AnswerFinder: Question Answering Technology

Diego Mollá

Centre for Language Technology
Department of Computer Science
Macquarie University

NGS Workshop
21-22 September 2006

Question Answering

What is Question Answering?

- Find the answer to an arbitrary question by searching text documents
- Finding answers *vs.* finding documents
- More NLP intensive than document retrieval

What Types of Questions does AnswerFinder handle?

- Fact-based questions
- Lists
- [Descriptions]
- [Summary-based]
- [Cross-lingual QA]

Question Answering

What is Question Answering?

- Find the answer to an arbitrary question by searching text documents
- Finding answers *vs.* finding documents
- More NLP intensive than document retrieval

What Types of Questions does AnswerFinder handle?

- Fact-based questions
- Lists
- [Descriptions]
- [Summary-based]
- [Cross-lingual QA]

Question Answering

What is Question Answering?

- Find the answer to an arbitrary question by searching text documents
- Finding answers *vs.* finding documents
- More NLP intensive than document retrieval

What Types of Questions does AnswerFinder handle?

- Fact-based questions
- Lists
- [Descriptions]
- [Summary-based]
- [Cross-lingual QA]

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Find Documents that Contain the Answer

- Recall more important than precision
- The answer is not in the question

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Identify Answer-looking Strings

- Entity types depend on the domain of the QA system
- Recall more important than precision
- Multi-label NER

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Analysing the Question

- Rule-based
- Machine learning

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Dependency Structures

- Syntactic similarity between question and answer
- Third-party parser (Connexor FDG)

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Logical Forms as Graphs

- Graph algorithms to find similarities between the question and the answer

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Learning the Location of the Answer

- Learn the connection between the question and the answer
- ML enables portability
  - Domains (restricted domain QA)
  - Language (planned cross-lingual QA system for next CLEF)

AnswerFinder Technology

What Technology does AnswerFinder Use?

- Document Retrieval
- Information Extraction
  - Named Entity Recognition
- Question Classification
- Sentence Representation
  - Syntactic Information
  - Graph-based Logical Forms
- Answer Extraction
  - Machine Learning
- Summarisation
  - Answer Composition

Finding a Complex Answer

- Long answers that require composing information from several documents
- Combining QA with summarisation
- Baseline system submitted to DUC

Thoughts about Next Generation Search

The Future of NGS

- Attempt to find the user needs
- Answer-oriented search
- Portability across domains
- Use of domain-specific information
  - Terminology
  - Ontologies