Esfinge at CLEF 2007: First Steps in a Multiple Question and Multiple Answer Approach

Luis Miguel Cabral, Luis Fernando Costa, Diana Santos
{Luis.M.Cabral,Luis.Costa,Diana.Santos}@sintef.no
Linguateca / SINTEF ICT
PB 124, Blindern NO-0314 Oslo, Norway

Esfinge on the Web: http://www.linguateca.pt/Esfinge/

### Esfinge Overview
- General domain question answering system, participation in QA@CLEF since 2004.
- The starting point was the redundancy-based architecture described in [Brill 2003].
- Exploiting the redundancy in the Web where Portuguese is one of the most used languages.
- For 2007: use of the M,N-O,P model for QA (multiple questions, multiple answers), anaphoric resolution, and Wikipedia.

### Answer Selection
- Merging different runs by analysing the result (3 to 6 runs)
- Number of times an answer occurs in the results
- The relevance of the support snippet regarding the question text
- Automatic Selection achieves 55% - 69% of the correct answers in the best possible selection
- Combination of several runs did not produce better results than some of the individual runs combined and therefore needs further attention

### Experiments
- Comparing runs with two different regular expression pattern generations
- Esfinge regular expression patterns
- PALAVRAS generated patterns
- Measuring the influence of the Web
- Web + Newspapers + Wiki
- Newspapers + Wiki
- Measuring the import of using Wikipedia
- Web + Newspapers + Wiki
- Web + Newspapers
- Using phrase patterns based on the syntactical analysis by PALAVRAS seems to slightly increase the performance
- Addition of Wikipedia brought 37.5% (9/24) increase in performance
- Use of the Web brought 32% (8/25) increase in performance

### Future Work
- Extend the question reformulation module
- Create expressions that include the answer expected position
- Use ontologies
- Use further syntactic analysis
- Create high precision filters for closed-class types of answers
- Improve choice among several answers
- Weigh redundancy by lexical counts in each document collection
- Use link structure of Wikipedia
- Do an empirical study for weighting alternative questions

### Anaphor Resolution
- Coping with five kinds of anaphoric expressions (see paper)
- Involving PALAVRAS, a broad-coverage parser for Portuguese [Bick 2000]
- Dealing with sentence arguments (object, subject, etc.) as candidates
- Shallow reference resolution algorithm: 71% accuracy in general (over 122 questions)
- Producing several questions from the original one
- As a by-product, producing "syntactically-based patterns" as an alternative to regular expression reformulation of Esfinge
- Example (3 questions out of 1):
  - Contra quem é que Steffi Graf não jogou nas semi-finais de Roland Garros?
  - Contra quem é que o pai de Steffi Graf não jogou nas semi-finais de Roland Garros?
  - Contra quem é que ela não jogou nas semi-finais de Roland Garros?

This module had relatively good results (although it had little influence in the final performance given the few cases it had to resolve)

### Results

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Right Answers (all questions)</th>
<th>Unsupported Answers</th>
<th>Incorrect Answers</th>
<th>Incorrect Answers +</th>
<th>Right Answers (all questions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R2007.P1.P2</td>
<td>12</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>R2007.P1.P2</td>
<td>11</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Web + News + Wild</td>
<td>13</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>News + Wiki</td>
<td>25</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Web + News</td>
<td>24</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Automatic selection 3-5</td>
<td>31</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Web + News + Wild</td>
<td>35</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Web + Wiki</td>
<td>23</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>News + News</td>
<td>19</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>Automatic selection F:10</td>
<td>34</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>Web + News</td>
<td>40</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Automatic selection 3-4</td>
<td>44</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>Automatic selection 3-4</td>
<td>44</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>Best selection 3-5</td>
<td>65</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Best selection in 2006</td>
<td>56</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>ESFINGE</td>
<td>56</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

i) Right answers (including NIL)  ii) Partial right answers on lists  iii) NIL right answers