Health Suggestions: A Chrome Extension to Help Laypersons Search for Health Information

Carla Teixeira Lopes[1,2(✉)] and Tiago Almeida Fernandes[1]

[1] Department of Informatics Engineering, Faculty of Engineering,
University of Porto, Porto, Portugal
`{ctl,ei11054}@fe.up.pt`
[2] INESC TEC, Porto, Portugal

Abstract. To help laypeople surpass the common difficulties they face when searching for health information on the Web, we built Health Suggestions, an extension for Google Chrome to assist users obtaining high-quality search results in the health domain. This is achieved by providing users with suggestions of queries formulated in different terminologies and languages. Translations of health expressions might not be obvious and queries in languages with a strong presence on the Web facilitate the access to high-quality health contents. On the other hand, the use of lay terminology might contribute to increase users' comprehension and medico-scientific terminology to obtain more detailed and technical contents. Results show a good acceptance of the suggestions, confirm the utility of a multilingual and multi-terminology approach and show its usefulness to more successful searches.

Keywords: Query suggestion · Health · Interactive IR · Multilingual

1 Introduction

Nowadays it is common to check the Internet for health-related information before consulting the doctor, 72 % of Internet users say they looked online for health information within the past year [1]. However, lack of scientific vocabulary or searching in languages with less information (e.g.: Portuguese compared to English) might limit users' access to health information on the Web. In the health domain, obtaining non-accurate contents might result in serious consequences. Since few users assess the quality of health-related web contents and a minority (41 %) checks their diagnosis with medical professionals [1], it is urgent to remove or reduce barriers in health information access.

Expectations on search results may differ depending on the user's knowledge. Although searching with medico-scientific terminology leads to more detailed contents, these may not be understandable for users with less health literacy. On the other hand, users with more expertise in a subject might prefer more scientific contents [3].

Health Suggestions assists laypeople searching for medical information, providing suggestions using lay and medico-scientific terminology related with user's original query. For now, we provide suggestions both in Portuguese and English despite Portuguese being only shown if it is the original query language. The assistance is made in search engines, the starting point for health information seeking for 77 % of people [1]. We are supporting: Google, Bing and Yahoo, having together a share of 89.9 % worldwide in 2016 [5]. We aim to provide laypeople with a mechanism for reaching higher-quality health contents fitted to users' health expertise.

2   Related Work

The impact of queries terminology on health information retrieval has already been studied. A study addressing the impact of lay and medico-scientific terminology in users with different levels of health literacy and topic familiarity [3], concluded that medico-scientific queries achieve higher precision. However, documents retrieved with these queries are less readable and less understood by users. Authors conclude that a medico-scientific query should be suggested whenever a person has the capacity to digest scientific documents. If the person has not enough health literacy, a lay query is preferred.

A different study evaluated the impact of translating a health query to English in users with different English proficiencies [2]. Authors concluded that non-English' speaking users having at least elementary English proficiency can benefit from this type of query translation.

3   Health Suggestions

We decided to implement Health Suggestions as a Google Chrome extension for two main reasons. First, since our focus is web information retrieval, we wanted to extend browser's functionality and reach all the main search engines. Second, we picked Chrome due to its current dominance as users' preferred browser. On May, 2016 it had a share of 71.4 % [8]. The extension is available on the Web[1].

Health Suggestions presents itself as a panel in Chrome's bottom right corner (Fig. 1) where several actions can be triggered, including: search for suggested queries; switch search engines; minimise/maximise the panel and close the panel. For example, if a suggestion is clicked a new search is performed for that suggestion, or if Bing's icon is clicked then the same search is performed in Bing.

To provide an easy access to all the above actions, the panel follows the user across the entire health-related search session. The panel is only visible when Health Suggestions finds a match for the query in our data structures. This way, the panel won't disturb the user in other types of searches. The extension has an options page (Fig. 2) which can be accessed by clicking the extension icon (blue heart) in the navigation bar. In this page the user can: turn the extension on/off;

---

[1] http://irlab.fe.up.pt/p/healthsuggestions.

Fig. 1. Suggestions' panel

Fig. 2. Options page

(dis)allow logging; opt for a local or remote database; specify queries' language or ask the extension to automatically detect it.

The system is divided in two distinct modules: suggestion engine and logging engine. The suggestion engine is responsible for generating the suggestions. It uses an English and Portuguese version of the Consumer Health Vocabulary (CHV), a vocabulary that links everyday words about health to technical terms used by health care professionals [6]. The Portuguese version was translated using the Google Translator API with an estimated accuracy of 84.2 %, notwithstanding we are continuously working to improve it [7]. The extension is standalone as the user can decide to use a local database instead of the remote one.

The architecture of the suggestion system is similar to the one used in a previous work [4] where more information is provided. In sum we created two inverted indices, one in each language, with the CHV. In the inverted index each stemmed term is associated with an inverse string frequency (isf) and a postings list, that is, the list of CHV strings where it appears. To determine the vocabulary of terms, CHV's strings were tokenized and stop words were removed. In the remaining terms, letters were reduced to lower case, the diacritics were removed and terms were also stemmed.

The steps to obtain suggestions are presented in Fig. 3. Steps in blue are performed in the extension while steps in red can be either performed in the extension or in the server. It can be either a function call, if in standalone mode, or a call to a REST API in the server. The majority of the work is accomplished on the extension in order to achieve a faster response time.
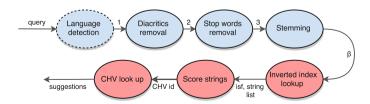

Fig. 3. Suggestion engine architecture

The extension was generated using a boilerplate[2] following the best practices for Google Chrome extensions development. The technologies involved are the ones used for front-end development: HTML, CSS and JavaScript. For the databases, we used WebSQL for storing the inverted indices and the CHV, we picked it because in the server this data is already stored in a SQL database. Local database updates are subject to extension updates. We have used the following third party libraries: Snowball.js[3]; franc.js[4]; sha1.js[5]; TimeMe.js[6]. The server provides a REST API. We have decided to use Django-Rest-Framework (Python) due to its ease of use and the possibility to obtain a working API vary fast. We chose PostgreSQL as our database management system.

In order to understand users' search process and to improve Health Suggestions we developed a logging system that tracks most of users' actions while performing health-related searches. The logs are anonymous and the user can disable them through the options page. Logging is focused on: query, search engine, search results, search engine's related queries and the suggestions provided by the Health Suggestions; time spent on SERP and Health Suggestions' panel; visited web pages: time on page and number of scroll events; clicks: on the extension's panel, search results or any page hyperlink; copy/cut events; find events. The logging engine is divided in two modules: the extension and the server. In the client we decided to use IndexedDB for its capacity to save JavaScript objects and its transaction support. In the server, logs are saved permanently in a PostgreSQL database.

4   Experiment

An user experiment was conducted to evaluate Health Suggestions. Four major research questions drove the experiment: (1) How are suggestions used? (2) Why are suggestions used? (3) How do users assess the utility of the suggestions provided by the system? (4) Do the suggestions lead to a more successful search?. For the fourth question, we considered the number of relevant documents saved by the users and their feeling of success with the task.

For the study, 36 students (30 female and 6 male) were recruited. Their average age is 21,9 (standard deviation 5,9). The majority of the participants search the web for 8 to 13 years (78 %). The search frequency for 19 % is daily and for 81 % it is more frequent than daily. In a scale from 1 to 5 about how often participants find what they are looking for, 4 participants (11 %) classified it as 3, 26 participants (72 %) as a 4 and 6 participants (17 %) as a 5.

---

[2] http://extensionizr.com.
[3] https://github.com/fortnightlabs/snowball-js.
[4] https://github.com/wooorm/franc.
[5] https://github.com/sytelus/CryptoJS.
[6] https://github.com/jasonzissman/TimeMe.js.

4.1 Setup

For the study we used two systems, one including Health Suggestions and the other one without it. Participants were divided in two groups. Half of them, the assisted group, used the system with the extension and the other group, the unassisted one, didn't have access to the suggestions. None of the participants was involved in the definition of the task. Participants were unaware of the fact that some received assistance and that some did not. No significant difference in average search experience was found between the groups, neither regarding the number of years searching the Web (10.9 years vs. 9.7 years, p-value = 0.2), neither regarding users' self assessment of success in web searching (both groups with a mode and median of 4).

Each participant was assigned 4 simulated work task situations. To define the simulated situations, we asked 20 persons with no medical expertise and spanning a wide range of ages and education levels to state the health topic for which they had most recently searched on the Web. Then, we randomly selected topics and created a scenario for each. The information situations were described to the users in Portuguese. The order of the tasks was rotated.

Users had to freely formulate 3 queries for each task and save the relevant documents from the top 10 results for each query. In the end of each search task, the users of the assisted group had to explain if and how they used the suggestions (clicked on them, used terms from one of the suggestions, used terms from several suggestions), why they were considered useful and were also asked to assess their utility.

4.2 Results

How are suggestions used? To answer this research question we analysed the behaviour of users in the assisted group. Suggestions were presented in 71 % of the issued queries. In 27 % of these cases users clicked on them, in 15 % have used one suggestion to extract terms that were then used in the following query and in 4 % of the cases have extracted terms from more than one suggestions.

Why are suggestions used? Users presented 5 main reasons for using suggestions. In 35 % of the cases, users simply said they used them because they presented synonymous. However, the most popular reason is presenting alternatives in medico-scientific terminology (37 %). English suggestions are also very valued (24 %). Although in a smaller scale, lay (3 %) terminology and the Portuguese language (1 %) were also reasons for using the suggestions.

How do users assess the utility of the suggestions provided by the system? We asked users to assess the utility of the provided suggestions in a scale of 1 (not useful) to 3 (useful). From the set of presented suggestions, 38 % were considered useful and 33 % partially useful. The remaining proportion (29 %) was considered not useful.

Do the suggestions lead to a more successful search? The average number of relevant documents saved by the users was significantly higher in the assisted group (16.3 versus 14.1, p-value = 0.017). Regarding users' assessment of task

success, in a scale of 1–5, we found that the unassisted group had a larger median (5) than the assisted group (4). However, this is not a significant difference.

5   Conclusions and Future Work

By providing both lay and medico-scientific suggestions, Health Suggestions support laypersons reaching information accordingly to their knowledge. Although preliminary, results show a good acceptance of the suggestions, confirm the rationale that is behind the development of this extension, that is, the utility of a multilingual and multi-terminology approach and show their usefulness in the retrieval of a larger number of relevant documents.

Our next step is to use the data logged in this user study, to deeply understand people interactions with the search panel and the given suggestions. We are also planning to analyse if the extension contributes to improve the knowledge users acquire during the health searching.

References

1. Fox, S., Duggan, M.: Health online 2013. Technical report, Pew Research Center's Internet & American Life Project, Washington, D.C., January 2013
2. Lopes, C.T., Ribeiro, C.: Measuring the value of health query translation: an analysis by user language proficiency. J. Am. Soc. Inform. Sci. Technol. **64**(5), 951–963 (2013). http://dx.doi.org/10.1002/asi.22812
3. Lopes, C.T., Ribeiro, C.: Effects of terminology on health queries: an analysis by user's health literacy and topic familiarity, vol. 39, Chap. 10, pp. 145–184. Emerald Group Publishing Limited (2015). http://www.emeraldinsight.com/doi/abs/10.1108/S0065-283020150000039013
4. Lopes, C.T., Ribeiro, C.: Effects of language and terminology on the usage of health query suggestions. In: Conference and Labs of the Evaluation Forum (2016)
5. NetMarketShare: Search Engine Market Share (2016). https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0&qpsp=2016&qpnp=1&qptimeframe=Y. Accessed 17 June 2016
6. NLM: 2012AA Consumer Health Vocabulary Source Information (2012). http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/CHV/index.html
7. Silva, A.C., Lopes, C.T.: Health translations: a crowdsourced, gamified approach to translate large vocabulary databases. In: CISTI 2016 - 11th Iberian Conference on Information Systems and Technologies (2016)
8. w3schools.com: Browser Statistics (2016). http://www.w3schools.com/browsers/browsers_stats.asp. Accessed 17 June 2016