A Hybrid Filtering Approach for Question Answering

Sisay Fissaha Adafre[1] and Josef van Genabith[2]

[1] National Center for Language Technology, School of Computing, DCU,
sadafre,josef@computing.dcu.ie,
[2] IBM CAS Dublin

Abstract

We describe a question answering system that took part in the bilingual CLEFQA task (German-English) where German is the source language and English the target language. We used the BableFish online translation system to translate the German questions into English. The system is targeted at *Factoid* and *Definition* questions. Our focus in designing the current system is on testing our online methods which are based on information extraction and linguistic filtering methods. Our system does not make use of precompiled tables or Gazetteers but uses Web snippets to rerank candidate answers extracted from the document collections. WordNet is also used as a lexical resource in the system.

Our question answering system consists of the following core components: Question Analysis, Passage Retrieval, Sentence Analysis and Answer Selection. These components employ various Natural Language Processing (NLP) and Machine Learning (ML) tools, a set of heuristics and different lexical resources. Seamless integration of the various components is one of the major challenges of QA system development. In order to facilitate our development process, we used the Unstructured Information Management Architecture (UIMA) as our underlying framework.

Question analysis consists of a question classification and a query generation component. We employed both machine-learning and rule-based methods for defining question classes. For the machine learning approach, we trained a classifier using the data set provided by Li and Roth [4].

Passage Indexing/Retrieval splits the CLEF documents into passages. The sentences in the passages are POS tagged and chunked, and indexed. **Sentence Analysis** involves running a named entity recogniser on the sentences returned by the retrieval module. The named entities constitute the candidate answers. The sentences are also parsed using a Lexical Functional Grammar(LFG)-based parser developed at Dublin City University [1, 3].

Answer Reranking reranks candidate answers based on different sources of evidence, such as syntactic similarity of the sentence with the question, proximity of query terms to the candidate answers, similarity of the semantic type of the candidate answer to the answer type, and centrality of the sentence with respect to a corpus of web snippets retrieved using the query terms extracted from the question.

Syntactic Similarity is measured in terms of the number of shared dependency relations between the sentence and the question. For this, we parsed both the questions and the sentences using a Lexical Functional Grammar(LFG)-based parser mentioned earlier. The system takes the output of a syntactic parser (Charniak parser) and generates an F-Structure, a labeled bilexical dependency graph [2]. The output can also be provided in the form of a set of dependency triples. We count how many dependency pairs are shared between the questions and answers, normalise the resulting value and add it to the overall score of the named entities extracted from the sentence.

Lexical/Corpus based evidences: Term Proximity is a measure based on counts query terms that appear in the vicinity of the candidate answer. **Type Filtering** uses *Wikipedia Category* and *WordNet Hierarchy* to compute semantic similarity between the expected answer type and the candidate answer. **Web based evidence** is used to find answers, and to rerank candidate answers from the *CLEF Collection* and *Wikipedia Collection*. The overall scores for reranking candidate sentences are computed as a sum of the *Retrieval scores*, *Syntactic similarity*, *Term Proximity*,*Type Filtering* and *Web-based evidence*.

Definition Questions: The system takes the *topic* generated by the question analysis module, and submits it as a query to the retrieval module. The returned passages and Wikipedia articles are split into sentences. Sentences that do not contain the topic are removed from the list. We assign more weight to sentences with copula verb constructions with the *topic* as a subject, e.g. TOPIC is .... Finally, the sentences are reranked using evidence obtained from the web.

Experimental Evaluation of QA@CLEF 2008 The system returned 16 *exact* answers (8 *Factoids* and 8 Definitions), and 25 correct answers counting *unsupported* answers. The web reranking component contributed significantly. Error analysis shows different sources of errors: *Translation*, *Questions classification*, *Named entity recognition*. Post CLEF evaluation showed that minor adjustment to the question classifier and correcting some technical errors (coding errors) improved the result significantly: the system returned 19 *exact* answers for factoid questions.

Further Work: The current system is limited in scope and mainly employs a combination of shallow linguistic analysis methods and machine learning techniques. Our future plan is to extend the scope of the system to include more questions types, and improve the methods that are already implemented. Specifically, we would like to extend the application of the LFG-based parser output to other components. Initial application of the parser output to the ML based Question classifier showed significant improvement. We are also planning to extend dependency triple based scoring method to include the full LFG-based parse output. Furthermore, we are working to bring in logic-based reasoning methods in the system. The approach drives logic-based representations of questions and candidate sentences based on the output of the LFG Parser. This will allow us to make inferences which form the basis for finding implicit relations between questions and answers. Coreference resolution is an important component of a QA system that is not well developed in our system. We will explore the application of the LFG parser for problem of coreference resolution.

References

1. Cahill, A., Burke, Odonovan, R., Riezler, S., Genabith, J., Way, A.: Wide-Coverage Statistical Parsing Using Automatic Dependency Structure Annotation. *Computational Linguistics.* 34(1), 81–124 (2008)
2. Eugene Charniak. A maximum-entropy-inspired parser. In *Proceedings of the first conference on North American chapter of the Association for Computational Linguistics*, pages 132–139, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
3. Ron Kaplan and Joan Bresnan. Lexical functional grammar, a formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations.* MIT Press, Cambridge, MA, pages 173281. 1982.
4. X. Li and D. Roth. Learning question classifiers, 2002. In *Proc. COLING 2002*