Chapter 25
Searching Health Information in Question-Answering Systems

María-Dolores Olvera-Lobo
CSIC, Spain & Universidad de Granada, Spain

Juncal Gutiérrez-Artacho
Universidad de Granada, Spain & Universidad Pablo de Olavide, Spain

ABSTRACT

Question-Answering Systems (QA Systems) can be viewed as a new alternative to the more familiar Information Retrieval Systems. These systems try to offer detailed, understandable answers to factual questions, in order to retrieve a collection of documents related to a particular search (Jackson & Schilder, 2005). The authors carry out a study to evaluate the quality and efficiency of open- and restricted-domain QA systems as sources for physicians and users in general through one monolingual evaluation and another multilingual. Their objective led them to use definition-type questions in order to evaluate QA systems and determine if they are useful to retrieve medical information. In addition, they analyze and evaluate the results obtained, and identify the source or sources used by the systems and their procedure (Olvera-Lobo & Gutiérrez-Artacho, 2010, 2011).

INTRODUCTION

The advent of the Web and its subsequent expansion have provided the general public with access to enormous volumes of information, offering unquestionable benefits. Nevertheless, this has also brought disadvantages such as overloads of information—which in this environment is even more acute—or the fact that much of the information is incorrect, incomplete, or inaccurate, whether intentionally so or not. Consequently, it becomes indispensable to develop tools and procedures that enable the user to acquire reliable information that is relevant for a particular consultation. This is the challenge that faces Information Retrieval (IR)

Information Retrieval is a discipline focused in the problems of information items' selection from a storage system in order to facilitate retrieval

DOI: 10.4018/978-1-4666-3986-7.ch025

for the users' needs (Salton, 1970). Traditionally, IR is understood as a fully automatic process that responds to a user query by examining a collection of documents and returning a sorted document list that should be relevant to the user requirements as expressed in the query (Baeza-Yates & Ribeiro-Nieto, 1999). Simply stated, it could be said that retrieval implies finding certain requested information in a storage system or database of information (Meadow, 1993). An optimal IR system recovers *all* the relevant documents (implying an exhaustive search, i.e. a high recall) and *only* the relevant documents (implying perfect accuracy, that is to say, a high precision). This traditional model involves many implied restrictions: a) the assumption that users want full-text documents, rather than answers, and that the query will satisfied with these documents; b) that the process is direct and unidirectional rather than interactive; c) and finally, that the query and document share the same language.

Information overload is felt more strongly on the Web than elsewhere. All too often a query made with a Web search tool (search engine or meta-search engine) results in the retrieval of too many pages—many of which are useless or irrelevant to the user. Question Answering systems (QA systems) are an evolutionary improvement in IR systems. As alternative traditional IR systems, they give correct and understandable answers to factual questions (Pérez-Coutiño et al., 2004)—rather than just offering a list of documents related to the search. The benefit is that users do not have to read whole documents to find the desired information. Therefore, professionals from various areas are beginning to recognize the usefulness of these systems, for quickly and effectively finding specialized information (Crouch *et al.*, 2005).

In recent years, some of the efforts to improve IR in the Web have focused on the design and development of the so-called QA systems. The development of the QA Systems gained strong impetus in the conference on information retrieval TREC (*Text REtrieval Conference*[1])—primarily beginning with TREC-8 (Vorhees, 1999)—which since 1992 has constituted an important international forum to unite and foment research in different areas of information retrieval.

We have carried out a research to evaluate the quality and efficiency of open- and restricted-domain QA systems as sources for physicians and users in general. Our objective led us to use definition-type questions in order to evaluate QA systems and determine if there are useful to retrieve medical information. Also we analyzed and evaluated the results obtained, and identified the source or sources used by the systems and their procedure (Olvera-Lobo & Gutiérrez-Artacho, 2010; Olvera-Lobo & Gutiérrez-Artacho, 2011). So we have carried out two evaluations, one in four monolingual QA systems and the other in a multilingual QA system.

BACKGROUND

According to a study by Ely et al. (2000), medical specialists invest an average of more than two minutes searching for information related to questions that arise and, despite the time taken up, adequate answers are often not found. In this sense, several works have demonstrated the confidence of medical specialists in the use of QA systems as a method of searching and retrieving specialized information (Lee et al., 2006; Yu & Kaufman, 2007). Patients have also increasingly consulted these systems, before and after seeing the doctor, to gather information on the nature of the illness, treatment recommendations, contraindications, etc. (Zweigenbaum, 2005)

QA Systems are designed to offer understandable responses to factual questions of specialized content rapidly and precisely in such a way that the user does not have to read the complete documents to satisfy a particular query. These systems begin with the user's question in order to construct coherent answers in everyday language (Costa & Santos, 2007).

Related Content