# Using Semantic Relations Extracted from Medline for Biomedical Question Answering

**Dimitar Hristovski[a], Thomas C Rindflesch[b]**

[a] *Institute of Biomedical Informatics, Faculty of Medicine Ljubljana, Slovenia*
[b] *National Library of Medicine, NIH, Bethesda, USA*

**Abstract and Objective**

*It is challenging for biomedical researchers to stay current with the literature in their field. Information retrieval systems are widely used; however, they return documents that have to be read by the user to extract relevant information. We propose a methodology for question answering implemented in a prototype tool which returns answers (known facts) first, and only later the documents from which the facts are extracted. Our question answering methodology is based on semantic relations extraction from Medline with the SemRep natural language processing system. The extracted relations are organized in a database and made available for searching through a Web-based tool. Our approach is able to provide answers for a wide array of questions that arise in clinical work and biomedical research.*

***Keywords:***

Text mining, Natural language processing, Medline, Question answering, Information extraction

## Introduction

The large size of the life sciences literature makes it difficult even for experts to absorb all the relevant knowledge in their field of interest. Sophisticated technologies are needed, and automatic text mining techniques are increasingly used to help assimilate online textual resources. The most widely used are information retrieval systems such as PubMed, which searches the Medline biomedical bibliographic databases. These systems are very efficient and robust. However, in response to a user's query they do not provide answers (or facts), but a set of documents (citations) that the user has to read in order to extract the required answers. For example, if a user wants an overview of the Parkinson's disease literature, a PubMed search will return tens of thousands of documents. If the user is interested in treatments for a disease, with some skill it is possible to specify an effective query, but the result will be a set of documents that have to be read.

Question answering systems, on the other hand, aim at providing answers (known facts). For the above example about the treatments of a disease, a question answering system would provide as answers particular drugs that are used to treat that disease.

## Methods

We extract semantic relations from a large Medline subset with the SemRep natural language processing system. SemRep is able to extract the following categories of semantic relations: genetic etiology (ASSOCIATED_WITH, PREDISPOSES, CAUSES), substance relations (INTERACTS_WITH, INHIBITS, STIMULATES, pharmacological effects (AFFECTS, DISRUPTS, AUGMENTS), clinical actions (ADMINISTERED_TO, TREATS, …), organism characteristics (LOCATION_OF, PART_OF, PROCESS_OF) and co-existence (CO-EXISTS_WITH). The extracted semantic relations have the form of Subject-Relation-Object (e.g. Levodopa-TREATS-Parkinson Disease or alpha-Synuclein-ASSOCIATED_WITH-Parkinson Disease). We store the relations in a database management system and provide a question answering tool that is based on searching the extracted relations. The question can be a Boolean query referring to any combination of subject, relation, and object. For example, the query "relation:TREATS AND obj_name:parkinson" corresponds to the question "What are the treatments for Parkinson's disease?" and returns as answers relations such as Deep brain stimulation-TREATS-Parkinson disease and Dopamine Agonists-TREATS-Parkinson disease. On request, the sentences from which the relations are extracted can be shown, as well as the Medline citations in which such sentences occur.

## Results

We used SemRep to process 43,369,616 sentences from 6,699,763 MEDLINE citations published between 1999 and the end of March 2009. 21,089,124 semantic relations were extracted.

## Conclusion

We propose a methodology and describe a tool for biomedical question answering which is able to provide answers to a wide array of questions and is a useful complement to existing information retrieval systems.