

## **Bridging Search Engines and Natural Language Applications: Early Results in Semantic Predicate Retrieval**

**Paul Ogilvie  
Carnegie-Mellon University  
USA**

Abstract:

Research in Information Retrieval has historically focused on supporting human users and their information needs. However, there has been growing interest in the use of search engines to support natural language applications. For example, question answering systems use search engines on a text corpus to retrieve passages that may help the system answer a natural language question. The question answering system then processes the passages in search of linguistic structures or other syntactic and semantic features that help the question answering system synthesize answers from the passages. Specific relationships of the keywords (or synonyms resulting from question analysis) with the linguistic structure in the passages may be crucial for the passage to be useful to the question answering system. Yet few search engines can represent this information in their indexes, query languages, or retrieval models.

We present a discussion of how linguistic annotations may be used within the Indri search engine to better support question answering systems. This talk also considers difficulties in working with annotations resulting from natural language processors, as these processing applications may make errors or return annotations that overlap with annotations resulting from other applications. We then motivate the task of semantic predicate retrieval for the support of question answering systems and present some results using synthetic queries. Queries using the structure and keywords of the desired result performs better in both precision and recall of sentences containing the desired semantic predicate than keyword queries alone. These experiments suggest that if the structure of the result and the relationship between the structure and the keywords is important for the notion of relevance, it is very important for the retrieval system to model this structure and allow query capabilities.

The Indri search engine is a part of the Lemur toolkit which is available at <http://www.lemurproject.org> .