Question answering

- Overview and task definition
- History
- Open-domain question answering
  - Basic system architecture
  - Predictive indexing methods
  - Pattern-matching methods
  - Advanced techniques

Question Answering

Text collection

answer + supporting text

What was the name of the enchanter played by John Cleese in the movie “Monty Python and the Holy Grail”?

Trivia: John Cleese was a Cornell A.D. White professor-at-large.

The answer (circa 2001)

History

- Closed-domain QA systems
  - LUNAR [Woods & Kaplan, 1977]
  - WOLFIE [Thompson & Mooney, 1998]
  - Q/A [Lehnert, 1978]

- Open-domain QA systems
Towards open-domain QA

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries and leading scientists from around the world.

“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, a tropical ecologist at the Smithsonian Institution and an Amazon specialist.

“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon, especially in Brazil, except bad news,” Lovejoy said in a recent interview.

Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil, but the forest also covers parts of the eight surrounding countries.

Lovejoy was one of the organizers of an unusual workshop held in mid-January in Manaus, Brazil, a sprawling city of 1 million people in the heart of the Amazon. It was the center of Brazil’s once-thriving rubber trade.

Question Answering

- Performance

<table>
<thead>
<tr>
<th>%correct</th>
<th>avg rank</th>
<th>1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC-8 (1999):</td>
<td>70%</td>
<td>1.4</td>
</tr>
<tr>
<td>TREC-9 (2000):</td>
<td>65%</td>
<td>1.7</td>
</tr>
<tr>
<td>TREC-10 (2001):</td>
<td>70%</td>
<td>1.3</td>
</tr>
<tr>
<td>TREC-11 (2002):</td>
<td>83%</td>
<td>1</td>
</tr>
<tr>
<td>TREC-12 (2003):</td>
<td>70%</td>
<td>1</td>
</tr>
<tr>
<td>TREC-13 (2004):</td>
<td>84%</td>
<td>1</td>
</tr>
</tbody>
</table>

- List questions

1915: List the names of chewing gums.

Stimorol  Orbit  Winterfresh  Double Bubble
Dirol  Trident  Spearmint  Bazooka
Doublemint  Dentynex  Freedent  Hubba Bubba
Juicy Fruit  Big Red  Chiclets  Nicorette

- Can’t just rely on a single document

- Performance

- TREC 2003: F .40
- TREC 2004: F .62
R/P/F-measure

- Recall
- Precision
- F-measure, F-score

Definition Questions

- Who is Colin Powell?
- What is mold?

- Hard to evaluate
  - Who is the audience?
  - Evaluation requires matching concepts in the desired response to concepts in a system response

  TREC 2003:
  - Audience: questioner is an adult, a native speaker of English, and an “average” reader of US newspapers
  - Results: $F .55$

Context Task

- Track a target discourse object through a series of questions

<table>
<thead>
<tr>
<th>21</th>
<th>Club Med</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.1</td>
<td>FACTOID</td>
</tr>
<tr>
<td>21.2</td>
<td>LIST</td>
</tr>
<tr>
<td>21.3</td>
<td>FACTOID</td>
</tr>
<tr>
<td>21.4</td>
<td>OTHER</td>
</tr>
</tbody>
</table>

- Performance
  - TREC 2004
    - Factoids: .84 initial; .74 non-initial
    - Lists: .62 $F$
    - Other: .46 $F$

Question answering

- Overview and task definition
- History
- Open-domain question answering

Basic system architecture

[Cardie et al., ANLP 2000]

- Predictive indexing methods
- Pattern-matching methods
- Advanced techniques
Document retrieval

- **Standard ad-hoc IR using full-text indexing**

- **Example QA system uses**
  - vector space model
  - text retrieval system: Smart
  - standard term-weighting strategies (tfidf)
  - no automatic relevance feedback
Baseline evaluation

- Document retrieval only
- Corpus
  - TREC-8 development corpus (38 questions)
  - TREC-8 test corpus (200 questions)

<table>
<thead>
<tr>
<th>Development (38)</th>
<th>Test (200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart Correct</td>
<td>Smart Correct</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>MAR = 3.33</td>
<td>MAR = 3.07</td>
</tr>
</tbody>
</table>

**MAR = Mean Answer Rank**

Baseline evaluation

- Smart performs better than its scores would suggest
  - Development corpus
    - For 18 of 38 questions, the answer appears in the top-ranked document
    - For 33 of 38 questions, the answer appears in one of the top 7 documents
    - For only 2 questions does Smart fail to retrieve the answer in one of the top 25 documents
  - Test corpus
    - Over half (110) of the questions are answered in the top-ranked document
    - Over 75% of the questions (155) are answered in the top 5 documents
    - 19 questions were not answered in the top 20 documents

System architecture: passage retrieval

Passage Retrieval
The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world. That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist. “It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon.”

Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil.

Which country has the largest part of the Amazon rain forest?

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries.

Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil, but the forest also

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries. Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil, but the forest also

Japan will not fund construction of the final segment of a controversial highway through the Amazon rain forest in Brazil. Sen. Bob Kasten of Wisconsin said Japan’s ambassador to the United States, Nobuo Matsunga, “assu
Evaluation: text summarization

<table>
<thead>
<tr>
<th></th>
<th>Development (38)</th>
<th>Test (200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>MAR</td>
</tr>
<tr>
<td>Smart</td>
<td>3</td>
<td>3.33</td>
</tr>
<tr>
<td>Text Summarization</td>
<td>4</td>
<td>2.25</td>
</tr>
</tbody>
</table>

MAR = Mean Answer Rank

Evaluation: passage retrieval

- **Summarization method can limit performance**
  - Development corpus
    - In only 23 of the 38 developments questions (61%) does the correct answer appear in the summary for one of the top \( k=7 \) documents
  - Test corpus
    - In only 135 of the 200 developments questions (67.5%) does the correct answer appear in the summary for one of the top \( k=6 \) documents

Linguistic filters

- 50 byte answer length effectively eliminates *how, why, and definition* questions
- almost all of the remaining question types are likely to have **noun phrases** as answers
  - development corpus: 36 of 38 questions have noun phrase answers
- consider adding at least a simple linguistic filter that considers only noun phrases as answer hypotheses

System architecture: linguistic filters

- Doc Retrieval
- Passage Retrieval
- Noun phrase filter
- Document collection
- Documents, text passages
- Answer hypotheses
- 1st Guess
- 2nd Guess
- 3rd Guess
- 4th Guess
- 5th Guess
- Linguistic filters
The noun phrase filter

Which country has the largest part of the Amazon rain forest?

ordered list of summary extracts

[The huge Amazon rain forest] is regarded as vital to [the global environment].

[Japan] will not fund [the construction] of [the final segment] of [a controversial highway] through [the Amazon rain forest] in [Brazil], according to [a senior Republican senator].

ordered list of NPs

answer hypotheses

10-word chunker

Question-Answering System

“Japan Brazil a new plan Amazon countries A section”

“Northwestern Brazil A plan the Amazon region”

“eight surrounding countries A Brazilian company”

“union leader his modest wooden house The people”

“defense Brazilian wealth the international market”

QA using the NP filter

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries. Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil, but the forest also

ordered list of summary extracts

[Cardie & Pierce ACL98]

Sentence reordering

Summaries

NP finder

Summaries

NPs

Guesses

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries. Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil, but the forest also

Japanese ambassador to the United States, Nobuo Matsunga, “assu

Evaluation: NP filter

<table>
<thead>
<tr>
<th></th>
<th>Development (38)</th>
<th>Test (200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>MAR</td>
<td>Correct</td>
</tr>
<tr>
<td>Smart</td>
<td>3</td>
<td>3.33</td>
</tr>
<tr>
<td>Text Summarization</td>
<td>4</td>
<td>2.25</td>
</tr>
<tr>
<td>TS + NPs</td>
<td>7</td>
<td>2.29</td>
</tr>
</tbody>
</table>

MAR = Mean Answer Rank

- Using NP finder of Cardie & Pierce (1998)
  - ~94% precision and recall on Wall Street Journal text
- How much does the (unnatural) NP “chunking” help?
  - Without it, only 1 and 20 questions answered for each corpus, respectively
  - NP filter is extracting good guesses, but better linguistic processing is needed to promote the best guesses to the top of the ranked guess list
Semantic class checking

- Approximate question type using question word
  - **Who** is the president of the U.S.?
    - person
  - **Which country** has the largest part of the Amazon rain forest?
    - country
  - **Where** is the Connecticut River?
    - state? county? country? location?
  - **What fabric** should one use to make curtains?
    - fabric???
- Check that head noun (i.e. the last noun) of answer NP is of the same type
  - a **man** = person
  - Massachusetts = state, location

Semantic type checking

- Use lexical resource to determine semantic compatibility
  - WordNet
- Proper names handled separately since they are unlikely to appear in WordNet
  - Small set (~20) rules

Evaluation: semantic class filter

<table>
<thead>
<tr>
<th></th>
<th>Development (38)</th>
<th>Test (200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>MAR</td>
</tr>
<tr>
<td>Smart</td>
<td>3</td>
<td>3.33</td>
</tr>
<tr>
<td>Text Summarization</td>
<td>4</td>
<td>2.25</td>
</tr>
<tr>
<td>TS + NPs</td>
<td>7</td>
<td>2.29</td>
</tr>
<tr>
<td>TS + NPs + Semantic Type</td>
<td>21</td>
<td>1.38</td>
</tr>
</tbody>
</table>

MAR = Mean Answer Rank

- Weak syntactic and semantic information allows large improvements
- Problems?
Sources of error

Development questions (38)
- Smart: 6%
- TS: 31%
- Ling Filters: 63%

Test questions (200)
- Smart: 40%
- TS: 24%
- Ling Filters: 36%