

# AnswerFinder: Question Answering and Related Tasks

Diego Mollá

Centre for Language Technology  
Department of Computer Science  
Macquarie University

NGS Workshop  
19-20 July 2007

# Programme

1 AnswerFinder

2 Technology Developed

# Who Are We? - AnswerFinder

- <http://www.ics.mq.edu.au/~diego/answerfinder/>
- Members:
  - 1 **Diego Mollá** (Project Director, Chief Investigator)
    - Design, Sentence Representation, QA Rules
  - 2 **Robert Dale** (Chief Investigator)
    - Generation, Summarisation
  - 3 **Menno van Zaanen** (Research Associate)
    - Implementation, Machine Learning
  - 4 **Luiz Augusto Pizzato** (PhD student)
    - Document Retrieval, Web
- Past members:
  - Ben Hutchinson
  - Mary Gardiner
  - Elena Akhmatova
  - Stephen Wan
  - Daniel Smith

# 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 are We Addressing?

- Fact-based questions
- Lists
- Cross-lingual
- Speech transcripts
- Query-driven summaries

# Question Answering

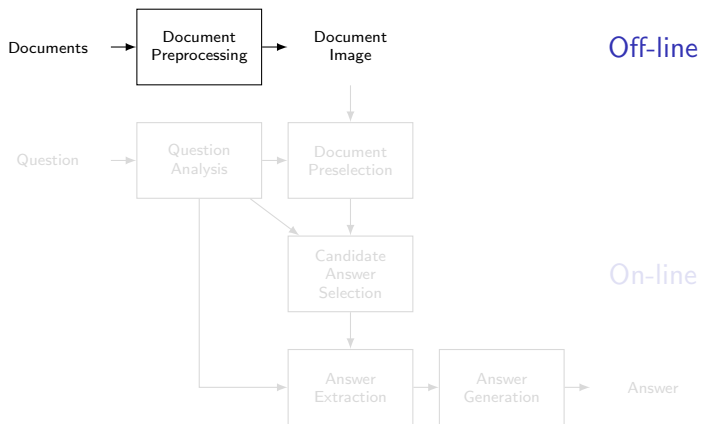
## What is Question Answering?

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

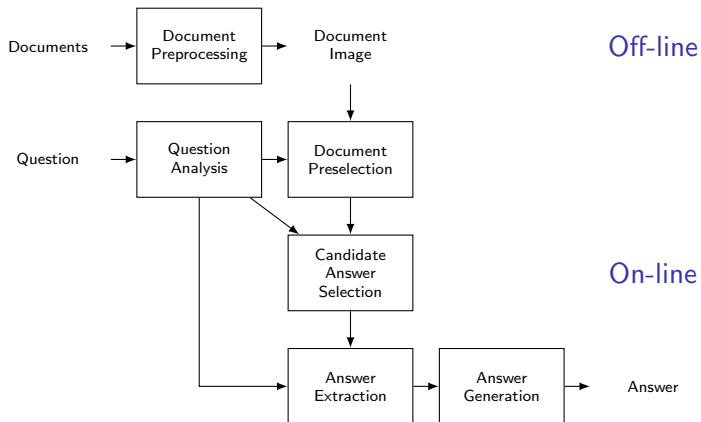
## What Types of Questions are We Addressing?

- Fact-based questions
- Lists
- Cross-lingual
- Speech transcripts
- Query-driven summaries

# Architecture



# Architecture



# Programme

1 AnswerFinder

2 Technology Developed

# Retrieval for QA

## Issues

- 1 Need to handle possibly large collections of documents
- 2 If the correct document is not selected, all other tasks are useless
- 3 Current IR methods aim at finding relevant documents but not documents that contain the answer
- 4 Even if a document is found, the answer must be easy to locate

## What we are Doing Here (Luiz)

- 1 Experimenting with using syntactic information and semantic role labeling

# Retrieval for QA

## Issues

- 1 Need to handle possibly large collections of documents
- 2 If the correct document is not selected, all other tasks are useless
- 3 Current IR methods aim at finding relevant documents but not documents that contain the answer
- 4 Even if a document is found, the answer must be easy to locate

## What we are Doing Here (Luiz)

- 1 Experimenting with using syntactic information and semantic role labeling

# Question Analysis

## Issues

- 1 Questions must be processed on-line
- 2 Questions are short
- 3 Need to identify the expected answer type

## Current Method

- Use surface patterns of regular expressions

## Eventually

- Learning by structure induction (Menno)

# Question Analysis

## Issues

- 1 Questions must be processed on-line
- 2 Questions are short
- 3 Need to identify the expected answer type

## Current Method

- Use surface patterns of regular expressions

## Eventually

- Learning by structure induction (Menno)

# Candidate Answer Selection

## Goal

Find all sentences (or text fragments) that contain the answer

## Our Current Method

- 1 Ignore all sentences that do not contain a string of the expected answer type
  - Use a named entity recogniser
- 2 Score remaining sentences using a combination of lexical, syntactic, and semantic features

## Our Sentence Scoring Method

Lexical Information Word overlap

Syntactical Information Grammatical relation overlap

Semantic Information Logical forms, logical graphs

# Candidate Answer Selection

## Goal

Find all sentences (or text fragments) that contain the answer

## Our Current Method

- 1 Ignore all sentences that do not contain a string of the expected answer type
  - Use a named entity recogniser
- 2 Score remaining sentences using a combination of lexical, syntactic, and semantic features

## Our Sentence Scoring Method

Lexical Information Word overlap

Syntactical Information Grammatical relation overlap

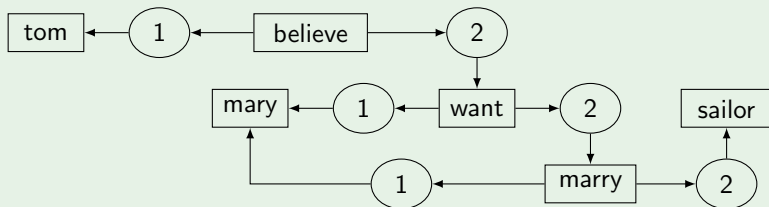
Semantic Information Logical forms, logical graphs

# A Logical Graph (Diego)

## Features

- Format inspired on Sowa's Conceptual Graphs
- Directed, bipartite graphs
  - Concepts
  - Relations
- Built from the output of Connexor FDG parser

*Tom believes that Mary wants to marry a sailor*



# Named Entity Recognition

## Issues

- 1 Current NER technology aims at trade-off between recall and precision
  - But for QA we prefer high recall
- 2 Usually best NE tag is given
  - But second best could be the right one

## Our Method (Diego, Menno, Daniel)

- 1 Use regular expressions and lists for the “easy” cases
- 2 Use ML to learn the “difficult” cases
- 3 Allow multiple tags

# Named Entity Recognition

## Issues

- 1 Current NER technology aims at trade-off between recall and precision
  - But for QA we prefer high recall
- 2 Usually best NE tag is given
  - But second best could be the right one

## Our Method (Diego, Menno, Daniel)

- 1 Use regular expressions and lists for the “easy” cases
- 2 Use ML to learn the “difficult” cases
- 3 Allow multiple tags

# Answer Extraction

## Issues

- 1 AE is by far the most difficult task
- 2 At this point we can use computationally-intensive NLP technology

## Our Method (Diego)

- 1 Combine information from NER module and graph rules
- 2 Graph rules are learned from graph representations

# Answer Extraction

## Issues

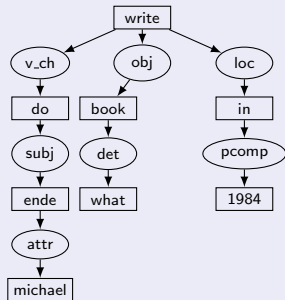
- 1 AE is by far the most difficult task
- 2 At this point we can use computationally-intensive NLP technology

## Our Method (Diego)

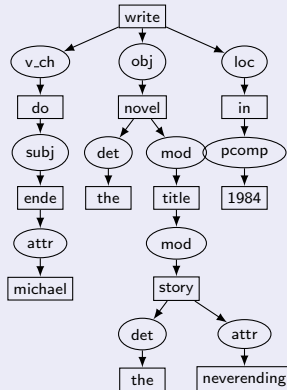
- 1 Combine information from NER module and graph rules
- 2 Graph rules are learned from graph representations

# Example of Answer Selection with Graph Rule

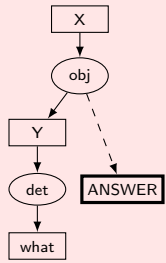
Q: What book did Michael Ende write in 1984?



A: In 1984 Michael Ende wrote the novel titled "The Neverending Story"

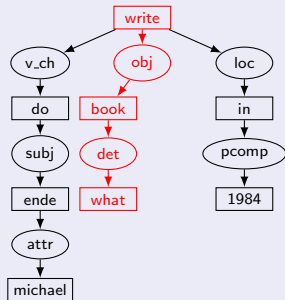


The Rule

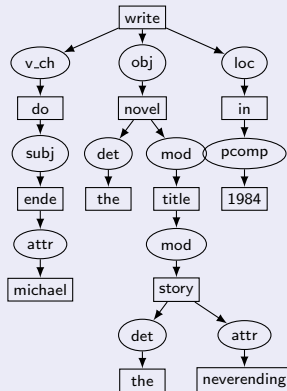


# Example of Answer Selection with Graph Rule

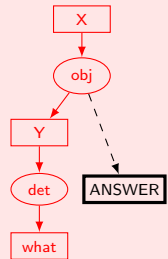
Q: What book did Michael Ende write in 1984?



A: In 1984 Michael Ende wrote the novel titled "The Neverending Story"

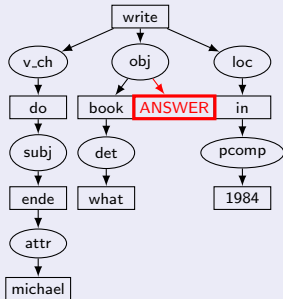


The Rule

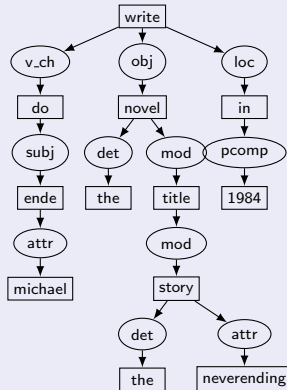


# Example of Answer Selection with Graph Rule

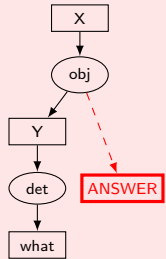
Q: What book did Michael Ende write in 1984?



A: In 1984 Michael Ende wrote the novel titled "The Neverending Story"

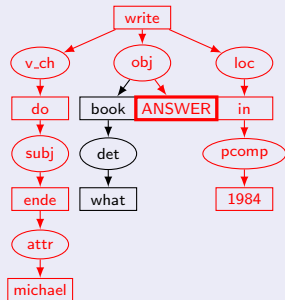


The Rule

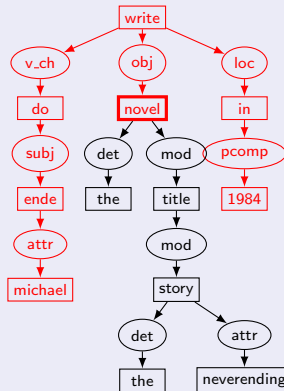


# Example of Answer Selection with Graph Rule

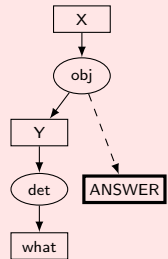
Q: What book did Michael Ende write in 1984?



A: In 1984 Michael Ende wrote the novel titled "The Neverending Story"

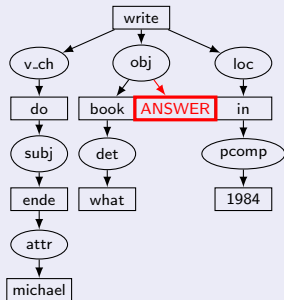


The Rule

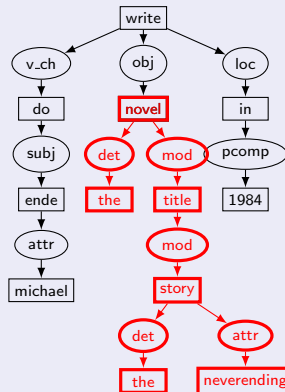


# Example of Answer Selection with Graph Rule

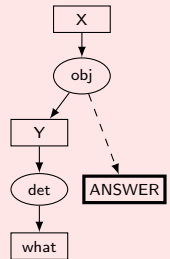
Q: What book did Michael Ende write in 1984?



A: In 1984 Michael Ende wrote the novel titled "The Neverending Story"



The Rule



# QA on Speech Transcripts

## Issues

- 1 Introduction of noise (false starts, unfinished sentences, etc)
- 2 Automated transcripts introduce recognition errors
- 3 Automated transcripts do not encode punctuation or capitalisation
- 4 Different genre

## Our Approach (Diego, Steve)

### Still in the early stages of development

- 1 Emphasis on correct NE recognition
  - ML using transcript corpora
- 2 Eventually, handle language model and dialogue

# QA on Speech Transcripts

## Issues

- 1 Introduction of noise (false starts, unfinished sentences, etc)
- 2 Automated transcripts introduce recognition errors
- 3 Automated transcripts do not encode punctuation or capitalisation
- 4 Different genre

## Our Approach (Diego, Steve)

### Still in the early stages of development

- 1 Emphasis on correct NE recognition
  - ML using transcript corpora
- 2 Eventually, handle language model and dialogue

# Cross-lingual QA

## Issues

- 1 Goal is to accept a question in source language (Dutch) and return an answer from documents in target language (English)
- 2 Translation errors
- 3 Translation of named entities

## Our Approaches (Menno in collaboration with Groningen)

- 1 Now: Translate question to target language and use regular QA methods
- 2 Eventually: Process the question in source language and pass the analysis to QA system
  - 1 Answer type
  - 2 Syntactic/semantic dependencies

# Cross-lingual QA

## Issues

- 1 Goal is to accept a question in source language (Dutch) and return an answer from documents in target language (English)
- 2 Translation errors
- 3 Translation of named entities

## Our Approaches (Menno in collaboration with Groningen)

- 1 **Now:** Translate question to target language and use regular QA methods
- 2 **Eventually:** Process the question in source language and pass the analysis to QA system
  - 1 Answer type
  - 2 Syntactic/semantic dependencies

# Query-driven Summaries

## Goal

- Process complex information needs that require combining information from various documents
- More complex than list, definition or description questions

## Example from DUC 2006

```
<topic>
<num> D0601A </num>
<title> Native American Reservation System - pros and cons </title>

<narr>
Discuss conditions on American Indian reservations or among Native American communities.
Include the benefits and drawbacks of the reservation system. Include
legal privileges and problems.
</narr>
</topic>
```

# Our Approach to Query-driven Summaries (Diego, Stephen)

## We are still in the early stages of development

- 1 Convert the question into a list of simple questions
- 2 Process each question independently through AnswerFinder
- 3 Take note of the answers and their locations in the source text
- 4 Combine the answers using summarisation techniques and sentence merging techniques

# That's all — questions?

<http://www.ics.mq.edu.au/~diego/answerfinder/>

- Text-based question answering
- Named-entity recognition for QA
- Retrieval for QA
- Graph representations of sentences
- Machine learning methods in various modules
- QA on speech transcripts
- Cross-lingual QA
- Query-driven summaries