

Context-Based Health Information Retrieval

Carla Teixeira Lopes

Department of Informatics Engineering
Faculty of Engineering, University of Porto
carla.lopes@fe.up.pt

July, 2013

Abstract

The Web has profoundly changed the way we access information. The increase in the amount of information available and the easy access to the Web have contributed to the popularity of search in this medium. While this is true in several domains, it is particularly significant in the health domain. In fact, search for health information is the third most popular online activity and it is performed by almost 3 out of 4 American Internet users. Besides its popularity, health web search is particularly important to inform health consumers, encouraging them to become more participatory in their health management. In the health domain, the context surrounding the search is extremely rich and we believe it can contribute to improve system's performance. Although information retrieval has made significant progresses supporting its retrieval methods solely on the query and document collection, it has been recognized that context should be explored. Context can, for example, be used to disambiguate a query, to retrieve documents adjusted to the expertise of the searcher or to adjust documents to the patient's medical record.

In this dissertation we investigate the effects of context features in consumer health information retrieval. In addition, we propose and evaluate strategies to use context features in query formulation support. The popularity of web search on the health domain made us focus on Web retrieval. We opt to concentrate on health consumers due to the lack of research focusing on this specific public. The thesis behind our work is that consumer health information retrieval is affected by context features, which can be used in query formulation support to improve retrieval performance.

In three exploratory studies, we analyze how a large number of features affect different outcomes of the retrieval process. Based on findings from these studies and on the importance of query formulation support in a domain where terminology can be a barrier and translations to other languages are not always obvious, we explored the impact of query translations in users with different characteristics. Based on the assumption that a query using a language that is popular on the Web may easily reach high-quality contents, one study analyzes the effects of translating a query to the English language in users with different levels of English proficiency. Findings show that users having, at least, elementary English proficiency benefit from English query suggestions. The other studies focus on query terminology translation, that is, translation between lay and medico-scientific terminology, considering users' health literacy and topic familiarity. Although several strategies have been previously proposed to

overcome the terminology gap between health consumers and web documents, none considers users health literacy and topic familiarity. Findings suggest that users with inadequate health literacy and users who are unfamiliar with the topic should be provided with recommendations of lay queries. On the other hand, users with higher health literacy or higher topic familiarity should be given alternative queries using medico-scientific terminology.

Based on the above findings, we developed a query suggestion system that, using domain information gathered from an existing consumer health vocabulary, identifies the medical concepts included in the query and returns four types of suggestions combining the Portuguese or English languages with the lay or medico-scientific terminologies. We found that suggestions offered by the system had a good acceptance, with English suggestions being preferred to Portuguese ones in *basic* and *proficient* users and medico-scientific suggestions being preferred to lay ones in higher levels of health literacy. We concluded that a retrieval system including the implemented suggestion strategy without any kind of personalization tends to be better than a system without suggestions with respect to precision, correctness of the resulting knowledge and also of its incorrectness. We also concluded that this system tends to be slightly worse in terms of motivational relevance. Of these, only the incorrectness difference is significant. This is extremely relevant in the health domain where incorrect information can have serious consequences. Moreover, we also found that the personalization of this system to users' English proficiency and health literacy, biasing users towards the suggestions more beneficial to them, outperforms the system without personalization, in terms of medical accuracy of the obtained knowledge.

Soon available at: http://www.carlalopes.com/pubs/lopes_PhD_2013.pdf.