Project Status

Group Members:

Immanuel Espiritu

Jayvee Febrer

John Amata

Due to difficulties in finding a project that encapsulates a research based software development project at a specific computer science discipline, we are now changing our topic. We have chosen the topic Question and Answering System that is aligned to Natural Language Processing discipline.   
  
These research aims to develop a Question and Answering system based on existing methodologies and finding a way to bridge into a working system for companies in aiding their human-computer interaction. Also to provide our own concept with the help of pre-existing methodologies.

For now, we are in the data gathering process. Here are some related literature we are currently studying in relation to our chosen topic.

**Related Literature**

**A Boolean Based Question Answering System:**

In recent years, the continuous research and development of the Internet boosts the need of network information resources. The network already became the information and data exchange platform. It is not easy for users to find useful information in such a large database. Thus, search engine was designed as the important tool for people to find the information they want. Using search engines to find information is the common method for most of the people in the world. The search engine works based on the key words provided by users and it returns millions of related webpages back to users (Gulli, A., and Signorini, A [2004]). However, among those webpages, some of the information is not useful, so users will waste their time reading lots of unrelated information and they need enough time to find out the direct information. There are many popular search engines being used today such as Google, Yahoo and Bing.

[**http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=6279&context=etd**](http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=6279&context=etd)

**START Question answering System (MIT)**

In 1993, the Question Answering System named START has been created in the Artificial Intelligence laboratory in MIT. This system is the first web-based Question Answering System. START can provide the accurate information to users and it can answer millions of English questions that include geographical questions, film related questions, celebrities’ questions and some definition questions. This system can be seemed as a combination system because it still remains two knowledge libraries: “START KB” and “Internet Public Library”. Thus, the system can use the data from these two libraries to find the candidate answers when users ask related questions. Otherwise, START will firstly analyze the questions, then find the candidate answer through the search engine and return the direct answers back to users. For example, when user asks: “Who was Bill Gates?”, the system will return back: “Cofounder, Microsoft. Born William H. Gates on October 28, 1955, Seattle, Washington.” The system will also return the link which is the original source contains the answer. If users want to find more information about the questions, they can simply click the links and view the information on the webpage.

<http://start.csail.mit.edu/index.php>

**WAND ALGORITHM**

Web search services process thousands of queries per second, and filter their answers from collections containing very large amounts of data. Fast response to queries is a critical service expectation. The well-known WAND processing strategy is one way of reducing the amount of computation necessary when executing such a query. The value of WAND has now been validated in a wide range of studies, and has become one of the key baselines against which all new top-k processing algorithms are benchmarked. However, most previous implementations of WAND-based retrieval approaches have been in the context of the BM25 Okapi similarity scoring regime. Here we measure the performance of WAND in the context of the alternative Language Model similarity score computation, and find that the dramatic efficiency gains reported in previous studies are no longer achievable. That is, when the primary goal of a retrieval system is to maximize effectiveness, WAND is relatively unhelpful in terms of attaining the secondary objective of maximizing query throughput rates. However, the BM-WAND algorithm does in fact help reducing the percentage of postings to be scored, but with additional computational overhead. We explore a variety of tradeoffs between scoring metric and processing regime and present new insight into how score-safe algorithms interact with rank scoring.

<http://www.culpepper.io/publications/pcm13-adcs.pdf>

**MULDER**

University of Washington created a Question Answering System named MULDER. It 10 is the first automatic Question Answering System worked base on Internet. This system does not apply any knowledge library and it searches the answers only used the information from Internet. For each question, MULDER will send back a list of candidate answers instead of only one answer. The system will use statistic method to add weights for each answer and the weight is called “confidence level”. For example, when user asks: “Who was the first American in space?”, MULDER will return a list of answers. Among those answers, the answer “Alan Shepard” has 70% confidence level and “John Glenn” has 15% confidence level. At the same time, the system will show the summary of content for those answers and the links as well.

**The future works in Question Answering Systems research area**

The research of Question Answering System has made lots of achievements by the involvement of lots of researchers. However, there still exists some problem that need the researchers to resolve. Most of the Question Answering System is a small 12 application system and the range of solutions provides by these system is limited. Some of the systems even need manual operation. From the current research results, the Question Answering System can be improved in the following aspects. Firstly, the improvement in the question processing step. This step should involve more techniques in Natural Language Processing to solve the syntax and semantic problems have generated. Secondly, the answer extraction technique have to be completed which aim to extract the important information and provide users with accuracy answers. Thirdly, the typing method for the question could be different. The question could be inputted with voice or graph.

[**http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=6279&context=etd**](http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=6279&context=etd)