





TAILORED RETRIEVAL OF HEALTH INFORMATION FROM THE WEB FOR FACILITATING COMMUNICATION AND EMPOWERMENT OF ELDERLY PEOPLE

Marco Alfano^{1,5}, Biagio Lenzitti², Davide Taibi³ and Markus Helfert⁴

¹Lero, Dublin City University, Dublin, Ireland

²Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italia

³Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Palermo, Italia

⁴Lero, Maynooth University, Maynooth, Co. Kildare, Ireland

⁵Anghelos Centro Studi sulla Comunicazione, Palermo, Italia

marco.alfano@lero.ie, biagio.lenzitti@unipa.it, davide.taibi@itd.cnr.it, markus.helfert@lero.ie

Keywords: E-Health, Patient Empowerment, Communication Process, Health Information Seeking, User Requirements, Structured Data on the Web.

Abstract: The empowerment of patients/citizens includes, as a basic step, the acquisition of health/medical information helping patients/citizens to understand medical conditions and therapies and act consequently. Nowadays, the main source for this type of information is the World Wide Web (or Web, for short). The retrieval of health information from the Web entails a communication process between patient/citizen and the Web. The information retrieved on the Web can facilitate (or complicate) the communication process between patients/citizens and healthcare professionals such as doctors. The communication process can be modelled by considering the following three levels: Syntactic-Technical, Semantic-Meaning, and Pragmatic-Effectiveness. An efficacious communication between two entities happens when all these communication levels are fully addressed. In the case of retrieval of health information from the Web, although a generic search engine is able to work at the syntactic-technical level, the semantic-meaning and pragmatic-effectiveness parts of the communication process are left to the user and this can present many difficulties, especially for elderly people. This work presents the characteristics and use of a meta-search engine, FACILE, that has been built in order to overcome the challenges confronted during the search process and fully comply with the three levels of the communication process. In particular, FACILE allows a patient/citizen to specify his/her information requirements in a simple way and retrieves the “right” amount of Web content and in a language that he/she can easily understand thus positively affecting his/her understanding and use of the found information and, consequently, the empowerment process.

1. INTRODUCTION

According to the World Health Organization (WHO), empowerment is “a process through which people gain greater control over decisions and actions affecting their health” (WHO, 1998). It includes, as a basic step, the acquisition of health/medical information that helps patients/citizens to understand, among others, medical conditions and therapies, acquire self-confidence to discuss them with medical professionals and, together, make the best-informed decisions (Akerkar & Bichile, 2004; Smith, 2004). The main source of health/medical information is,

nowadays, the World Wide Web (or Web, for short) with the number of Web health information seekers that have been steadily increasing over the years (Pew Research Center, 2013; Taylor, 2010). Search engines are more and more used as the main tool to provide Web information. However, generic search engines do not make any distinction among the users and overload them with the amount of information. On the other hand, specific search engines mostly work on medical literature. Finally, specialized

¹ <https://orcid.org/0000-0001-7200-9547>

² <https://orcid.org/0000-0003-2664-7788>

³ <https://orcid.org/0000-0002-0785-6771>

⁴ <https://orcid.org/0000-0001-6546-6408>

websites (e.g., WebMD⁵, MedlinePlus⁶, Health on Net Foundation Select⁷, MDConsult⁸) are usually not free and contain focused information built by hand. The use of a search engine for Web information retrieval entails a communication process between the user (e.g., a patient) and the search engine. It affects, among others, the amount, understanding and use of the found information. Moreover, the found information may, in turn, facilitate or complicate the communication process between patients and healthcare professionals such as doctors (Smith, 2004).

Communication processes have been modelled in various ways in the past decades. A very famous communication model (if not the most famous) is the one introduced by Morris in relation to its theory of signs (Morris, 1938). It is made up of three levels, i.e., *syntactic*, *semantic* and *pragmatic*. It is paired up by the communication model introduced by Shannon and Weaver (Shannon, 1949) that tackles the problems of communication at the *technical*, *semantic* and *effectiveness* levels. The connection of the two models, as shown in the next section, results in the following three levels:

- Pragmatic - Effectiveness;
- Semantic - Meaning;
- Syntactic - Technical.

Although a complete communication process should take place in order to allow patients/citizens to fully understand and efficaciously use the found information, a quick analysis of the three-level communication model shows how a generic search engine is only able to retrieve information from the Web working at the syntactic-technical level, leaving the semantic and pragmatic part of the communication process in charge of the user. Elderly people, in particular, may have big difficulties in expressing their requirements, and understanding and using the received information, as expected at the semantic and pragmatic communication levels. This prevents them, for example, from having a true understanding of their medical conditions (semantic-meaning level) and acquire the self-confidence to communicate with their doctors and make a shared and informed decision (pragmatic-effectiveness level). Since a generic search engine is not able to work at the three communication levels, it does not constitute a real aid in the process of empowering a patient/citizen, especially an elderly one.

This work presents the characteristics and use of a meta-search engine, FACILE, that has been created in

order to satisfy the user information needs and overcome the challenges confronted during the search process. In particular, FACILE allows a user to specify his/her information requirements in a simple way. It retrieves tailored Web information by exploiting the Web semantic capabilities provided by health-lifesci.schema.org structured data. It provides the “right” amount of Web content, without overwhelming the user, and in a language that he/she can easily understand thus positively affecting his/her understanding of conditions and treatment alternatives and facilitating, ultimately, his/her empowerment process.

The paper is organized as follows. Section 2 illustrates the problems and challenges presented by the retrieval of Web health information in relation to the expected communication process, together with the motivation of the present work. Section 3 presents the characteristics and implementation details of the FACILE meta engine that overcomes the problems and challenges presented in Section 2. Section 4 presents the use and experimental results of FACILE. Sections 5 and 6 present a discussion of the obtained results and some conclusions.

2. BACKGROUND AND MOTIVATION

“Engaging and empowering people & communities” constitutes the first of the five strategies of the “Framework on integrated people-centred health services” of WHO (WHO, 2016). It calls for a paradigm shift on the relation between patients/citizens and health. In fact, empowered 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 (European Health Parliament, 2017; Alfano et al., 2019a; Alfano et al., 2019b; Bodolica et al., 2019; Bravo et al., 2015, Cerezo et al., 2016; Fumagalli et al., 2015).

As seen in Section 1, 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 (Pew Research Center, 2013; Taylor, 2010; UK national statistics, 2010; Instituto Nacional de Estadística, 2010). Search

⁵ <http://www.webmd.com/>

⁶ <http://www.nlm.nih.gov/medlineplus/>

⁷ <http://www.hon.ch/>

⁸ <http://www.mdconsult.com/>

engines are the main tools used to retrieve information from the Web (Pletneva, 2011; Roberts, 2017). 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 or of poor quality. Moreover, the Web is full of information not easily understandable since users, such as patients/citizens, lack a specific expertise in the health domain. Although, some of these problems might be overcome with the advanced features of a search engine, generic users, and mainly elderly people, do not usually have the skills required to use such features and avoid these problems. Specialized search engines, on the other hand, mainly work on medical literature and result quite complex for generic users, and especially the elderly ones. Finally, specialized health/medical websites are mainly built by hand so presenting a limited and often outdated amount of information (compared to what is available on the Web). Moreover, they are often not free.

The retrieval of health information from the Web usually occurs through a communication process between a user (an elderly patient/citizen in our case) and a search engine. Notice that a “complete” communication process entails different levels of communication. Many communication models exist in the literature (Hahn & Paynton, 2014; Cherry, 1966; Johnson & Klare, 1961). One of the most famous and utilized is the one introduced by Charles W. Morris in its theory of signs related to Semiotics (Morris, 1938). It is made up of three levels:

- *Syntactic* (signs and their relation to other signs);
- *Semantic* (relations between signs and designata);
- *Pragmatic* (signs in relation to their users).

Morris’ model has been used for creating communication models, such as the one introduced by Paul Watzlawick et al. in their work on the pragmatics of human communication (Watzlawick et al. 1967). They divide the study of human communication in three sectors: a *syntactic* one, that deals with the *information transmission*, a *semantic* one that deals with *meaning*; and a *pragmatic* one that deals with *behaviour*. Similarly, Shannon and Weaver, in their mathematical theory of communication (Shannon, 1949), tackle the communication problem by using a slightly different terminology:

- *Technical*: How accurately can the information be transmitted?
- *Semantic*: How precisely is the meaning conveyed?

- *Effectiveness*: How effectively does the received information affect behaviour?

Finally, Carlile, in his work (Carlile, 2004), creates a sort of equivalence between Morris’ model and Shannon and Weaver’s one—and then of its terms—to create a framework that can assist to manage knowledge across different parties.

As seen in the Introduction, this work deals with tailored retrieval of health information for user understanding, use and empowerment and consequent improvement in the communication/interaction with medical professionals. With a specific focus on elderly people, the objective is to improve the overall communication between a user and the search engine so that he/she can easily express his/her requirements through everyday language and obtain the “right” and easy-to-understand amount of information.

To this end, we consider a communication process that involves language but it broadens from just the structure, meaning, or use of language, as outlined in Morris’ model, to Shannon and Weaver’s practical concerns about what is required for effective communication across domains. Fig. 1 reports the three-levels communication model and shows an equivalence of the terms used at the three levels and the addition of the term *meaning* at the semantic level (Watzlawick et al. 1967).

The analysis of the communication model of Fig. 1 shows that a generic search engine only works at the syntactic-technical level by retrieving Web page addresses (URLs) based on the keyword(s) specified by the user. It has not been designed to understand the user specific requirements (in the user 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). As a consequence, non-medical experts, and especially elderly people, can have 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) and then are unable to make shared and informed decisions (pragmatic level).



Fig. 1. Three-levels communication model.

A generic search engine, because of its generality in the search of Web information, does not help much in the process of empowering patients/citizens, especially elderly ones. Thus, we have thought of creating a meta-search engine that allows a user to specify his/her information requirements in a simple way and provides the “right” amount of Web content (without overwhelming the user) and in a language that he/she can easily understand. This fully complies with all three levels of the communication model and provides a practical help to the empowerment process.

The next sections present the characteristics and use of such a meta-search engine, FACILE, that has been built in order to satisfy the user information needs and overcome the challenges confronted during the search process.

3 A META ENGINE FOR TAILORED RETRIEVAL OF HEALTH INFORMATION

FACILE is a meta search engine specifically designed to facilitate the empowerment process of patients/citizens through the acquisition of knowledge online. Its objective is to support users in the health information seeking process on the Web according to their specific requirements.

The identification of the main requirements of the health information seekers on the Web has been carried out in (Alfano et al. 2019a) by analysing the works presented in (Pletneva et al. 2011; Banna et al., 2016; Roberts, 2017; Pian et al. 2017; Pang et al., 2015; Keselman, 2008). This literature review, although limited, has consistently shown the following main requirements of health information seekers:

- *Language complexity;*
- *Information classification/customization;*

⁹ <https://schema.org/>

¹⁰ <https://health-lifesci.schema.org/>

- *Information quality* (mainly intended as information trustworthiness).

FACILE has been developed, based on these requirements by exploiting the semantic features of the Web and in particular those related to structured data and *schema.org*⁹ with particular reference to its health-lifesci extension. The use of structured data on the Web is increasing over the years and the exploitation of structured data to collect information from the Web, in different sectors, has proven to be effective (Dietze et al. 2017).

3.1 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 health information seekers. To this end, we have exploited the semantic information available in the World Wide Web and, in particular, that provided by *schema.org*, an initiative funded by some major Web players, that aims to create, maintain, and promote schemas for structured data on the Internet. For the scope of the present work, we consider the *health-lifesci* extension¹⁰ that contains 93 types, 175 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 initiative¹¹. The Web Data Commons (WDC) (Meusel, 2014) 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)¹². The data released in November 2018 have been used in this work. The whole dataset contains about 2.5 billion pages and about 37.1% of them contain structured data.

The dataset dump, used in our study, consists of 31.5 billion RDF n-quads¹³. 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 only the subset containing types, properties and enumeration values of the health-lifesci.schema.org extension.

¹¹ <http://webdatacommons.org/>

¹² <http://commoncrawl.org>

¹³ <https://www.w3.org/TR/n-quads/>

- *MedicalCode*, provides a code for a medical entity.
- *MedicalClinic*, indicates a hospital or a medical school.
- *MedicalSpecialty*, indicates a specific branch of medical science or practice.

We have then taken the audience elements, i.e., *Patient*, *Clinician* and *MedicalResearcher* for the language complexity, the above *health.lifesci.schema.org* elements for the classification/customization of information and some *schema.org* elements related to data provenance (e.g., author and publisher) for the quality information (because the *health-lifesci* extension does not present such elements). This has brought us to create a mapping between the user

requirements and the *schema.org* elements for the two user categories. This mapping expands the one presented in (Alfano et al., 2019a) and it 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. This, in turn, improves their empowerment level and allows a better communication/interaction with the medical professionals.

Table 1. Mapping between *schema.org* elements and user requirements for the two user categories.

	Language Complexity	Information Classification	Information Quality
Non-expert	Audience: - <i>Patient</i>	Type of document: - <i>MedicalWebPage</i> - <i>WebPage</i> Topic Classification: - <i>MedicalCondition</i> - <i>MedicalTherapy</i> - <i>Drug</i> - <i>MedicalCode</i> - <i>MedicalClinic</i> - <i>MedicalSpecialty</i>	Reliability: - <i>author</i> , - <i>publisher</i> , - <i>lastRevised</i> , - <i>datePublished</i>
Expert	Audience: - <i>Clinician</i> - <i>Medical Researcher</i>	Type of document: - <i>ScholarlyArticle</i> - <i>MedicalWebPage</i> Topic Classification: - <i>MedicalCondition</i> - <i>MedicalTherapy</i> - <i>Drug</i> - <i>MedicalCode</i> - <i>MedicalClinic</i> - <i>MedicalSpecialty</i>	Reliability: - <i>author</i> , - <i>publisher</i> , - <i>lastRevised</i> , - <i>datePublished</i>

3.3 FACILE Implementation

The mapping between user requirements and *schema.org* elements, shown in the previous section, has been used to build a meta search engine, FACILE, that provides the different audience types with the proper Web contents in terms of language complexity, information quality and information classification. It expands an initial version of the system that only takes into account the language

complexity user requirement (Alfano et al, 2019c). The FACILE meta search engine is available at the address

<http://www.math.unipa.it/simplehealth/faciles>

and Fig. 3 shows its user interface.



Fig. 3. FACILE user interface.

The user interface includes a simple text input, similar to that of a generic search engine, where the user can insert the term(s) to be searched. Moreover, the user can decide to filter the results in order to get more focused information and not be overwhelmed with the amount of information that generic search engines provide. This is simply done through the two switch buttons shown in Fig. 3. The first switch button, “Specify the research”, acts on the classification user requirement. It allows to search for either medical “Web Page” or for “Scholarly article” (Fig. 4). Medical Web pages usually present a mixed language and contain different types of information making them suitable to different target audiences. Scholarly articles are mostly research papers with a more technical language and mainly targeted to medical experts. The second switch button, “Show Quality Information”, acts on the quality user requirement and provides information such as the last time the Web page has been reviewed, the publication date and the author or the publisher of the page (when available). This switch button does not filter out results, but it shows additional information useful to evaluate the information trustworthiness. When selecting “Web Page”, with the first switch button, some sub-filters appear in order to allow a classification of the Web pages in terms of the required type of medical information (Fig. 4).



Fig. 4. Web page sub-filters.

The Web Page sub-filters can be activated through checkboxes that can be easily selected by a user with no specific knowledge (such as an elderly patient), because they indicate common terms in the health domain. More than one checkbox can be selected each time and the filtering will be performed using the mapping of Table 1 and providing the following information:

- **Condition:** It will present the pages that contain a description of a medical condition (searched keyword) and the values of the properties related to *MedicalCondition* schema.org element.
- **Therapy:** It will present the pages that contain the therapy for a medical condition (searched keyword) and the values of the properties related to *MedicalTherapy* schema.org element.
- **Drug:** It will present the pages that contain the medicines related to a medical condition (searched keyword) and the values of the properties related to *Drug* schema.org element.
- **Code:** It will present the pages that contain the code of a medical condition (searched keyword) and the values of the properties related to *MedicalCode* schema.org element. The code, together with the coding system, can be used to look on specialized website to find a specific condition/part of the body/therapy/drug, or other useful information.
- **Institution:** It will present the pages that contain the institutions (e.g., hospitals) that deal with a medical condition (searched keyword) and the values of the properties related to *MedicalClinic* schema.org element.
- **Specialist:** It will present the pages that contain the values of the properties related to

MedicalSpecialty schema.org element among which the medical specialists that deal with a medical condition (searched keyword).

- **Audience:** It will indicate the target audience of the Web page (patient/clinician/medical researcher) and the values of the properties related to *MedicalAudience* schema.org element.

Finally, when a user is interested in finding more technical information, he/she can select “Scholarly Article”. He/she will also have the possibility to specify a year in order to get the articles that have been reviewed that year or later (Fig. 5).



Fig. 5. Scholarly Article sub-filters.

4 EXPERIMENTAL USE

FACILE, as shown in the previous section, has been designed and implemented to be used by different user typologies, i.e., medical experts and non-experts. Nevertheless, in relation to what discussed previously, we have executed some tests to evaluate FACILE effectiveness and usefulness for its use with non-experts and mainly with elderly people.

Preliminarily, notice that FACILE can be used as a generic search engine by inserting any keyword on the text input (Fig. 3) and receiving at most the top 50 results. Notice that a user usually analyses the first 25-30 results when using a generic search engine such as Google™ (Alfano et al. 2019b). Thus, from this point of view, the user is not penalized by using FACILE even though it is not using the whole Web but only the part that contains *schema.org* structured data. Thus, if we search for the *hepatitis* keyword, FACILE will provide the first 50 URLs that present the higher number of *hepatitis* occurrences (Fig. 6)



Fig. 6. First 10 results of generic search for *hepatitis* keyword.

The effectiveness of FACILE can be appreciated as soon as the “Web page” filter and some of its sub-filters are used. In the example of Fig. 7, the “Condition” and “Code” sub-filters are checked. In this case FACILE provides just ten results so reducing the number of pages to be analysed and focusing to the Web content of interest. This allows an elderly patient, for example, to save a great deal of time by just focusing on the specific type of information he/she has requested (condition in this case). Moreover, FACILE provides some information directly in the response page such as the describe condition (Specialty: Gastroenterologic and Name: Cirrhosis) that can be used for further investigation if of interest. Also the code identifier (e.g., CUI in the UMLS coding system) and code value (e.g., C0019167 that corresponds to *Hepatitis B e Antigens*), can be used to have a unique reference of the information that is being read and that can be used for further discussion with a doctor or in a hospital. Notice that we plan to implement the translation of the code value in its corresponding term so to add this information directly in the response window.



Fig. 7. First results of search for *hepatitis* keyword with Condition and Code sub-filters.

If the user enables the “Show Quality Information?” switch button, the quality information, such as “Date published” or “Last reviewed”, will be shown (Fig. 8) and the user will have the possibility to choose, for

example, the most recent information. If the author or publisher information are present, the user will have the possibility to analyse the information only if he/she trusts the source it is coming from.



Fig. 8. First results of search for *hepatitis* keyword with Quality filter activated.

If the user wants to find pages that match his/her medical literature level, he/she will check the “Audience” sub-filter. In this case the target audience of the page will be shown (wherever present - Fig. 9) and the user will be able to select the most suitable pages. For example, in the case of Fig. 9, an elderly patient would not analyse the pages whose audience is “clinician” so avoiding to waste time with Web content that is too difficult to be understood.



Fig. 9. First results of search for *hepatitis* keyword with Audience sub-filter activated.

Notice that in (Alfano et al., 2019c) we have evaluated the language familiarity of Web pages targeted to different audience types. This has been done by computing the “term familiarity index” of a word (i.e., number of results provided by the Google search engine, Kloehn, et al., 2018; Leroy, et al., 2012) and then computing the language familiarity of a Web page as the average of the term familiarity indexes of its words. The results clearly show that, on average, the Web pages targeted to patients have a much higher language familiarity, and thus a simpler terminology than the Web pages targeted to clinicians or medical researchers.

For sake of completeness of all the filtering possibilities, we can finally assume that a medical expert is looking for information on *hepatitis*. In this case, given his/her skills, he/she will probably search for scholarly articles obtaining a result such as the one reported in Fig. 10 that provides more technical information.



Fig. 10. First results of search for *hepatitis* keyword with “Scholarly article” filter activated.

5 DISCUSSION

Although, the development of FACILE is still at prototypal stage and more experiments are undoubtedly needed, its principles and practical use, presented in the previous sections, show how FACILE complies with the three-level communication model presented in Section 2 (Fig. 1) allowing a non-expert user (e.g., an elderly patient) to specify his/her requirements in simple language, translating into structured data elements for retrieving health information from the Web, and providing the user with the “right” amount of reliable and simple information that he/she can easily understand and employ in his/her process of empowerment and in subsequent communication/interaction with medical professionals. The compliance of Facile with the three levels of the communication model can be further analysed as follows:

1. **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 semantic Web) and returns the requested information as any generic search engine does.
2. **Semantic (Meaning):** The translation capability of FACILE (mapping model) allows the user to specify his/her requirements in simple terms and translates them into *schema.org* elements that FACILE uses in order to extract the appropriate information from the Web (communication phase: user to FACILE). Moreover, the ability of FACILE, to retrieve Web pages that present different language complexity levels, allows the user to choose the pages whose language can be easily understood (communication phase: FACILE to user).
3. **Pragmatic (Effectiveness):** The response presented by FACILE has a pragmatic impact on the user (mainly a non-expert one) in terms of focused and reliable Web results (by avoiding to spend time analysing the large amount of results that a generic search engine provides), the specific information provided in the response page, and the language simplicity. The consequence is that a user 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, improves his/her ability to understand his/her medical conditions and increases his/her self-confidence to communicate with medical professionals to make informed decisions.

Notice that the process of finding the “right” information may be iterative. Although, FACILE has the objective of immediately providing the user with the desired information, its flexibility and easiness of use allows the user to perform further searches when needed in order to get further information. For example, first he/she might need to understand a medical condition/disease and then find a therapy and the medicines for it. The user might also want to remove the filtering information, at some stage, in order to have the possibility of analysing more pages at the same time. Having already analysed the more focused information (and related Web pages), he/she will have the ability and confidence to expand his/her exploration by quickly analysing the remaining information (or part of it). 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, the 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 act upon them. Notice that, although, most of the time, users search for health information on the Web before or after having consulted a medical professional as a doctor, the ideal case would see a doctor accompanying the patient in his/her navigational path on the Web so to help him/her to find the most suitable and reliable medical information. Of course, this is not always possible, due to time constraints. Nevertheless, we are working with medical professionals (doctors and nurses) to understand (and implement as much as possible) their advices to patients for navigating the Web in search of “good” health/medical information.

6 CONCLUSIONS

In this work, we have presented a three-levels communication model that allows a full interaction of a patient/citizen with a search engine and then an efficacious retrieval of health information from the Web. We have shown as a generic search engine only reaches the first level of the communication model because it does not allow the user to specify his/her requirements and thus it does not provide focused information that the user can easily and promptly understand and use.

We have then presented FACILE, a meta search engine, that allows a user to specify his/her information requirements in a simple way and maps them to *schema.org* structured elements. It then retrieves the “right” and simple Web content without overwhelming the user. This positively affects his/her understanding and use and, as a consequence, the empowerment process.

The principles and first experimental results are satisfying and show FACILE potentialities even though, the used dataset (created with the 2018 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 2017 and 2016 datasets (the three datasets present differences) so to have more data to experiment with and, hopefully, more results.

A deeper analysis also needs to be executed in order to better understand the mapping between the user requirements and the *schema.org* elements, as in the case of the information quality. Furthermore, although FACILE is simple and intuitive, we are in

the process of running some tests with elderly people to evaluate their engagement level in using FACILE and understand what are the aspects that need to be improved. An evaluation of the reached empowerment level is also important and we are in the process of running a randomized control trial with patients with congenital heart diseases. After using FACILE, the reached health literacy and empowerment levels will be measured.

As a future work, we plan to simplify the text of Web pages with complex terminology (Alfano et al, 2018; Alfano et al., 2015) so that also non-experts can have access to this kind of information. Finally, we want to create a visual framework (Alfano et al., 2016) that uses the retrieving capabilities of FACILE and allows easy creation of advanced health services for elderly people such as virtual assistants.

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 Irish Software Research Centre (www.lero.ie).

We would like to thank Allan Gomes, of the University of Nantes, for the technical support in developing the FACILE search engine.

REFERENCES

- Alfano, M., Lenzitti, B., Taibi, D., Helfert, M., (2019a). Provision of tailored health information for patient empowerment: An initial study. In *CompSysTech '19 Proceedings of the 20th International Conference on Computer Systems and Technologies* (pp. 213-220).
- Alfano M., Lenzitti B., Taibi D., Helfert M. (2019b) ULearn: Personalized Medical Learning on the Web for Patient Empowerment. In: Herzog M., Kubincová Z., Han P., Temperini M. (eds) *Advances in Web-Based Learning – ICWL 2019*. ICWL 2019. Lecture Notes in Computer Science, vol 11841. Springer, Cham.
- Alfano, M., Lenzitti, B., Taibi, D., Helfert M., (2019c). Facilitating access to health Web pages with different language complexity levels. *Proc. of the 5th Inter. Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2019)*, 2-4 May 2019, Heraklion-Crete.
- Alfano, M., Lenzitti, B., Lo Bosco, G., and Taibi, D., 2018. Development and Practical Use of a Medical Vocabulary-Thesaurus-Dictionary for Patient Empowerment. *Proc. of ACM International Conference on Computer Systems and Technologies (CompSysTech'18)*, Ruse.
- Alfano, M., Lenzitti, B., Lo Bosco, G., and Taibi, D., 2016. A Framework for Opening Data and Creating Advanced Services in the Health and Social Fields. *Proc. of ACM International Conference on Computer Systems and Technologies (CompSysTech'16)*, Palermo.
- Alfano, M., Lenzitti, B., Lo Bosco, G., and Perticone, V., 2015. An Automatic System for Helping Health Consumers to Understand Medical Texts, *Proc. of HEALTHINF 2015*, Lisbon, pp. 622-627.
- Akerkar, S., & Bichile, L., 2004. Health Information on the Internet: Patient Empowerment or Patient Deceit? *Indian Journal of Medical Sciences*, 58(8). Pp. 321-326.
- Banna, S., Hasan, H. & Dawson, P., 2016. Understanding the diversity of user requirements for interactive online health services. *International Journal of Healthcare Technology and Management*, 15(3).
- Bodolica, V., & Spraggon, M. (2019). Toward patient-centered care and inclusive health-care governance: a review of patient empowerment in the UAE. *Public Health*, 169(971), 114–124.
- Bravo, P., Edwards, A., Barr, P. J., Scholl, I., Elwyn, G., & McAllister, M. (2015). Conceptualising patient empowerment: A mixed methods study. *BMC Health Services Research*, 15(1), 1–14.
- Carlile P (2004) Transferring, Translating and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries, *Organization Science*, 15(5), pp. 555-68.
- Cerezo, P. G., Juvé-Udina, M. E., & Delgado-Hito, P. (2016). Concepts and measures of patient empowerment: A comprehensive review. *Revista Da Escola de Enfermagem*.
- Cherry, C. (1966). *On human communication: a review, a survey, and a criticism*. Cambridge, Mass, M.I.T. Press.
- Dietze S., Taibi D., Yu R., Barker P., d'Aquin M., 2017. Analysing and Improving Embedded Markup of Learning Resources on the Web. *Proc. of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*, Switzerland, 283-292.
- European Health Parliament. (2017). *PATIENT EMPOWERMENT AND CENTREDNESS*.
- Fumagalli, L. P., Radaelli, G., Lettieri, E., Bertele', P., & Masella, C. (2015). Patient Empowerment and its neighbours: Clarifying the boundaries and their mutual relationships. *Health Policy*. 119(3), 384–394.
- Hahn, L. K. & Paynton, S. T. 2014. *Survey of Communication Study*. https://en.wikibooks.org/wiki/Survey_of_Communication_Study.
- Health promotion glossary. Geneva: World Health Organization; 1998.

- Instituto Nacional de Estadística. 2010. Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares.
- Johnson, F.C., Klare, G.R. 1961. General Models of Communication Research: A Survey of the Developments of a Decade, *Journal of Communication*, Volume 11, Issue 1, March 1961, Pages 13–26.
- Keselman, A., Logan, R., Smith, C. A., Leroy, G., & Zeng-Treitler, Q. (2008). Developing informatics tools and strategies for consumer-centered health communication. *Journal of the American Medical Informatics Association : JAMIA*, 15(4), 473–483.
- Kloehn, N. et al., 2018. Improving consumer understanding of medical text: Development and validation of a new subsimplify algorithm to automatically generate term explanations in English and Spanish. *Journal of Medical Internet Research*, 20(8).
- Leroy, G. et al., 2012. Improving perceived and actual text difficulty for health information consumers using semi-automated methods. *AMIA Annual Symposium Proceedings*. pp.522–31.
- Morris, C.W. 1938. Foundations of the theory of signs. *International encyclopedia of unified science*, vol. 1, no. 2. The University of Chicago Press, Chicago.
- Meusel, R., Petrovski, P., and Bizer, C. 2014. The WebDataCommons Microdata, RDFa and Microformat Dataset Series. *Proc. of the 13th International Semantic Web Conference (ISWC14)*, Springer-Verlag New York, USA, 277-292.
- Pang, P. C.-I.; Verspoor, K.; Pearce, J.; Chang, S. 2015. Better Health Explorer: De-signing for Health Information Seekers. In OzCHI '15 Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (pp. 588–597).
- Pletneva, N., Vargas, A. & Boyer, C., 2011. Requirements for the general public health search. *Khresmoi Public Deliverable D8.1.1*.
- Pew Research Center, 2013. Health online 2013, <http://www.pewinternet.org/2013/01/15/health-online-2013/>.
- Pian, W., C.S.G. Khoo, J. Chi. 2017. Automatic classification of users' health information need context: Logistic regression analysis of mouse-click and eye-tracker data. *Journal of Medical Internet Research*, 19(12).
- Roberts, T. 2017. Searching the Internet for Health Information: Techniques for Patients to Effectively Search Both Public and Professional Websites. *SLE Workshop at Hospital for Special Surgery Tips For Evaluating the Quality of Health*, 1–12.
- Shannon, C., W. Weaver. 1949. *The Mathematical Theory of Communications*. University of Illinois Press, Urbana, IL.
- Smith, T. 2004. Exploring the characteristics of active health seekers, the thinking behind patient preferences, and the implications for patient-professional relationships. *Quality and Safety in Health Care*, 13(6), 474–477.
- Taylor, H. 2010. HI-Harris-Poll-Cyberchondrics. Harris Interactive. <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/>.
- UK national statistics, 2010. Statistical bulletin: Internet Access 2010. *Office for National Statistics*. 27 Aug 2010.
- Watzlawick, P., Beavin, J. B., & Jackson, D. D. 1967. *Pragmatics of human communication: a study of interactional patterns, pathologies, and paradoxes*. Norton, New York.
- World Health Organization (WHO). 2016. Framework on integrated, people-centred health services: Report by the Secretariat. *World Health Assembly*, (A69/39), 1–12.
- World Health Organization (WHO). 1998. Health promotion glossary. <https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf>.