The MEANING approach for automatic acquisition of sense examples.

Montse Cuadros, Jordi Atserias, Mauro Castillo
TALP Research Center, UPC
Jordi Girona, 1-3 Edifici Omega 305
08034 Barcelona
Spain,
{cuadros,batalla,castillo}@lsi.upc.edu

German Rigau
IXA Group, EHU
Euskal Herriko Unibertsitatea
08034 Donostia
Spain,
{rigau}@si.ehu.edu

1 ExRetriever

ExRetriever characterises each sense of a word as a specific query. This is automatically done by using a particular query construction strategy, which is defined a priori by an expert. Each different strategy can take into account the information related to words and available into a lexical knowledge base in order to automatically generate the set of queries.

The current version of ExRetriever is able to use different lexical databases (through the MCR) and different corpora (SemCor, BNC, the Web, etc.) through a common API.

In order to easily implement different query construction strategies, ExRetriever has been powered with a declarative language. This language allows the manual definition of complex query construction strategies and it is briefly described in the following section.

2 Experiments

Within the framework of the MEANING project, both direct and indirect evaluation experiments of the ExRetriever performance have been designed.

Using ExRetriever on SemCor we can perform detailed micro-analysys on the data available. That is, we can easily perform many adjustments for building queries and filtering appropriately those unwanted examples, balancing the trade-off between coverage and precision.

Each one of such experiments consists of applying a particular query construction strategy to a set of 73 English words from Senseval-2 lexical sample task. The resulting specific queries (one for each sense word) automatically generated by applying each strategy have been tested against Semcor. Due to the small size of Semcor, specific queries are likely to produce poor recall. However, Semcor is the unique sense tagged resource providing large quantities of examples for all-words.

Six different query construction strategies have been tested, some of them inspired in those used in the literature, Lea (Leacock et al. 98), Mol1-3 (Mihalcea & Moldovan 99), Mea1-2 (Fernández et al. 04).

3 Results

<table>
<thead>
<tr>
<th>Q</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>NSen</th>
<th>WSC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lea1</td>
<td>98.84</td>
<td>3.66</td>
<td>7.06</td>
<td>23254</td>
<td>23</td>
</tr>
<tr>
<td>Mol1</td>
<td>99.35</td>
<td>4.72</td>
<td>9.01</td>
<td>3241</td>
<td>10</td>
</tr>
<tr>
<td>Mol3</td>
<td>8.12</td>
<td>26.11</td>
<td>12.39</td>
<td>7611</td>
<td>47</td>
</tr>
<tr>
<td>Mea1</td>
<td>9.28</td>
<td>24.38</td>
<td>13.44</td>
<td>9490</td>
<td>54</td>
</tr>
<tr>
<td>Mea2</td>
<td>10.69</td>
<td>26.28</td>
<td>15.20</td>
<td>17171</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 1: Overall figures

Table 1 shows the overall figures for each query when applied to the total 73 words of the test set. WSC* (Cuadros et al. 04) is a measure of the coverage of the senses of the word.

References


