

In Search of the Why

Developing a system for answering *why*-questions

Suzan Verberne
Radboud University Nijmegen
s.verberne@let.ru.nl

Abstract

The problem of automatically answering natural language questions by pinpointing exact answers in a large text (web) corpus has been studied since the mid 1990s. Most research has been directed at answering factoid questions: questions that expect a short, clearly identifiable answer; usually a named entity such as a person name, location or year. In this dissertation, we have focused on the problem of answering *why*-questions (*why*-QA). *Why*-questions require a different approach than factoid questions because their answers tend to be longer and more complex. Our main research question was: “What are the possibilities and limitations of an approach to *why*-QA that uses linguistic information in addition to text retrieval techniques?”

We first experimented with a simple bag-of-words approach on a set of open-domain *why*-questions and Wikipedia as answer corpus. With Lemur as retrieval engine and TF-IDF as ranking model, we were able to retrieve a correct answer passage in the top-10 for 45% of the questions. The most important limitation of the bag-of-words approach for *why*-QA is that the structure of the questions and the candidate answers is not taken into account. We studied a number of levels of linguistic information on both the side of the question and the side of the answer passage in order to find out which type of information is the most important for answering *why*-questions.

We implemented a re-ranking module that incorporates knowledge about the syntactic structure of *why*-questions and the document context of the answers. With this module, we were able to improve significantly over the already quite reasonable bag-of-words baseline. After we optimized the feature combination in a learning-to-rank set-up, our system reached an MRR of 0.35 with a success@10 score of 57%. These scores were reached with only eight overlap features, one of which was the baseline ranker TF-IDF and the others were based on linguistic information (e.g. question focus, cue words and WordNet Similarity) and document structure (e.g. document title and the position of the answer passage in the document).

For solving the remaining 43% of the questions, we found that more is needed than classic NLP. Our conclusion is that *why*-QA deserves renewed attention from the field of artificial intelligence.

The dissertation is available online at <http://lands.let.ru.nl/~sverbern/>