Towards Semantic-based Overlap Measures for Question Answering

Diego Mollá-Aliod
10 Dec. 2003

Architecture of a QA System

- Question Analysis
  - Question classification
    - Find answer type and focus
  - Indexing, Document Preselection
    - IR techniques
    - Bag of Words
- Filtering
  - IE techniques
  - NE recognition
- Scoring
  - Pattern matching
  - Syntax, semantics

Grammatical Relations

- Carroll, Briscoe, Sanfilippo “Parser evaluation: a survey and a new proposal”, Proceedings LREC98
- Purpose
  - Compare the output of different parsers
- How
  - Use relations between words (subject, object, etc)
    - Inspired in dependency theory
  - Hierarchy of grammatical relations
    - To accommodate different parser granularities
Grammatical Relations (Carroll et al. 1998)

Parser Accuracy

Corpus and methodology: Carroll and Briscoe

<table>
<thead>
<tr>
<th></th>
<th>With Link Grammar</th>
<th>With Conexor FDG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>SUBJ</td>
<td>50.3%</td>
</tr>
<tr>
<td></td>
<td>OBJ</td>
<td>48.5%</td>
</tr>
<tr>
<td></td>
<td>XCOMP</td>
<td>62.2%</td>
</tr>
<tr>
<td></td>
<td>MOD</td>
<td>57.2%</td>
</tr>
<tr>
<td>Average</td>
<td>54.6%</td>
<td>74.6%</td>
</tr>
<tr>
<td>Recall</td>
<td>SUBJ</td>
<td>39.1%</td>
</tr>
<tr>
<td></td>
<td>OBJ</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>XCOMP</td>
<td>32.1%</td>
</tr>
<tr>
<td></td>
<td>MOD</td>
<td>53.7%</td>
</tr>
<tr>
<td>Average</td>
<td>43.7%</td>
<td>59.7%</td>
</tr>
</tbody>
</table>

QA with Grammatical Relations

- A man named Richard Sears has been playing a joke on shoppers.
  (detmod man a),
  (subj name man), (dobj name richard_sears),
  (detmod joke a), (subj play man), (aux play have),
  (aux play be),
  (ncmod shopper play on), (dobj play joke)
- Who played a joke on shoppers?
  (subj play who), (dobj play joke),
  (ncmod shopper play on), (detmod joke a)

Flat Logical Forms

- Avoid the use of nested expressions
  - Conjunction of predicates with all variables existentially bounded
  - Relify objects, events, and properties
- Also called minimal logical forms because they encode minimal information for the task of question answering
  - Incremental extensibility
QA with Flat Logical Forms

A man named Richard Sears has been playing a joke on shoppers.

holds(v_010), object('man',v_02,[v_x2]),
evt('name',v_23,[v_x3,v_x2,v_x4]),
object('play_on',v_e5,[v_x2,v_x10,v_x12]),
object('richard_sears',v_04,[v_x4]),
object('joke',v_010,[v_x10]),
object('shopper',v_012,[v_x12])

Who played a joke on shoppers?

holds(v_e2), object('who',v_01,[v_x1]),
evt('play_on',v_e2,[v_x1,v_x4,v_x6]),
object('joke',v_04,[v_x4]), object('shopper',v_06,[v_x6])

Corpus Used — Sample of file rm4-16.txt

1989 Remedia Publications, Comprehension/5Wts-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?

The Evaluation — Document Set

- 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)

Corpus Used — Sample of file rm4-16.txt.ne

1989 Remedia Publications, Comprehension/5Wts-4
Watch Out for Sears!
(<ENAMEX TYPE="PERSON">Sears</ENAMEX>)
(<ENAMEX TYPE="LOCATION">North Redwood</ENAMEX>, <ENAMEX TYPE="LOCATION">Minn.</ENAMEX>.
<TM>September, 1889</TM>) 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 <ENAMEX TYPE="MONETARY">95 cents</ENAMEX>.

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?
Corpus Used – Sample of file rm4-16.txt.snra

1989 Remedia Publications, Comprehension/SWs-4
Watch Out for Sears!
<ANSQ1>: (North Redwood, Minn., September, 1889) <ANSQ1>: <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, lots 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?

Our QA Prototype

- Question Analysis
  
<table>
<thead>
<tr>
<th>Regex</th>
<th>Expected Answer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Who</td>
<td>person, organization</td>
</tr>
<tr>
<td>*What</td>
<td>any</td>
</tr>
<tr>
<td>*When</td>
<td>date, time</td>
</tr>
<tr>
<td>*Where</td>
<td>location</td>
</tr>
<tr>
<td>*Why</td>
<td>any</td>
</tr>
</tbody>
</table>

- Filtering  
  - No expected answer: penalization of -100

- Scoring  
  - Overlap of GramReps or FLFs

Architecture of Our QA System

The Evaluation

- Mean Reciprocal Rank (MRR)  
  - Used in past QA tracks of TREC  
  - Examine the 5 highest-ranking answer candidates of each question  
  - \[ RR = \frac{1}{r_1 + \frac{1}{r_2} + \frac{1}{r_3} + \frac{1}{r_4} + \frac{1}{r_5}}, \text{ or } 0 \]  
  - \[ RR = \frac{1}{r_1 + \frac{1}{r_2} + \frac{1}{r_3} + \frac{1}{r_4} + \frac{1}{r_5}}, \text{ or } 0 \]  
  - MRR = mean of all RR
Best and Worst Cases

- The overlap scores in the GramRel and MLFs are very low
- The right answer is competing with other sentences

What does the Sears ad offer?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Sentence</th>
<th>Score</th>
<th>Overlap</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1989 Remedia Publications, Comprehension/SW&lt;32&gt;s-4</td>
<td>1</td>
<td>compound_noun(v,x2,v,x3)</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>(North Redwood, Minn)</td>
<td>1</td>
<td>compound_noun(v,x2,v,x3)</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>Not long ago, he ran an ad in some newspapers in small towns</td>
<td>1</td>
<td>object(adv,c7,[v,x7])</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>The ad showed a drawing of lovely furniture</td>
<td>1</td>
<td>object(adv,c2,[v,x2])</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>The ad said the furniture was for sale</td>
<td>1</td>
<td>object(adv,c2,[v,x2])</td>
<td>yes</td>
</tr>
</tbody>
</table>

Combining Scores

- Weight on Grammatical Relations
  Score = 27 x GRO + 9 x FO + 3 x DO + WO
- Weight on Flat Logical Forms
  Score = 27 x FO + 9 x GRO + 3 x DO + WO

WO = word form overlap
DO = dependency overlap (using the output of Conexor FDG)
FO = overlap of FLFs
GRO = overlap of GramRel
Conclusions

- Overlap of FLFs is better than overlap of GramRelS
- The impact of the parser for QA is small, but...
- Overlap of word forms is better!
- Questions
  - The Reading Comprehension corpus is too simple?
  - The parser and semantic interpreter introduces too many errors?
  - More complex overlap measures are perhaps better?
  - Logical forms are not required for simple (fact-based) questions?

Epilogue — Using Another Corpus

- TREC 2003 Passage Task
  - Document set: 3Gb of newswire text
  - Factoid questions (used in TREC 2002)
  - Retrieve a 250-word passage containing the answer
- Results:

<table>
<thead>
<tr>
<th>Run</th>
<th>Formula</th>
<th>Without</th>
<th>With NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>answfind1</td>
<td>3wo+gro</td>
<td>16.8%</td>
<td>19.1%</td>
</tr>
<tr>
<td>answfind2</td>
<td>9wo+3gro+fo</td>
<td>16.8%</td>
<td>19.3%</td>
</tr>
<tr>
<td>answfind3</td>
<td>9wo+3fo+gro</td>
<td>15.6%</td>
<td>18.4%</td>
</tr>
</tbody>
</table>

QA with Grammatical Relations — In Reality

- A man named Richard Sears has been playing a joke on shoppers.
  (xmod _ man name), (detmod _ man a),
  (subj name man), (dobj name richard _),
  (detmod _ joke a), (subj sear man _), (subj play sear _),
  (aux _ play have), (aux _ play be), (nmod _ play on),
  (xcomp _ play joke)
- Who played a joke on shoppers?
  (subj play who _), (dobj play joke _), (nmod _ play on),
  (detmod _ joke a)

QA with Flat Logical Forms — In Reality

- A man named Richard Sears has been playing a joke on shoppers.
  holds(v_o10, object('man',v_o2,[v_x2]),
  evt('name',v_e3,[v_X3,v_x4,v_x2]),
  object('joke_on',v_o10,[v_e5,v_x12]),
  object('richard',v_o4,[v_x4]), evt('sear',v_e5,[v_x2]),
  object('shopper',v_o12,[v_x12])
- Who played a joke on shoppers?
  holds(v_e2, object('who',v_o1,[v_x1]),
  evt('play_on',v_e2,[v_x1,v_x4,v_x6]),
  object('joke',v_o4,[v_x4]), object('shopper',v_o6,[v_x6])