**University of Alicante at WiQA 2006**

**Antonio Toral Ruiz, Georgiana Puşcaşu, Lorena Moreno Montagudo, Rubén Izquierdo Beviá, Estela Saquete Boro**

Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante, Spain
{atalor.georgia,loren, ruben, stela}@dlsi.ua.es

**Abstract**

The approach we adopted for solving this task uses Information Retrieval, query expansion by feedback, relevance and novelty re-ranking, as well as temporal ordering. Our system has participated both in the Spanish and English monolingual tasks. For each of the two participations the results are promising because, by employing a language independent approach, we obtain scores above the average. Moreover, in the case of Spanish, our result is very close to the best achieved score.

---

**Introduction**

Why Wikipedia is important for NLP?
- Size (and growing pace)
- General domain coverage
- Multilingualism
- More "structured" than raw corpora

**WiQA**

CLEF-2006 Pilot Task
Discover related entries in Wikipedia
Find info relevant for a topic not yet present in its Wikipedia entry

---

**Procedure**

**Information Retrieval (IR)**
- Xapian: probabilistic open source library
- Feedback-driven strategy with query expansion

**Post processing**
- Eliminate snippets from the topic entry
- Eliminate snippets from entries linked to the topic entry

**Relevance & novelty re-ranking**
- Measure the relevance/novelty of retrieved snippets with respect to the topic entry on the basis of novel NEs, temporal expressions, nouns and verbs

**Temporal ordering**
- Interchange snippets containing temporal expressions so that the result reflects their temporal order

*only for English

---

**Results**

**ENGLISH**

<table>
<thead>
<tr>
<th>Run ID</th>
<th>Avg yield</th>
<th>MRR</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.98</td>
<td>0.53</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>2.63</td>
<td>0.52</td>
<td>0.32</td>
</tr>
</tbody>
</table>

**SPANISH**

<table>
<thead>
<tr>
<th>Run ID</th>
<th>Avg yield</th>
<th>MRR</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>1.52</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>med</td>
<td>2.46</td>
<td>0.52</td>
<td>0.32</td>
</tr>
<tr>
<td>max</td>
<td>3.38</td>
<td>0.59</td>
<td>0.37</td>
</tr>
</tbody>
</table>

---

**Conclusions**

Results above average & close to best for Spanish

Study effects of novelty/relevance re-ranking & temporal ordering (results have not been those expected)

Further investigate this topic departing from our IR approach