

# Applying Biomedical Ontologies on Semantic Query Expansion

Andre Bechara, Maria Luiza M. Campos, Vanessa Braganholo  
Informatics Graduate Program (PPGI), Federal University of Rio de Janeiro, Brazil

## Abstract

*This poster presents an ongoing work on using biomedical ontologies to improve efficiency on information retrieval.*

## Introduction

The interpretation of a question (or information need) depends, among other things, of a series of lexical-semantic relations that complement and help the cognitive process of answering that information need. Despite this fact, currently used information retrieval mechanisms take few advantages of the semantic interpretation of users' information needs (usually specified through keywords). In most of the cases, those mechanisms are based on keyword matching, and thus are excessively dependant on the query and document terms.

There are several past results<sup>1,2</sup> showing that, in general, information retrieval based on domain knowledge decreases the accuracy of keyword based search engines. We believe this approach deserves further discussion and experimentation, looking for more strong evidences that these negative results can really be generalized. Moreover, there are some questions left unanswered by previous work that our experiment is addressing:

(i) Using a scientific ontology, with formal construction and maintenance processes, such as the OBO ontologies, would produce better results? (ii) Are there more efficient query expansion techniques using available domain knowledge? (iii) Is a scientific ontology complete enough to fulfill the information retrieval researchers' needs, in general?

## Semantic Query Expansion

To try to answer some of these questions, we run a query expansion experiment using the Gene Ontology (GO) as domain knowledge. As the document repository, we used an extraction of 10 years of Pub Med publications (from 1994 to 2004), which contains approximately 4.6 millions of documents. This dataset is a test-collection used by the information retrieval community, called Genomic TREC.

## Results

To evaluate our ontology-based semantic query expansion technique, we measured the effectiveness of the information retrieval mechanism with and without expansion. In a nutshell, the average result showed an increase of 28% on synonyms relations and a small decrease on other relations.

Our results show a lot of consistence with past related work. In fact, if the expansion strategy does not selectively choose when and how to expand, only synonym relations are worth to be used. However, looking further, it is possible to find several opportunities to try other expansion strategies. For example, the problem with query expansion using generalization/specialization relationships is that, if it is always applied, the bad results are more frequent than the good ones. But, if the strategy is to be selective on when to use these relations for expansion, the increasing on accuracy can be outstanding. As shown by our experiment, there was a query with 98% increment on effectiveness.

## Conclusion

We strongly believe that it is premature to assume that semantics-based query expansion is, in general, a recall-enhancing, precision-degrading technique. Our experiments suggest that by using scientific based ontologies (like OBO ontologies) with formal relations, it is possible to increase both recall and precision. Our group is currently revising this first experiment towards a better semantic query expansion strategy.

## Acknowledgements

This work was partially funded by CAPES and CNPq research grants 311454/2006-2, 306889/2007-2 and 484713/2007-8.

## References

1. Fox E. Lexical relations enhancing effectiveness of information retrieval systems. SIGIR Forum, New York, v.15, n.3, p.5-3.
2. Voorhees E. Query expansion using lexical-semantic relations. In: ACM SIGIR conference on research and development in information retrieval, Proceedings, Dublin:17, p.61-69, 1994