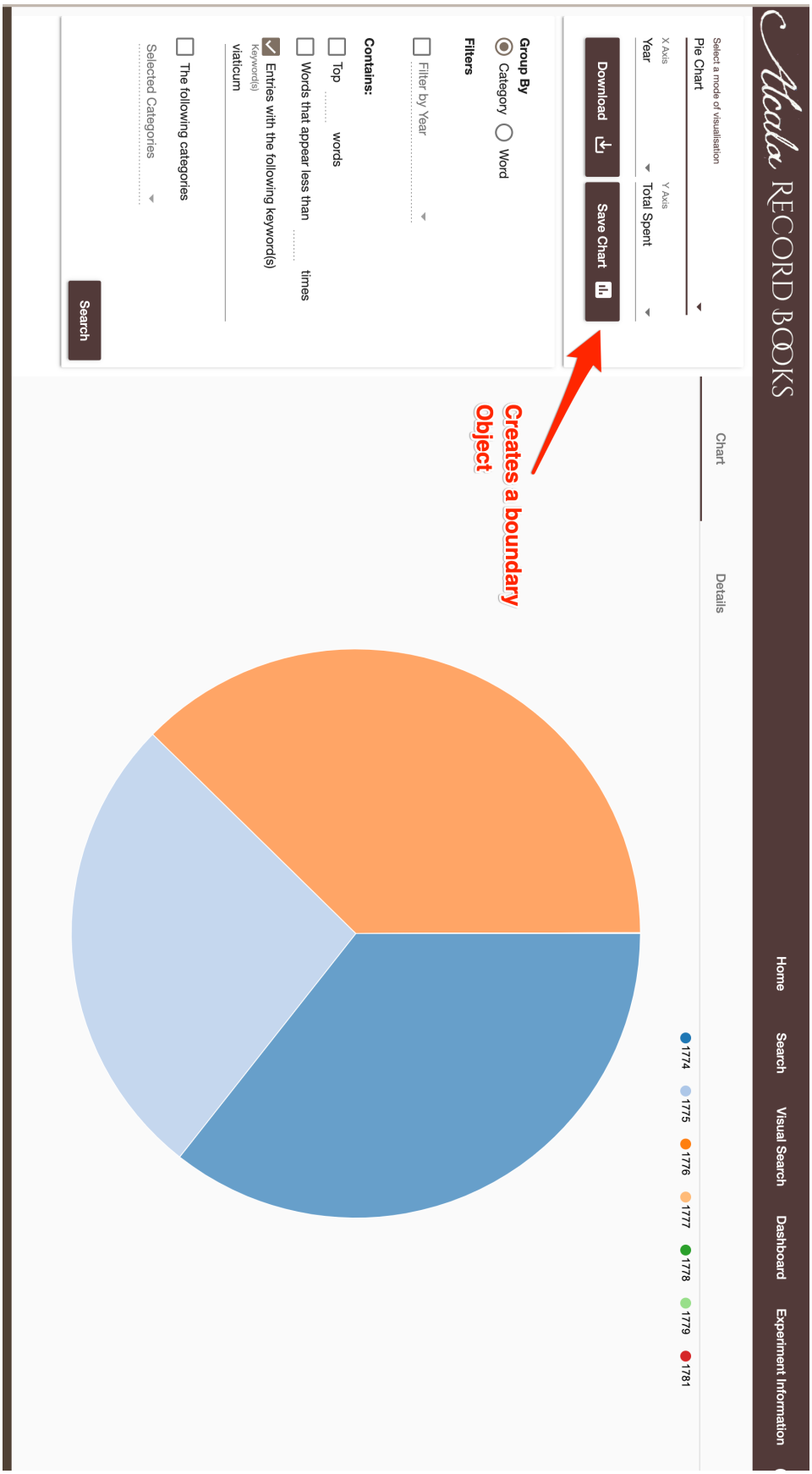


FIGURE E.12: Annotated screen grab highlighting the ability to create a boundary object from a visualisation search as seen in Figure 4.7

FIGURE E.13: Annotated screen grab highlighting the ability to create a boundary object from a page view as seen in Figure 4.8

FIGURE E.14: Screen grab from the Alcalá Record Books annotated to highlight the options for changing the x and y axis of the chart as seen in Figure 4.10

E.5 Chapter 5

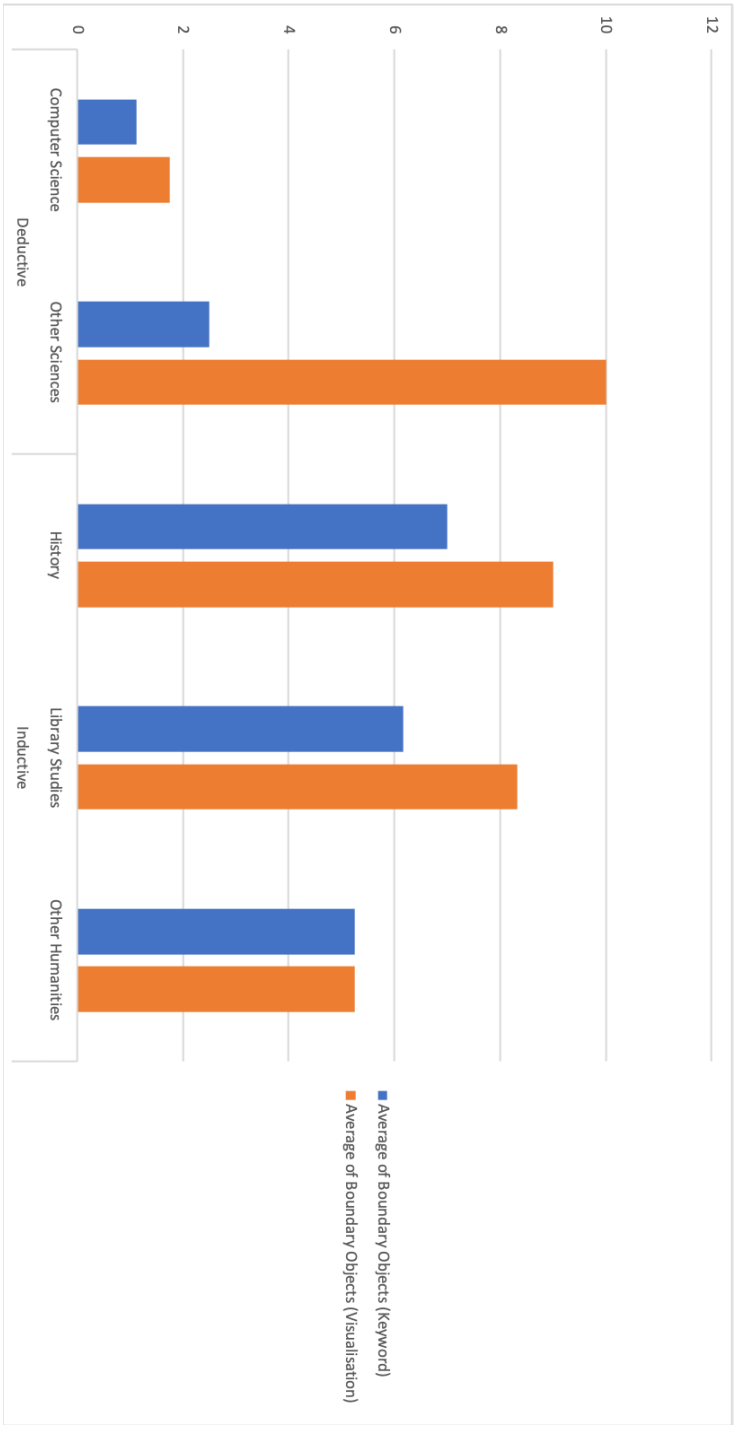


FIGURE E.15: Graph showing the change in the average number of boundary objects created during each search type across research area and training type as seen in Figure 5.14

E.6 Chapter 6

No figures to show for this chapter.

Glossary

activity theory Proposed by Alexei Leont'ev in the late 1970s, Leont'ev theorised that individuals occasionally participated in actions that did not themselves satisfy a need, but would instead lead to the eventual satisfaction of a need. He categorised activities as those acts which *immediately* satisfy a need and actions as those acts which satisfy an *eventual* need. Thus activities can consist of one or more actions. 21–23, 28, 29, 75

attention The ability of the cognitive system to maintain focus on a particular task or tasks. 44–47, 49, 59, 85, 86

behaviourism One of the theories of learning, which stipulates that learning is an outgrowth of behaviour, and it is therefore the responsibility of the teacher to construct situations that reinforce proper behaviour in order for the student to learn. 13–15

boundary object "[Artefacts] that aim to bridge concurrent cognitive models through abstraction from all domains of partners" [73, p. 396]. These objects are used to represent constructed knowledge created by the user as a result of her interaction with the DRE.. xvi, xvii, xix, 22, 23, 95, 113, 114, 116, 117, 119, 121, 124, 126, 130, 132–135, 137–139, 146, 147, 154, 157, 161, 383

central executive The subsystem in Baddeley's model of working memory that serves as a central processor for the various other subsystems, manages attention, and communicates with the long-term store to commit and retrieve information. xiii, 5–12, 44, 45, 70, 87, 156–158

close reading A method of consuming text that involves a meticulous analysis of the text, situated within its whole, and understood within a broad social or cultural context. It often involves the analysis of plot, theme, and genre. 35–38, 40, 41, 49–52, 62, 64–67, 70, 71

cognitive overload Occurs when the reader is presented with too much information and their capacity to process and comprehend is overloaded. 43, 45, 49, 57, 68, 71

cognitivism One of the theories of learning, which stipulates that learning is a result of cognitive structures in the brain and focuses on systems of attention and working memory. Information is often reduced to *chunks* in order to facilitate

working memory processing and builds upon a student's prior knowledge. Like behaviourism, cognitivism maintains that knowledge is absolute. 13–16

constructivism One of the theories of learning, which builds upon the ideas of cognitivism but states that knowledge is *not* absolute and instead is filtered through the lens of experience. 13, 16, 28

continuous reading A type of reading that involves reading from beginning to end. It is often juxtaposed against other discontinuous methods of reading such as hyperreading. 34, 35, 41

digital reading A method of reading that takes place in a digital environment and can employ other non-traditional reading approaches besides that of close reading. 41, 45

discontinuous reading A method of reading text that does *not* involve reading from start to finish. This type of reading is most common now in hypertexts but is also engaged in traditional print texts as well. 34, 35, 41, 68, 71

distant reading An approach to reading proposed by Moretti that leverages data visualisations in order to visualise various nodes and connections present within a text. 35, 36, 44, 49–52, 64, 70, 71

engagement An extension of immersion, describing the mechanism through which the user experiences immersion. 84–86

episodic buffer The subsystem in Baddeley's model of working memory that acts as a temporary store for information combined from the visuospatial sketchpad and the phonological loop. 6–11, 44, 45, 156–158

gestalt principle Also known as the Gestalt Principles of Visual Perception, these 7 principles describe how humans interpret visual information. 52, 61, 62, 64–66, 70, 71

Gestalt Principles of Visual Perception see gestalt principles. 52, 71

hyperreading The process of reading hypertexts, which often involves moving from linked topic to linked topic in discontinuous fashion. 35, 40–43, 49, 68, 71

hypertext Text that leverages markup (typically HTML) in order to piece together related content. 34, 35, 40–45

immersion The subjective experience of the user within a software ecosystem wherein she becomes absorbed in a set of tasks and activities that ultimately lead to the satisfactory completion of a goal. 82–89, 96, 97

inner scribe The component of the visuospatial sketchpad that acts as a process for rehearsal, alteration, and manipulation of visual information stored in the visual cache. 7, 10

Interaction Design The study of how users interact with systems and the methods used to facilitate and improve this process. 82

Keyword In Context A type of concordance line, a KWIC module will provide a snippet of text surrounding a particular keyword in order to show the relevance of the keyword in the search result as it relates to the text of the object returned in the result.. 121

long-term memory The aspect of memory where data is stored over long periods of time. Unlike other aspects of the memory system, long-term memory is seen as nearly infinite in its storage capacity. 9, 45, 89, 118, 156, 158

lucene index A type of full-text search implemented by the Lucene java library. The primary goal of this type of index is to increase performance when attempting to search across large sections of text (such as books or manuscripts).. 121

macroanalysis Proposed by Matthew Jockers as a different interpretation of distant reading, macroanalysis argues for a large scale analysis of the text (as opposed to a "reading" of the text). 50, 52

network analysis A type of textual analysis that attempts to describe the relationships between networked parts (typically actors within the text, places, or other objects, etc.). xiv, 61, 66

on::hover A visual event that occurs within a UI when the user hovers the mouse over a particular element on the screen. xiv, 48, 60, 67

phonological loop The subsystem in Baddeley's model of working memory that is responsible for managing auditory input and processing. xiii, 5–12, 44, 45, 156–158

presence The subjective feeling of the user being present, or existing within, a software system or dynamic virtual world. 83–85

reading The process of consuming and understanding written text. 33–38, 40–42, 44, 45, 50–52, 66, 71

SAS A model of an executive function of attentional control, SAS explains attention through a series of learned actions that are then formed into scripts that can be applied to various situations. 9

selective attention The process of focusing one's attention via specific stimuli. 46, 49

sentiment analysis A type of textual analysis that attempts to quantify the emotional valence of the tone of the text. 61, 64, 65

signal detection The ability to discern between information-relevant patterns and visual "noise". 46

split-attention effect Occurs when the learner's attention must be split between multiple focus points in order to comprehend the material at hand. 47, 48, 63, 70

TF-IDF an abbreviation which stands for "(T)erm (F)requency, (I)nverse (D)ocument (F)requency", TF-IDF is a statistical measure used to determine the importance of words within a corpus. It is a standard statistical measure used in both document classification and text mining [341]. 120

topic modelling A type of textual analysis that uses various algorithms to generate a group of topics that are related to the content of the text. 61

vector-space analysis A type of textual analysis that uses an algorithm to determine the syntactic relationship between words and generate a plot of the generated relationship(s). xiv, 61, 63, 64, 70

visual attention The process of selective attention whereby attention is focused via visual means or through visual stimuli. 46, 47

visual cache The component of the visuospatial sketchpad that acts as a passive store for information related to static visual patterns. 7, 10

visual search The process through which attention is focused with the goal of focusing on a particular element within the visual field. 47, 49, 63

visuospatial sketchpad The subsystem of Baddeley's model of working memory responsible for translating visual and spatial information. xiii, 5–11, 44, 45, 68, 69, 71, 156–158

working memory The memory system posited by Baddeley and Hitch to better describe the underlying process of short-term memory. 5–7, 9–12, 28, 29, 36, 43–45, 71, 86, 87, 118, 156–158

Bibliography

[1] Jeff Johnson. *Designing with the Mind in Mind: A Simple Guide to Understanding User Interface Design Rules*. 1st. Amsterdam: Morgan Kaufmann, 2010. URL: http://ebookcentral.proquest.com/lib/nuim/reader.action?docID=535170 (visited on 08/02/2017).

[2] Peter Stallybrass. "Books and Scrolls: Navigating the Bible". In: *Books and Readers in Early Modern England: Material Studies*. Ed. by Jennifer Andersen and Elizabeth Sauer. Philadelphia, PA: University of Pennsylvania Press, 2002, pp. 42–79.

[3] D. C. Greetham. *Textual Scholarship: An Introduction*. New York, NY: Garland Publishing, 1994. ISBN: 0-8153-0058-1.

[4] Franco Moretti. *Distant Reading*. Verso, 2013. ISBN: 978-1-78168-084-1.

[5] Matthew L. Jockers. *Macroanalysis: Digital Methods & Literary History*. Chicago, IL: University of Illinois Press, 2013. ISBN: 978-0-252-07907-8.

[6] Johanna Drucker. *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press, 2014. ISBN: 978-0-674-72493-8.

[7] Daniel Schacter, Daniel Gilbert, Daniel Wegner, et al. *Psychology: Second European Edition*. 2nd. New York, NY: Palgrave, 2016. ISBN: 978-1-137-40674-3.

[8] R. C. Atkinson and R. M. Shiffrin. "Human Memory: A Proposed System and its Control Processes". In: *Psychology of Learning and Motivation* 2 (1968), pp. 89–195. DOI: 10.1016/S0079-7421(08)60422-3. (Visited on 04/10/2018).

[9] Michael W. Eysenck and Mark T. Keane. *Cognitive Psychology: A Student's Handbook*. 7th. London, UK: Psychology Press, 2015. ISBN: 978-1-84872-416-7.

[10] George Sperling. "The Information Available in Brief Visual Presentations". In: *Psychological Monographs: General and Applied* 74.11 (1960), pp. 1–29. DOI: 10.1037/h0093759. (Visited on 04/11/2018).

[11] Tracy D. Tomlinson, David E. Huber, Cory A. Rieth, et al. "An Interference Account of Cue-Independent Forgetting in the No-Think Paradigm". In: *Proceedings of the National Academy of Sciences* 106.37 (Sept. 2009), pp. 15588–15593. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0813370106. (Visited on 04/12/2018).

[12] Karl R. Gegenfurtner and George Sperling. "Information Transfer in Iconic Memory Experiments." In: *Journal of Experimental Psychology: Human Perception and Performance* 19.4 (1993), pp. 845–866. ISSN: 1939-1277, 0096-1523. DOI: 10.1037/0096-1523.19.4.845. (Visited on 07/22/2019).

[13] Alan D. Baddeley and Graham Hitch. "Working Memory". In: *Psychology of Learning and Motivation*. Vol. 8. Stirling, Scotland: University of Stirling, 1974, pp. 47–89. DOI: 10.1016/S0079-7421(08)60452-1. (Visited on 04/11/2018).

[14] Alan Baddeley. "The Episodic Buffer: A New Component of Working Memory?" In: *Trends in Cognitive Sciences* 4.11 (Nov. 2000), pp. 417–423. ISSN: 1364-6613. DOI: 10.1016/S1364-6613(00)01538-2. (Visited on 10/14/2017).

[15] Alan Baddeley, Richard Allen, and Graham Hitch. "Investigating the Episodic Buffer". In: *Psychologica Belgica* 50.3-4 (Oct. 2010). ISSN: 2054-670X. DOI: 10.5334/pb-50-3-4-223. (Visited on 04/10/2018).

[16] Bradley R. Buchsbaum. "The Role of Consciousness in the Phonological Loop: Hidden in Plain Sight". In: *Frontiers in Psychology* 4 (Aug. 2013). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00496. (Visited on 04/17/2018).

[17] Alan D Baddeley. "Working Memory: Theories, Models, and Controversies". In: *Annual Review of Psychology* 63 (Jan. 2012), pp. 1–29. DOI: 10.1146/annurev-psych-120710-100422. (Visited on 10/02/2019).

[18] Robert H. Logie. *Visuo-Spatial Working Memory*. Hove, UK: Lawrence Erlbaum Associates, 1995. ISBN: 0-86377-107-6.

[19] Alan D. Baddeley, Richard J. Allen, and Graham J. Hitch. "Binding in Visual Working Memory: The Role of the Episodic Buffer". In: *Neuropsychologia* 49.6 (May 2011), pp. 1393–1400. ISSN: 0028-3932. DOI: 10.1016/j.neuropsychologia.2010.12.042. (Visited on 10/14/2017).

[20] Donald A. Norman and Tim Shallice. "Attention to Action: Willed and Automatic Control of Behavior". In: *Consciousness and Self Regulation: Advances in Research*. Ed. by R.J. Davidson, G.E. Schwartz, and D. Shapiro. Vol. IV. New York, NY: Plenum Press, 1986. URL: https://www.researchgate.net/publication/202165686 (visited on 07/12/2018).

[21] Alan Baddeley. "Exploring the Central Executive". In: *The Quarterly Journal of Experimental Psychology A* 49.1 (Feb. 1996), pp. 5–28. ISSN: 14640740, 02724987. DOI: 10.1080/027249896392784. (Visited on 10/14/2017).

[22] Nicholas C. Burbules. "Rhetorics of the Web: Hyperreading and Critical Literacy". In: *Page to Screen: Taking Literacy into the Electronic Era*. Ed. by Ilana Synder. New York, NY: Taylor & Francis, 1998, pp. 102–122. ISBN: 978-0-415-17465-7.

[23] Anne Treisman. "Feature Binding, Attention and Object Perception". In: *Philosophical Transactions: Biological Sciences* 353.1373 (1998), pp. 1295–1306. ISSN: 0962-8436. URL: https://www.jstor.org/stable/56884 (visited on 03/19/2019).

[24] John Sweller and Paul Chandler. "Why Some Material Is Difficult to Learn". In: *Cognition and Instruction* 12.3 (1994), pp. 185–233. URL: https://www-jstor-org.jproxy.nuim.ie/stable/3233760 (visited on 10/02/2019).

[25] Akira Miyake, Naomi P. Friedman, Michael J. Emerson, et al. "The Unity and Diversity of Executive Functions and Their Contributions to Complex 'Frontal Lobe' Tasks: A Latent Variable Analysis". In: *Cognitive Psychology* 41.1 (Aug. 2000), pp. 49–100. ISSN: 0010-0285. DOI: 10.1006/cogp.1999.0734. (Visited on 04/09/2018).

[26] Akira Miyake and Naomi P. Friedman. "The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions". In: *Current Directions in Psychological Science* 21.1 (2012), pp. 8–14. ISSN: 0963-7214. URL: http://www.jstor.org/stable/23213113 (visited on 04/09/2018).

[27] M. Daneman and P. M. Merikle. "Working Memory and Language Comprehension: A Meta-Analysis". eng. In: *Psychonomic Bulletin & Review* 3.4 (Dec. 1996), pp. 422–433. ISSN: 1069-9384. DOI: 10.3758/BF03214546.

[28] Kate Cain, Jane Oakhill, and Peter Bryant. "Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills". In: *Journal of Educational Psychology* 96.1 (2004), pp. 31–42. ISSN: 1939-2176(Electronic),0022-0663(Print). DOI: 10.1037/0022-0663.96.1.31.

[29] Henk J Haarmann, Eddy J Davelaar, and Marius Usher. "Individual Differences in Semantic Short-Term Memory Capacity and Reading Comprehension". In: *Journal of Memory and Language* 48.2 (Feb. 2003), pp. 320–345. ISSN: 0749-596X. DOI: 10.1016/S0749-596X(02)00506-5. (Visited on 09/12/2019).

[30] Suzan Nouwens, Margriet A. Groen, and Ludo Verhoeven. "How Working Memory Relates to Children's Reading Comprehension: The Importance of Domain-Specificity in Storage and Processing". In: *Reading and Writing* 30.1 (2017), pp. 105–120. ISSN: 0922-4777. DOI: 10.1007/s11145-016-9665-5. (Visited on 09/12/2019).

[31] Kate Cain. "Individual Differences in Children's Memory and Reading Comprehension: An Investigation of Semantic and Inhibitory Deficits". eng. In: *Memory (Hove, England)* 14.5 (July 2006), pp. 553–569. ISSN: 0965-8211. DOI: 10.1080/09658210600624481. (Visited on 09/12/2019).

[32] Micaela E. Christopher, Akira Miyake, Janice M. Keenan, et al. "Predicting Word Reading and Comprehension with Executive Function and Speed Measures Across Development: A Latent Variable Analysis". eng. In: *Journal of Experimental Psychology. General* 141.3 (Aug. 2012), pp. 470–488. ISSN: 1939-2222. DOI: 10.1037/a0027375.

[33] George Siemens. "Connectivism: A Learning Theory for the Digital Age". In: *International Journal of Instructional Technology and Distance Learning* 2.1 (2005), pp. 3–10. DOI: 10.3109/0142159X.2016.1173661. (Visited on 10/02/2019).

[34] David Kolb. *Experiential Learning: Experience as the Source of Learning and Development*. 2nd edition. Englewood Cliffs, NJ: Prentice Hall, 2015. ISBN: 978-0-13-389240-6.

[35] Mariane Hedegaard. "A Cultural-Historical Approach to Learning in Classrooms". In: *Outlines* 1 (2004), pp. 21–34. URL: http://www.outlines.dk/contents/Outlines041/Hedegaard04.pdf (visited on 07/22/2019).

[36] John B. Watson. "Psychology as the Behaviorist Views It". In: *Psychological Review* 20.2 (1913), pp. 158–177. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0074428. (Visited on 10/02/2019).

[37] Daniel T. Willingham. *Cognition: The Thinking Animal*. 3rd edition. Upper Saddle River, NJ: Pearson Education, 2007. ISBN: 978-0-13-173688-7.

[38] William Bechtel, Adele Abrahamsen, and George Graham. "The Life of Cognitive Science". In: *A Companion to Cognitive Science*. Ed. by William Bechtel and George Graham. Blackwell Companions to Philosophy 13. Malden, MA USA: Blackwell Publishing, 1998.

[39] B.F. Skinner. *About Behaviorism*. New York: Vintage, 1974. ISBN: 978-0-394-71618-3.

[40] B.F. Skinner. "Why I Am Not a Cognitive Psychologist". In: *Behaviorism* 5.2 (1977), pp. 1–10. URL: https://www.jstor.org/stable/27758892 (visited on 04/10/2019).

[41] Owen Conlan. "The Multi-Model, Metadata Driven Approach to Personalised eLearning Services". Doctoral Thesis. Dublin, Ireland: Trinity College Dublin, 2005. URL: http://hdl.handle.net/2262/77608 (visited on 04/01/2018).

[42] Agnès Aflalo. *The Failed Assassination of Psychoanalysis: The Rise and Fall of Cognitivism*. London, UK: Routledge, 2018. ISBN: 978-1-78241-292-2. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=4530424 (visited on 04/22/2019).

[43] John R. Anderson, Daniel Bothell, Michael D. Byrne, et al. "An Integrated Theory of the Mind". In: *Psychological Review* 111.4 (Oct. 2004), pp. 1036–1060. ISSN: 0033-295X. DOI: 10.1037/0033-295X.111.4.1036. (Visited on 04/26/2019).

[44] John R. Anderson, Jon M. Fincham, Yulin Qin, et al. "A Central Circuit of the Mind". In: *Trends in Cognitive Sciences* 12.4 (Apr. 2008), pp. 136–143. ISSN: 1364-6613. DOI: 10.1016/j.tics.2008.01.006. (Visited on 04/26/2019).

[45] Zoltan Torey and Daniel C. Dennett. *The Crucible of Consciousness: An Integrated Theory of Mind and Brain*. Cambridge, MA: MIT Press, 2009. ISBN: 978-0-262-25518-9. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=3339021 (visited on 04/26/2019).

[46] William R. Uttal. *Dualism: The Original Sin of Cognitivism*. Mahwah, NJ: Routledge, 2004. ISBN: 978-1-4106-1123-9. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=234252 (visited on 04/22/2019).

[47] Elspeth McKay, John Lenarcic, and Lawrence Tomei, eds. *Macro-Level Learning through Massive Open Online Courses (MOOCs): Strategies and Predictions for the Future*. Advances in Educational Technologies and Instructional Design. IGI Global, 2015. ISBN: 978-1-4666-8324-2. DOI: 10.4018/978-1-4666-8324-2. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8324-2 (visited on 04/22/2019).

[48] O. A. B. Hassan. "Learning Theories and Assessment Methodologies – An Engineering Educational Perspective". In: *European Journal of Engineering Education* 36.4 (Aug. 2011), pp. 327–339. ISSN: 0304-3797. DOI: 10.1080/03043797.2011.591486. (Visited on 04/22/2019).

[49] Kaya Yilmaz. "The Cognitive Perspective on Learning: Its Theoretical Underpinnings and Implications for Classroom Practices". In: *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 84.5 (Aug. 2011), pp. 204–212. ISSN: 0009-8655. DOI: 10.1080/00098655.2011.568989. (Visited on 04/22/2019).

[50] Tonya Rutherford-Hemming. "Simulation Methodology in Nursing Education and Adult Learning Theory". In: *Adult Learning* 23.3 (Aug. 2012), pp. 129–137. ISSN: 1045-1595. DOI: 10.1177/1045159512452848. (Visited on 04/22/2019).

[51] Dennis W. Cheek. *Thinking Constructively About Science, Technology, and Society Education*. eBook. Albany, NY USA: SUNY Press, 1992. ISBN: 978-0-7914-0939-8.

[52] David H. Jonassen, Kyle L. Peck, and Brent G. Wilson. *Learning with Technology: A Constructivist Perspective*. Upper Saddle River, NJ: Prentice Hall, 1999. ISBN: 0-13-271891-X.

[53] Roger C. Schank. "Goal-Based Scenarios". In: *Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson*. Ed. by Roger C. Schank and Ellen Langer. Kindle. Hillsdale, N.J: Psychology Press, 1994, pp. 1–32.

[54] Diana Laurillard. *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*. 2nd edition. London, UK: Routledge, 2002. ISBN: 0-415-25678-X.

[55] Brett K. Hayes, Evan Heit, and Haruka Swendsen. "Inductive Reasoning". en. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 1.2 (Mar. 2010), pp. 278–292. ISSN: 1939-5078, 1939-5086. DOI: 10.1002/wcs.44. (Visited on 08/24/2019).

[56] Jennifer Edmond. "How Scholars Read Now: When the Signal Is the Noise". In: *Digital Humanities Quarterly* 12.2 (2018). URL: http://www.digitalhumanities.org/dhq/vol/12/2/000388/000388.html (visited on 09/21/2019).

[57] John Guillory. "How Scholars Read". In: *ADE Bulletin* 146.10 (2008), pp. 8–17. URL: https://www.ade.mla.org/bulletin/article/ade.146.8 (visited on 09/21/2019).

[58] Sarah E. Kersh and Chelsea Skalak. "From Distracted to Recursive Reading: Facilitating Knowledge Transfer through Annotation Software". In: *Digital Humanities Quarterly* 12.2 (Sept. 2018). ISSN: 1938-4122. URL: http://www.digitalhumanities.org/dhq/vol/12/2/000387/000387.html (visited on 10/02/2019).

[59] Joshua L. Mann. "How Technology Means: Texts, History, and Their Associated Technologies". In: *Digital Humanities Quarterly* 12.3 (Sept. 2018). ISSN: 1938-4122. URL: http://www.digitalhumanities.org/dhq/vol/12/3/000398/000398.html (visited on 10/02/2019).

[60] Kathleen Yancey, Liane Robertson, and Kara Taczak. *Writing Across Contexts: Transfer, Composition, and Sites of Writing*. English. 1st edition. Logan: Utah State University Press, 2014. ISBN: 978-0-87421-937-1.

[61] Bridget Almas, Emad Khazraee, Matthew Thomas Miller, et al. "Manuscript Study in Digital Spaces: The State of the Field and New Ways Forward". In: *Digital Humanities Quarterly* 12.2 (July 2018). ISSN: 1938-4122. URL: http://www.digitalhumanities.org/dhq/vol/12/2/000374/000374.html (visited on 10/02/2019).

[62] P. N. Johnson-Laird. "Deductive Reasoning". In: *Annual Review of Psychology* 50 (1999), pp. 109–35. DOI: 10.1146/annurev.psych.50.1.109. (Visited on 08/24/2019).

[63] Samia Khan. "New Pedagogies on Teaching Science with Computer Simulations". In: *Journal of Science Education and Technology* 20.3 (2011), pp. 215–232. ISSN: 1059-0145. URL: http://www.jstor.org/stable/41499394 (visited on 09/22/2019).

[64] Sanwar Ali. "Effective Teaching Pedagogies for Undergraduate Computer Science". English. In: *Mathematics and Computer Education; Old Bethpage* 39.3 (2005), pp. 243–257. ISSN: 07308639. URL: http://search.proquest.com/docview/235807504/abstract/909B45E9D5304BB6PQ/1 (visited on 09/22/2019).

[65] Hong Zhu. *Software Design Methodology: From Principles to Architectural Styles*. Oxford, UK: Elsevier Science & Technology, 2005. ISBN: 978-0-08-045496-2. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=269543 (visited on 09/22/2019).

[66] John Allen Robinson. *Software Design for Engineers and Scientists*. Oxford, UK: Elsevier Science & Technology, 2004. ISBN: 978-0-08-047440-3. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=288912 (visited on 09/22/2019).

[67] Michael Jackson. *Software Requirements and Specifications: A Lexicon of Practice, Principles and Prejudices*. 1st edition. Reading, MA: Addison-Wesley Professional, 1995. ISBN: 978-0-201-87712-0.

[68] Hugh G. Gauch Jr. *Scientific Method in Brief*. New York, NY: Cambridge University Press, 2013. ISBN: 978-1-107-30708-7. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=1113122 (visited on 09/22/2019).

[69] Aleksei Nikolaevich Leont'ev. *Activity and Consciousness*. Trans. by Marie J. Hall. Web. Pacifica, CA: Prentice Hall, 1978. URL: https://www.marxists.org/archive/leontev/works/activity-consciousness.pdf (visited on 09/28/2018).

[70] Lev S. Vygotsky. *Thought and Language*. Trans. by Eugenia Hanfmann, Gertrude Vakar, and Alex Kozulin. Revised and Expanded. Cambridge, MA: MIT Press, 2012. ISBN: 978-0-262-51771-3.

[71] Wolff-Michael Roth and Yew-Jin Lee. "'Vygotsky's Neglected Legacy': Cultural-Historical Activity Theory". In: *Review of Educational Research* 77.2 (2007), pp. 186–232. ISSN: 00346543, 19351046. URL: http://www.jstor.org.jproxy.nuim.ie/stable/4624893 (visited on 10/02/2019).

[72] Victor Kaptelinin and Bonnie A. Nardi. *Acting with Technology: Activity Theory and Interaction Design*. Cambridge, MA: MIT Press, 2006. ISBN: 978-0-262-11298-7.

[73] Paolo Diviacco and Alessandro Busato. "Maps, Graphs, and Annotations as Boundary Objects in Knowledge Networks, Distributed Cognition and Collaborative E-Research". In: *Collaborative Knowledge in Scientific Research Networks*. Ed. by Paolo Diviacco, Peter Fox, Cyril Pshenichny, et al. Advances in Knowledge Acquisition, Transfer, and Management. Hershey, PA USA: IGI Global, 2015, pp. 387–408. URL: 10.4018/978-1-4666-6567-5.ch010 (visited on 02/06/2019).

[74] Yrjo Engeström. *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*. 2nd Edition. Cambridge: Cambridge University Press, 2014. ISBN: 978-1-107-64010-8. URL: https://doi.org/10.1017/CBO9781139814744 (visited on 10/03/2018).

[75] Helen Beetham. "Designing for Active Learning in Technology-Rich Contexts". In: *Rethinking Pedagogy for a Digital Age Designing for 21st Century Learning*. Ed. by Helen Beetham and Rhona Sharpe. 2nd edition. New York, NY: Taylor & Francis, 2013, pp. 31–48. ISBN: 978-0-415-53996-8.

[76] Andrew M Colman. *Motivation*. Oxford, 2015. URL: http://www.oxfordreference.com/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681-e-5239 (visited on 11/07/2017).

[77] A. H. Maslow. "A Theory of Human Motivation". In: *Psychological Review* 50.4 (July 1943), pp. 370–396. ISSN: 0033-295X. DOI: 10.1037/h0054346. (Visited on 11/14/2014).

[78] William Joseph Harrigan and Michael Lamport Commons. "Replacing Maslow's Needs Hierarchy With an Account Based on Stage and Value". In: *Behavioral Development Bulletin* 20.1 (Apr. 2015), pp. 24–31. DOI: 10.1037/h0101036. (Visited on 11/23/2018).

[79] Chatura Pulasinghage. "Employee Motivation: What Factors Motivate Employees to Work in Nongovernmental Organizations (NGO) in Sri Lanka: A Study According to Maslow's Hierarchy of Needs Model". In: *International Journal of Interdisciplinary Social Sciences* 5.4 (2010), pp. 197–211. DOI: 10.18848/1833-1882/CGP/v05i04/51652. (Visited on 10/23/2019).

[80] Robert Urwiler. "The IT Value Hierarchy: Using Maslow's Hierarchy of Needs as a Metaphor for Gauging the Maturity Level of Information Technology Use Within Competitive Organizations". In: *Information Systems Management* 25.1 (2008), pp. 83–88. DOI: 10.1080/10580530701777206. (Visited on 11/23/2018).

[81] Don Gorman. "Maslow's Hierarchy and Social and Emotional Wellbeing". In: *Aboriginal & Islander Health Worker Journal* 34.1 (Jan. 2010), pp. 27–29. URL: https://eprints.usq.edu.au/6830/ (visited on 11/23/2018).

[82] Robert J. Zalenski and Richard Raspa. "Maslow's Hierarchy of Needs: A Framework for Achieving Human Potential in Hospice". In: *Journal of Palliative Medicine* 9.5 (Oct. 2006), pp. 1120–1127. DOI: 10.1089/jpm.2006.9.1120. (Visited on 11/23/2018).

[83] Matthew D. Lieberman. *Social: Why Our Brians Are Wired to Connect*. New York, NY: Crown Publishers, 2013. ISBN: 978-0-307-88909-6.

[84] Pamela Brown Rutledge. "Social Networks: What Maslow Misses". In: *Psychology Today* (Nov. 2011). URL: http://www.jlampl.net/SocialNetworksMaslow.pdf (visited on 01/09/2019).

[85] Donald Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York, New York, USA: Basic Books, 2004. ISBN: 978-0-465-05136-6.

[86] Albert Bandura. "Self-Efficacy Mechanism in Human Agency." In: *American Pscyhologist* 37.2 (1982), pp. 122–147. DOI: 10.1037/0003-066X.37.2.122. (Visited on 01/14/2019).

[87] Susan Bergin, Aidan Mooney, John Ghent, et al. "Using Machine Learning Techniques to Predict Introductory Programming Performance". In: *International Journal of Computer Science and Software Engineering* 4.12 (Dec. 2015), pp. 323–328. URL: http://ijcsse.org/published/volume4/issue12/p2-V4I12.pdf (visited on 07/25/2017).

[88] Bong Gee Jang, Kristin Conradi, Michael C. McKenna, et al. "Motivation". In: *Reading Teacher* 69.2 (Sept. 2015), pp. 239–247. DOI: 10.1002/trtr.1365. (Visited on 11/07/2017).

[89] Ulrich Schiefele and Andreas Krapp. "Topic Interest and Free Recall of Expository Text". In: *Learning and Individual Differences* 8.2 (1996), pp. 141–160. ISSN: 1041-6080. DOI: 10.1016/S1041-6080(96)90030-8. (Visited on 11/08/2017).

[90] Jacqueline S. Eccles. "Subjective Task Value and the Eccles et al. Model of Achievement-Related Choices". In: *Handbook of Competence and Motivation*. Ed. by Andrew J. Elliot and Carol S. Dweck. 1st edition. New York, NY: The Guilford Press, 2005, pp. 105–121. ISBN: 978-1-59385-606-9. URL: http://www.rcgd.isr.umich.edu/garp/articles/eccles05d.pdf (visited on 11/08/2017).

[91] Dale H. Schunk and Barry J. Zimmerman. "Influencing Children's Self-Efficacy and Self-Regulation of Reading and Writing Through Modeling". In: *Reading & Writing Quarterly* 23.1 (Jan. 2007), pp. 7–25. ISSN: 1057-3569. DOI: 10.1080/10573560600837578. (Visited on 11/08/2017).

[92] James W. Chapman and William E. Tunmer. "Development of Young Children's Reading Self-Concepts: An Examination of Emerging Subcomponents and Their Relationship with Reading Achievement". In: *Journal of Educational Psychology* 87.1 (Mar. 1995), pp. 154–167. ISSN: 0022-0663. DOI: 10.1037/0022-0663.87.1.154. (Visited on 11/08/2017).

[93] Carole Ames and Jennifer Archer. "Achievement Goals in the Classroom: Students' Learning Strategies and Motivation Processes". In: *Journal of Educational Psychology* (1988), pp. 260–267. DOI: 10.1037/0022-0663.80.3.260. (Visited on 10/02/2019).

[94] Carol S. Dweck, Jennifer A. Mangels, and Catherine Good. "Motivational Effects on Attention, Cognition, and Performance". In: *Motivation, Emotion, and Cognition: Integrative Persepctives on Intellectual Functioning and Development*. Ed. by David Yun Dai and Robert J. Sternberg. Mahwah, NJ USA: Lawrence Erlbaum Associates, 2004, pp. 41–56.

[95] Paul R. Pintrich. "Multiple Goals, Multiple Pathways: The Role of Goal Orientation in Learning and Achievement". In: *Journal of Educational Psychology* 92.3 (Sept. 2000), pp. 544–555. ISSN: 0022-0663. DOI: 10.1037/0022-0663.92.3.544. (Visited on 11/08/2017).

[96] Nicole Lang Beebe and Jan Guynes Clark. "A Hierarchical, Objectives-Based Framework for the Digital Investigations Process". In: *Digital Investigation* 2.2 (June 2005), pp. 147–167. ISSN: 1742-2876. DOI: 10.1016/j.diin.2005.04.002. (Visited on 11/16/2017).

[97] Marc Prensky. "Digital Natives, Digital Immigrants". In: *On the Horizon* 9.5 (Oct. 2001), pp. 1–6. URL: https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf (visited on 09/12/2017).

[98] Diana Oblinger. "Boomers, Gen-Xers, and Millennials: Understanding the "New Students"". In: *Educause Review* 38.4 (2003), pp. 37–47. URL: https://er.educause.edu/-/media/files/article-downloads/erm0342.pdf (visited on 06/20/2019).

[99] Marc Prensky. "Listen to the Natives". In: *Educational Leadership* 63.4 (2006), pp. 8–13. ISSN: 0013-1784. URL: http://www.ascd.org/ASCD/pdf/journals/ed_lead/el200512_prensky.pdf (visited on 10/02/2019).

[100] Chris Jones and Simon Cross. "Is There a Net Generation Coming to University?" In: *Dreams Begins Responsibility: Choice Evidence and Change*. Manchester, UK], 2009, pp. 10–20. URL: https://repository.alt.ac.uk/645/1/ALT-C_09_proceedings_090806_web_0299.pdf (visited on 06/20/2019).

[101] Mai Neo, Ken Tse-Kian Neo, and Sally Thian-Li Lim. "Designing a Web-Based Multimedia Learning Environment with Laurillard's Conversational Framework: An Investigation on Instructional Relationships". In: *Turkish Online Journal of Educational Technology - TOJET* 12.3 (July 2013), pp. 39–50. ISSN: 1303-6521. URL: https://eric.ed.gov/?id=EJ1016940 (visited on 06/20/2019).

[102] Gwendolyn Angela Lawrie, Lisbeth Grøndahl, Simon Boman, et al. "Wiki Laboratory Notebooks: Supporting Student Learning in Collaborative Inquiry-Based Laboratory Experiments". In: *Journal of Science Education and Technology* 25.3 (June 2016), pp. 394–409. ISSN: 1573-1839. DOI: 10.1007/s10956-016-9601-0. URL: https://doi.org/10.1007/s10956-016-9601-0 (visited on 06/20/2019).

[103] Jörgen Holmberg. "Applying a Conceptual Design Framework to Study Teachers' Use of Educational Technology". In: *Education and Information Technologies* 22.5 (Sept. 2017), pp. 2333–2349. ISSN: 1573-7608. DOI: 10.1007/s10639-016-9536-3. URL: https://doi.org/10.1007/s10639-016-9536-3 (visited on 06/20/2019).

[104] John Tooby and Leda Cosmides. "On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation". In: *Journal of Personality* 58.1 (Mar. 1990), pp. 17–67. DOI: 10.1111/j.1467-6494.1990.tb00907.x. (Visited on 10/10/2017).

[105] Edward O. Wilson. *On Human Nature*. Revised. Cambridge, MA: Harvard University Press, 2004. ISBN: 978-0-674-01638-5.

[106] Kurt Goldstein. *Human Nature: In the Light of Psychopathology*. Cambridge, MA: Harvard University Press, 1940. ISBN: 978-0-674-18735-1.

[107] Nancy K. Schlossberg. "A Model for Analyzing Human Adaptation to Transition". In: *The Counseling Psychologist* 9.2 (1981), pp. 2–18. DOI: 10.1177/001100008100900202. (Visited on 10/10/2017).

[108] Colin W.G. Clifford, Michael A. Webster, Garrett B. Stanley, et al. "Visual Adaptation: Neural, Psychological, and Computational Aspects". In: *Vision Research* 47.25 (Nov. 2007), pp. 3125–3131. DOI: 10.1016/j.visres.2007.08.023. (Visited on 10/10/2017).

[109] Gillian Rhodes, Tamara L. Watson, Linda Jeffery, et al. "Perceptual Adaptation Helps Us Identify Faces". In: *Vision Research* 50.10 (May 2010), pp. 963–968. DOI: 10.1016/j.visres.2010.03.003. (Visited on 10/10/2017).

[110] Gordon M. Redding and Benjamin Wallace. "Generalization of Prism Adaptation". In: *Journal of Experimental Psychology: Human Perception and Performance* 32.4 (2006), pp. 1006–1022. DOI: 10.1037/0096-1523.32.4.1006. (Visited on 09/16/2017).

[111] Gordon M. Redding and Benjamin Wallace. *Adaptive Spatial Alignment*. Mahwah, NJ USA: Lawrence Erlbaum Associates, 1997. ISBN: 978-0-8058-2395-0.

[112] Stephen E. Palmer. "Perceptual Organization in Vision". In: *Steven's Handbook of Experimental Psychology*. Ed. by Hal Pashler, Steven Yantis, Douglas Medin, et al. 3rd. New York, NY: John Wiley & Sons, Ltd, 2002, pp. 177–234. ISBN: 978-0-471-21442-7. DOI: 10.1002/0471214426. (Visited on 10/02/2019).

[113] George P. Landow. *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization*. 3rd. Baltimore, MD: Johns Hopkins University Press, 2006. ISBN: 978-0-8018-8257-9.

[114] Ferris Jabr. "The Reading Brain in the Digital Age: The Science of Paper Versus Screens". In: *Scientific American* (Apr. 2013). URL: https://www.scientificamerican.com/article/reading-paper-screens/ (visited on 10/02/2019).

[115] Anne Niccoli. "Paper or Tablet? Reading Recall and Comprehension". In: *Educause Review* (Sept. 2015). URL: https://er.educause.edu/articles/2015/9/paper-or-tablet-reading-recall-and-comprehension (visited on 10/10/2017).

[116] Stanislas Dehaene and Gareth Cook. "Your Brain on Books". In: *Scientific American* (Nov. 2009). URL: https://www.scientificamerican.com/article/your-brain-on-books/ (visited on 10/10/2017).

[117] Richard C. Anderson, Elfrieda H. Hiebert, Judith A. Scott, et al. "Becoming a Nation of Readers: The Report of the Commission on Reading". In: *Education and Treatment of Children* 11.4 (1988), pp. 389–396. URL: https://www-jstor-org.jproxy.nuim.ie/stable/42899086 (visited on 10/02/2019).

[118] Peter Freebody and Allan Luke. "Literacies Programs: Debates and Demands in Cultural Context". In: *Prospect: An Australian Journal of TESOL* 5.3 (1990), pp. 7–16. URL: http://eprints.qut.edu.au/49099/1/DOC090312.pdf (visited on 09/17/2017).

[119] Anne Mangen and Adriaan van der Weel. "The Evolution of Reading in the Age of Digitisation: An Integrative Framework for Reading Research". In: *Literacy* 50.3 (Sept. 2016), pp. 116–124. DOI: 10.1111/lit.12086. (Visited on 09/25/2017).

[120] Lois Ruth Harris, Christina Ruth Davidson, and Kerry Therese Aprile. "Understanding Teacher Aides' Definitions of Reading: Implications for Classroom Practice". In: *The Australian Educational Researcher* 42.5 (2015), pp. 627–644. DOI: 10.1007/s13384-015-0181-4. (Visited on 09/17/2017).

[121] James P. Gee. "What Is Literacy?" In: *Rewriting Literacy: Culture and the Discourse of the Other*. Ed. by Candace Mitchell and Kathleen Weiler. New York, NY: Bergin & Garvey, 1991, pp. 3–11. ISBN: 978-0-89789-228-5.

[122] Alan Baddeley. "Working Memory and Language: An Overview". In: *Journal of Communication Disorders* 36.3 (May 2003), pp. 189–208. DOI: 10.1016/S0021-9924(03)00019-4. (Visited on 10/13/2017).

[123] Alan D Baddeley, Neil Thomson, and Mary Buchanan. "Word Length and the Structure of Short-Term Memory". In: *Journal of Verbal Learning and Verbal Behavior* 14.6 (Dec. 1975), pp. 575–589. URL: https://doi.org/10.1016/S0022-5371(75)80045-4 (visited on 10/23/2019).

[124] Douglas Fisher and Nancy Frey. *Close Reading and Writing from Sources*. Newark, DE: International Reading Association, 2014. ISBN: 978-0-87207-158-2.

[125] Annette Frederico. *Engagements with Close Reading*. New York, NY: Routledge, 2016. ISBN: 978-0-415-74802-5.

[126] Sven Birkerts. *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. eBook. Farour, Straus and Giroux, 2006. ISBN: 978-0-86547-957-9. URL: https://www.scribd.com/read/182557342/The-Gutenberg-Elegies-The-Fate-of-Reading-in-an-Electronic-Age (visited on 09/17/2017).

[127] Helen Nicholson. "Close Reading". In: *Research in Drama Education: The Journal of Applied Theatre and Performance* 22.2 (2017), pp. 183–185. DOI: 10.1080/13569783.2017.1309738. (Visited on 09/17/2017).

[128] Jane Austen. *The Annotated Pride & Prejudice*. Ed. by D. M. Shapard. 2nd. New York, NY: Random House, 2012. ISBN: 978-0-307-95090-1.

[129] Jane Austen. *The Annotated Sense and Sensibility*. Ed. by David M. Shapard. 1st edition. New York, NY: Anchor Books, 2011. ISBN: 978-0-307-39076-9.

[130] CCSSI. *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. 2010. URL: http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf (visited on 02/12/2019).

[131] Kathleen A. Hinchman and David W. Moore. "Close Reading: A Cautionary Interpretation". In: *Journal of Adolescent & Adult Literacy* 56.6 (Mar. 2013), pp. 441–450. URL: https://www-jstor-org.jproxy.nuim.ie/stable/41827875 (visited on 10/02/2019).

[132] Anne Mangen, Bente R Walgermo, and Kolbjørn Brønnick. "Reading Linear Texts on Paper Versus Computer Screen: Effects on Reading Comprehension". In: *International Journal of Educational Research* 58 (2013), pp. 61–68. DOI: 10.1016/j.ijer.2012.12.002. (Visited on 12/11/2017).

[133] Annie Piolat, Jean-Yves Roussey, and Olivier Thunin. "Effects of Screen Presentation on Text Reading and Revising". In: *International Journal of Human-Computer Studies* 47.4 (Oct. 1997), pp. 565–589. DOI: 10.1006/ijhc.1997.0145. (Visited on 12/11/2017).

[134] Maria Giulia Cataldo and Jane Oakhill. "Why are Poor Comprehenders Inefficient Searchers? An Investigation into the Effects of Text Representation and Spatial Memory on the Ability to Locate Information in Text". In: *Journal of Educational Psychology* 92.4 (2000), pp. 791–799. DOI: 10.1037//0022-0663.92.4.791. (Visited on 10/02/2019).

[135] Eugene B Zechmeister and Jack McKillip. "Recall of Place on the Page". In: *Journal of Educational Psychology* 63.5 (Oct. 1972), pp. 446–453. DOI: 10.1037/h0033249. (Visited on 10/02/2019).

[136] Rakefet Ackerman and Morris Goldsmith. "Metacognitive Regulation of Text Learning: On Screen Versus On Paper". In: *Journal of Experimental Psychology: Applied* 17.1 (2011), pp. 18–32. DOI: 10.1037/a0022086. (Visited on 10/02/2019).

[137] Sarah Lichtenstein, Baruch Fischhoff, and Lawrence D. Phillips. "Calibration of Probabilities: The State of the Art to 1980". In: *Judgment under Uncertainty: Heuristics and Biases*. Ed. by Daniel Kahneman, Paul Slovic, and Amos Tversky. Cambridge: Cambridge University Press, 1982, pp. 306–334. URL: https://doi.org/10.1017/CBO9780511809477.023 (visited on 07/18/2017).

[138] Liana Suantak, Fergus Bolger, and William R. Ferrell. "The Hard–Easy Effect in Subjective Probability Calibration". In: *Organizational Behavior and Human Decision Processes* 67.2 (Aug. 1996), pp. 201–221. DOI: https://doi.org/10.1006/obhd.1996.0074. (Visited on 09/24/2017).

[139] Cheryl Brown and Laura Czerniewicz. "Debunking the 'Digital native': beyond digital apartheid, towards digital democracy". In: *Journal of Computer Assisted Learning* 26 (Sept. 2010), pp. 357–369. DOI: 10.1111/j.1365-2729.2010.00369.x.

[140] David White, Lynn Silipigni Connaway, Donna Lanclos, et al. *Evaluating Digital Services: A Visitors and Residents Approach*. Research Report. OCLC Research, 2014. URL: https://www.oclc.org/research/publications/2014/evaluating-digital-services.html (visited on 12/19/2019).

[141] Shane White, Stephen Garton, Graham White, et al. *Digital Harlem: Everday Life 1915-1930*. Sydney, Australia, 2017. URL: http://digitalharlem.org/ (visited on 08/03/2017).

[142] Lynne G. Duncan, Sarah P. McGeown, Yvonne M. Griffiths, et al. "Adolescent Reading Skill and Engagement with Digital and Traditional Literacies as Predictors of Reading Comprehension". In: *British Journal of Psychology* 107.2 (May 2016), pp. 209–238. DOI: 10.1111/bjop.12134. (Visited on 09/24/2017).

[143] Julie Coiro, Michele Knobel, Colin Lankshear, et al. "Central Issues in New Literacies and New Literacies Research". In: *Handbook of Research on New Literacies*. Ed. by Julie Coiro, Michele Knobel, Colin Lankshear, et al. New York, NY: Taylor & Francis, 2008, pp. 1–20. ISBN: 978-0-8058-5652-1.

[144] John Palfrey and Urs Gasser. *Born Digital: Understanding the First Generation of Digital Natives*. New York, NY: Basic Books, 2008. ISBN: 978-0-465-01856-7.

[145] Mark Bauerlein. "The Resistance to 21st-Century Reading". In: *Reading at a Crossroads? Disjunctures and Continuities in Current Conceptions and Practices*. Ed. by Rand J. Spiro, Michael DeSchryver, Michelle Schira Hagerman, et al. Kindle. New York, NY: Taylor & Francis, 2015. ISBN: 978-0-415-89169-1.

[146] N. Katherine Hayles. *How We Think: Digital Media and Contemporary Technogensis*. Chicago, IL: University of Chicago Press, 2012. ISBN: 978-0-226-32142-4.

[147] Naomi Baron. *Words Onscreen: The Fate of Reading in a Digital World*. Kindle. New York, NY: Oxford University Press, 2015. ISBN: 978-0-19-931576-5.

[148] R. Nageswari, B. Ravikumar, and T. Stephen Jayamani. "Building Lateral Thinking Strategies to Impart English Speaking Skills". In: *Pertanika Journal of Social Scienes & Humanities* 24.2 (June 2016), pp. 687–700. URL: http://www.pertanika.upm.edu.my/Pertanika%20PAPERS/JSSH%20Vol.%2024%20(2)%20Jun.%202016/08%20JSSH%20Vol%2024%20(2)%20Jun%202016_pg687-700.pdf (visited on 09/25/2017).

[149] Tamer S. Hamza and Doaa K. Hassan. "Consequential Creativity: Student Competency and Lateral Thinking Incorporation in Architectural Education". In: *International Journal of Technology and Design Education* 26.4 (Nov. 2016), pp. 587–612. DOI: 10.1007/s10798-015-9321-4. (Visited on 09/25/2017).

[150] Kathryn Coates and Andy Ellison. *Introduction to Information Design*. eBook. London: Laurence King Publishing Ltd., 2014. ISBN: 978-1-78067-325-7. URL: http://ebookcentral.proquest.com/lib/nuim/reader.action?docID=1876178 (visited on 09/25/2017).

[151] Yuejiao Zhang. "Wiki Means More: Hyperreading in Wikipedia". In: *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia* (2006), pp. 23–26. DOI: 10.1145/1149941.1149946. (Visited on 10/02/2019).

[152] Esther Uso-Juan and M Noella Ruiz-Madrid. "Reading Printed Versus Online Texts. A Study of EFL Learners' Strategic Reading Behavior". In: *International Journal of English Studies* Vol. 9.2 (2009), pp. 59–79. DOI: 10.6018/ijes.9.2.90751. (Visited on 10/02/2019).

[153] Loretta F Kasper. "Interactive Hypertext and the Development of ESL Students' Reading Skills". In: *The Reading Matrix* 3.3 (2003). URL: http://www.readingmatrix.com/articles/kasper/index2.html (visited on 08/03/2017).

[154] Arthur R. Cohen, Ezra Stotland, and Donald M. Wolfe. "An Experimental Investigation of Need for Cognition". In: *Journal of Abnormal and Social Psychology* 51.2 (1955), pp. 291–294. DOI: 10.1037/h0042761. (Visited on 09/26/2017).

[155] John T. Cacioppo, Richard E. Petty, Jeffrey A. Feinstein, et al. "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition". In: *Psychological Bulletin* 119.2 (Mar. 1996), pp. 197–253. DOI: 10.1037/0033-2909.119.2.197. (Visited on 09/26/2017).

[156] Heidar Mokhtari, Mohammad-Reza Devarpanah, Mohammad-Hossein Dayyani, et al. "Students' Need for Cognition Affects Their Information Seeking Behavior". In: *New Library World* 114.11/12 (2013), pp. 542–549. DOI: 10.1108/NLW-07-2013-0060. (Visited on 09/26/2017).

[157] Phillip L. Ackerman and Eric D. Heggestad. "Intelligence, Personality, and Interests: Evidence for Overlapping Traits". In: *Psychological Bulletin* 121.2 (Mar. 1997), pp. 219–245. DOI: 10.1037/0033-2909.121.2.219. (Visited on 09/26/2017).

[158] Sophie von Stumm, Benedikt Hell, and Tomas Chamorro-Premuzic. "The Hungry Mind: Intellectual Curiosity Is the Third Pillar of Academic Performance". In: *Perspectives on Psychological Science* 6.6 (Nov. 2011), pp. 574–588. URL: http://www.jstor.org/stable/41613532 (visited on 09/26/2017).

[159] Sophie von Stumm and Phillip L. Ackerman. "Investment and Intellect: A Review and Meta-Analysis". In: *Psychological Bulletin* 139.4 (July 2013), pp. 841–869. DOI: 10.1037/a0030746. (Visited on 09/26/2017).

[160] Franzis Preckel, Heinz Holling, and Miriam Vock. "Academic Underachievement: Relationship with Cognitive Motivation, Achievement Motivation, and Conscientiousness". In: *Psychology in the Schools* 43.3 (Feb. 2006), pp. 401–411. DOI: 10.1002/pits.20154. (Visited on 09/26/2017).

[161] Julia Grass, Alexander Strobel, and Anja Strobel. "Cognitive Investments in Academic Success: The Role of Need for Cognition at University". In: *Frontiers in Psychology* 8.790 (May 2017), pp. 1–9. DOI: 10.3389/fpsyg.2017.00790. (Visited on 09/26/2017).

[162] Maria Sicilia, Salvador Ruiz, and Jose L. Munuera. "Effects of Interactivity in a Web Site: The Moderating Effect of Need for Cognition". In: *Journal of Advertising* 34.3 (2005), pp. 31–45. URL: http://www.jstor.org.jproxy.nuim.ie/stable/4189307 (visited on 09/26/2017).

[163] John Sweller, Jeroen J G van Merrienboer, and Fred G W C Paas. "Cognitive Architecture and Instructional Design". In: *Educational Psychology Review* 10.3 (1998), pp. 251–296.

[164] Ton de Jong. "Cognitive Load Theory, Educational Research, and Instructional Design: Some Food for Thought". In: *Instructional Science* 38.2 (Aug. 2009), pp. 105–134. DOI: 10.1007/s11251-009-9110-0. URL: 2%20oct%202019.

[165] Diana DeStefano and Jo-Anne LeFevre. "Cognitive Load in Hypertext Reading: A Review". In: *Computers in Human Behavior* 23.3 (May 2007), pp. 1616–1641. DOI: 10.1016/j.chb.2005.08.012. (Visited on 10/02/2019).

[166] Neil Burgess and Graham J. Hitch. "A Revised Model of Short-Term Memory and Long-Term Learning of Verbal Sequences". In: *Journal of Memory and Language*. Special Issue on Memory Models 55.4 (Nov. 2006), pp. 627–652. ISSN: 0749-596X. DOI: 10.1016/j.jml.2006.08.005. (Visited on 10/14/2017).

[167] Valerie K. Sims and Mary Hegarty. "Mental Animation in the Visuospatial Sketchpad: Evidence from Dual-Task Studies". In: *Memory & Cognition* 25.3 (May 1997), pp. 321–332. ISSN: 0090-502X, 1532-5946. DOI: 10.3758/BF03211288. (Visited on 10/14/2017).

[168] Shinmin Wang and Susan E. Gathercole. "Working Memory Deficits in Children with Reading Difficulties: Memory Span and Dual Task Coordination". In: *Journal of Experimental Child Psychology* 115.1 (May 2013), pp. 188–197. ISSN: 0022-0965. DOI: 10.1016/j.jecp.2012.11.015. (Visited on 10/14/2017).

[169] Andy V. Pham and Ramzi M. Hasson. "Verbal and Visuospatial Working Memory as Predictors of Children's Reading Ability". In: *Archives of Clinical Neuropsychology* 29.5 (Aug. 2014), pp. 467–477. ISSN: 0887-6177. DOI: 10.1093/arclin/acu024. (Visited on 10/14/2017).

[170] Deny Menghini, Alessandra Finzi, Giovanni Augusto Carlesimo, et al. "Working Memory Impairment in Children With Developmental Dyslexia: Is it Just a Phonological Deficity?" In: *Developmental Neuropsychology* 36.2 (Jan. 2011), pp. 199–213. ISSN: 8756-5641. DOI: 10.1080/87565641.2010.549868. (Visited on 10/14/2017).

[171] Jan L. Plass, Dorothy M. Chun, Richard E. Mayer, et al. "Cognitive Load in Reading a Foreign Language Text with Multimedia Aids and the Influence of Verbal and Spatial Abilities". In: *Computers in Human Behavior* 19 (Jan. 2003), pp. 221–243. DOI: 10.1016/S0747-5632(02)00015-8. (Visited on 09/05/2017).

[172] Richard E. Mayer. *Multimedia Learning*. 2nd. Cambridge: Cambridge University Press, 2009.

[173] William James. *The Principles of Psychology*. New York, NY: Henry Holt & Co, 1890.

[174] Timothy J. Buschman and Sabine Kastner. "From Behavior to Neural Dynamics: An Integrated Theory of Attention". In: *Neuron* 88.1 (Oct. 2015), pp. 127–144. DOI: 10.1016/j.neuron.2015.09.017. (Visited on 10/02/2019).

[175] Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes, et al. "Information-Seeking, Curiosity, and Attention: Computational and Neural Mechanisms". In: *Trends in Cognitive Sciences* 17.11 (Nov. 2013), pp. 585–593. ISSN: 1364-6613. DOI: 10.1016/j.tics.2013.09.001. (Visited on 10/14/2017).

[176] Jacqueline Gottlieb, Mary Hayhoe, Okihide Hikosaka, et al. "Attention, Reward, and Information Seeking". In: *Journal of Neuroscience* 34.46 (Nov. 2014), pp. 15497–15504. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3270-14.2014. (Visited on 10/14/2017).

[177] Stephanie A. McMains and Sabine Kastner. "Visual Attention". In: *Encyclopedia of Neuroscience*. Ed. by Marc D. Binder, Nobutaka Hirokawa, and Uwe Windhorst. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 4296–4302. ISBN: 978-3-540-29678-2. DOI: 10.1007/978-3-540-29678-2_6344. (Visited on 10/02/2019).

[178] Michael I. Posner, Charles R. R. Syner, and Brian J. Davidson. "Attention and the Detection of Signals". In: *Journal of Experimental Psychology: General* 109.2 (1980), pp. 160–174. DOI: 10.1037/0096-3445.109.2.160. (Visited on 10/02/2019).

[179] William R. Newman. *The Chymistry of Isaac Newton*. Indianapolis, IN, Feb. 2006. URL: http://webapp1.dlib.indiana.edu/newton/ (visited on 08/03/2017).

[180] Anne M Treisman and Garry Gelade. "A Feature Integration Theory of Attention". In: *Cognitive Psychology* 12 (1980), pp. 97–136.

[181] Wolfgang Schnotz and Christian Kürschner. "A Reconsideration of Cognitive Load Theory". In: *Educational Psychology Review* 19.4 (Sept. 2007), pp. 469–508. URL: https://link-springer-com.jproxy.nuim.ie/article/10.1007%2Fs10648-007-9053-4 (visited on 10/02/2019).

[182] Matthew Davies, Tim Hitchcock, and Robert Shoemaker. *Locating London's Past*. Hertfordshire, United Kingdom, Dec. 2011. URL: https://www.locatinglondon.org/index.html (visited on 08/03/2017).

[183] Ziming Liu. "Reading Behavior in the Digital Environment: Changes in Reading Behavior Over the Past Ten Years". In: *Journal of Documentation* 61.6 (Dec. 2005), pp. 700–712. DOI: 10.1108/00220410510632040. (Visited on 09/28/2017).

[184] Anne Mangen. "The Digitization of Literary Reading: Contributions from Empirical Research". In: *Orbis Litterarum* 71.3 (2016), pp. 240–262. DOI: 10.1111/oli.12095. (Visited on 10/14/2017).

[185] Gregory Crane. "What Do You Do with a Million Books?" In: *D-Lib magazine* 12.3 (2006). URL: http://www.dlib.org/dlib/march06/crane/03crane.html (visited on 10/02/2019).

[186] Franco Moretti. *Graphs, Maps, Trees: Abstract Models for a Literary Theory*. Paperback. London: Verso, 2007. ISBN: 978-1-84467-026-0.

[187] Alan Bullock and Stephen Trombley, eds. *The New Fontana Dictionary of Modern Thought*. Third. London: HarperCollins, 1999. ISBN: 978-0-00-686383-0.

[188] Columbia University Press. *The Columbia Encyclopedia*. 5th. Boston: Houghton Mifflin Harcourt, 1993. ISBN: 978-0-395-62438-8.

[189] Tania Schlatter and Deborah Levinson. *Visual Usability: Principles and Practices for Designing Digital Applications*. Waltham, MA: Morgan Kaufmann, 2013. ISBN: 978-0-12-398536-1.

[190] Jennifer Turner and Jessica Schomberg. "Inclusivity, Gestalt Principles, and Plain Language in Document Design". In: *In the Library with the Lead Pipe* (June 2016). URL: https://doaj.org/article/c60693f7e0444c0ab0b0410defb795c3 (visited on 10/02/2019).

[191] Alan Liu, Rama Hoetzlein, and Rita Raley. *RoSE*. Santa Barbara, CA, 2012. URL: http://rose.english.ucsb.edu/ (visited on 08/03/2017).

[192] Susan Schreibman, Emma Clarke, Vinayak Das Gupta, et al. *The Letters of 1916*. Maynooth, Co. Kildare, Ireland, Mar. 2016. URL: http://www.letters1916.ie (visited on 05/02/2017).

[193] Susan Schreibman, Vinayak Das Gupta, and Neale Rooney. "Notes from the Transcription Desk: Visualising Public Engagement". In: *English Studies* 0.0 (July 2017), pp. 1–20. DOI: 10.1080/0013838X.2017.1333754.

[194] IBM. *IBM Logo*. CTB19. July 2015. URL: https://www-03.ibm.com/press/us/en/photo/20190.wss (visited on 04/16/2018).

[195] Stefan Müller, Carsten Dilba, and Rashid-S. Pegah. *Borchward's Journey*. Berlin, 2015. URL: http://quellen-perspectivia.net/en/borchward/start.

[196] Deborah Leem, Jenny Thomas, Barbara McGillivray, et al. *London Smells*. London, 2017. URL: http://londonsmells.co.uk/ (visited on 08/21/2017).

[197] David Kelly, Justin Tonra, and Lindsay Ann Reid. *Personæ: A Character-Visualisation Tool for Dramatic Texts*. Galway, July 2016. URL: http://davidkelly.ie/projects/personae/ (visited on 08/21/2017).

[198] Scott Nesbit. *Visualizing Emancipation*. Athens, GA, 2012. URL: http://www.americanpast.org/emancipation/ (visited on 08/09/2017).

[199] Robyn Adams. *The Diplomatic Correspondence of Thomas Bodley, 1585-1597*. London, July 2011. URL: http://www.livesandletters.ac.uk/bodley/bodley.html (visited on 04/01/2019).

[200] Daniel D. Lee and H. Sebastian Seung. "Learning the Parts of Objects by Non-Negative Matrix Factorization". In: *Nature (London)* 401.6755 (Oct. 1999), pp. 788–791. DOI: 10.1038/44565. (Visited on 08/08/2017).

[201] Aakanksha Sharaff and Naresh Kumar Nagwani. "Email Thread Identification Using Latent Dirichlet Allocation and Non-Negative Matrix Factorization Based Clustering Techniques". In: *Journal of Information Science* 42.2 (2016), pp. 200–212. DOI: 10.1177/0165551515587854. (Visited on 08/08/2017).

[202] V. Paul Pauca, Farial Shahnaz, Michael W. Berry, et al. "Text Mining Using Non-Negative Matrix Factorizations". In: *Proceedings of the 2004 SIAM International Conference on Data Mining*. Lake Buena Vista, FL: SIAM, Apr. 2004, pp. 452–456. DOI: 10.1137/1.9781611972740.45. (Visited on 08/08/2017).

[203] Xu Wei, Liu Xin, and Gong Yihong. "Document Clustering Based on Non-Negative Matrix Factorization". In: *SIGIR '03 Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*. Toronto, Canada: ACM, Aug. 2003, pp. 267–273. DOI: 10.1145/860435.860485. (Visited on 08/08/2017).

[204] Emily Barry. *Applying Data Science to the Supreme Court: Topic Modeling Over Time with NMF (and a D3.JS Bonus)*. July 2016. URL: http://www.emilyinamillion.me/blog/2016/7/13/visualizing-supreme-court-topics-over-time (visited on 08/08/2017).

[205] Rebecca Welzenbach. *ECCO-TCP: Eighteenth Century Collections Online*. Blog. 2017. URL: http://www.textcreationpartnership.org/tcp-ecco/ (visited on 08/08/2017).

[206] Ryan Heuser. *Word Vectors in the Eighteenth Century, Episode 4: Semantic Networks*. Blog. Sept. 2016. URL: http://ryanheuser.org/word-vectors-4/ (visited on 08/08/2017).

[207] *Focus Group B: Understanding the Use of Visualisations and Digital Scholarly Projects in Academic Research*. Dec. 2016.

[208] Matthew Jockers. *Syuzhet Package*. Apr. 2016. URL: https://cran.rstudio.com/web/packages/syuzhet/vignettes/syuzhet-vignette.html (visited on 08/21/2017).

[209] Stéphane Bouchard, Jeanne Talbot, André-Anne Ledoux, et al. "Presence is Just an Illusion: Using fMRI to Locate the Brain Area Responsible to the Meaning Given to Places". In: *Annual Review of Cybertherapy and Telemedicine 2010*. Ed. by Brenda K. Wiederhold, Giuseppe Riva, and Sun I. Kim. Vol. 154. Amsterdam: IOS Press, 2010, pp. 193–196. ISBN: 978-1-60750-561-7.

[210] Lee Hadlington. "Cognitive Factors in Online Behaviour". In: *Cyberpsychology*. Ed. by Alison Attrill. New York, NY: Oxford University Press, 2015, pp. 249–267. ISBN: 978-0-19-871258-9.

[211] Genevieve Marie Johnson and Julia Ann Johnson. "Internet Use and Complex Cognitive Processes". In: *Proceedings of the IADIS International Conference on e-Society 2008*. Ed. by Piet Kommers and Pedro Isaías. Algarve, Portugal: Elsevier, Apr. 2008, pp. 83–90. ISBN: 978-972-8924-55-3.

[212] Kaveri Subrahmanyam, Patricia Greenfield, Robert Kraut, et al. "The Impact of Computer Use on Children's and Adolescents' Development". In: *Journal of Applied Developmental Psychology* 22.1 (Feb. 2001), pp. 7–30. DOI: 10.1016/S0193-3973(00)00063-0. (Visited on 09/11/2017).

[213] Gary W. Small, Teena D. Moody, Prabba Siddarth, et al. "Your Brain on Google: Patterns of Cerebral Activation During Internet Searching". In: *The American Journal of Geriatric Psychiatry* 17.2 (Feb. 2009), pp. 116–126. DOI: 10.1097/JGP.0b013e3181953a02. (Visited on 09/11/2017).

[214] Alan Cooper, Robert Reiman, and David Cronin. *About Face: The Essentials of Interaction Design*. 3rd. Indianapolis, IN: Wiley Publishing, 2007. ISBN: 978-0-470-08411-3.

[215] Donald Norman. "Human-Centered Design Considered Harmful". In: *Interactions* 12.4 (July 2005), pp. 14–19. URL: https://doi.org/10.1145/1070960.1070976 (visited on 04/05/2017).

[216] Carla B. Zoltowski, William C. Oakes, and Andmonica E. Cardella. "Students' Ways of Experiencing Human-Centered Design". In: *Journal of Engineering Education* 101.1 (2012), pp. 28–59. ISSN: 1069-4730. DOI: 10.1002/j.2168-9830.2012.tb00040.x. (Visited on 05/29/2017).

[217] Rex Hartson and Pardha S. Pyla. *The UX Book*. Process and Guidelines for Ensuring a Quality User Experience. Waltham, MA USA: Morgan Kaufmann, 2012. ISBN: 978-0-12-805342-3.

[218] Phillippe N. Tobler, Christopher D. Florillo, and Wolfram Schultz. "Adaptive Coding of Reward Value by Dopamine Neurons". In: *Science* 307.5715 (Mar. 2005), pp. 1642–1645. URL: http://www.jstor.org/stable/3841781 (visited on 06/05/2017).

[219] Pascale Waeltl, Anthony Dickinson, and Wolfram Schultz. "Dopamine Responses Comply with Basic Assumptions of Formal Learning Theory". In: *Nature (London)* 412.6842 (July 2001), pp. 43–48. DOI: 10.1038/35083500. (Visited on 06/05/2017).

[220] Christopher Bergland. *The Neurochemicals of Happiness*. Nov. 2012. URL: https://www.psychologytoday.com/blog/the-athletes-way/201211/the-neurochemicals-happiness (visited on 06/05/2017).

[221] Stan Ruecker, Milena Radzikowska, and Stéfan Sinclair. *Visual Interface Design for Digital Cultural Heritage*. A Guide to Rich-Prospect Browsing. Burlington: Ashgate Publishing, 2011. ISBN: 978-1-138-25030-7.

[222] William Lidwell, Kritina Holden, and Jill Butler. *Universal Principles of Design: 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Teach through Design.* Kindle. Beverly, MA: Rockport Publishers, 2003. ISBN: 978-1-59253-587-3.

[223] Joseph S. Valacich, D. Veena Parboteeah, and John D. Wells. "The Online Consumer's Hierarchy of Needs". In: *Communications of the ACM* 50.9 (Sept. 2007), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/1284621.1284624. (Visited on 11/14/2017).

[224] Yi-Hsuen Shih and Min Liu. "The Importance of Emotional Usability". In: *Journal of Educational Technology Systems* 36.2 (Mar. 2008), pp. 203–218.

[225] Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994. ISBN: 978-0-399-13894-2.

[226] Baba Shiv, George Loewenstein, Antoine Bechara, et al. "Investment Behavior and the Negative Side of Emotion". In: *Psychological Science* 16.6 (2005), pp. 435–439. ISSN: 0956-7976. URL: http://www.jstor.org/stable/40064245 (visited on 11/29/2017).

[227] Kurosu Masaaki and Kaori Kashimura. "Apparent Usability vs. Inherent Usability: Experimental Analysis on the Determinants of the Apparent Usability". In: *CHI '95 Conference Companion on Human Factors in Computing Systems*. Denver, CO USA: ACM Press, May 1995, pp. 292–293. DOI: 10.1145/223355.223680. (Visited on 04/28/2017).

[228] Noam Tractinsky. "Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues". In: *CHI '97 Processings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. Atlanta, GA USA: ACM Press, Mar. 1997, pp. 115–122. DOI: 10.1145/258549.258626. (Visited on 04/28/2017).

[229] Alison Attrill. "Age Versus Goal-Directed Internet Use". In: *Cyberpscyhology*. Ed. by Alison Attrill. New York, NY: Oxford University Press, 2015, pp. 29–48. ISBN: 978-0-19-871258-9.

[230] Giuseppe Carenini. "An Analysis of the Influence of Need for Cognition on Dynamic Queries Usage". In: *CHI '01 Extended Abstracts on Human Factors in Computing Systems*. Seattle, WA USA: ACM Press, 2001, pp. 383–384. DOI: 10.1145/634067.634293. (Visited on 04/29/2017).

[231] Yong Zhang. "Responses to Humorous Advertising: The Moderating Effect of Need for Cognition". In: *Journal of Advertising* 25.1 (1996), pp. 15–32. URL: http://www.jstor.org/stable/4188995 (visited on 04/29/2017).

[232] Harsha Gangadharbatla. "Facebook Me: Collective Self-Esteem, Need to Belong, and Internet Self-Efficacy as Predictors of the Igeneration's Attitudes Toward Social Networking Sites". In: *Journal Of Interactive Advertising* 8.2 (2008), pp. 5–15. DOI: 10.1080/15252019.2008.10722138. (Visited on 04/29/2017).

[233] Terry Daughtery, Matthew Eastin, and Harsha Gangadharbatla. "eCRM: Understanding Internet Confidence and the Implications for Customer Relationship Management". In: *Advances in Electronic Marketing*. Hershey, PA USA: IDI Global, 2005, pp. 67–82.

[234] Donna L. Hoffman and Thomas P. Novak. "A New Marketing Paradigm for Electronic Commerce". In: *The Information Society: An International Journal* 13.1 (1997), pp. 43–54. DOI: 10.1080/019722497129278. (Visited on 04/29/2017).

[235] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Updated. New York, NY: The Free Press, 2017. ISBN: 978-0-262-53348-5.

[236] Bob G. Witmer and Michael J. Singer. "Measuring Presence in Virtual Environments: A Presence Questionnaire". In: *Presence: Teleoperators and Virtual Environments* 7.3 (June 1998), pp. 225–240. ISSN: 1054-7460, 1531-3263. DOI: 10.1162/105474698565686. (Visited on 03/15/2019).

[237] Dominic Arsenault. "Dark Waters: Spotlight on Immersion". In: *GAMEON-NA International Conference Proceedings*. Ghent: Eurosis, 2005, pp. 50–52. URL: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/13052 (visited on 03/11/2019).

[238] Alison McMahan. "Immersion, Engagement, and Presence: A Method for Analyzing 3-D Video Games". In: *The Video Game Theory Reader*. Ed. by Mark J. P. Wolf and Bernard Perron. New York, NY: Routledge, 2003, pp. 67–86.

[239] Laura Ermi and Frans Mäyrä. "Fundamental Components of the Gameplay Experience: Analysing Immersion". In: *Worlds in Play: International Perspectives on Digital Games Research*. Ed. by Jennifer Jenseon and Suzanne de Castell. New York, NY: Peter Lang Publishing, 2005, pp. 15–27.

[240] Richard J. Gerrig. *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Abingdon, Oxon: Routledge, 2018. ISBN: 978-0-300-05434-7.

[241] Mel Slater. *A Note on Presence Terminology*. 2003. URL: http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm (visited on 03/10/2019).

[242] Ernest Adams. *Fundamentals of Game Design*. 3rd edition. USA: Pearson Education, 2014. ISBN: 978-0-321-92967-9.

[243] Marie-Laure Ryan. *Narrative as Virtual Reality 2: Revisiting Immersion and Interactivity in Literature and Electronic Media*. Baltimore, MD: John Hopkins Univeristy Press, 2015.

[244] Marie-Laure Ryan. "Interactive Narrative, Plot Types, and Interpersonal Relations". In: *Interactive Storytelling*. Ed. by Ulrike Spierling and Nicolas Szilas. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, pp. 6–13. ISBN: 978-3-540-89454-4.

[245] Niels Christian Nilsson, Rolf Nordahl, and Stefania Serafin. "Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence". en. In: *Human Technology* 12.2 (Nov. 2016), pp. 108–134. ISSN: 17956889. DOI: 10.17011/ht/urn.201611174652. (Visited on 03/11/2019).

[246] Jesse Schell. *The Art of Game Design: A Book of Lenses*. 2nd edition. Burlington, MA USA: Morgan Kaufmann, 2015. ISBN: 978-1-4665-9864-5.

[247] Dario D. Salvucci and Niels A. Taatgen. "Threaded Cognition: An Integrated Theory of Concurrent Multitasking." en. In: *Psychological Review* 115.1 (2008), pp. 101–130. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/0033-295X.115.1.101. (Visited on 03/19/2019).

[248] Mary L. Courage, Aishah Bakhtiar, Cheryll Fitzpatrick, et al. "Growing Up Multitasking: The Costs and Benefits for Cognitive Development". In: *Developmental Review*. Special Issue: Living in the "Net" Generation: Multitasking, Learning, and Development 35 (Mar. 2015), pp. 5–41. ISSN: 0273-2297. DOI: 10.1016/j.dr.2014.12.002. (Visited on 03/19/2019).

[249] Cal Newport. *Digital Minimalism: Choosing a Focused Life in a Noisy World*. New York: Penguin Random House LLC, 2019. ISBN: 978-0-525-54287-2.

[250] Oxford Reference, ed. *Affective response*. Oxford, 2019. DOI: 10.1093/oi/authority.20110803095354414. (Visited on 03/11/2019).

[251] Suning Zhu, Ashish Gupta, David Paradice, et al. "Understanding the Impact of Immersion and Authenticity on Satisfaction Behavior in Learning Analytics Tasks". In: *Information Systems Frontiers* (June 2018). ISSN: 1572-9419. DOI: 10.1007/s10796-018-9865-4. URL: https://doi.org/10.1007/s10796-018-9865-4 (visited on 03/07/2019).

[252] Sasha A. Barab, Kurt D. Squire, and William Dueber. "A Co-Evolutionary Model for Supporting the Emergence of Authenticity". In: *Educational Technology, Research and Development* 48.2 (2000), p. 37. URL: https://search.proquest.com/docview/218050622?accountid=12309 (visited on 03/22/2019).

[253] Stijn Van Laer and Jan Elen. "In Search of Attributes that Support Self-Regulation in Blended Learning Environments". English. In: *Education and Information Technologies; New York* 22.4 (July 2017), pp. 1395–1454. ISSN: 13602357. DOI: 10.1007/s10639-016-9505-x. (Visited on 03/22/2019).

[254] Pan Liu, Simon Rigoulot, and Marc D. Pell. "Cultural Immersion Alters Emotion Perception: Neurophysiological Evidence from Chinese Immigrants to Canada". In: *Social Neuroscience* 12.6 (Nov. 2017), pp. 685–700. ISSN: 1747-0919. DOI: 10.1080/17470919.2016.1231713. (Visited on 03/25/2019).

[255] Larry Cahill and James L. McGaugh. "Mechanisms of Emotional Arousal and Lasting Declarative Memory". In: *Trends in Neurosciences* 21.7 (July 1998), pp. 294–299. ISSN: 0166-2236. DOI: 10.1016/S0166-2236(97)01214-9. (Visited on 03/25/2019).

[256] Jessie Pallud. "Impact of Interactive Technologies on Stimulating Learning Experiences in a Museum". In: *Information & Management* 54.4 (June 2017), pp. 465–478. ISSN: 0378-7206. DOI: 10.1016/j.im.2016.10.004. (Visited on 03/25/2019).

[257] Jennifer StGeorge, Allyson Holbrook, and Robert Cantwell. "Affinity for Music: A Study of the Role of Emotion in Musical Instrument Learning". In: *International Journal of Music Education* 32.3 (Aug. 2014), pp. 264–277. DOI: 10.1177/0255761413491178. (Visited on 03/25/2019).

[258] Reinhard Pekrun, Thomas Goetz, Wolfram Titz, et al. "Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research". In: *Educational Psychologist* 37.2 (June 2002), pp. 91–105. ISSN: 0046-1520. DOI: 10.1207/S15326985EP3702_4. (Visited on 03/25/2019).

[259] Zeying Wan, Yinglei Wang, and Nicole Haggerty. "Why People Benefit from E-Learning Differently: The Effects of Psychological Processes on E-Learning Outcomes". In: *Information & Management* 45.8 (Dec. 2008), pp. 513–521. ISSN: 0378-7206. DOI: 10.1016/j.im.2008.08.003. (Visited on 03/25/2019).

[260] Aleck C. H. Lin, Walter D. Fernandez, and Shirley Gregor. "Understanding Web Enjoyment Experiences and Informal Learning: A Study in a Museum Context". In: *Decision Support Systems*. 1) Computational Approaches to Subjectivity and Sentiment Analysis 2) Service Science in Information Systems Research : Special Issue on PACIS 2010 53.4 (Nov. 2012), pp. 846–858. ISSN: 0167-9236. DOI: 10.1016/j.dss.2012.05.020. (Visited on 03/25/2019).

[261] Alice Hall. "Perceptions of the Authenticity of Reality Programs and Their Relationships to Audience Involvement, Enjoyment, and Perceived Learning". In: *Journal of Broadcasting & Electronic Media* 53.4 (Nov. 2009), pp. 515–531. ISSN: 0883-8151. DOI: 10.1080/08838150903310468. (Visited on 03/25/2019).

[262] Roy F. Baumeister, Jessica L. Alquist, and Kathleen D. Vohs. "Illusions of Learning: Irrelevant Emotions Inflate Judgments of Learning". en. In: *Journal of Behavioral Decision Making* 28.2 (2015), pp. 149–158. ISSN: 1099-0771. DOI: 10.1002/bdm.1836. (Visited on 03/25/2019).

[263] David Finch, Melanie Peacock, Donna Lazdowski, et al. "Managing Emotions: A Case Study Exploring the Relationship Between Experiential Learning, Emotions, and Student Performance". In: *The International Journal of Management Education* 13.1 (Mar. 2015), pp. 23–36. ISSN: 1472-8117. DOI: 10.1016/j.ijme.2014.12.001. (Visited on 03/25/2019).

[264] Dianne Cyr. "Modeling Web Site Design Across Cultures: Relationships to Trust, Satisfaction, and E-Loyalty". In: *Journal of Management Information Systems* 24.4 (Apr. 2008), pp. 47–72. ISSN: 0742-1222, 1557-928X. DOI: 10.2753/MIS0742-1222240402. (Visited on 03/25/2019).

[265] Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. "On-Line Trust: Concepts, Evolving Themes, A Model". In: *International Journal of Human-Computer Studies*. Trust and Technology 58.6 (June 2003), pp. 737–758. ISSN: 1071-5819. DOI: 10.1016/S1071-5819(03)00041-7. (Visited on 03/25/2019).

[266] Gerald L. Lohse and Peter Spiller. "Electronic Shopping". In: *Commun. ACM* 41.7 (July 1998), pp. 81–87. ISSN: 0001-0782. DOI: 10.1145/278476.278491. (Visited on 03/25/2019).

[267] Jakob Nielsen, Rolf Molich, Carolyn Snyder, et al. *SearcH: E-Commerce User Experience*. Tech. rep. Vol 5. John Wiley & Sons, 2000, pp. 1–23.

[268] Bruce B. Frey. *Self-efficacy*. Mar. 2019. URL: https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation (visited on 03/25/2019).

[269] Jasmin C. Hutchinson, Todd Sherman, Nevena Martinovic, et al. "The Effect of Manipulated Self-Efficacy on Perceived and Sustained Effort". In: *Journal of Applied Sport Psychology* 20.4 (Oct. 2008), pp. 457–472. ISSN: 1041-3200. DOI: 10.1080/10413200802351151. URL: https://doi.org/10.1080/10413200802351151 (visited on 03/25/2019).

[270] "A Practical Perspective on the Self-Efficacy of Middle-Managers: Strategies to Bolster Task Accomplishment". In: *Development and Learning in Organizations: An International Journal* 17.4 (Aug. 2003), pp. 4–6. ISSN: 1477-7282. DOI: 10.1108/14777280310485337. (Visited on 03/25/2019).

[271] Mitchell Hamilton, Velitchka D. Kaltcheva, and Andrew J. Rohm. "Social Media and Value Creation: The Role of Interaction Satisfaction and Interaction Immersion". In: *Journal of Interactive Marketing* 36 (Nov. 2016), pp. 121–133. ISSN: 1094-9968. DOI: 10.1016/j.intmar.2016.07.001. (Visited on 03/22/2019).

[272] Kathryn Sutherland, Marilyn Deegan, Elena Pierazzo, et al. *Jane Austen's Fiction Manuscripts*. Oxford, 2019. URL: https://janeausten.ac.uk/index.html (visited on 04/01/2019).

[273] Jane H. Ohlmeyer, Micheál Ó Siochrú, Thomas Bartlett, et al. *1641 Depositions*. Dublin, 2010. URL: http://1641.tcd.ie/index.php (visited on 04/01/2019).

[274] Saraswathi Bellur and S. Shyam Sundar. "Talking Health With a Machine: How Does Message Interactivity Affect Attitudes and Cognitions?: Interactivity as Conversation". en. In: *Human Communication Research* 43.1 (Jan. 2017), pp. 25–53. ISSN: 03603989. DOI: 10.1111/hcre.12094. (Visited on 04/02/2019).

[275] Sheizaf Rafaeli. "Interactivity: From New Media to Communication". In: *Advancing Communication Science: Merging Mass and Interpersonal Prcoesses*. Ed. by Robert P. Hawkings, John M. Wiemann, and Pingree Suzanne. Newbury Park, CA: SAGE Publications, Inc., 1988, pp. 110–134.

[276] Paula Findlen, Dan Edelstein, and Nicole Coleman. *Mapping the Republic of Letters*. Stanford, CA, 2013. URL: http://republicofletters.stanford.edu/index.html (visited on 04/01/2019).

[277] Mark Stansbury and David Kelly. *Earlier Latin Manuscripts*. Galway, Ireland, 2016. URL: https://elmss.nuigalway.ie (visited on 04/01/2019).

[278] Elton Barton, Stefan Bouzarovski, Chris Pelling, et al. *Hestia*. Milton Keynes, 2014. URL: https://hestia.open.ac.uk/ (visited on 04/01/2019).

[279] Patricia O Connell. *The Irish College at Alcalá de Henares: 1649-1785*. Dublin, Ireland: Four Courts Press, 1997. ISBN: 978-1-85182-345-1.

[280] Thomas O'Connor. "The Domestic and International Roles of Irish Overseas Colleges, 1590-1800". In: *College Communities Abroad: Education, Migration and Catholicism in Early Modern Europe*. Ed. by Liam Chambers and Thomas O'Connor. Manchester, UK: Manchester University Press, 2018.

[281] Micheline Kerney Walsh. "The Irish College of Alcalá de Henares". In: *Seanchas Ardmhacha: Journal of the Armagh Diocesan Historical Society* 11.2 (1985), pp. 247–257. ISSN: 0488-0196. DOI: 10.2307/29742353. (Visited on 07/17/2019).

[282] John Keating, Aja Teehan, Damien Gallagher, et al. "A Digital Edition of a Spanish 18th Century Account Book: User Driven Digitisation". In: *Jahrbuch für Computerphilologie* 10 (2010), pp. 169–188. URL: http://eprints.maynoothuniversity.ie/2650/ (visited on 04/15/2018).

[283] Shane McGarry. *The Alcalá Record Books*. Maynooth, IE, 2018. URL: http://jkeatingsrv.cs.nuim.ie/ (visited on 07/17/2019).

[284] Carlos E. Otero. *Software Engineering Design: Theory and Practice*. Boca Raton, FL: CRC Press, 2012. ISBN: 978-1-4398-5168-5.

[285] Susan Schreibman, Meredith Dabek, Noel Groome, et al. *The Woodman Diary*. Maynooth, Co. Kildare, Ireland, June 2015. URL: http://dhprojects.maynoothuniversity.ie/woodman/ (visited on 04/26/2017).

[286] Alex Rüger and Henk Wals. *Vincent Van Gogh The Letters*. Amsterdam, 2009. URL: http://vangoghletters.org/vg/letters.html (visited on 07/17/2019).

[287] Richard L. Moreland, John M. Levine, and Melissa L. Wingert. "Creating the Ideal Group: Composition Effects at Work". In: *Understanding Group Behaviour: Small Gorup Processes and Interpersonal Relations*. Ed. by Erich Witte and James H. Davis. eBook. Vol. 2. New York, NY: Psychology Press, 2009, pp. 11–36.

[288] John Pruitt and Tamara Adlin. *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. San Fransico, CA: Morgan Kaufmann, 2006.

[289] A. Gersie. *Reflections on Therapeutic Storymaking: The Use of Stories in Groups*. LCCB: lc97183739. London: Jessica Kingsley Publishers, 1997. ISBN: 978-1-85302-272-2.

[290] Kim Dent-Brown and Michael Wang. "Developing a rating scale for projected stories." In: *Psychology and psychotherapy* 77.Pt 3 (Sept. 2004), pp. 325–333.

[291] Maynooth University. *Maynooth University Research Ethics Policy*. 2015. URL: https://www.maynoothuniversity.ie/sites/default/files/assets/document/Research%20Ethics%20Policy_0.pdf%20August%202015_0.pdf (visited on 07/19/2019).

[292] Alireza Moghaddam. "Coding Issues in Grounded Theory". In: *Issues in Educational Research* 16.1 (2006), pp. 52–66. URL: http://iier.org.au/iier16/moghaddam.html (visited on 08/28/2019).

[293] Juliet Corbin and Anselm Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 4th. USA: SAGE Publications, Inc., 2015. ISBN: 978-1-4129-9746-1.

[294] David Eltis and Allen Tullos. *Slave Voyages: The Trans-Atlantic Slave Trade Database*. Atlanta, GA USA, 2008. URL: https://www.slavevoyages.org/ (visited on 07/23/2019).

[295] Shane McGarry. *Designing from a Narrative: Leveraging the 6-Part Story Method as a Focus Group Activity to Facilitate User-Focused Design Discussions*.

[296] Eric Gamma, Richard Helm, Ralph Johnson, et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series 8. London: Addison-Wesley, 1995. ISBN: 978-0-201-63361-0.

[297] L.D. Rosen, K. Whaling, L.M. Carrier, et al. "The Media and Technology Usage and Attitudes Scale: An Empirical Investigation". In: *Computers in human behavior* 29.6 (Nov. 2013), pp. 2501–2511. ISSN: 0747-5632. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4338964/ (visited on 07/31/2019).

[298] John J. Shaughnessy, Eugene B. Zechmeister, and Jeanne S. Zechmeister. *Research Methods in Psychology*. 10th. New York, NY: McGraw Hill Education, 2015. ISBN: 978-0-07-782536-2.

[299] Bettina Laugwitz, Theo Held, and Martin Schrepp. "Construction and Evaluation of a User Experience Questionnaire". In: *USAB*. 2008. DOI: 10.1007/978-3-540-89350-9_6. (Visited on 10/02/2019).

[300] Sharon Lawner Weinberg and Sarah Knapp Abramowitz. *Statistics Using SPSS: An Integrative Approach*. 2nd. New York, NY: Cambridge University Press, 2008. ISBN: 978-1-107-46122-2.

[301] David S. Moore, George P. McCabe, and Bruce A. Craig. *Introduction to the Practice of Statistics*. 9th edition. New York, NY: W.H. Freeman and Company, 2016. ISBN: 978-1-319-01338-7.

[302] William H. Kruskal and W. Allen Wallis. "Use of Ranks in One-Criterion Variance Analysis". In: *Journal of the American Statistical Association* 47.260 (Dec. 1952), pp. 583–621. ISSN: 0162-1459. DOI: 10.1080/01621459.1952.10483441. (Visited on 08/18/2019).

[303] Frank Wilcoxon. "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. ISSN: 0099-4987. DOI: 10.2307/3001968. (Visited on 08/18/2019).

[304] G. E. P. Box and Paul W. Tidwell. "Transformation of the Independent Variables". In: *Technometrics* 4.4 (1962), pp. 531–550. ISSN: 0040-1706. DOI: 10.2307/1266288. (Visited on 08/12/2019).

[305] Barbara G. Tabachnick and Linda S. Fidell. *Using Multivariate Statistics*. 6th edition. Essex, UK: Pearson Education Limited, 2014. ISBN: 978-0-205-84957-4.

[306] David W. Jr. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied Logisitc Regression*. 3rd. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, 2013. ISBN: 978-0-470-58247-3.

[307] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Construction of a Benchmark for the User Experience Questionnaire (UEQ)". In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4.4 (2017), p. 40. ISSN: 1989-1660. DOI: 10.9781/ijimai.2017.445. (Visited on 08/12/2019).

[308] Russell T. Warne. "A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists". In: *Practical Assessment, Research & Evaluation* 19.17 (2014), pp. 1–10. URL: https://pareonline.net/pdf/v19n17.pdf (visited on 10/23/2019).

[309] Harold Hotelling. "The Generalization of Student's Ratio". In: *The Annals of Mathematical Statistics* 2.3 (Aug. 1931), pp. 360–378. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177732979. (Visited on 08/18/2019).

[310] Shane A. McGarry. *Keyword Rseponse Transcription*. 2019.

[311] Pamela Pavliscak. *Data-Informed Product Design*. Sebastopol, CA: O'Reilly Media, 2015. URL: https://www.oreilly.com/library/view/data-informed-product-design/ (visited on 08/20/2019).

[312] Jonathan St. B. T. Evans, Stephen E. Newstead, and Ruth M.J. Byrne. *Human Reasoning: The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum Associates, 1993. ISBN: 978-0-86377-314-3.

[313] Irving M. Copi, Carl Cohen, and Daniel E. Flage. *Essentials of Logic*. 2nd. Abingdon, Oxon: Routledge, 2007. ISBN: 978-0-13-238034-8.

[314] Shane A. McGarry. *Visualisation Response Transcription*. 2019.

[315] Stijn de Rammelaere, Els Stuyven, and André Vandierendonck. "Verifying Simple Arithmetic Sums and Products: Are the Phonological Loop and the Central Executive Involved?" en. In: *Memory & Cognition* 29.2 (Mar. 2001), pp. 267–273. ISSN: 1532-5946. DOI: 10.3758/BF03194920. (Visited on 08/24/2019).

[316] Ilona Friso-van den Bos, Sanne H. G. van der Ven, Evelyn H. Kroesbergen, et al. "Working Memory and Mathematics in Primary School Children: A Meta-Analysis". In: *Educational Research Review* 10 (Dec. 2013), pp. 29–44. ISSN: 1747-938X. DOI: 10.1016/j.edurev.2013.05.003. (Visited on 08/24/2019).

[317] Eva Van de Weijer-Bergsma, Evelyn H. Kroesbergen, and Johannes E. H. Van Luit. "Verbal and Visual-Spatial Working Memory and Mathematical Ability in Different Domains Throughout Primary School". In: *Memory & Cognition* 43.3 (2015), pp. 367–378. ISSN: 0090-502X. DOI: 10.3758/s13421-014-0480-4. (Visited on 08/24/2019).

[318] Dale Shaffer, Wendy Doube, and Juhani Tuovinen. "Applying Cognitive Load Theory to Computer Science Education". In: *Proceedings on the Joint Conference EASE & PPIG*. Ed. by M. Petre and D. Budgen. Keele, UK, 2003, pp. 333–346. URL: https://www.researchgate.net/publication/250790986_Applying_Cognitive_Load_Theory_to_Computer_Science_Education (visited on 08/24/2013).

[319] Yvette Shen. "Placing Graphic Design at the Intersection of Information Visualization Fields". In: *Digital Humanities Quarterly* 12.4 (2018). URL: http://www.digitalhumanities.org/dhq/vol/12/4/000406/000406.html (visited on 09/13/2019).

[320] Matthew A. Vetter. "Archive 2.0: What Composition Students and Academic Libraries Can Gain from Digital-Collaborative Pedagogies". In: *Composition Studies; Chicago* 42.1 (2014), pp. 35–53, 157. ISSN: 1534-9322. URL: http://search.proquest.com/docview/1613617049/abstract/CF244CB0A91841ACPQ/1 (visited on 09/13/2019).

[321] Philippa Lyon, Patrick Letschka, Tom Ainsworth, et al. "Drawing Pedagogies in Higher Education: the Learning Impact of a Collaborative Cross-Disciplinary Drawing Course". In: *International Journal of Art & Design Education* 37.2 (May 2018), pp. 221–232. ISSN: 14768062. DOI: 10.1111/jade.12106. (Visited on 09/13/2019).

[322] Anna Maldoni. "A Cross-Disciplinary Approach to Embedding: A Pedagogy for Developing Academic Literacies". In: *Journal of Academic Language & Learning* 11.1 (2017), A104–A124. URL: http://journal.aall.org.au/index.php/jall/article/viewArticle/422 (visited on 09/13/2019).

[323] Kirsten Zimbardi and Paula Myatt. "Embedding Undergraduate Research Experiences within the Curriculum: A Cross-Disciplinary Study of the Key Characteristics Guiding Implementation". In: *Studies in Higher Education* 39.2 (Feb. 2014), pp. 233–250. ISSN: 0307-5079. DOI: 10.1080/03075079.2011.651448. (Visited on 09/13/2019).

[324] W. Fukuoka and J. H. Spyridakis. "Japanese Readers' Comprehension of and Preferences for Inductively Versus Deductively Organized Text". In: *IEEE Transactions on Professional Communication* 43.4 (Dec. 2000), pp. 355–367. DOI: 10.1109/47.888811. (Visited on 10/02/2019).

[325] Sepideh Yalda, Sepideh Yalda, Gary Zoppetti, et al. "Interactive Immersion Learning". In: *Bulletin of the American Meteorological Society* 93.12 (Dec. 2012), pp. 1811–1813. ISSN: 0003-0007. DOI: 10.1175/BAMS-93-12-1811. (Visited on 09/12/2019).

[326] Erin L. Robinson. "Immersion Learning in Social Work Education: A Pedagogical Tool for Enriching Knowledge and Practice Skills Among BSW Students". In: *Journal of Teaching in Social Work* 38.5 (Oct. 2018), pp. 536–550. ISSN: 0884-1233. DOI: 10.1080/08841233.2018.1516712. (Visited on 09/12/2019).

[327] Rosário Couto Costa. "The Place of the Humanities in Today's Knowledge Society". In: *Palgrave Communications* 5.1 (Apr. 2019), pp. 1–5. ISSN: 2055-1045. DOI: 10.1057/s41599-019-0245-6. (Visited on 10/02/2019).

[328] Lars Erik Udsen and Anker Helms Jørgensen. "The Aesthetic Turn: Unravelling Recent Aesthetic Approaches to Human-Computer Interaction". In: *Digital Creativity* 16.4 (Jan. 2005), pp. 205–216. ISSN: 1462-6268. DOI: 10.1080/14626260500476564. (Visited on 09/05/2019).

[329] John Unsworth. "Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?" In: London, UK: Kings College, May 2000. URL: http://people.virginia.edu/~jmu2m//Kings.5-00/primitives.html (visited on 09/13/2019).

[330] Carole L. Palmer, Lauren C. Teffeau, and Carrie M. Pirmann. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Library Report. Dublin, OH: OCLC Research, 2009, p. 57. URL: https://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf (visited on 09/13/2019).

[331] Lorna Hughes, Panos Constantopoulos, and Costis Dallas. "Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines". In: *A New Companion to Digital Humanities*. 2nd ed. Blackwell Companions to Literature and Culture 93. Malden, MA USA: John Wiley & Sons, 2016, pp. 192–213.

[332] Terras Warwick, Julianne Nyhan, Melissa Terras, et al. *Digital Humanities in Practice*. London, UK: Facet Publishing, 2012. ISBN: 978-1-85604-905-4. URL: http://ebookcentral.proquest.com/lib/nuim/detail.action?docID=1167435 (visited on 09/13/2019).

[333] Flavio Gortana, Franziska von Tenspolde, Daniela Guhlmann, et al. "Off the Grid: Visualizing a Numismatic Collection as Dynamic Piles and Streams". eng. In: *Open Library of Humanities* 4.2 (Oct. 2018), p. 30. ISSN: 2056-6700. DOI: 10.16995/olh.280. (Visited on 09/05/2019).

[334] Katie Chenoweth, Alexander Baron-Raiffe, and Rebecca Sutton Koeser. *Derrida's Margins*. 2018. URL: http://derridas-margins.princeton.edu (visited on 09/05/2019).

[335] Courtney Rivard, Taylor Arnold, and Lauren Tilton. "Building Pedagogy into Project Development: Making Data Construction Visible in Digital Projects". In: *Digital Humanities Quarterly* 013.2 (Aug. 2019). ISSN: 1938-4122. URL: http://www.digitalhumanities.org/dhq/vol/13/2/000419/000419.html (visited on 09/13/2019).

[336] John Ousterhout. *A Philosophy of Software Design*. 1st. Palo Alto, CA: Yaknyam Press, 2018. ISBN: 978-1-73210-220-0.

[337] Damian Rouson, Jim Xia, and Xu Xiaofeng. *Scientific Software Design: The Object-Oriented Way*. New York, NY: Cambridge University Press, 2011. ISBN: 978-0-521-88813-4.

[338] Andreea Molnar and Patty Kostkova. "Learning Through Interactive Digital Narratives". In: *Interative Digital Narrative: History, Theory and Practice*. Ed. by Hartmut Koenitz, Gabriele Ferri, Mads Haahr, et al. Routledge Studies in European Communication Research and Education. New York, NY: Routledge, 2015, pp. 200–210.

[339] Arvind Satyanarayan and Jeffrey Heer. "Authoring Narrative Visualizations with Ellipsis". In: *Computer Graphics Forum* 33.3 (2014), pp. 361–370. ISSN: 1467-8659. DOI: 10.1111/cgf.12392. (Visited on 09/05/2019).

[340] Marc Cajolet, Sarah Bloomer, and Alexandra Stevens. "Designing Around the Whole Story with User Narratives". In: *UX Magazine* 1295 (Aug. 2014). URL: https://uxmag.com/articles/designing-around-the-whole-story-with-user-narratives (visited on 10/02/2019).

[341] *Tf-idf :: A Single-Page Tutorial - Information Retrieval and Text Mining*. URL: http://www.tfidf.com/ (visited on 07/23/2019).