KIPO’s MT Activity & IP5 Mutual MT Project

August 30, 2009

Young-Pyo Kim
Technical Cooperation Division, KIPO
1. KIPO’s MT Activity: *What the KIPO is doing...*

- History
- Machine Translation Services
- Quality Improvement
- Future Plan

2. IP5 Mutual MT Project

- Overview
- Technical Implementation
- Short-term Tasks
KIPO’s MT Activity:

What the KIPO is doing…
Historical Review of KIPO’s MT Services

- **2000: Introduction of a Japanese to Korean MT system**
  - Rule-based engine
  - Technical terms dictionary

- **2005: Introduction of a Korean to English MT system**
  - Pattern-based engine
  - 300,000 pattern DB, technical terms dictionary

- **2008: Introduction of an English to Korean MT system**
  - Pattern-based engine
  - 300,000 pattern DB, technical terms dictionary
KIPO’s MT Activities for Patent Documents

- KIPO internal search system integrated with MT engine
  - E2K MT system to translate US, EP, PCT patent applications written in English
  - J2K MT system to translate JP patent applications

- K2E, E2K, J2K MT services through the KIPRIS, an online patent search system that is available to the general public

- K-PION, a Korean to English MT service currently being used by 33 Patent Offices

- TDA (Trilateral Document Access) system integrated with a Korean to English MT engine for the translation of KR publications, examination reports, and notifications published by KIPO
MT Quality as Evaluated by Patent Attorneys

MT Quality Evaluation

- Evaluation period: 1~2 Times / Year
- Evaluators: 5 patent attorneys from the U.S. and Korea
- Evaluation method: randomly selected 200 sentences for evaluation

☞ The highest and the lowest scores were excluded in order to obtain an evaluation result that is not biased.
### MT Quality as Evaluated by Patent Attorneys

#### Standard of Score: criteria established by KIPO

<table>
<thead>
<tr>
<th>Score</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>5(100%)</td>
<td>Entire sentence is correctly translated, but may not be perfect</td>
</tr>
<tr>
<td>4(80%)</td>
<td>One or two words are translated incorrectly, but the essential meaning of the sentence is preserved</td>
</tr>
<tr>
<td>3(60%)</td>
<td>Keywords are correctly translated, so the sentence is partially understandable, but the essential meaning of the sentence is not preserved</td>
</tr>
<tr>
<td>2(40%)</td>
<td>One or more words or phrases are correctly translated but keywords are not translated and so the essential meaning of the sentence is not preserved</td>
</tr>
<tr>
<td>1(20%)</td>
<td>Missing sentence or translation different from original text in content</td>
</tr>
<tr>
<td>0(0%)</td>
<td>Absolutely incomprehensible</td>
</tr>
</tbody>
</table>
## MT Quality as Evaluated by Patent Attorneys

### History of MT Quality Evaluation

<table>
<thead>
<tr>
<th>Type</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2E</td>
<td>54.25%</td>
<td>76.5%</td>
<td>79.3%</td>
<td>80.6%</td>
<td>80.4%</td>
<td>81.3%</td>
</tr>
<tr>
<td>E2K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81.3%</td>
<td>82.1%</td>
</tr>
<tr>
<td>J2K</td>
<td></td>
<td>97.1%</td>
<td>96.8%</td>
<td>96.9%</td>
<td></td>
<td>97.4%</td>
</tr>
</tbody>
</table>

2005: Specialized MT system customized for patent domain

2009: MT Dictionary Upgrade
MT Quality Improvement Activities of KIPO

- **MT system customized for patent domain** (‘04~’07)
  - MT engine was originally developed by ETRI and then made more advanced by Siriussoft (2004 - 2005)
  - Established 3.2 million technical terms (2004 - 2005)
  - Established unique sentence patterns in each of the fields of patent documents, such as Title, Abstract, Claim, and Description (2005 – 2007)
  - Optimize the dictionary, sentence pattern for the examination reports (2006- 2007)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent Documents</td>
<td>54.25%</td>
<td>80.60%</td>
<td>26.35%</td>
</tr>
</tbody>
</table>
MT Quality Improvement Activities of KIPO

- **MT Dictionary Upgrade (‘08~’09)**
  - K2E: 300,000 terms incorporated into the MT dictionary
  - E2K: 300,000 terms incorporated into the MT dictionary
  - J2K: 300,000 terms incorporated into the MT dictionary

<table>
<thead>
<tr>
<th>Service type</th>
<th>~2007</th>
<th>2008</th>
<th>2009</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2E</td>
<td>3,200,000</td>
<td>300,000</td>
<td></td>
<td>3,500,000</td>
</tr>
<tr>
<td>E2K</td>
<td>3,000,000</td>
<td></td>
<td>300,000</td>
<td>3,300,000</td>
</tr>
<tr>
<td>J2K</td>
<td>1,200,000</td>
<td></td>
<td>300,000</td>
<td>1,500,000</td>
</tr>
</tbody>
</table>
① MT Quality Improvement & Feedback

- Improving translation quality by making good use of user feedbacks (unregistered word, typos, grammatical errors)
② Automatic Extraction of New Terms

- Terms which do not exist in the dictionary are automatically retrieved from patent specifications.
- Such new terms are then translated by expert human translators.
- The pairs of new terms and their corresponding translations are then added to the dictionary.

③ Regular/Irregular Spell Check

- New terms are spell-checked before they are added to the dictionary
- Existing terms are also irregularly spell-checked.
④ Quality Assessment
– Once or Twice a year

Regular System Update
– MT dictionary - *Monthly*
– MT engine - *Quarterly*
Future Plan

- Implementation of K-PION services for trademark examination history/reports and design examination history/reports (2009)


- Implementation of a foreign language translation service for KIPO examiners (2010~)
IP5 Mutual MT Project
Overview of the IP5 Mutual MT Project

Goal: To offer high quality machine translation services

Collaborative activities

Improving the quality of MTs
- Upgrading MT systems using translation error reports
- Preventing errors in original documents
- Developing direct mutual MT systems between offices

Harmonization of MT services
- Harmonization of service contents (i.e., decision on what exactly should be translated)
- Free text translation service
Technical Implementation

➢ Quality Improvement of Machine Translations
  – Upgrade MT systems through feedbacks obtained from the review of machine-translated documents (Short-term Task)
  – Control errors in original documents
  – Exchange information and share know-how on MT
  – Develop direct mutual MT systems (under agreements between the offices involved)

➢ Harmonization of MT services
  – Harmonize the service contents of the different MT services
  – Implement a free text-translation service
IP5 Short-term Tasks – *what is next?*

- Resources required for the review of sample sentences and the upgrading of MT systems
  - quality standard and guideline for conducting reviews
  - error description method
  - number of sample sentences to pick
  - sentence sampling method
  - method and frequency of MT system updates
Thank you!

E-mail: zedero@kipo.go.kr