Open-Domain Question Answering

Mark Andrew Greenwood

Question answering aims to develop techniques that can go beyond the retrieval of relevant documents in order to return exact answers to natural language questions, such as *"How tall is the Eiffel Tower?"*, *"Which cities have a subway system?"*, and *"Who is Alberto Tomba?"*. Answering natural language questions requires more complex processing of text than employed by current information retrieval systems. A number of question answering systems have been developed which are capable of carrying out the processing required to achieve high levels of accuracy. However, little work has been reported on techniques for quickly finding exact answers.

This thesis investigates a number of novel techniques for performing open-domain question answering. Investigated techniques include: manual and automatically constructed question analysers, document retrieval specifically for question answering, semantic type answer extraction, answer extraction via automatically acquired surface matching text patterns, principled target processing combined with document retrieval for definition questions, and various approaches to sentence simplification which aid in the generation of concise definitions.

The novel techniques in this thesis are combined to create two end-to-end question answering systems which allow answers to be found quickly. AnswerFinder answers factoid questions such as *"When was Mozart born?"*, whilst Varro builds definitions for terms such as *"aspirin"*, *"Aaron Copland"*, and *"golden parachute"*. Both systems allow users to find answers to their questions using web documents retrieved by Google™. Together these two systems demonstrate that the techniques developed in this thesis can be successfully used to provide quick effective open-domain question answering.