CENTRE FOR
LANGUAGE
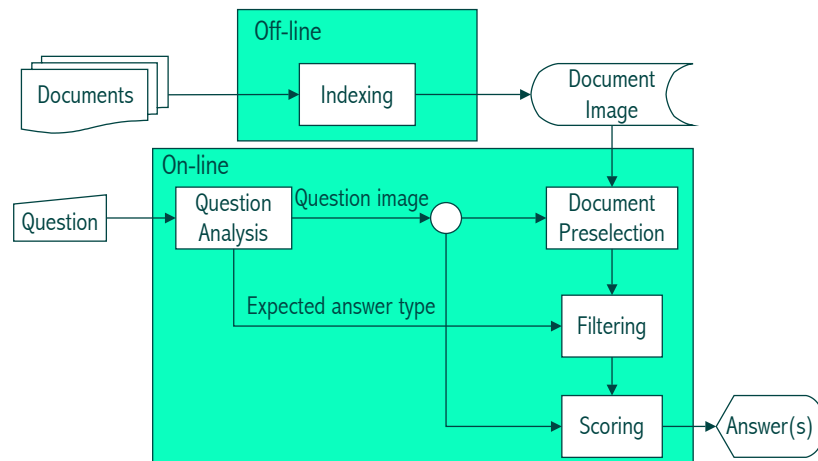**TECHNOLOGY**
MACQUARIE UNIVERSITY ~ SYDNEY

CLT's Participation in TREC-QA 2003
A Proposal

Diego Mollá-Aliod

5 May 2003

Outline

- Architecture of our QA system
- TREC-QA 2003
- A Proposal

Architecture of a QA System

The Reading Comprehension Corpus

- Remedia Publications' Reading Comprehension Tests
- Levels 2, 3, 4, and 5
- Every document contains 5 questions (*who, what, when, where, why*)
- The answer is always in the text
- Files:
  - Original text (*.txt)
  - Coreferences (*.txt.coref)
  - Named Entities (*.txt.ne)
  - Answers marked-up (*.txt.snra)
  - Answers extracted (*.txt.wdra.xml)

Sample of file rm4-16.txt

1989 Remedia Publications, Comprehension/5Ws-4
Watch Out for Sears!

(North Redwood, Minn., September, 1889) A man named Richard Sears has been playing a joke on shoppers.

Sears likes to sell items by mail. Not long ago, he ran an ad in some newspapers In small towns. The ad showed a drawing of lovely furniture. There was a sofa and two chairs.

The ad said the furniture was for sale. It said the pieces were made of fine metal frames and were beautiful to see. The ad said that for a short time only, these chairs would be shipped to all who paid 95 cents.

This message sounded too good to be true. Still, a lot of people sent in their money. Imagine their surprise when they received the furniture. The furniture was made for a doll house! They were tiny pieces.

Some people complained. That's when Sears showed them the tiny print in his ad.

In very small letters, he had included the word "miniature." That means the furniture was not full size. Sears says he did this to get attention.

1. Who played a joke on shoppers?
2. What does the Sears ad offer?
3. When did Sears play this joke?
4. Where is the word "miniature"?
5. Why did Sears play this joke?

Sample of file rm4-16.txt.ne

1989 Remedia Publications, Comprehension/5Ws-4
Watch Out for <ENAMEX TYPE="PERSON">Sears</ENAMEX>!

(<ENAMEX TYPE="LOCATION">North Redwood</ENAMEX>, <ENAMEX TYPE="LOCATION">Minn.</ENAMEX>, <TIMEX TYPE="DATE">September, 1889</TIMEX>) A man named <ENAMEX TYPE="PERSON">Richard Sears</ENAMEX> has been playing a joke on shoppers.

<ENAMEX TYPE="PERSON">Sears</ENAMEX> likes to sell items by mail. Not long ago, he ran an ad in some newspapers In small towns. The ad showed a drawing of lovely furniture. There was a sofa and two chairs.

The ad said the furniture was for sale. It said the pieces were made of fine metal frames and were beautiful to see. The ad said that for a short time only, these chairs would be shipped to all who paid <NUMEX TYPE="MONEY">95 cents</NUMEX>.

This message sounded too good to be true. Still, a lot of people sent in their money. Imagine their surprise when they received the furniture. The furniture was made for a doll house! They were tiny pieces.

Some people complained. That's when <ENAMEX TYPE="PERSON">Sears</ENAMEX> showed them the tiny print in his ad.

In very small letters, he had included the word "miniature." That means the furniture was not full size. <ENAMEX TYPE="PERSON">Sears</ENAMEX> says he did this to get attention.

1. Who played a joke on shoppers?
2. What does the <ENAMEX TYPE="PERSON">Sears</ENAMEX> ad offer?
3. When did <ENAMEX TYPE="PERSON">Sears</ENAMEX> play this joke?
4. Where is the word "miniature"?
5. Why did <ENAMEX TYPE="PERSON">Sears</ENAMEX> play this joke?

Sample of file rm4-16.txt.snra

1989 Remedia Publications, Comprehension/5Ws-4
Watch Out for Sears!
<ANSQ3>(North Redwood, Minn., September, 1889)</ANSQ3> <ANSQ1>A man named Richard Sears has been playing a joke on shoppers.</ANSQ1>

Sears likes to sell items by mail. <ANSQ3>Not long ago, he ran an ad in some newspapers In small towns.</ANSQ3> The ad showed a drawing of lovely furniture. There was a sofa and two chairs.

<ANSQ2>The ad said the furniture was for sale.</ANSQ2> It said the pieces were made of fine metal frames and were beautiful to see. <ANSQ2>The ad said that for a short time only, these chairs would be shipped to all who paid 95 cents.</ANSQ2>
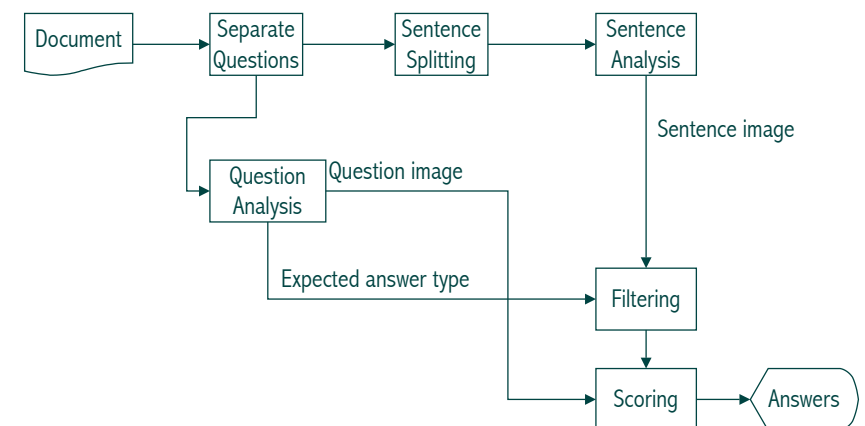
This message sounded too good to be true. Still, a lot of people sent in their money. Imagine their surprise when they received the furniture. The furniture was made for a doll house! They were tiny pieces.

Some people complained. That's when Sears showed them the tiny print in his ad.

In very small letters, he had included the word "miniature." That means the furniture was not full size. <ANSQ5>Sears says he did this to get attention.</ANSQ5>

1. Who played a joke on shoppers?
2. What does the Sears ad offer?
3. When did Sears play this joke?
4. Where is the word "miniature"?
5. Why did Sears play this joke?

A QA system for the RC Corpus

Document → Separate Questions → Sentence Splitting → Sentence Analysis

Separate Questions → Question Analysis

Sentence Analysis → Sentence image → Filtering

Question Analysis → Question image

Question Analysis → Expected answer type → Filtering
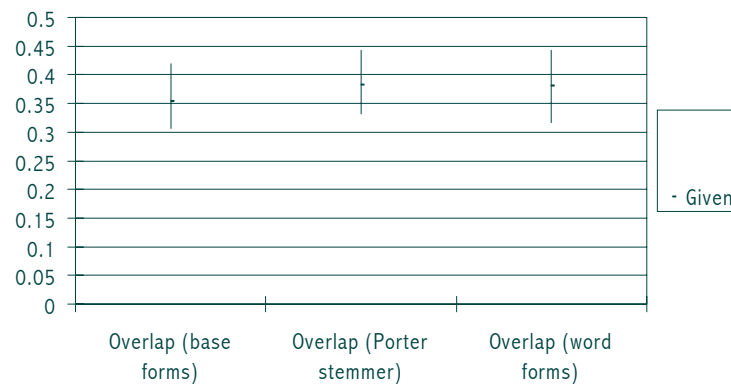
Filtering → Scoring → Answers

Possible Scoring Modules

- Word Overlap
- Dependency Overlap
- Grammatical Relation Overlap
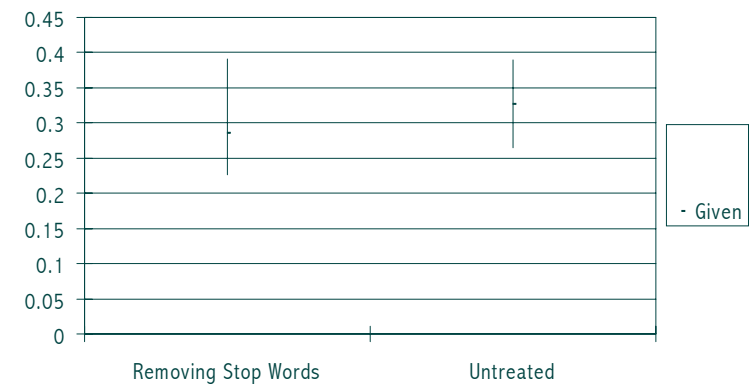- MLF Overlap
- A combination of the above
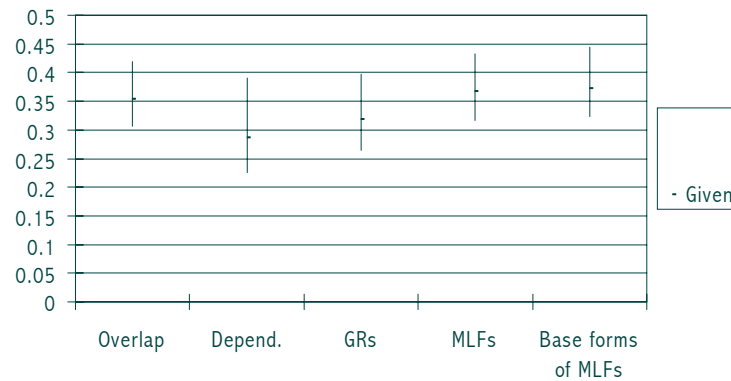
New Results (average of all levels)
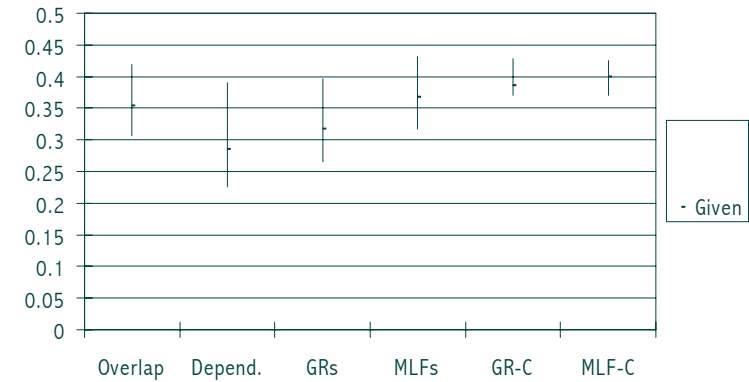
The Effect of Stemming

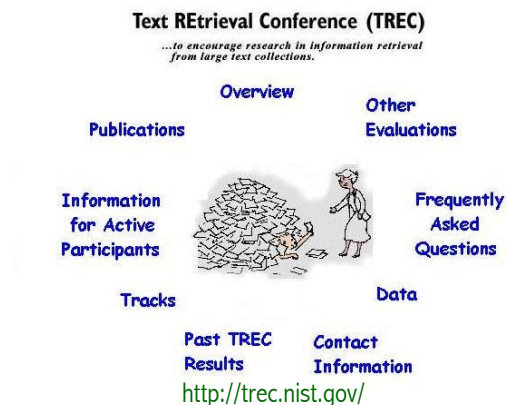## Stop Words or No Stop Words in the Dependencies?

What is the impact of the semantic dependences?

Chart with y-axis from 0 to 0.5 (increments of 0.05). X-axis categories: Overlap, Depend., GRs, MLFs, Base forms of MLFs. Legend: - Given

---

Outline

- Architecture of our QA system
- TREC-QA 2003
- A Proposal

Text REtrieval Conference

Text REtrieval Conference (TREC)
...to encourage research in information retrieval from large text collections.

Overview
Publications
Other Evaluations
Information for Active Participants
Frequently Asked Questions
Tracks
Data
Past TREC Results
Contact Information

http://trec.nist.gov/

- NIST = National Institute of Standards and Technology

TREC-QA 2003 Specifications ...

- 500 questions in total
- Three different types of questions
  - factoid questions
  - list questions (25-50 questions)
  - definition questions (25-50 questions)
- Factoid and list questions must produce exact answers
  - What about definition questions?
- Different evaluation criterion for each question type

... TREC-QA 2003 Specifications

- A single combined task
  - 500 questions
  - The question set will *not* indicate the question type
- All runs must be fully automatic
- The system cannot be modified after the training phase
- Each question must be processed from the same machine state

TREC-QA 2003 Corpus

- Data: AQUAINT corpus
  - Newswire text in English
    - Xinhua News Service (People's Republic of China)
    - New York Times News Service
    - Associated Press Worldstream News Service
  - Over 3 Gb when uncompressed
  - SGML tagged

Corpus — Sample

```
<DOC>
<DOCNO> NYT19980601.0001 </DOCNO>
<DOCTYPE> NEWS STORY </DOCTYPE>
<DATE_TIME> 1998-06-01 00:02 </DATE_TIME>
<HEADER>
A7753 &Cx1f; taf-z
u a &Cx13;  &Cx11;  BC-OBIT-LENIHAN-NYT &LR;     06-01
    0290
</HEADER>
<BODY>
<SLUG> BC-OBIT-LENIHAN-NYT </SLUG>
<HEADLINE>
KENNETH J. LENIHAN, SOCIOLOGIST WHO STUDIED CAUSES
    OF RECIDIVISM,
DIES AT 69
</HEADLINE>
   (sw)
  By WOLFGANG SAXON
  c.1998 N.Y. Times News Service
  (TEXT)
```

```
<P>
   NEW YORK _ Kenneth Joseph Lenihan, a New York research
sociologist who helped refine the scientific methods used in
criminology, died May 25 at his home in Manhattan. He was 69.
</P>
<P>
   The cause was a heart attack, his family said.
</P>
<P>
   Lenihan retired in 1995 as an associate professor of
     sociology
at John Jay College of Criminal Justice. He had joined the faculty
in 1980, after earlier stints as a researcher at Columbia
University's Bureau of Applied Social Research, the Vera
     Institute
of Justice in New York and the Bureau of Social Science
     Research in
Washington.
</P>
<P>
   He brought his expertise to the study of recidivism rates
     among
```

Corpus — Sample

```
<P>
   A large research project, it measured whether and how
     giving
jobs or money to recently released offenders would affect the
chances of their becoming repeaters. That project and further
studies formed the basis of a standard work in the field, which
     he
wrote with P. Rossi and D. Berk, ``Money, Work and Crime''
(Academic Press, 1980).
</P>
<P>
   Lenihan was born in Queens, and graduated from
     Columbia's School
of General Studies in 1960. He also earned his M.A. and
     Ph.D. in
sociology at Columbia, the latter in 1974.
</P>
<P>
   Lenihan is survived by two sons, Andrew of Miami, and
     William of
Manhattan; a daughter, Joan Lenihan of Seattle; four sisters,
Eileen McEwan of Houston, Moira Earhart of North Carolina,
```

```
</P>
</TEXT>
</BODY>
<TRAILER>
NYT-06-01-98 0002EDT &QL;
</TRAILER>
</DOC>
```

Tags Used in the Corpus

- apw
  - (DOC (DOCNO) (DOCTYPE*) (DATE_TIME*) (HEADER) (BODY (SLUG) (HEADLINE*) (TEXT (SUBHEAD*))) (TRAILER))
- nyt/1998:
  - (DOC (DOCNO) (DOCTYPE) (DATE_TIME) (HEADER) (BODY (SLUG*) (HEADLINE*) (TEXT (ANNOTATION*))) (TRAILER))
- xie/* (all years):
  - (DOC (DOCNO) (DATE_TIME) (BODY (HEADLINE) (TEXT)))

Factoid Questions in TREC 2002

- *What is the longest river in the United States?*
  - Correct and exact answers:
    - *Mississippi*
    - *the Mississippi*
    - *the Mississippi River*
    - *Mississippi River*
    - *mississippi*
  - Incorrect or inexact answers:
    - *At 2,348 miles the Mississippi River is the longest river in the US.*
    - *2,348 miles; Mississippi*
    - *the river Mississippi*
    - *Missipp*
    - *Missouri*

List Questions

- TREC 2002:
  - Each answer must contain one exact answer
  - The target number of answers is indicated in the question
  - *Name 9 countries that import Cuban sugar*
    - *China and Russia* is not one exact answer

- TREC 2003:
  - The target number of answers will not be indicated
  - When is a question a list question?

Definition Questions

- The evaluation of definition questions was problematic in TREC 2001
  - What is a good definition?
  - Context is needed
- No definition questions were used in TREC 2002
  - An AQUAINT pilot evaluation was made
- Proposal for TREC 2003:
  - A "typical user" will be assumed for the evaluation of definition questions
  - The answer is a list of "information nuggets"
  - Evaluate in terms of recall and precision

Deadlines

- The deadlines for 2002 were:
  - July 25: Testing questions and top ranked documents available
    - System to be frozen before downloading the questions
    - Possible to arrange a different (earlier) time
  - August 1: Submission of results to NIST
    - One week of time
  - October 1: Evaluated results from NIST
 - October 30: Conference papers to NIST
  - November 19-22: Conference

Outline

- Architecture of our QA system
- TREC-QA 2003
- A Proposal

The Proposal

- No time or resources to build a well-engineered system
- Implement a simple and functional system that filters out irrelevant text
- Experiment on overlap scores
  - dependencies, GRs, MLFs, …
- How to extract the <u>exact</u> answer?

Architecture of the System
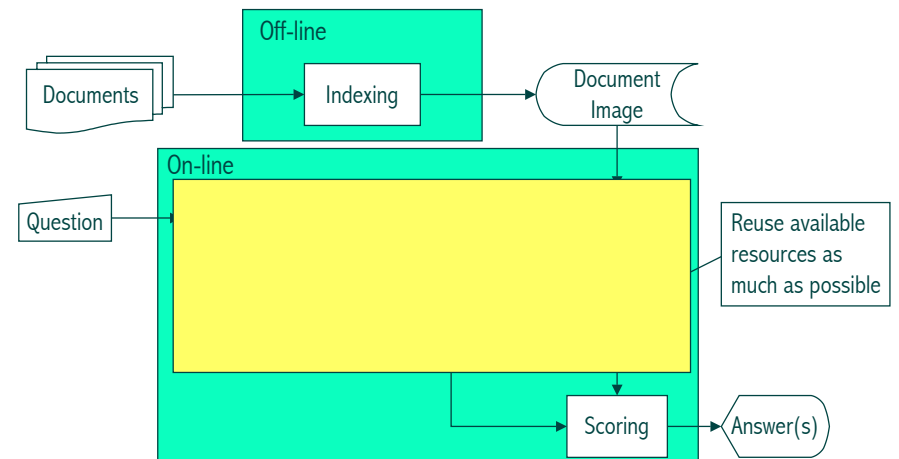
Phases

1. Indexing
   - Off-line stage
   - Extract dependency lists, GRs, and FLFs
2. QA Framework
   - Reuse as many resources as possible
   - Apply for MURDG?
     - Question classification
     - NE recogniser
3. IR System
   - Use documents provided by organisers
4. Filtering system, scoring modules
   - Basically, same as used in RC experiments
5. Extraction of the exact answer

Extraction of the exact answer

- Even if the source text is recognised, we need a good answer extraction module
- A string that contains more than strictly necessary makes the answer incorrect
  - *the river Mississipi*

- How to achieve exact answers?
  - Use the information encoded in the FLFs to determine the head of the answer
  - Expand the string according to the NE recogniser
  - Use heuristics when the above fails