Semantic Annotation for Multilingual Search

Shibamouli Lahiri

(shibamouli@cse.psu.edu)
Semantic Annotation

• Tags describe the content/meaning/sense of a particular piece of text, e.g. a sentence, a paragraph or a document

• Tags are given in natural language, so they need disambiguation (“jaguar” – an animal or a car)
Multilingual Access Modalities to Legal Resources Based on Semantic Disambiguation
(Peruginelli and Francesconi, 2009)

• Categorization of metadata (tags)
• Metadata-based, keyword-based and category-based querying
• Query language-independent metadata (document title)
• English is pivot language
Multilingual Access Modalities to Legal Resources Based on Semantic Disambiguation (Peruginelli and Francesconi, 2009)
Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval (Volk et al., 2002)

- English – German parallel corpus of 9000 documents

- Concept annotation with EuroWordNet

- Concept annotation with Medical Schema
Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval (Volk et al., 2002)

<table>
<thead>
<tr>
<th>Method</th>
<th>mAveP</th>
<th>Rel. Docs Retr.</th>
<th>AvP 0.1</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE2EN-DE-token</td>
<td>0.0512</td>
<td>66</td>
<td>0.1530</td>
<td>0.1160</td>
</tr>
<tr>
<td>DE2EN-MT-PT2001</td>
<td>0.1184</td>
<td>376</td>
<td>0.3382</td>
<td>0.2520</td>
</tr>
<tr>
<td>DE2EN-EWN</td>
<td>0.0090</td>
<td>111</td>
<td>0.0311</td>
<td>0.0160</td>
</tr>
<tr>
<td>DE2EN-UMLS</td>
<td>0.1620</td>
<td>366</td>
<td>0.3724</td>
<td>0.2800</td>
</tr>
<tr>
<td>DE2EN-MeSH</td>
<td>0.1699</td>
<td>304</td>
<td>0.3888</td>
<td>0.2600</td>
</tr>
<tr>
<td>DE2EN-Semrel</td>
<td>0.0229</td>
<td>23</td>
<td>0.0657</td>
<td>0.0480</td>
</tr>
<tr>
<td>DE2EN-all-combined</td>
<td>0.1774</td>
<td>404</td>
<td>0.3872</td>
<td>0.2720</td>
</tr>
<tr>
<td>DE2EN-SimThes</td>
<td>0.2290</td>
<td>409</td>
<td>0.4492</td>
<td>0.3640</td>
</tr>
<tr>
<td>DE2EN-SimThes+all-comb.</td>
<td>0.2955</td>
<td>518</td>
<td>0.5761</td>
<td>0.4600</td>
</tr>
</tbody>
</table>

Table 3: Results of the cross-language runs: German queries and English documents
Other Pointers

- The TRUST project (English, French, Italian, Polish, Portuguese)

- Workshop on Multilingual Semantic Web (WWW 2010)
References

