A System Framework for Concept- and Credibility-Based Multimedia Retrieval

5 authors, including:

Ralf Bierig
National College of Ireland

44 PUBLICATIONS   414 CITATIONS

SEE PROFILE

Mihai Lupu
Research Studios Austria

148 PUBLICATIONS   1,190 CITATIONS

SEE PROFILE

Allan Hanbury
TU Wien

304 PUBLICATIONS   5,808 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

KHRESMOI View project

MUCKE EU-IST View project

A System Framework for Concept- and Credibility-Based Multimedia Retrieval

Ralf Bierig
Vienna University of
Technology
Favoritenstr. 9-11 / 188
1040 Vienna, Austria
bierig@ifs.tuwien.ac.at

Cristina Serban
Alexandru Ioan Cuza
University of Iasi
Iasi, Romania
cristina.serban@
info.uaic.ro

Alexandra Siriteanu
Alexandru Ioan Cuza
University of Iasi
Iasi, Romania
alexandra.siriteanu@info.uaic.ro

Mihai Lupu
Vienna University of
Technology
Favoritenstr. 9-11 / 188
1040 Vienna, Austria
lupu@ifs.tuwien.ac.at

Allan Hanbury
Vienna University of
Technology
Favoritenstr. 9-11 / 188
1040 Vienna, Austria
hanbury@ifs.tuwien.ac.at

ABSTRACT

We present a multimedia retrieval system framework that incorporates components for processing multimedia content in different modes and languages. The framework provides concept-based information retrieval facilities that applies credibility information for result re-ranking. The architecture combines both a direct user interface and a batched evaluation interface for reproducible research in multimedia IR. The demo presents a preliminary version of the system framework and shows a use case based on the ImageCLEF 2011 Wikipedia test collection.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing methods*

General Terms

Concept Learning, Credibility, Experimentation, Framework

1. INTRODUCTION

Over the past 10 years, the growing social web has established an impressive selection of very large and complex data sets. The success of this movement is omnipresent: There are at least 200 social networks[1] currently active that attract 100s of millions of visitors monthly. Wikipedia as one major example is not only the most comprehensive encyclopedia,

---

[1] http://en.wikipedia.org/wiki/List_of_social_networking_websites

but has demonstrated to match the quality of commercially edited counterparts (e.g. Britannica) [8]. People spend considerable amounts of time on social networks where they contribute and shape the information landscape of the future. Flickr draws nearly 70 million visitors monthly to its collaborative image collection and 30 billion pieces of content are shared on Facebook every month [6]. The knowledge contained in these massive data networks is impressive and could be harvested and put to use for a wide range of subject areas. Although research has already been directed toward automatic information mining from these rich sources, the problem of knowledge extraction from multimedia content remains difficult. The main challenges are the heterogeneity of the data, the scalability of the processing methods and the reliability of the results. The CHIST-ERA MUCKE project focuses on these areas and aims to create an architecture and testbed that integrates and evaluates a scalable range of methods to overcome these challenges. In this demo, we present a preliminary prototype of the system framework that highlights the overall architecture and shows a use case of its application based on the ImageCLEF 2011 Wikipedia test collection. The demo provides a user interface where the general searcher can retrieve images from the forementioned collection based on their underlying concepts and results are ranked by their credibility properties. We equally address the researcher as a user of the system by offering a batch interface that allows for automatic TREC-style evaluations. Our aim is that by adding the notion of concept and credibility to multimedia search, we can help people to extract content that was formerly hidden and, in addition, rank the importance of results based on an alternative feature. The next section describes work that is closely related. We present the main aspects of our approach in Section 2 and describe the system architecture for our demo in Section 3 together with an overview of how researchers and users work with it and how new test collections are integrated. The ImageCLEF 2011 collection is applied as an example collection for our demo. We elaborate on our plans for the future in Section 4.

2. EXTENDING MULTIMEDIA RETRIEVAL

This section describes our focus on multimedia retrieval — mainly by extending traditional text and image indexing with a combined multi-modal concept index and using credibility as an alternative for re-ranking search results.

2.1 Concept Indexing

Automatically estimating similarity between texts is an old problem. Purely data driven methods, such as Latent Semantic Analysis (LSA) and Explicit Semantic Analysis (ESA) [2, 7] were proposed and perform well. Multimedia document similarity, on the other hand, is more difficult due to their much larger conceptual spaces. Methods that exploit low level features, such as colour and texture [9], and bag-of-visual-words [1] have problems capturing and describing the semantic content of images adequately. Recent studies apply object recognition that rely on invariant local descriptors [5]. Recently, the use of a large number of semantic concept detectors is proposed for multimedia retrieval [3, 4]. An important innovation of MUCKE will be the proposition of a multimedia concept similarity framework for reliable concept recognition that will combine text and image information within a single framework. Using LSA as a baseline for our proposed framework, we address the limitations of LSA being based too narrowly on bag-of-word representations and ignoring important linguistic phenomena in the underlying content. The distinction between text and image concepts takes into account their different nature. Whereas textual concepts are often high level (e.g. a car), visual concepts tend to focus on very specific low level attributes (e.g. metallic texture of a car). Combining both flavors of concepts in a common framework make them accessible for indexing and searching.

2.2 Credibility

An IR system may be viewed in four coarse-grained levels: a user, the content, a ranking or a retrieval system, and a user interface. The credibility model used in MUCKE distinguishes four aspects: *expertise*, *trustworthiness*, *quality*, and *reliability*, with the following definitions:

- *expertise* denotes the knowledge ability to provide truthful information;

- *trustworthiness* denotes the estimated intent to provide truthful information;

- *quality* denotes the ability to convey the truthfulness of the information provided;

- *reliability* denotes the consistency of the object's ability to provide information.

Each of the four above-mentioned aspects may be applied to each of the four levels where the issue of credibility may appear. In doing a credibility assessment, while it would be desirable to consider the system as a whole, in practice we may choose to focus on a particular aspect and a particular level of credibility. For instance, in the case of multimedia streams consisting of images and tags associated with them, we may focus on the credibility of the user by evaluating the quality of the tags the user created thus assessing the credibility of the images based on the user creating their tags. The aim of the system framework therefore is to support multimedia retrieval focused on the credibility of its user base, the content it retrieves, the algorithms that are used to retrieve and rank its content, and the user interface.

3. SYSTEM ARCHITECTURE

We describe the system architecture with respect to the overarching principles that drove our design process and then highlight the individual components. We then focus on how the framework satisfies different kinds of users and briefly describe how content collections are integrated.

3.1 Architectural Principles

In the context of credibility and concept-based multimedia, as described above, the framework offers a set of architectural principles that make it appealing for the research community:

- *Highly modularized and configurable* components allow researchers to apply the various parts of the framework as independent variables for their experiments.

- *Evaluation-based* research is supported by the ability to evaluate diverse settings as variables using automatic batch execution.

- *Reproducible research* is fostered by capturing configurations of key components as part of the evaluation and storing it for documentation and reuse. Researchers can retrieve not only the results but also the settings that generated the results.

- The framework is *extensible* in terms of

  - *new test collections* to better evaluate existing methods in different content settings and

  - *new logic* (e.g. new indexing and search methods) to widen the scope of retrieval and explore new algorithms and approaches.

3.2 Framework Components

The components of the system framework are depicted in Figure 1 and represent the conceptual elements of the architecture:

- The *concept* package describes the concept model. Every indexed document consists of a list of Field objects. There are three types of Field objects: TextField, TagField and ImageField corresponding to the different types of data a document can contain: text, tags and images. By processing these fields, a list of concept objects is obtained. We make a distinction between two types of concepts: textual and visual. The distinction is motivated by their different nature. Whereas textual concepts tend to be high level (e.g. tree, house or car), visual concepts are often treated as low level features (e.g. rectangular shapes, red color or wood surface pattern). Combining both kinds of concepts in a single framework provides the advantage of combining both levels of granularity and taking advantage from each level of detail. To extract concepts from fields, three document field types (text, tag and image) are applied that may have alternative implementations. This leads to a potentially large variety of ways with which concepts can be extracted from multimedia content.
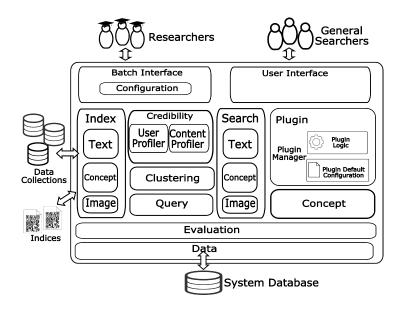
Figure 1: Components of the Concept and Credibility-Based Multimedia IR Framework

- The *credibility* package provides an interface for the four different types of credibility described in Section 2.2. The system framework currently encapsulates two of the four types of credibility. First, the credibility of the user who produced the content (a user who created an image on Flikr) is maintained by the user profiler. Second, the objective credibility of the content (e.g. the source, language or link structure of a document) that is generated by the content profiler. The credibility of the ranking or a retrieval system and the credibility of the user interface is currently not addressed but interfaces are provided to add these elements in the future.

- The *index* package offers interfaces to a text index, an image index and a concept index. The text index is represented by a classic Lucene[2] inverted text index where indexed documents are represented by a list of tokenized text fields. The image index is implemented using Lucene Lire[3] that incorporates image features (e.g. color distributions). The concept index uses the concept package to transform text, tags and images into text concepts (TConcepts) and visual concepts (VConcepts) before storing them into a concept-based index storage that is made available for concept search by the search package.

- The *search* package provides search functionality for queries using text, images and concepts in any combination. For every search, the query text, the image ID or a list of concepts will be provided along with an optional list. Already existing results may be applied as a filter to process the final list of results. The search is separated from index and query to keep the framework flexible and because it is generally required to involve a combination of several indices, either in parallel or sequentially.

- The *clustering* package provides tools for merging results into ranked lists, basically tools for re-ranking search results of different modalities while considering the similarity and the credibility of the content.

- The *query* package processes input strings or input image collections into unified query objects that consist of fields. There are three kinds of fields representing the three types of queries the framework handles: text, tags and image. These fields allow queries to be arbitrary combinations of different multimodal elements that can be used interchangeably within the framework.

- The *batch interface* provides batch-access to the framework where evaluations can be configured and executed by researchers. Currently, this interface is controlled by configuration files, one for each type of evaluation run. The *user interface* package provides a Google-like web interface for general users to query for images.

- The *plugin* package enables researchers to inject additional functionality — in particular configuration and logic relating to additional IR test collections. This is key for an extensible system architecture. A plugin optionally relates to a collection outside the code base (via a link) and may also generate own indices. At the minimum, it has a configuration that defines the location of collections and indices and provides the parametrization of additional operations. Plugins, once added, can then be used for batched IR evaluation runs in any combination and with diverse parameterizations within the framework.

- The *evaluation* package deals with the evaluation environment (i.e. the evaluation parameters and the evaluation results) and provides an interface to common IR metrics (e.g. p@10, MAP and R-Precision). Evaluation environments can be expanded with plugins and are stored with the data package.

---

[2]http://lucene.apache.org/core/
[3]http://www.lire-project.net/

- The *data* package provides a data model for the system framework that is stored in a database. Besides keeping a history of evaluation results it also stores the parametric settings of the evaluation runs that produced the results. This is an attempt to not only capture results but also to keep information about how they were generated, hence creating a more holistic and reproducible research landscape.

3.3 User Access

The system framework serves two kinds of users, researchers and general users, and provides user interfaces for these two users groups. An instance of the standard user interface is provided in Figure 2 that shows a minimalistic Google-like image search interface for the general searcher. The
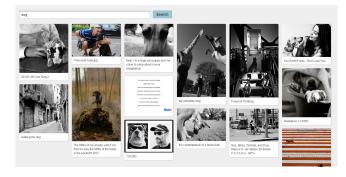

Figure 2: Standard Multimedia User Interface for image search

much more complex batch interface is provided by the means of configuration files that instruct the system framework to perform automatic evaluations on test collections and store both settings and results in the system database through the evaluation package and the data package.

3.4 Content Integration

The system framework provides general support for integrating a wide range of different IR test collections. Alternative test collections can be included as plugins with additional logic and adapted configuration parameters. These plugins define the location and structure of new data sets and inject new logic for indexing and searching. The aim of the framework is to provide a comprehensive set of functionality that minimizes the need for programming, and to reduce the integration work to an adapting of plugin parameters (e.g. by pointing to the test collection or by configuring existing indexing and search parameters of the core API). The demo applies the ImageCLEF 2011 Wikipedia test collection which consists of 237,434 wikipedia images dumped from September 2009 with a wide topical range that are associated with unstructured and noisy text in German, French and English. The collection provides 50 topics with binary relevance judgements. This collection represents a sound example for a use case as it provides collaboratively created multimodal data that crosses three common languages and provides low level text and image features that can be used for concept detection, indexing and search.

4. FUTURE WORK

The framework will be improved and extended throughout the project lifecycle as a platform that hosts our research on multimedia IR as it is needed by all MUCKE project partners. We plan to mature it to a quality where it can be released as an open source project for the benefit of the IR community. Future work will focus on the improvement and extension of the text and visual concept detection components that directly support improved indexing and searching capabilities, and the inclusion of alternative models of credibility that support all four credibility types. Furthermore, we plan to turn the framework into a service architecture that would allow it to work in distributed environments.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision, ECCV*, 1:22, 2004.

[2] E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 1606–1611, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[3] A. Hauptmann, R. Yan, W.-H. Lin, M. Christel, and H. Wactlar. Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. *IEEE Transactions on Multimedia*, 9(5):958–966, 2007.

[4] L.-J. Li, H. Su, Y. Lim, and L. Fei-Fei. Objects as attributes for scene classification. In K. Kutulakos, editor, *Trends and Topics in Computer Vision*, volume 6553 of *Lecture Notes in Computer Science*, pages 57–69. Springer Berlin Heidelberg, 2012.

[5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[6] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute, 2011.

[7] K. Radinsky, E. Agichtein, E. Gabrilovich, and S. Markovitch. A word at a time: computing word relatedness using temporal semantic analysis. In *WWW*, pages 337–346, 2011.

[8] N. J. Reavley, A. J. Mackinnon, A. J. Morgan, M. Alvarez-Jimenez, S. E. Hetrick, E. Killackey, B. Nelson, R. Purcell, M. B. Yap, A. F. Jorm, et al. Quality of information sources about mental disorders: a comparison of wikipedia with centrally controlled web and printed sources. *Psychological medicine*, 42(8):1753, 2012.

[9] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.