

Title: Using Concepts and Entailment for Passage Retrieval from Biomedical Literature

Abstract:

Scientists in healthcare and biomedical informatics have never had as much information available in electronic form as they do today. The increased variety of approaches for information retrieval and extraction offer the potential to combine different techniques and provide scientists with new ways of accessing information; but the real contribution to e-science will occur when the next information tools are consistent with the workflows used by scientists in a specific discipline. One area that holds much promise is recent work that focuses on retrieving relevant passages and entities from an article, rather than an entire document.

In this presentation, we will explore the degree to which a concept representation and methods that recognize textual entailment will aid in passage retrieval performance. Our approach combines concepts from the Unified Medical Language System (UMLS) with a syntax representation that has shown success in recognizing textual entailment. We use the Genomics TREC collection of 160,000 documents and 50 topics that biologists considered important. The standard measures of precision (the proportion of accurately retrieved passages divided by the number of passages retrieved), and recall (the percentage of accurately retrieved passages divided by the total number of relevant passages) and mean average precision (MAP) (the percentage of correctly retrieved passages at each level of recall) are used to evaluate results. One of the motivations behind this work is the historically low precision values measured in passage retrieval, and the need to investigate alternative approaches that increase relative precision. In addition, the scope of the UMLS enables us to explore the impact of different vocabularies on performance and may inform both manual and automated methods of ontology construction.

Citation:

Blake,C. and Nassar, N.(2007) Using Concepts and Entailment for Passage Retrieval from Biomedical Literature, Abstract in 2007 Microsoft eScience Workshop at RENCI, Chapel Hill, NC.