EQUAL – Encyclopaedic QA for Lists

Iustin Dornescu

Research Group in Computational Linguistics,
University of Wolverhampton, UK

October 2, 2009
Complex question answering on Wikipedia.  
50 multilingual list topics such as:

**GC-2009-06**  
Which Dutch violinists held the post of concertmaster at the Royal Concertgebouw Orchestra in the twentieth century?

**GC-2009-34**  
What eight-thousanders are at least partially in Nepal?
Encouraging results

We need something radically different to standard textual QA
Outline

1. Semantic QA: my vision for the future

2. EQUAL – Implementation details
# Semantic QA rEvolution

<table>
<thead>
<tr>
<th>Textual QA</th>
<th>vs.</th>
<th>web of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>[web of] documents</td>
<td></td>
<td>document retrieval</td>
</tr>
<tr>
<td>keywords &amp; co-occurrence</td>
<td></td>
<td>search for facts</td>
</tr>
<tr>
<td>textual snippet</td>
<td></td>
<td>concepts &amp; relations</td>
</tr>
<tr>
<td>gazetteers</td>
<td></td>
<td>graph patterns</td>
</tr>
<tr>
<td>text with entities</td>
<td></td>
<td>RDF Data</td>
</tr>
<tr>
<td>lexical semantics</td>
<td></td>
<td>entities with text</td>
</tr>
</tbody>
</table>

**Semantic QA**
Semantic QA representation

- represent Wikipedia using a semantic graph: entities characterised by types, attributes & properties, and connected to each other by relations

- a question is a composition of **constraints** about:
  - entities
  - their types
  - their properties
  - their relations

- finding answers means performing some actions from a given set (that the system understands)
Semantic QA

Dev. topic
Which football players from Brazil play in clubs from the Iberian Peninsula

1. identify Brazilian footballers
2. identify the club each of them plays for
3. test if the club is located in the Iberian Peninsula
Analysis and Feedback

A useful QA system must use the meaning, not the words:

**Analysis Phase**
Understand the information need behind the question
- detect different *ambiguity* sources
- create a *semantic question interpretation* for each ranking

**Feedback Phase**
Interact with the user
- allow user to disambiguate
- generate justification
- active learning
Outline

1. Semantic QA: my vision for the future
2. EQUAL – Implementation details
Analysis 1: EAT

a) Use an ad-hoc classifier for chunk delimiters: the first chunk with a plural noun contains the EAT;

List the *Italian places* where ...
Which *countries* have...
Name *Romanian poets* who...
Which *Dutch violinists* who...
In which *European countries* is...

b) Find the best matching Wikipedia Category.
Map the remaining chunks to semantic constraints:

a) entity: ..“are at least partly in [Nepal]”

b) category: ..“play in [Spanish football clubs]”

c) property: ..“with a population larger than 100,000 people ”

d) temporal: ..“in the twentieth century”

e) geographic: ..“are at least partly in [Nepal]”
The actual implementation of a semantic constraint.

- infobox attributes
- categories
- definition
- external datasources
- article text (NLP)
Results

- EQUAL: 813 answers – 385(correct), 105(unjustified), 323(incorrect)
- $P = 47.35\% \ (60.27\%)$
- $R^P = 27.20\% \ (34.62\%)$
- Total Answers: 1415