Characterizing Health-Related Information Needs of Domain Experts

Eya Znaidi\(^1\), Lynda Tamine\(^1\), Cecile Chouquet\(^2\), and Chiraz Latiri\(^3\)

\(^1\) IRIT, University of Toulouse
\(^2\) Institute of Mathematics, University of Toulouse
\(^3\) Computer Sciences Department, Faculty of Sciences of Tunis

Abstract. In information retrieval literature, understanding the users’ intents behind the queries is critically important to gain a better insight of how to select relevant results. While many studies investigated how users in general carry out exploratory health searches in digital environments, a few focused on how are the queries formulated, specifically by domain expert users. This study intends to fill this gap by studying 173 health expert queries issued from 3 medical information retrieval tasks within 2 different evaluation campaigns. A statistical analysis has been carried out to study both variation and correlation of health-query attributes such as length, clarity and specificity of either clinical or non clinical queries. The knowledge gained from the study has an immediate impact on the design of future health information seeking systems.

Keywords: Health Information Retrieval, Information Needs, Statistical Analysis.

1 Introduction

It is well known in information retrieval (IR) area that expressing queries that accurately reflect the information needs is a difficult task either in general domains or specialized ones and even for expert users [14,17]. Thus, the identification of the users’ intention hidden behind queries that they submit to a search engine is a challenging issue. More specifically, according to the Pew Internet and American Life Project, health-related queries are increasingly expressed by a wide range of age groups with a variety of backgrounds [23]; consumer health information through online environments support a variety of needs including the promotion of health and wellness, use of health care services, information about disease and conditions, and information about medical tests, procedures and treatment. Unfortunately, it reveals from the literature that despite of the diversity of the available health IR systems and the diversity of the used information sources, users still felt in retrieving relevant information that meet their specific mental needs [221]. To answer this issue, several studies focused on the analysis of health searchers’ behaviour, including attitudes, strategies, tasks and queries [11,16,18]. These studies involved large numbers of subjects within general web search settings, with uncontrolled experimental conditions, making it difficult to generalize their findings to expert searches involved by physicians.
Moreover, most of these studies focused on search behaviour through search strategies and tactics. Unlike previous work, we address more specifically in this paper, domain expert health search through the analysis of query attributes namely length, specificity and clarity using appropriate proposed measures built according to different sources of evidence. For this aim, we undertake an in-depth statistical analysis of queries issued from IR evaluation campaigns namely Text REtrieval Conference (TREC)\(^1\) and Conference and Labs of the Evaluation Forum (CLEF)\(^2\) devoted for different medical tasks within controlled evaluation settings. Our experimental study includes a statistical pair-wise attribute correlation analysis and a multidimensional analysis across tasks.

The remainder of this paper is structured as follows. Section 2 presents related work on health information searching. Section 3 details the query attributes and section 4 describes the tasks and query collections analysed in the study. In section 5 we present and discuss the results analysis. Finally, section 6 summarizes the study findings, highlights design implications and concludes the paper.

2 Related Work

The increasing amount of health information available from various sources such as government agencies, non-profit and for-profit organizations, internet portals etc. presents opportunities and issues to improve health care information delivery for medical professionals \([1]\), patients and general public \([10]\). One critical issue is the understanding of users’ search strategies and tactics for bridging the gap between their intention and the delivered information. To tackle this problem, several studies investigated mainly the analysis of consumer health information behaviour in one side and their query formulations in the other side. Regarding consumer’s health information behavior, several aspects have been investigated such as: (1) pattern of health information searching \([16]\): findings highlight in general that health IR obey to a trial-and-error process, or can be viewed as a serie of transitions between searching and browsing, (2) access results \([16]\): studies revealed that the majority of users access to top documents in the ranked outcome list of results, (3) goals, motivation and emotions particularly in social environments \([13]\): the authors emphasize that motivation is the main factor leading to the success or failure of health searches. More close to our work, the second category of research focused on query formulation issues by analysing query attributes such as length and topics. Several studies \([11,14,20]\) highlighted that queries are short containing less than 3 terms with an average of 2 terms. For instance authors in \([20]\) studied health related information searches on MedlinePlus and hospitals and revealed that queries lengths were in the range 1-3. The same general finding has been reported in \([11]\) regarding queries submitted to Healthlink on the basis of 377000 queries issued from search logs. \([14]\) reported quite analogous results from health web searches studies. Through other observations at the topic level \([8,19,22]\), where topics where

\(^1\) http://trec.nist.gov/
\(^2\) http://www.clef-initiative.eu/