Intelligent Agent Systems Lab, Institute of Information Science, Academia Sinica

IASL Korean-Chinese CLIR System
Query Translation CLIR System based on Bilingual Dictionary and Co-occurrence Method
Yu-Chun Wang, Cheng-Wei Lee, Richard Tzong-Han Tsai, Wen-Lian Hsu*
hsu@iis.sinica.edu.tw
Academia Sinica, Taiwan, R.O.C.

We propose an architecture for retrieving Chinese documents based on Korean queries in NTCIR CLIR K-C Task. Our system uses a bilingual dictionary to perform query translation. We expand our bilingual dictionary by extracting words and their translations from the Wikipedia site, an online encyclopedia. To resolve the problem of translating Western people's names into Chinese, we propose a transliteration mapping method. We translate queries from Korean query to Chinese by using a co-occurrence method.

**Architecture**

**Query Processing**

- **Rule-based Term Processing**
- **KLT Term Extractor**

**Term Translation**

- **Bilingual Dictionary**
- **Daum K-C Dictionary**
- **CKIP Autotag**
- **CIRB Index**

**Retrieval**

- **Lucene IR Engine**

**Query Translation**

**Bilingual Dictionary**

We use the free online Korean-Chinese bilingual dictionary provided by the Daum Korean web site.

**Wikipedia**

We use Wikipedia to expand our dictionaries for the proper nouns. The following is the procedure.

1. Send Korean Terms to Korean Wikipedia
2. Find the Inter-language Links in Korean Wikipedia Pages
3. Follow the Inter-language Link to Chinese Wikipedia

**Person Name Translation**

- **그린스펀**
  - Korean Transliteration of the Name "Greenspan"
- **Greenspan**
  - Original English Name
- **CNA English-Chinese Transliteration Table**
  - 葛林斯潘
  - Chinese Transliteration of the Name "Greenspan"

**Term Disambiguation**

Many different Chinese loanwords have the same pronunciation when written in the Hangul alphabet.

<table>
<thead>
<tr>
<th>Mutual Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{MI score}(t_{c_i}, Q) = \sum_{x \in Q} \sum_{y \in t_{c_i}} \frac{Pr(t_{c_i}) Pr(t_{x_i})}{Pr(t_{c_i}) Pr(t_{x_i})}$</td>
</tr>
</tbody>
</table>

**Performance**

<table>
<thead>
<tr>
<th>Run</th>
<th>Rigid</th>
<th>Relax</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>R-prec</td>
<td>MAP</td>
</tr>
<tr>
<td>IASL-K-C-T-01</td>
<td>0.1118</td>
<td>0.1420</td>
</tr>
<tr>
<td>IASL-K-C-D-01</td>
<td>0.1022</td>
<td>0.1331</td>
</tr>
</tbody>
</table>

**Acknowledgement**

We would like to thank CKIP for providing us AutoTag for Chinese word segmentation.