

Conceptual Guidance in Information Retrieval

Yoon-Ho Seol, MS, Stephen B. Johnson, PhD and James J. Cimino, MD

Department of Medical Informatics,
Columbia University College of Physicians and Surgeons, New York, New York

INTRODUCTION

As more medical knowledge is available on computers, there is a need for facilities that provide efficient and effective access to those collections of digital information. In order to access digital information, users must express their information needs explicitly. However, it is often difficult for users to express their information needs clearly enough to retrieve relevant information from a computer system.

In order to help users express their information needs, many systems suggest additional terms using a thesaurus or similar component. Such systems interact with users by suggesting a set of relevant query terms in the hope that new terms might improve the recognition of users' information needs and thus lead to better retrieval performance.

However, these term-based query approaches are limited in that the users are not able to describe the relationships among terms to clarify their needs. For example, it is not possible to express whether a user is interested in drugs that treat a disease or drugs that cause a disease. To support this, the system requires a deeper conceptual framework that captures a more complete context of users' information needs.

PROPOSED APPROACH

Our approach is based on a knowledge base that contains the patterns of information needs, called generic queries. Conceptual guidance is provided with a question-oriented interaction based on the integration of multiple query contexts; namely, *application context*, *clinical context*, and *document context*. All contexts are based on a conceptual graph model and are represented with XML. A framework of the system is depicted in Figure 1; and the dotted box is the focus of our research.

The *clinical context*, information about the patient in question, is used to help shape the retrieval process of the generic query. This context is also used in post-processing of retrieved documents to provide a result that is more relevant to the patient case than a result created in the absence of such information.

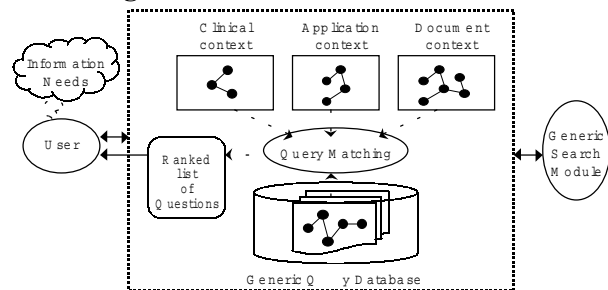


Fig. 1. Conceptual Guidance in Information Retrieval

The *application context*, the specific task the user is doing at the moment that an information need arises, is used to predict the kind of information need, thus allowing us to try to match typical needs with the specific instance.

The *document context* takes a previous search result (if any) into account to generate the next set of clinical questions. This cyclic operation allows a user to express information need and search the documents in an interactive manner.

A collection of generic queries is a key component in a question-based interaction. Previous analysis identified that clinical queries can be represented as conceptual graphs, using the UMLS knowledge base. We are exploring methods to collect generic queries from the medical literature by identifying patterns in a semantically enriched representation of documents where the relationships of terms are identified.

Pertinent clinical questions are suggested to the user based on similarity measures between the query contexts and the generic queries. Several algorithms are being explored to develop similarity metrics by using graph matching and distance heuristics.

DISCUSSION

A graph model based approach for conceptual guidance is proposed and the prototype is being developed to demonstrate the potentials of our approach. A formal evaluation is needed to validate our approach in terms of acquisition, representation, and manipulation of generic queries to help users express their information needs more quickly and completely.