



On-Line Retrieval of Health Information Based on Language Complexity, Information Customization and Information Quality

Marco Alfano^{3,4}(✉) , Biagio Lenzitti¹ , Davide Taibi² , and Markus Helfert³ 

¹ Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy
biagio.lenzitti@unipa.it

² Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Palermo, Italy
davide.taibi@itd.cnr.it

³ Lero, Maynooth University, Maynooth, Co. Kildare, Ireland
{marco.alfano, markus.helfert}@lero.ie

⁴ Anghelos Centro Studi sulla Comunicazione, Palermo, Italy

Abstract. A patient, nowadays, acquires on-line health information mainly by means of a search engine. Generic search engines have been shown to be limited and unsatisfactory, at times, because of their generic searches that overload users with the amount of results. Moreover, they are not able to provide customized information to different types of users. At the same time, specific search engines mostly work on medical literature and provide extracts from medical journals that are mainly useful for medical researchers and experts. As a consequence, the found health information may or may not help a user (mainly a non-expert one) for a full comprehension of what he/she is looking for (e.g., a condition, a therapy or a drug). This may negatively affect his/her empowerment process and the interaction with healthcare professionals such as doctors. This work presents a custom search engine, FACILE, that allows to overcome the limitations discussed above by facilitating finding and comprehension of on-line information and thus positively affecting the empowerment process and communication/interaction with healthcare professionals. A user, both a medical expert or non-expert, can specify his/her information requirements in terms of language level, specific information searched, and information quality. FACILE will provide the links to the web pages that comply with such requirements. In particular, FACILE will provide links to web pages with different language levels (simple or complex) and, for each link, the response page will show the requested specific information (condition, therapy, drug, etc.) and a graphical indication (0–5 stars) of the information quality.

Keyword: Digital health patient empowerment · Patient-doctor communication · Health information seeking · User requirements · Structured data · Search engine

1 Introduction

Health empowerment of people and communities is part of the “Framework on integrated people-centred health services” of WHO [1] and drives towards a paradigm shift

© Springer Nature Switzerland AG 2021

M. Ziefle et al. (Eds.): ICT4AWE 2020, CCIS 1387, pp. 1–20, 2021.

https://doi.org/10.1007/978-3-030-70807-8_1

on the relation between patients/citizens and health. Empowered people/patients have the necessary knowledge, skills, attitudes and self-awareness about their condition to understand their lifestyle and treatment options, make informed choices about their health and have control over the management of their condition/health in their daily life [2–9].

The acquisition of medical/health information is a basic step in the empowerment process and the main source of health/medical information is, nowadays, the Web [10–13]. Search engines are the main tools used to retrieve information from the Web [14, 15]. However, generic search engines do not make any distinction among the users and overload them with a huge amount of information that is often outdated and of poor quality. Moreover, the Web is full of information not easily understandable by users such as patients/citizens because they lack a specific expertise in the health domain. Specialized search engines (PubMed¹ or Cochrane Library²) mainly work on medical literature and are quite complex for generic users, and especially the elderly ones. Finally, specialized health/medical websites (e.g., WebMD³, MedlinePlus⁴, or Health on Net Foundation Select⁵) are mainly built by hand so presenting a limited and often outdated amount of information (compared to what is available on the Web). In addition, they are often not free.

The retrieval of health information from the Web requires a communication process between a user and a search engine. Notice that a “complete” communication process, usually, entails different levels of communication. Many communication models exist in the literature and a very famous model (if not the most famous) is the one introduced by Morris in relation to his theory of signs [16]. It is made up of three levels, i.e., *syntactic*, *semantic*, and *pragmatic* and it has been used in several works dealing with human communication [17–19]. Shannon and Weaver looked at other aspects of communication and presented a mathematical theory of communication that is focused on information transmission [20]. Even though they were mainly dealing with the technical aspects of communication, they introduced other two levels above the *technical* one, i.e., the *semantic* and *effectiveness* levels, that are influenced by the *technical* level.

This *syntactic-semantic-pragmatic* communication model can be used, in principle, for human-to-human communication (e.g., Patient-Doctor) whereas the *technical-semantic-effectiveness* communication model can be used for human-to-machine communication (e.g., Patient-Web), as shown in Fig. 1. Unfortunately, these communication models are not fully used for both Patient-Doctor communication (very often doctors use a technical language that is too complex for patients) and Patient-Web communication (a search engine only works at the technical level by retrieving web-page addresses based on user keywords). Even worse, the two communication processes often aim to opposite directions. In fact, when a patient looks for medical information on the Web before seeing a doctor, this often leads to the infamous quarrel between patients and doctors during the visit [21]. The Patient-Doctor and Patient-Web communication processes are,

¹ <https://pubmed.ncbi.nlm.nih.gov/>.

² <https://www.cochranelibrary.com/>.

³ <https://www.webmd.com/>.

⁴ <https://medlineplus.gov/>.

⁵ <https://www.hon.ch/HONselect/index.html>.

in this case, contrasting each other and, surely, are not coherently contributing to the overall comprehension process of health conditions, treatments, etc., by patients.

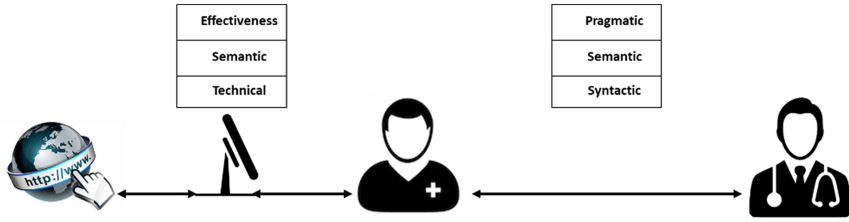


Fig. 1. Patient-Web and Patient-Doctor communication models.

Interestingly enough, the two models have been connected by [22] and an equivalence between the terms at the three levels has been established, in practice. We can then consider an *integrated* communication model that presents the following levels (with the *semantic* term that has been associated with *meaning* as indicated in [23]):

- **Pragmatic-Effectiveness:** How effectively does the received information affect behaviour?
- **Semantic-Meaning:** How precisely is the meaning conveyed?
- **Syntactic-Technical:** How accurately can the information be transmitted?

The objective is then to allow Patient-Web and Patient-Doctor communication to harness the three levels of the communication model and to aim to the same direction for the benefit of the person/patient. As seen above, the communication/interaction between a human user and a generic search engine only involves the syntactic-technical level by retrieving Web-page addresses (URLs) based on the keyword(s) specified by the user. A generic search engine has not been designed to understand user specific requirements (in his/her own language) and, thus, it is only able to provide the user with generic information leaving him/her with the task of selecting, understanding and using the retrieved information (semantic and pragmatic part of the communication process). Therefore, non-medical experts can be overwhelmed with the results and experience great difficulties in the comprehension and use of the found information. This, in turn, reflects on their ability to have a “true” two-way communication with their doctors because, for example, they do not have a complete understanding of their medical conditions (semantic level)—when they do not misunderstand them—and then they are unable to make shared and informed decisions (pragmatic level).

This work presents the characteristics and use of a custom search engine, FACILE, that has been created to satisfy the user information needs and overcome the communication challenges confronted during the search process. FACILE allows a user to specify his/her information requirements in a simple way. It, then, retrieves tailored Web information by exploiting the Web semantic capabilities provided by schema.org structured data. By doing so, FACILE provides the “right” amount of Web content, without overwhelming the user, and in a language that he/she can easily understand. This positively affects person/patient comprehension of conditions, treatment alternatives, etc.

and, ultimately, facilitates, his/her empowerment process and communication with medical professionals. The paper is organized as follows. Section 2 presents the principles and main characteristics of FACILE. Section 3 presents the implementation details and use of FACILE. Section 4 presents some experimental results and a discussion of the obtained results. Finally, Sect. 5 presents some conclusions and future work.

Some of the information presented in this paper is based on a previous work [24]. The present paper, however, extends the previous study by adopting a newer dataset, from 2019, of structured data and extracting a higher number of health-lifesci.schema.org and schema.org quadruples so to provide the user with a richer set of information. Moreover, the mapping model of FACILE has been improved to separate the types of users (medical experts and non-experts) and provide them with differentiated information in terms of language level and specific information. Furthermore, the FACILE interface has been completely redesigned in terms of usability and the customized information, as well as the quality information, is shown to favor user comprehension. Finally, FACILE principles and characteristics have been revised and are explained in more details and some new experimental results are presented to back FACILE effectiveness.

2 FACILE Principles and Characteristics

FACILE is a custom search engine specifically designed to facilitate the search of health information online by allowing users to specify their requirements during the search. In what follows we present its main principles and characteristics.

2.1 Identification of Requirements of On-Line Health Information Seekers

The identification of the main user requirements when searching for health information on the Web has been carried out in [4] by analysing the works presented in [14, 15, 25–29]. This literature review, although limited, has consistently shown the following main requirements by on-line health information seekers:

- Language complexity;
- Information classification/customization;
- Information quality (mainly intended as information trustworthiness).

The FACILE search engine has been developed, based on these requirements by exploiting the semantic features of the Web and particularly those related to structured data and schema.org with particular reference to its health-lifesci extension.

2.2 Use of schema.org and Health-Lifesci Structured Data

As said above, we have investigated how to leverage structured data to find suitable Web pages that satisfy the requirements of on-line health information seekers. To this end, we have exploited the semantic information available in the Web and, in particular, the one provided by schema.org⁶, an initiative funded by some major Web players, that

⁶ <https://www.schema.org/>.

aims to create, maintain, and promote schemas for structured data on the Internet. For the scope of the present work, we have considered the health-lifesci⁷ extension that presently contains 80 types, 162 properties and 125 enumeration values related to the health/medical field.

We have performed an analysis of the health-lifesci elements using the data made available by the Web Data Commons initiative. The Web Data Commons (WDC) [30] contain all Microformat, Microdata and RDFa (Resource Description Framework in Attributes) data extracted from the open repository of Web crawl data named Common Crawl (CC). The data used in this work have been released in November 2019. The whole dataset contains about 2.5 billion pages and about 38.1% of them contain structured data.

The dataset dump, used in our study, consists of 44.2 billion RDF quadruples. These are sequences of RDF terms in the form {s, p, o, u}, where {s, p, o} represents a statement consisting of subject, predicate, object, while u represents the URI of the document from which the statement has been extracted. From the whole dataset, we have selected the subset containing one or more of the 367 elements (types, properties, and enumeration values) of health-lifesci.schema.org. The obtained subset contains 125,920,198 quadruples.

As a first step, the extracted quadruples have been directly inserted into a DB table with a size of around 16 GB. Next, in order to make the consultation of this table faster and more efficient, an indexing process has been carried out by replacing the four elements of the quadruple with the indexes of four indexed tables, one for each element.

2.3 Mapping Health Information Seeker Requirements to schema.org Elements

When speaking of on-line health information seekers, we can, mainly, consider two classes of users:

- Non experts (e.g., patients or citizens);
- Experts (e.g., physicians or medical researchers).

These two categories have different requirements, that can be connected to the language complexity and the other user requirements presented above. It is, then, important to understand which schema.org elements can be mapped to the user requirements to use the structured data present in the Web and deliver tailored information to the user.

Language Complexity. In relation to the language-complexity requirement, health-lifesci.schema.org includes the *MedicalAudience* element that indicates whether the content is more suitable for a non-expert (*Patient* type) or an expert (*Clinician* and *MedicalResearcher* types) [31]. A preliminary analysis of the dataset containing the health.life-sci.schema.org quadruples has shown that only a small part (around 5%) contains one or more *Patient*, *Clinician* or *MedicalResearcher* schema.org elements. Thus, by taking only those URLs, we would limit the search because most web pages containing health/medical schema.org structured information would not be considered. Moreover, the fact that a web page contains a *Patient* schema.org element does not

⁷ <https://health-lifesci.schema.org/>.

necessarily indicate that those pages have a simpler language but only that they are targeted to patients (probably because of the page content). The same applies to the web pages that contain *Clinician* or *MedicalResearcher* schema.org elements that do not necessarily indicate the language used is complex.

In another work [31], we have measured the language complexity levels of the URLs containing *Patient*, *Clinician* or *MedicalResearcher* elements and generalized the analysis to URLs that do not contain any of these elements. To this end, we have analysed the English and non-empty web pages (around 50%) of the *Patient*, *Clinician* and *MedicalResearcher* subsets and, for each web page, we have computed its language complexity level. This has been done by computing the ‘term familiarity index’ [31–33] of each term found in the page (number of Google results) and then calculating the average for the page.

The experimental results, presented in [31], show that the web pages targeted to *Patient*, present, on average, a much higher term familiarity index whereas the web pages targeted to *Clinician* and *MedicalResearcher* present, on average, a lower term familiarity index. This kind of equivalence has convinced us to provide non-experts with URLs presenting a high term familiarity index (and thus a simpler terminology) and experts with URLs presenting a low term familiarity index (and thus a more complex terminology). By doing so, we are able to rank all URLs on the base of the FACILE language level, i.e., simple for non-experts and complex for experts. Moreover, coherently with the *MedicalAudience* schema.org representation, URLs containing a *Patient* schema.org element will be ranked higher for non experts whereas URLs containing *Clinician* or *MedicalResearcher* schema.org elements will be ranked higher for experts.

Information Customization. In order to proceed with a customization of information for the users, we have preliminarily investigated what type of health/medical information is primarily searched on the Web, such as medical conditions, therapies and drugs [10, 11, 14]. We have then verified that this type of information appears amongst the most recurring health-lifesci.schema.org elements to have enough data to be processed by FACILE and provide an effective information classification/customization. At the end of this process, we have selected the following health.life-sci.schema.org elements for creating the classification of Web pages (the definitions are taken from <https://health-lifesci.schema.org/>):

- *MedicalCondition*, indicates any condition of the human body that affects the normal functioning of a person, whether physically or mentally.
- *MedicalTherapy*, indicates a medical intervention designed to prevent, treat, and cure human diseases and medical conditions.
- *Drug*, indicates a chemical or biologic substance, used as a medical therapy, that has a physiological effect on an organism.
- *MedicalClinic*, indicates a hospital or a medical school.
- *MedicalCode*, provides a code for a medical entity.
- *MedicalScholarlyArticle*, indicates a medical article.

Note that the information on conditions, therapies, drugs, and hospitals can be of interest to both non-experts and experts whereas the information on codes and medical articles are more technical and so mainly indicated for medical experts. Therefore, we have decided to consider the first four elements for non-experts and all six elements for experts.

Information Quality. The health-lifesci.schema.org elements, being focused on the health/medical field, did not provide us with elements that can be immediately used for evaluating the information quality. Instead, by analyzing the literature [34–36], we found that the more general schema.org was able to provide some of those elements. We have selected *author*, *publisher*, *reviewedBy* and *recognizingAuthority* schema.org elements as an indication of reliability and *dateCreated*, *dateModified*, *datePublished* and *lastReviewed* schema.org elements as an indication of timeliness.

Note that, in principle, on-line health information seekers have the same requirements on the quality information irrespective of them being medical experts or not. Consequently, we have selected the same schema.org elements for providing information quality to both non-experts and experts.

Mapping Table. The above considerations about language complexity, information customization and information quality have led us to create a mapping between the user requirements and the schema.org elements for the non-expert and expert user categories. This mapping expands the one presented in [4, 24] and is reported in Table 1.

These schema.org elements are used by FACILE to retrieve Web pages and extract information based on the user specific requirements. As we will show in the next sections, by using FACILE, users can easily and quickly find the right amount of information that is reliable and, in a language suitable to their health literacy level, in full compliance with the three levels of the communication model presented above.

Table 1. Mapping between user requirements and schema.org elements for the “non-expert” and “expert” user categories.

	Language complexity	Information customization	Information quality
Non-expert	Simple - <i>High Term Familiarity</i>	- <i>MedicalCondition</i> - <i>MedicalTherapy</i> - <i>Drug</i> - <i>MedicalClinic</i>	Reliability: - <i>author</i> - <i>publisher</i> - <i>reviewedBy</i> - <i>recognizingAuthority</i> Timeliness: - <i>dateCreated</i> - <i>dateModified</i> - <i>datePublished</i> - <i>lastReviewed</i>

(continued)

Table 1. (continued)

	Language complexity	Information customization	Information quality
Expert	Complex - <i>Low Term Familiarity</i>	- <i>MedicalCondition</i> - <i>MedicalTherapy</i> - <i>Drug</i> - <i>MedicalClinic</i> - <i>MedicalCode</i> - <i>MedicalScholarlyArticle</i>	Reliability: - <i>author</i> - <i>publisher</i> - <i>reviewedBy</i> - <i>recognizingAuthority</i> Timeliness: - <i>dateCreated</i> - <i>dateModified</i> - <i>datePublished</i> - <i>lastReviewed</i>

3 FACILE Implementation and Use

The mapping between user requirements and schema.org elements, shown in the previous Section, has been used to build FACILE in order to provide experts and non-experts with the proper Web contents in terms of language complexity, information customization and information quality. It completes the previous versions of the system, the first one considering only the language level [31] and the second one taking into account only the information customization and quality [24]. Moreover, the algorithm used to provide the custom information has been completely revised with respect to the previous versions. Fig. 2 shows the user interface of the FACILE search engine for non-experts (a) and experts (b). FACILE is available at the address <https://cohealth.lero.ie/facile>.

The user interface includes a simple text input, like that of a generic search engine, where the user can insert the term(s) to be searched. Moreover, the user can filter the results to get more focused information by specifying his/her requirements in terms of language level, specific information required and information quality. The *simple* language level is recommended for non-experts whereas the *complex* language level is recommended for experts. Non-experts can select specific information by checking one or more of the four items reported in Fig. 2a whereas experts can choose one or more of the six items reported in Fig. 2b. Finally, the “Show Quality Information” switch, acts on the quality user requirement, for both types of users, and provides a measure of the quality of the web pages by evaluating the schema.org parameters reported in Table 1 (when available). In what follows we provide further details on the implementation and use of FACILE.

3.1 Language Level and Ranking Algorithm

After having inserted the quadruples into the database, as seen in Sect. 2.2, the English web pages have been analysed. For each web page, the *Term Frequency* of each word (not considering the stop words, i.e., the most common words) has been calculated together with its *Term Familiarity* (by using a Google™ API which returns the number of google results for a term). Depending on the selected language level, simple or complex, FACILE ranks the found URLs according to the following ranking formulas:

(a)

(b)

Fig. 2. FACILE user interface.

Simple Language.

$$R = \alpha * (Term_Frequency/Max_Term_Frequency) + (1 - \alpha) * (Page_Familiarity_Index)/Max_Familiarity_Index \quad (1)$$

Complex Language.

$$R = \alpha * (Term_Frequency/Max_Term_Frequency) + (1 - \alpha) * (1 - Page_Familiarity_Index / Max_Familiarity_Index) \quad (2)$$

Where:

- *Term_Frequency* is the number of occurrences of the keyword(s) in the page divided by the document length;
- *Max_Term_Frequency* is the maximum of *Term_Frequency* of the keyword(s) in all found Web pages;
- *Page_Familiarity_Index* is the page familiarity, i.e., the mean of the term familiarity indexes of the Web page;
- *Max_Familiarity_Index* is the maximum page familiarity of all found Web pages.
- α allows to differently weigh the term frequency and page familiarity.

Notice that we use the *Term Frequency* to consider that web pages have different lengths and then the number of occurrences of a term can greatly vary. Notice, also, that the simple-language formula is a weighted combination of the *Term-Frequency* percentage and *Term-Familiarity* percentage because we want to find meaningful pages (i.e., with a high number of occurrences of the searched item) but with the simplest language. The complex-language formula is a weighted combination of the *Term-Frequency* percentage and $(1 - \textit{Term-Familiarity})$ percentage because we want to find meaningful pages (i.e., with a high number of occurrences of the searched item) but with the most complex/technical language.

We have made some preliminary experiments with the weight and found out that a value of $\alpha = 0.4$ provided us with the best results in terms of correspondence between the intended audience and the provided Web pages. For example, Fig. 3 shows the first results of FACILE for the “asthma” term when the user chooses the “simple” language level and Fig. 4 shows the first results of FACILE for the “asthma” term when the user chooses the “complex” language level.

Term(s) to search: asthma

Search

Indicate your requirements:

Language Level

- Simple
- Complex

Specific Information

- Condition
- Therapy
- Drug
- Institution

Show Quality Information?

58 Pages Found

Uri
https://www.verywellhealth.com/asthma-overview-4014760
https://www.emedicinehealth.com/occupational_asthma/article_em.htm
https://www.verywellhealth.com/asthma-signs-and-symptoms-200643?utm_source=...
https://www.verywellhealth.com/can-my-period-worsen-my-asthma-200880
https://www.dupixenthcp.com/
https://www.verywellhealth.com/cromolyn-sodium-and-nedocromil-what-you-need-...
https://www.verywellhealth.com/what-is-advaair-diskus-2011159
https://www.drugs.com/ppa/beclomethasone-oral-inhalation.html
https://www.verywellhealth.com/expiratory-grunting-sighing-and-asthma-26345-...
https://www.drmcguokin.com/your-body-will-tell-you-what-it-likes-and-does-n-...
https://www.medicinenet.com/script/main/art.asp?articlekey=197495
https://www.drugs.com/cons/theo-time.html
https://www.drugs.com/ppa/terbutaline.html
https://www.wakehealth.edu/Condition/a/Asthma
https://www.verywellhealth.com/what-causes-gerd-1741914

Fig. 3. FACILE results for “simple” language level.

Term(s) to search: asthma

Search

Indicate your requirements:

Language Level

- Simple
- Complex

Specific Information

- Condition
- Therapy
- Drug
- Institution
- Code
- Medical Article

Show Quality Information?

58 Pages Found

Uri
https://www.medicinenet.com/script/main/alphaidx.asp?p=h_26
https://www.medicinenet.com/script/main/alphaidx.asp?p=s_26
https://www.medicinenet.com/script/main/art.asp?articlekey=205666
https://www.medicinenet.com/script/main/alphaidx.asp?p=n_26
https://www.medicinenet.com/doctors/drug_allergies/7bca2b16-8f5a-4a91-b7e8-...
https://www.medicinenet.com/doctors/indoor_allergens/6d308599-690d-404f-b2bf-...
https://www.medicinenet.com/doctors/indoor_allergens/6d9ee6b6-d445-40a1-b8f-...
https://www.medicinenet.com/doctors/drug_allergies/1f1129239-c466-423a-9b8f-...
https://www.medicinenet.com/doctors/indoor_allergens/e07df9d9-90f3-42a4-916-...
https://www.medicinenet.com/doctors/mold_exposure/2d7da749-36b4-4697-9e28-f-...
https://www.medicinenet.com/doctors/mold_exposure/24c88db9-b05a-de11-86ed-0-...
https://www.medicinenet.com/doctors/indoor_allergens/b02f17ce-e9a7-4f65-9a3-...
https://www.medicinenet.com/doctors/mold_exposure/55e1d433-fd2a-45b7-a4fe-4-...
https://www.medicinenet.com/doctors/drug_allergies/9aa0dbbe-49ee-47a6-9043-...
https://www.medicinenet.com/doctors/indoor_allergens/bbfd3f55-2bf6-477c-be0-...
https://www.medicinenet.com/doctors/indoor_allergens/05362154-f52-4f69-9ec-...

Fig. 4. FACILE results for “complex” language level.

3.2 Information Customization

As shown in Fig. 2, the user has the possibility to select which specific information he/she is interested in through checkboxes that can be easily selected by a user with no specific knowledge, because they indicate common terms in the health domain. More than one checkbox can be checked at the same time. FACILE will extract the custom information by using the mapping of Table 1 and will provide the following results:

- **Condition (for Non-experts and Experts):** It will present the pages that contain a description of a medical condition and the values of the properties related to the *MedicalCondition* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *name*, *alternateName*, and *description* properties.
- **Therapy (for Non-experts and Experts):** It will present the pages that contain information about a therapy and the values of the properties related to the *MedicalTherapy* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *name*, *alternateName*, *description*, *potentialAction*, and *contraindication* properties.
- **Drug (for Non-experts and Experts):** It will present the pages that contain the information about a medicine and the values of the properties related to the *Drug* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *name*, *nonProprietaryName*, *activeIngredient*, *description*, and *prescriptionStatus* properties.
- **Institution (for Non-experts and Experts):** It will present the pages that contain the institutions (e.g., hospitals) that deal with a medical condition and the values of the properties related to the *MedicalClinic* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *alternateName*, *description*, *hasMap*, *image*, *telephone*, and *email* properties.
- **Code (for Experts):** It will present the pages that contain the code of a medical condition and the values of the properties related to the *MedicalCode* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *name*, *alternateName*, *description*, *code*, *codeValue*, and *codingSystem* properties. The *code*, together with the *codingSystem*, can be used to look on specialized websites to find a specific condition/part of the body/therapy/drug, or other useful information.
- **MedicalScholarlyArticle (for Experts):** It will present the pages that contain information about medical scholarly articles and the values of the properties related to the *MedicalScholarlyArticle* schema.org element. In order not to overwhelm the user with information on the response page, FACILE will only show the values of the *description*, *publicationType*, *url*, *abstract*, *author*, *creator*, *dateCreated*, *datePublished*, *editor*, *headline*, and *sameAs* properties.

Fig. 5 shows the first results of FACILE for the “asthma” term (with simple language) when the user selects the Condition, Drug and Institution fields of the specific information.

Term(s) to search

Search

Indicate your requirements:

Language Level

- Simple
- Complex

Specific Information

- Condition
- Therapy
- Drug
- Institution

Show Quality Information?

58 Pages Found

Uri
https://www.verywellhealth.com/asthma-overview-4014760
https://www.emedicinehealth.com/occupational_asthma/article_em.htm
https://www.verywellhealth.com/asthma-signs-and-symptoms-200643?utm_source=...
https://www.verywellhealth.com/can-my-period-worsen-my-asthma-200880
https://www.dupixenthcp.com/
<p>Condition</p> <p>name: Atopic Dermatitis, Asthma</p>
<p>Drug</p> <p>activeingredient: dupilumab</p> <p>description: DUPIXENTU00AE (dupilumab) is indicated for the treatment of patients aged 12 years and older with uncontrolled moderate-to-severe atopic dermatitis and for your moderate-to-severe asthma patients aged 12+ with an eosinophilic phenotype or with oral corticosteroid dependent asthma. Serious adverse reactions may occur. Please see Important Safety Information and full PI on website.</p> <p>prescriptionStatus: Available by prescription only.</p> <p>description: DUPIXENTU00AE (dupilumab) is indicated for the treatment of patients aged 12 years and older with uncontrolled moderate-to-severe atopic dermatitis and for your moderate-to-severe asthma patients aged 12+ with an eosinophilic phenotype or with oral corticosteroid dependent asthma. Serious adverse reactions may occur. Please see Important Safety Information and full PI on website.</p>
https://www.verywellhealth.com/cromolyn-sodium-and-nedocromil-what-you-need-...
https://www.verywellhealth.com/what-is-advaiv-diskus-201159
https://www.drugs.com/ppa/beclomethasone-oral-inhalation.html
<p>Drug</p> <p>name: beclomethasone</p> <p>prescriptionStatus: http://schema.org/PrescriptionOnly</p>
https://www.verywellhealth.com/expiratory-grunting-sighing-and-asthma-26345-...
https://www.dmcguckin.com/your-body-will-tell-you-what-it-likes-and-does-n-...
<p>Institution</p> <p>alternateName: Chiropractic</p> <p>description: Chiropractor Valparaiso offering chiropractic and natural health solutions. Call or email us today to find out more! On-site X-ray facilities.</p> <p>image: https://cdn.vortiaa.com/childsites/uploads/3089/files/logo.png</p>

Fig. 5. FACILE results (simple language) with specific information checked.

Fig. 6 shows the first results of FACILE for the “asthma” term (with complex language) when the user selects the Medical-Article field of the specific information.

Term(s) to search

Search

Indicate your requirements:

Language Level

- Simple
- Complex

Specific Information

- Condition
- Therapy
- Drug
- Institution
- Code
- Medical Article

Show Quality Information?

58 Pages Found

Results
https://www.medicinenet.com/script/main/artifact.asp?art_58
https://www.medicinenet.com/script/main/artifact.asp?art_58
https://www.medicinenet.com/script/main/artifact.asp?art_58
https://www.medicinenet.com/script/main/artifact.asp?art_58
https://www.medicinenet.com/diseases/allergies/food/11-2534-4491-2148-...
https://www.medicinenet.com/doctors/indoo_allergies/5c30596-000c-4041-52b-...
https://www.medicinenet.com/doctors/indoo_allergies/59e205b-0445-4014-b3f-...
https://www.medicinenet.com/doctors/indoo_allergies/7113229-c408-423a-963f-...
https://www.medicinenet.com/doctors/indoo_allergies/e072968-9054-4244-419-...
https://www.medicinenet.com/doctors/mood_allergies/2f197-70-324-40913ac3f4-...
https://www.medicinenet.com/doctors/mood_exposure/24c3329-6254-9e11-55ac-6-...
https://www.medicinenet.com/doctors/indoo_allergies/020717c-4974-4853-9a3-...
https://www.medicinenet.com/doctors/mood_exposure/55a12433-423a-4507-a14e-4-...
https://www.medicinenet.com/doctors/indoo_allergies/5ea2506a-48ea-4725-4243-...
https://www.medicinenet.com/doctors/indoo_allergies/3a03395b-2386-477c-84d-...
https://www.medicinenet.com/doctors/indoo_allergies/20202154-202-10948c-...
https://www.medicinenet.com/doctors/indoo_allergies/81c3a7f-4994-4858-914-...
https://www.medicinenet.com/doctors/indoo_allergies/4e026a35-2395-4e4d-e431-...
https://www.medicinenet.com/doctors/indoo_allergies/5e44a81-489a-401a-844c-...
https://www.medicinenet.com/doctors/indoo_allergies/5207177-4854-4e4d-b3f-...
https://www.medicinenet.com/doctors/indoo_allergies/1c33459-38af-495e-495b-...
https://www.medicinenet.com/doctors/indoo_allergies/0302c67-4510-4e4b-891f-...
https://www.medicinenet.com/doctors/indoo_allergies/10a53c-10314111-aa63-...
https://www.medicinenet.com/doctors/indoo_allergies/9b32493-c034-e811-0500-...
https://www.medicinenet.com/script/main/artifact.asp?artidkey=197465
https://www.emedicinehealth.com/drug-salmeterol_inhalation/article_em.htm
https://www.mayoclinic.org/diseases-conditions/wheezing-induced-asthma/symptom
https://www.verywellhealth.com/asthma-overview-4014760
https://doctors.amtallhealth.org/asthma-shah-alk-village-allergy-immunology
https://www.verywellhealth.com/can-my-period-worsen-my-asthma-200880
https://www.drugs.com/rm/abucetol.html
https://www.verywellhealth.com/what-is-advaiv-diskus-201159
https://www.wakehealth.edu/Condition/a/asthma
https://www.emedicinehealth.com/drug/beclomethasone_oral_inhalation_pi/asthma.htm
https://medicine.yale.edu/pediatrics/profile/asthma_mech

Medical Article

description: Shen, A, Walker DM. The impact of race and ethnicity on care in the pediatric emergency department. *Paediatr Respir Rev* 2010; 22:284-8.

url: https://doi.org/10.1016/j.prr.2010.05.001

publicationType: Journal Article

pubmed: https://pubmed.ncbi.nlm.nih.gov/20428601/

description: Tenra A, Langham M, Norrish C, Saracco K, Chen L. Remember the saphorous: ultrasound mediated first responder life resuscitation of drowned victims in young children. *Pediatric Emergency Care* 2011; 27:1213-4.

url: https://www.ncbi.nlm.nih.gov/pubmed/22124202

description: Tenra A, Hasan A, Langham M, Saracco K, Chen L. Diagnosis of intubation by physical-mechanic bronchograms in the emergency department. *Annals Of Emergency Medicine* 2012; 60:264-8.

url: https://www.ncbi.nlm.nih.gov/pubmed/22268602

Fig. 6. FACILE results (complex language) with specific information checked.

3.3 Information Quality

The quality of the web pages provided by FACILE is evaluated in terms of reliability and timeliness by checking the presence of the schema.org elements reported in Table 1. To provide the user with a value and a visual indication about the information quality of the page, we have thought to provide each element with a weight that considers its relative importance. Considering some analysed literature [34–36], we have decided to assign different weights to the elements for a total of 10. As a first step, we have decided to assign 2/3 of the weights to Reliability and 1/3 of the weights to Timeliness because we want to value more reliability over timeliness. As a second step, we have assigned the following weights to the information quality schema.org elements:

- Reliability:
 - *author* 1
 - *publisher* 1.5
 - *reviewedBy* 1.5
 - *recognizingAuthority* 2.5 (Tot. 6.5)
- Timeliness:
 - *dateCreated* 0.5
 - *dateModified* 1.5
 - *datePublished* 0.5
 - *lastReviewed* 1 (Tot. 3.5)

Thus, for each web page, FACILE checks the presence of one or more quality-information elements and sums up the related weights. In order to provide the user with a direct visual information FACILE divides the result by two, rounds it to the whole or the half, and transforms the obtained number in stars so to provide each page with a number of stars that ranges between 0 and 5, with steps of half a star. Fig. 7 shows the first results of FACILE for the “asthma” term (with simple language and specific information checked) when the “Show Quality Information” switch is turned on.

The screenshot shows the FACILE search interface. On the left, there is a search bar with the term 'asthma' and a 'Search' button. Below it, there are filters for 'Language Level' (Simple selected, Complex unselected) and 'Specific Information' (Condition selected, Therapy unselected, Drug unselected, Institution unselected). A 'Show Quality Information?' switch is turned on. On the right, the search results are displayed, showing 58 pages found. The first five results are visible, each with a URL, a title, and a quality score represented by stars (e.g., 4 stars, 3 stars, 4 stars, 4 stars, 4 stars).

Fig. 7. FACILE results with “show quality information” switched on.

Notice that by clicking on the quality stars, the specific information-quality data will be shown (Fig. 8) so that the user will have the possibility to directly analyse them.

The screenshot shows the FACILE search interface. On the left, there is a search bar with 'asthma' entered and a 'Search' button. Below it, there are options to 'Indicate your requirements:' for 'Language Level' (Simple selected, Complex unselected) and 'Specific Information' (Condition, Therapy, Drug, Institution, all selected). A 'Show Quality Information?' toggle is also visible. On the right, the search results are displayed, showing 58 pages found. Three results are visible, each with a quality score indicated by stars (4, 3, and 4 stars respectively) and detailed metadata including author, date published, last reviewed, publisher, and reviewer.

Fig. 8. Quality information details obtained by clicking on the quality stars.

4 Experimental Results and Discussion

FACILE, as shown in the previous Sections, has been designed and implemented to be used by different user typologies, i.e., medical experts or non-experts and we have executed some tests to evaluate its effectiveness and usefulness.

FACILE can be used as any other search engine by inserting any keyword(s) on the text input (Fig. 2). It will then rank the results in terms of the language level: simple (by default), as shown in Fig. 3, or complex, as shown in Fig. 4. For example, the first result of Fig. 3 presents a term familiarity of 3.53 million whereas the first result of Fig. 4 presents a term familiarity of 0.88 million. This clearly indicates how FACILE can separate pages with simple language from the ones with complex language. It already shows how FACILE outperforms a generic search engine in terms of providing web results with different language levels since a generic search engine is not able to provide such kind of responses.

The effectiveness of FACILE can further be appreciated when the specific-information fields are checked. In the example of Fig. 5, the condition, drug and institution fields are checked and this information is immediately shown in the response page for the web pages that contain it. This allows a user to save time by immediately finding the specific information he/she is looking for without the need of going through the content of the web pages. He/she will examine the complete content of the web pages only if he/she needs further information. In the example of Fig. 6, an expert user has checked the Medical-Article field and immediately gets a list of scientific articles with the related links. In this case, the user does not even need to browse the web page because the links to the articles, on the response page, allow him/her to directly go the articles of interest.

If the user switches on the “Show Quality Information” button, the quality of the web page is visually shown as a number of stars (0–5, Fig. 7), immediately providing the user with an information on the page trustworthiness. Nevertheless, if the user wants more detailed information about the quality, this information will be shown by clicking on the quality stars (Fig. 8). Notice that, the fact that a web page has a small number of stars (or no stars at all), does not necessarily imply that the information quality of the page is low and that the page is not trustworthy, but only that the chosen schema.org elements are not present. Thus, in general, a user is encouraged first to look at the web pages with a high number of stars, because that means that the page is of good quality, and then at the web pages with a low number of stars if they present other elements of interest.

To complete the evaluation of FACILE, we have run some experiments that compute the term familiarity of the web pages with simple and complex language levels. We have used six of the most searched diseases in Google™⁸ and, in particular, three common terms (diabetes, depression, and anxiety) and three less common terms (yeast infection, psoriasis and lyme disease). The first ten results of each search have been considered and, for each group, the average familiarity index has been computed. Fig. 9 shows the results of this analysis (term familiarity is expressed in million). As expected, the web pages with a simple language level present a higher term familiarity than the web pages with the complex language level. More interesting is the fact that there is a clear separation between the familiarity indexes of the two categories even though the familiarity indexes of the web pages with simple-language levels are quite constant and the familiarity indexes of the web pages with complex-language levels present more differences. This appears more evident for the familiarity indexes of the less common terms (more technical) that present, in general, a higher term familiarity. This can be attributed to the fact that the common terms have a higher number of web pages and then a clearer separation between pages targeted to medical experts and non-experts whereas the less common terms have a lower number of web pages and they appear more targeted to non-experts (considering their higher familiarity indexes). Further experiments are undoubtedly needed to better understand FACILE behavior in these cases.

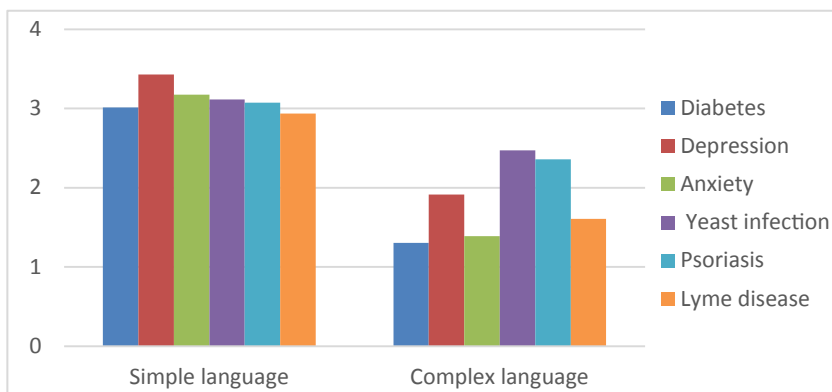


Fig. 9. Average familiarity indexes (in million) of most searched health diseases in Google™.

Although FACILE can still be improved and more experiments are undoubtedly needed, its principles and practical use show how it complies with the three-level communication model presented in the Introduction. In particular:

Syntactic-Technical: FACILE presents the same retrieval capabilities as a generic search engine and, as such, it allows the user to search for health information on the Web (although it restrains the search to the Web structured data) and provides the requested information as any generic search engine does.

⁸ <https://www.beckershospitalreview.com/quality/most-commonly-googled-diseases.html>.

Semantic-Meaning: The translation capability of FACILE (mapping model) allows the user to specify his/her requirements in simple terms. FACILE “understands” such requirements and translates them into schema.org elements to extract the proper information from the Web (communication phase: from user to FACILE). Moreover, the ability of FACILE, to retrieve Web pages with different language levels, allows the user to choose the pages with the preferred level (communication phase: from FACILE to user).

Pragmatic-Effectiveness: The response presented by FACILE has a pragmatic impact on the user by providing understandable, focused, and reliable Web results. The consequence is that a user (mainly non-expert/patient) is, overall, greatly facilitated in finding, understanding, and using health/medical information on the Web and then in his/her empowerment process. This, among others, facilitates the comprehension of his/her medical conditions and greatly increases his/her ability to communicate with medical professionals and make informed decisions. Moreover, the ability of FACILE to provide focused and customized information can also facilitate the task of medical professionals to accompany patients in their search of on-line health information.

Overall, FACILE provides the user with the possibility of filtering and re-ranking the Web results according to his/her specific requirements but it leaves the user fully in charge of his/her navigational path on the Web. In this way, a user can freely and simply choose what he/she needs in terms of health/medical information, in any moment, so to achieve his/her empowerment objectives and better communicate/interact with medical professionals. This allows, overall, to have an integrated communication with regards to both Patient-Web and Patient-Doctor communication/interaction (Fig. 10) that is aligned and overcomes, in practice, the well-known quarrel between doctors and patients discussed in the Introduction.

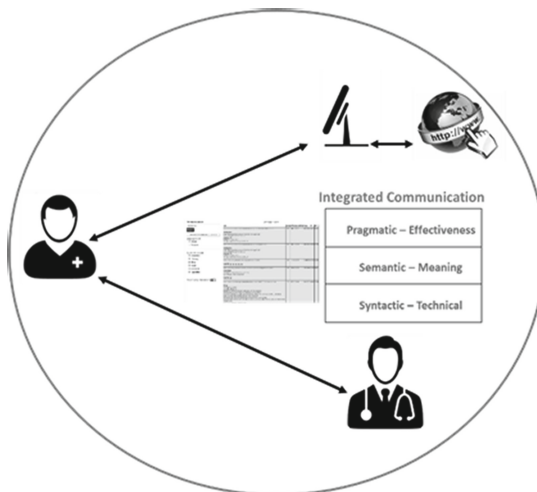


Fig. 10. FACILE integrated communication model.

Of course, an application that helps patient-doctor communication at the semantic level (helping mutual understanding) would be highly desirable. To this end, we have developed the SIMPLE application that translates the medical/technical terms of any text into lay terms, adding a short and simple explanation [37–39]. This helps users, mainly non-medical experts, to understand medical texts (e.g., medical reports) and then to better communicate with medical professionals. Notice that SIMPLE fully complies with the three-level communication model because it facilitates mutual understanding between patients and doctors (semantic level) and helps them in defining and reaching a specified objective (pragmatic level). The joint use of FACILE and SIMPLE allows, overall, a person/patient to communicate in the same (simple) way with both the Web, through FACILE, and a doctor, through SIMPLE, by fully exploiting the integrated communication model for human-to-machine and human-to-human communication.

5 Conclusions and Future Work

This work has presented a custom search engine, FACILE, that allows users (both medical experts and non-experts) to retrieve health information from the Web on the base of their specific requirements in terms of language level, information customization and information quality. FACILE maps those requirements into schema.org structured elements and retrieves the “right” and simple Web content without overwhelming the user as a generic search engine often does. This positively affects user understanding and use of health information and, therefore, the empowerment process and communication with healthcare professionals. Moreover, FACILE fully complies with the “integrated” three-levels communication model that allows a full interaction of a person/patient with both a machine (e.g., search engine) and another human (e.g., doctor).

The principles and first experimental results are satisfying and show FACILE potentialities even though, the used dataset (created with the 2019 structured data of Web Data Commons) has proven, sometimes, too limited in terms of provided results. Thus, we are in the process of adding the datasets of the previous years, such as the ones of 2018 and 2017, because the various datasets present some differences in terms of URLs and thus provide more data to experiment with. Furthermore, although the current version of FACILE is simple and intuitive and it has been widely improved with respect to its previous versions, we are in the process of running some tests with non-experts, such as elderly people, to evaluate their engagement level and understand the aspects that need further improvement. An evaluation of the reached empowerment level through FACILE is also important and we are in the process of running randomized control trials with patients with chronic conditions. The reached health literacy and empowerment levels will be measured because of the use of FACILE.

As a future work, we plan to analyse in more details the health-lifesci.schema.org elements to evaluate whether further customized information can be provided to the user but being always careful not to overwhelm him/her. We also plan to improve the information quality algorithm by considering the dates the web pages have been created, modified, published, and so on, beside just evaluating the presence of the related information. We also want to further develop the integrated communication model for improving the human-to-human and human-to-machine communication processes that underlie patient

empowerment. For example, we plan to integrate the FACILE and SIMPLE applications to allow a person to use a single interface to find health information on-line and simplify it. Finally, we want to create a visual framework [40] that uses the retrieving capabilities of FACILE and translating capabilities of SIMPLE and allows easy creation of advanced health services, such as virtual assistants, for elderly people/patients.

Acknowledgements. This work was partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754489 and by Science Foundation Ireland grant 13/RC/2094 with a co-fund of the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero, the Science Foundation Ireland Research Centre for Software, www.lero.ie.

References

1. World Health Organization (WHO). Framework on integrated, people-centred health services: Report by the Secretariat. World Health Assembly, (A69/39), pp. 1–12 (2016)
2. World Health Organization (WHO). (1998. Health promotion glossary. <https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf>
3. European Health Parliament. Patient Empowerment and Centredness (2017)
4. Alfano, M., Lenzitti, B., Taibi, D., Helfert, M.: Provision of tailored health information for patient empowerment: an initial study. In: Proceedings of the 20th International Conference on Computer Systems and Technologies (CompSysTech 2019). Association for Computing Machinery, New York, NY, USA, pp. 213–220 (2019). <https://doi.org/10.1145/3345252.3345301>
5. Herzog, M.A., Kubincová, Z., Han, P., Temperini, M. (eds.): Advances in web-based learning – ICWL 2019. LNCS, vol. 11841. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-35758-0>
6. Bodolica, V., Spraggon, M.: Toward patient-centered care and inclusive health-care governance: a review of patient empowerment in the UAE. *Public Health* **169**(971), 114–124 (2019)
7. Bravo, P., Edwards, A., Barr, P.J., Scholl, I., Elwyn, G., McAllister, M.: Conceptualising patient empowerment: a mixed methods study. *BMC Health Serv. Res.* **15**(1), 1–14 (2015)
8. Cerezo, P.G., Juvé-Udina, M.E., Delgado-Hito, P.: Concepts and measures of patient empowerment: a comprehensive review. *Revista Da Escola de Enfermagem* **50**(4), 667–674 (2016)
9. Fumagalli, L.P., Radaelli, G., Lettieri, E., Bertere, P., Masella, C.: Patient empowerment and its neighbours: clarifying the boundaries and their mutual relationships. *Health Policy* **119**(3), 384–394 (2015)
10. Pew Research Center. Health online (2013). <https://www.pewinternet.org/2013/01/15/health-online-2013/>
11. Taylor, H.: HI-Harris-Poll-Cyberchondrics. Harris Interactive (2010). <https://theharrispoll.com/the-latest-harris-poll-measuring-how-many-people-use-the-internet-to-look-for-information-about-health-topics-finds-that-the-numbers-continue-to-increase-the-harris-poll-first-used-the-word-cyberch/>
12. UK national statistics. Statistical bulletin: Internet Access. Office for National Statistics, 27 August 2010
13. Instituto Nacional de Estadística. Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares (2010)

14. Pletneva, N., Vargas, A., Boyer, C.: D8.1.1. Requirements for the general public health search. Khresmoi Public Deliverable (2011)
15. Roberts, T.: Searching the internet for health information: techniques for patients to effectively search both public and professional websites. In: SLE Workshop at Hospital for Special Surgery Tips For Evaluating the Quality of Health, pp. 1–12 (2017)
16. Morris, C.W.: Foundations of the Theory of Signs. International Encyclopedia of Unified Science, vol. 1, no. 2. The University of Chicago Press, Chicago (1938)
17. Hahn, L.K., Paynton, S.T.: Survey of Communication Study (2014). https://en.wikibooks.org/wiki/Survey_of_Communication_Study
18. Cherry, C.: On Human Communication: A Review, A Survey, and A Criticism. M.I.T. Press, Cambridge (1966)
19. Johnson, F.C., Klare, G.R.: General models of communication research: a survey of the developments of a decade. *J. Commun.* **11**(1), 13–26 (1961)
20. Shannon, C., Weaver, W.: The Mathematical Theory of Communications. University of Illinois Press, Urbana (1949)
21. Smith, T.: Exploring the characteristics of active health seekers, the thinking behind patient preferences, and the implications for patient-professional relationships. *Qual. Safety Health Care* **13**(6), 474–477 (2004)
22. Carlile, P.: Transferring, translating and transforming: an integrative framework for managing knowledge across boundaries, organization. *Science* **15**(5), 555–568 (2004)
23. Watzlawick, P., Beavin, J.B., Jackson, D.D.: Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes. Norton, New York (1967)
24. Alfano, M., Lenzitti, B., Taibi, D., Helfert, M.: Tailored retrieval of health information from the web for facilitating communication and empowerment of elderly people. In: Proceedings of the 6th International Conference on Information and Communication Technologies for Ageing Well and e-Health - Volume 1: ICT4AWE, pp. 205–216 (2020). <https://doi.org/10.5220/0009576202050216> ISBN 978-989-758-420-6
25. Banna, S., Hasan, H., Dawson, P.: Understanding the diversity of user requirements for interactive online health services. *Int. J. Healthc. Technol. Manage.* **15**(3), 253–271 (2016)
26. Pian, W., Khoo, C.S.G., Chi, J.: Automatic classification of users' health information need context: logistic regression analysis of mouse-click and eye-tracker data. *J. Med. Internet Res.* **19**(12), e424 (2017)
27. Pang, P.C.-I., Verspoor, K., Pearce, J., Chang, S.: Better health explorer: designing for health information seekers. In: OzCHI 2015 Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, pp. 588–597 (2015)
28. Keselman, A., Logan, R., Smith, C.A., Leroy, G., Zeng-Treitler, Q.: Developing informatics tools and strategies for consumer-centered health communication. *J. Am. Med. Inf. Assoc. JAMIA* **15**(4), 473–483 (2008)
29. Ardito, S.C.: Seeking consumer health information on the internet. *Online Searcher* **37**(4), 1–5 (2013). <https://www.infotoday.com/OnlineSearcher/Articles/Medical-Digital/Seeking-Consumer-Health-Information-on-the-Internet-90558.shtml>
30. Meusel, R., Petrovski, P., Bizer, C.: The webdatacommons microdata, RDFa and microformat dataset series. In: Mika, P. (ed.) The Semantic Web – ISWC 2014. LNCS, vol. 8796, pp. 277–292. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11964-9_18
31. Alfano, M., Lenzitti, B., Taibi, D., Helfert M.: Facilitating access to health Web pages with different language complexity levels. In: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2019), 2–4 May 2019, Heraklion-Crete (2019)
32. Kloehn, N., et al.: Improving consumer understanding of medical text: development and validation of a new subsimplify algorithm to automatically generate term explanations in English and Spanish. *J. Med. Internet Res.* **20**(8), e10779 (2018)

33. Leroy, G. et al.: Improving perceived and actual text difficulty for health information consumers using semi-automated methods. In: AMIA Annual Symposium Proceedings, pp. 522–531 (2012)
34. World Health Organization (WHO). Improving Data Quality: A Guide for Developing Countries, pp. 1–74 (2003). https://www.wpro.who.int/publications/docs/improving_data_quality.pdf
35. Eysenbach, G., Powell, J., Kuss, O., Sa, E.-R.: Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA J. Am. Med. Assoc.* **287**(20), 2691–2700 (2002). <https://doi.org/10.1001/jama.287.20.2691>
36. Jadad, A.R., Gagliardi, A.: Rating health information on the internet. *JAMA* **279**(8), 611 (1998)
37. Alfano, M., Lenzitti, B., Lo Bosco, G., Muriana, C., Piazza, T., Vizzini, G.: Design, development and validation of a system for automatic help to medical text understanding. *Int. J. Med. Inf.* (2020a). <https://doi.org/10.1016/j.ijmedinf.2020.104109>.
38. Alfano, M., Lenzitti, B., Lo Bosco, G., Taibi, D.: Development and Practical use of a medical vocabulary-thesaurus-dictionary for patient empowerment. In: Proceedings of ACM International Conference on Computer Systems and Technologies (CompSysTech 2018), Ruse (2018)
39. Alfano, M., Lenzitti, B., Lo Bosco, G., Peticone, V.: An automatic system for helping health consumers to understand medical texts. In: Proceedings of HEALTHINF 2015, Lisbon, pp. 622–627 (2015)
40. Alfano, M., Lenzitti, B., Lo Bosco, G., and Taibi, D.: A framework for opening data and creating advanced services in the health and social fields. In: Proceedings of ACM International Conference on Computer Systems and Technologies (CompSysTech 2016), Palermo (2016)