Automated Question Answering Over the Web: An Adaptive Search and Retrieval Strategy

ABSTRACT

The problem of efficiently finding answers to natural language questions over the web has gained much attention. Currently, useful experimental models for implementing question answering work well only for smaller, specific collections of documents and/or they only handle short, single factoid-type questions. Other more generally focused models retrieve and re-rank only a set of documents most likely to contain an answer. These approaches rely on only a few specific strategies to implement question answering. A more comprehensive and dynamic model of question answering may provide better performance for both retrieving candidate answer pools and extracting specific answers.

Such a new model will be designed that efficiently combines automatic question reformulation, search strategy selection, query expansion, and answer extraction/pooling techniques. The system will automatically learn question transformation, adapting queries for the most popular web search engines, and training itself on collections of question answer pairs, such as FAQs. Questions will be matched against automatically learned question types and reformulated into queries based on answer phrases likely to appear within a document containing the answer. The semantic answer type will also be determined based on the question type and used to recognize potential answers. During system training, the top ranked documents retrieved by the search engine will be examined for an appropriate answer. User answer-acceptance feedback will be collected to re-rank document entries and/or refine new queries as necessary during live runs.

INTRODUCTION

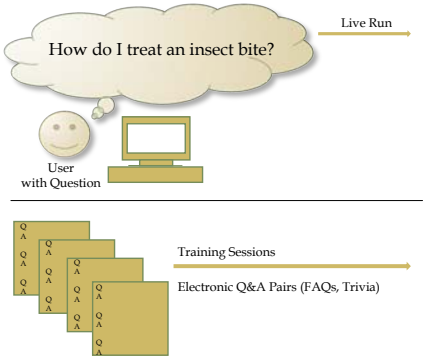
Current search sources on the web targeting natural language queries (NLQ) such as AskJeeves.com (now Ask.com) use databases of pre-compiled answers, meta-searching, and other proprietary methods (Agichtein, Lawrence and Gravano 2004). Historically, other search sources facilitate interaction with human experts such as AskMe.com and Google Answers. Also, web search engines such as Google and Yahoo typically treat NLQs as a list of terms and retrieve documents similar to those terms. Unfortunately, NLQs contain a connection between terms that is lost when treated as either a simple list or a weighted list of terms (Croft, Turtle and Lewis 1991). Interestingly, each web search engine returns different documents with different ranking, even by using the same query (Agichtein, Lawrence and Gravano 2004). However, when considering question answering, documents with the best answers may only contain a few terms from the original query, and therefore, be ranked low by the search engine.

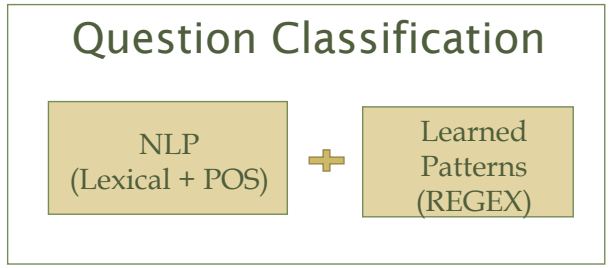
STATE OF THE ART

Google, the highest ranking web search engine, has incorporated the use of online dictionaries and encyclopedias which help with factoid-type questions. It can also use simple part of speech (POS) changes to important terms. However, the results begin to contain less acceptable, or precise, answers the more complex or lengthy the questions become. In addition, FAQFinder is an important system of note. It uses a vector-space information retrieval (IR) engine to retrieve a list of relevant FAQ files (a small, specific domain of documents), attempting to locate an equivalent question. FAQFinder also uses WordNet for query expansion. Research has shown that more precision may be attained over the web by treating queries more like questions (Agichtein, Lawrence and Gravano 2004).

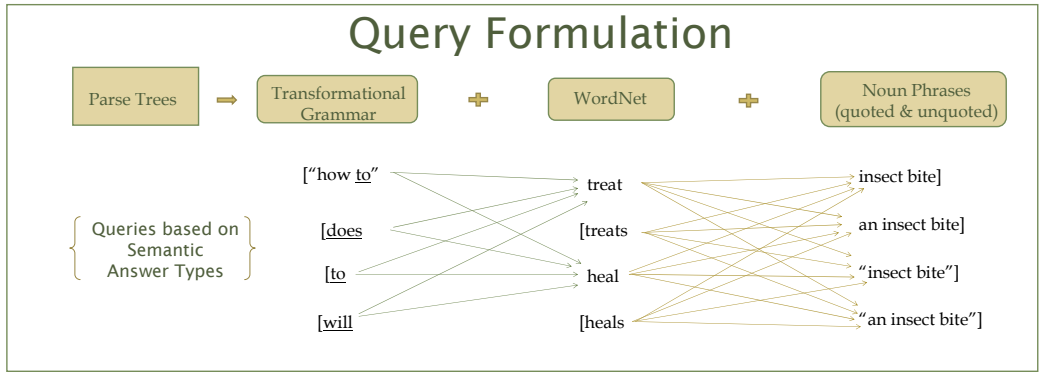
One question answering system, Mulder (Kwok, Etzioni and Weld 2001), answers questions such as "Who was the first American in space?" and runs on top of a web search engine (a large, general domain of documents). Mulder uses natural language processing (NLP) to classify both question types and answer types. After the initial NLP, Mulder formulates several queries. General queries include using the most important keywords from the original question combined with lexical and semantic processing, and the most specific include derived noun phrases and quoted phrases of varying word length.

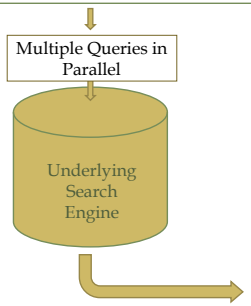
Another question answering system, Tritus (Agichtein, Lawrence and Gravano 2000), uses a similar technique as Mulder's quoted phrases. It automatically learns transformation rules based on finding phrase patterns from a training collection of question answer pairs (FAQs, Trivia Q&A). Tritus builds its queries based on these patterns and makes adjustments based on rule performance. Documents returned are judged for the likelihood of containing an answer via an IDF-based measure. However, Mulder completes the process by extracting answers using a word distance and clustering approach matched against the earlier predicted answer type.

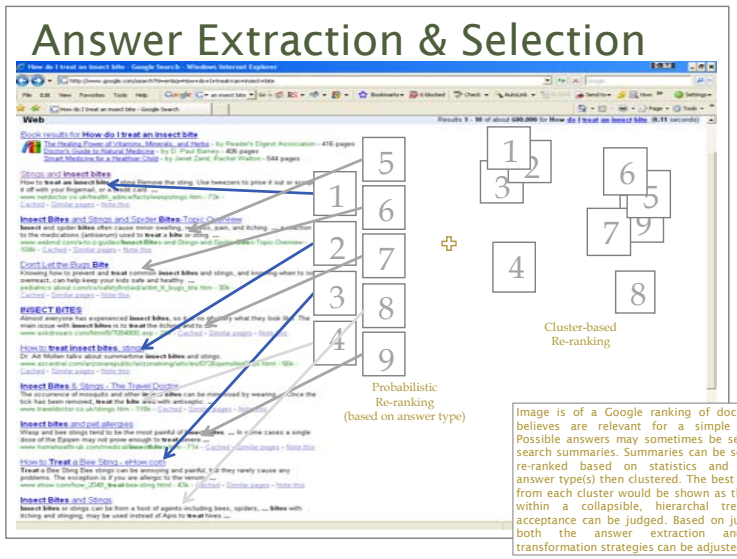




Parse Trees with Links







PROBLEMS

- Mulder's method of query transformation is semi-automated as it depends on human predetermined rules for transforming the question and works for factoid-type questions.
- Tritus attempts only to return documents that are likely to contain the answer, either for the user's review or as input for an actual answer extraction system. In addition, Tritus does not incorporate NLP when transforming its questions.
- A hybrid question answering system should be created and tested to determine whether known working ideas and certain opposing views can be integrated in a novel, performance improving way.

RESEARCH QUESTIONS

1. Can NLP of questions combined with pattern recognition from training collections be used to improve the ability to learn question transformations for web search engines?
2. Does the routing of particular learned question transformations to specific web search engines enhance efficiency?
3. Subsequently, can the question answering system adapt to ineffective searches by choosing question transformations that provide varying degrees of precision/recall?
4. How does combining NLP and REGEX pattern recognition from training collections increase the precision/recall of definite answer results?
5. Can user answer-acceptance feedback/verification be used to improve the simple answer recognition algorithms currently in use? Clustering and probabilistic approaches yield different types of documents (Croft and Thompson, 1984).
6. Subsequently, can this feedback be used to improve the learning of question classification and their respective answer types?

PRINCIPAL ELEMENTS

Major Model Components

- The underlying web search engines to be used are: Google, MSN (powered by Windows Live Search), Ask.com, and Yahoo.
- Hybrid question answering system: NLP and Pattern Recognition learning subsystem for transforming questions, UI for user answer-acceptance feedback/verification to adjust both question transformation and answer extraction
- FAQs and other question answer pairs (Internet Public Library) for training the hybrid system
- Pre-screened, consistent TREC document collections to control evaluation of question transformation performance

Participants

- Beginner to advanced information retrieval users for answer-acceptance feedback
- Training judges for answer verification

Quinsulon L. Israel
1st Year Ph.D. Student
qi23@drexel.edu
The iSchool at Drexel
Energizing the Infosphere