CONTROLLED NATURAL LANGUAGE QUERIES TO WEB DATABASES

Tamás Mészáros, meszaros@mit.bme.hu*
Tadeusz Dobrowiecki, dobrowiecki@mit.bme.hu*
Margit Kiss, kiss.margit@webit.hu**
** Department of Measurement and Information Systems, BUTE*
*** Faculty of Humanities, ELTE*

Authors present a method to query databases and document storage systems using natural languages. Based on this method a prototype application was developed that provides a Web-based natural language query interface to a database of Hungarian nouns and their semantic arguments. The basis of the method is a controlled natural language query, the prototype system utilizes Artificial Intelligence, XML and Web technologies to implement it in the selected application field.

To provide an efficient and easy access to electronically stored information is a complex task for computer science that became everyone's problem by the dawn of the World Wide Web. The most commonly used techniques for this task were developed in library science and belong to the field of information retrieval. These methods use only a shallow understanding of the stored information and are not capable of solving the task efficiently. Due to this more and more emphasis is put on developing methods for storage and retrieval that are based on the deeper semantic understanding of the texts (W3C Semantic Web, new query languages, etc.). These methods, however, are still under development and in some areas even the proper theoretical results are missing.

At the Department of Measurement and Information Systems (Budapest University of Technology and Economics) authors analyzed the general characteristics of information retrieval tasks and systems, and developed new methods to enhance their effectiveness. They also investigated the possible application of Artificial Intelligence techniques in this field, namely natural language understanding and the application of ontologies. The goal was to incorporate these methods in information retrieval systems in a way that they increase the effectiveness of these systems, simplify their user interfaces, but don't affect their performance and usability.

The main idea of the presented system is the application of so-called controlled natural languages during the query. This method is based on natural language techniques, but it simplifies their task – and allows their successful application in the targeted systems – by putting constraints on the natural language their process. A controlled language can be more easily processed by computers, the disambiguation is simpler task this way, and the query can be easily translated into machine-readable form (into SQL in our case). On the other hand this method allows the end user to present the query using natural language making thus the application of complex and not-well-understood web forms obsolete.

In the full paper authors present the technical details of the Web system, and they also elaborate the characteristics of controlled natural language queries. The presented system provides a natural language query interface to a Hungarian noun and semantic arguments database created at the ELTE Humanities Faculty, which is also discussed in detail.