Information Processing and Management 53 (2017) 332–350

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

On the impact of domain expertise on query formulation, relevance assessment and retrieval performance in clinical settings

Lynda Tamine*, Cecile Chouquet

Paul Sabatier University, 118 route de Narbonne, Toulouse 31062 Cedex 9, France

ARTICLE INFO

Article history:
Received 28 July 2016
Revised 15 November 2016
Accepted 17 November 2016

Keywords:
Medical information retrieval
Domain expertise
Query formulations
Relevance assessment
Retrieval performance

ABSTRACT

The large volumes of medical information available on the web may provide answers for a wide range of users attempting to solve health-related problems. While experts generally utilize reliable resources for diagnosis search and professional development, novices utilize different (social) web resources to obtain information that helps them manage their health or the health of people who they care for. A diverse number of related search topics address clinical diagnosis, advice searching, information sharing, connecting with experts, etc. This paper focuses on the extent to which expertise can impact clinical query formulation, document relevance assessment and retrieval performance in the context of tailoring retrieval models and systems to experts vs. non-experts. The results show that medical domain expertise 1) plays an important role in the lexical representations of information needs; 2) significantly influences the perception of relevance even among users with similar levels of expertise and 3) reinforces the idea that a single ground truth does not exist, thereby leading to the variability of system rankings with respect to the level of user's expertise. The findings of this study presents opportunities for the design of personalized health-related IR systems, but also for providing insights about the evaluation of such systems.

1. Introduction

Several studies (Fox & Duggan, 2013; Fox, 2011) have clearly shown that people, both experts (e.g., physicians and nurses) and novices (e.g., patients and their family), have strong desires for medical information. Regardless of the domain expertise of users seeking information, medical and health-search have been acknowledged as a complex search tasks leading to search failures or biases (Ely et al., 2005, 2007; Roberts, Simpson, Demner-Fushman, Voorhees, & Hersh, 2015; White & Horvitz, 2015). Even if it appears that the effectiveness of specialized search engines within the medical domain is not significantly higher than the effectiveness of general web search engines (Bin & Lun, 2001), several previous studies have revealed that significant differences between them in search intents may be linked to the information resources being used (Choudhury, Morris, & White, 2014; Natarajan, Stein, Jain, & Elhadad, 2010; Zhang & Fu, 2011):

- *General web resources*: This category of resources includes resources indexed by general web search tools and social platforms not particularly devoted to or certified for health concerns, thereby leading to general web searching (which is

* Corresponding author.
E-mail addresses: tamine@irit.fr, lechani@irit.fr (L. Tamine), cecile.chouquet@math.univ-toulouse.fr (C. Chouquet).

http://dx.doi.org/10.1016/j.ipm.2016.11.004
0306-4573/

different from a vertical search). Searching the web for health-related information has been acknowledged as a frequent activity of a wide variety of users (Fox & Duggan, 2013; Spink et al., 2004; Zhang & Fu, 2011). The web is used for addressing a wide range of search topics, such as those concerning (Choudhury et al., 2014; Eysenbach, Powell, & Englesakis, 2004; Spink et al., 2004; White & Horvitz, 2009; Zhang & Fu, 2011): 1) general health, drug and dosing, and disease management (searching for rare diseases or updates on common diseases); 2) (differential) diagnosis or referral guidelines; 3) professional development; 4) personal and opinion-oriented goals (personalized healthy lifestyle information such as diet, nutrition, and sexual health information); 5) advice (e.g., advice after being dissatisfied with professional care); 6) information sharing (e.g., with doctors and/other patients) ; 7) people with similar conditions on social platforms; and 8) connecting with experts.

- *Clinical information resources*: This category of resources is used within a domain-specific or vertical search, including 1) electronic health records (EHRs) that are used by medical professionals and 2) medical scientific reviews or content from certified health and medical sites that are both used by experts (e.g., clinicians) and non-experts (novices) for different purposes. Expert clinical information searches are generally performed by clinicians under the Evidence-Based Medicine (EBM) approach (Sackett, 1997) as the basis for clinical decisions that better suit the patient under consideration. In contrast, non-expert clinical information searches are completed to help patients and their representatives to better understand their own health conditions or conditions of people they care for. Searching for clinical information is also a common pursuit. A previous study showed, for example, that 1.8 billion clinical searches were conducted on PubMed in 2011 (NLM, 2012); another previous study showed that one-third of PubMed users is not medical experts (Lacroix & Mehnert, 2002).
  Early studies (Ely et al., 2000; Pratt & Wasserman, 2000) proposed a general classification for search topics hidden behind clinical queries that are clearly less diversified than are health-related searches performed on general web resources. In Pratt and Wasserman (2000), the authors classified clinical queries that were addressed to MEDLINE into 10 category topics, including prevention, risk factors, diagnosis, symptoms, treatments and side effects.

In this paper, clinical information search is specifically investigated, the performance of which remains questionable and subject to numerous issues (Cohen, Stavri, & Hersh, 2004; Francke, Smit, & de Veer, 2008; Natarajan et al., 2010; Suominen et al., 2013; White & Horvitz, 2015). These issues mainly arise from the following: 1) the complexity of expressing precise, context-specific clinical queries that better facilitate the identification of the relevant evidence and 2) the lack of a higher level expertise that can be used to perform evidence appraisal. Thus, we argue that an ideal clinical search engine should exploit information nuggets from both the query and the domain expertise level of the user to accurately identify clinically relevant information. Achieving this requires a deep understanding of the key differences that exist between expert-based and non-expert-based clinical information searches. To the best of our knowledge, how expert-clinical queries differ from non-expert queries is not well established in the literature; furthermore, the differences in the relevance assessment provided by either experts or non-experts and their impact on system ranking stability have not thoroughly investigated. With this in mind, we attempt to investigate the differences, commonalities, and relationships between expert-based and novice-based clinical searches. We focus on: 1) the query formulation in terms of length, domain-specificity and difficulty attributes, acknowledged as being important factors that could contribute to search success/failure (Ely et al., 2005, 2007; Tamine, Chouquet, & Palmer, 2015); 2) the relevance assessment in terms of difficulty and related reasons, relevance agreement between assessors, time spent to assess relevance and 3) the relationship between user's expertise level and retrieval effectiveness with respect to his relevance assessment. We conducted our study by assigning search tasks to experts and novices via two distinct crowdsourcing platforms allowing to recruit the two categories of clinical information seekers (experts/novices). To design reliable simulated clinical search tasks, we used the medical cases provided within major medical IR evaluation tracks namely the TREC[1] Filtering (Robertson & Hull, 2000) and the CLEF[2] e-Health (Suominen et al., 2013) with related search contexts.

The remainder of this paper is structured as follows. In Section 2 we describe research work related to the effects of domain expertise on search and relevance assessment and those related to crowdsourced user studies. To put this work in context, the findings are reported for both cross-domain expertise and specific medical domain expertise. Section 3 announces the research questions and then describes the studies that we perform in order to identify the commonalities and the differences between expert-based search and novice-based search within the medical domain, including query formulation, relevance assessment and retrieval performance. In Section 5 we report the findings of our studies based on quantitative and qualitative analysis. Section 6 discusses the results and highlights the study implications. Section 7 concludes this article.

2. Related work

2.1. On the influence of domain expertise on information search: query formulation, search behavior and search difficulty

Based on intensive research work that has been performed in information science, researchers agree that information seeking and retrieval are perceived as cognitive activities constrained by several contextual factors used for reducing the

[1] Text REtrieval Conference.
[2] Conference and Labs of the Evaluation Forum.