Current Status of MT application in JPO

August 30, 2009

Japan Patent Office

Susumu BANI
Contents

- History of Patent Information Web Service
- IP Information Dissemination
- IPDL (Industrial Property Digital Library)
- AIPN (Advanced Industrial Property Network)
- Approaches for Improved Translation Result
- Optimization Plan
- Future Image of Machine Translation
History of Patent Information Web Service

- Apr. 1996: JPO Official Website launched
- Mar. 1999: “Industrial Property Digital Library (IPDL)” started on the website
- Mar. 2000: “Full-text Translation” service released
- Feb. 2003: The Official Website renewed
- Oct. 2004: IPDL transferred to National Center for Industrial Property Information and Training (INPIT) Website
JPO : Policy Making on IP Information Dissemination

INPIT : Provision of Basic Primary Information

Databases with Machine Translation

**IPDL (Industrial Property Digital Library)**
Offering 71.0 million IP information with search function in Japanese and MT English to public users.

**AIPN (Advanced Industrial Property Network)**
Offering patent application information and dossiers in MT English to foreign IP offices.
IPDL (Industrial Property Digital Library)

**English Databases on IPDL**

1. Patent Databases
   - **DB with Machine Translation Service**
     - Patent & Utility Model Gazette DB
     - Patent & Utility Model Concordance
     - FI/F-term Search
   - **Human Translation Database**
     - Patent Abstracts of Japan (PAJ)

2. Design Database
   - **DB with Machine Translation Service**
     - Design Gazette DB

3. Trademark Database
   - **Human Translation Database**
     - Japanese Trademark Database
     - Japanese Figure Trademarks
     - Japanese Well-Known Trademark List of Goods and Services
Full-text Patent Publication

- All information (except drawings) is translated by machine translation system.

CLAIMS

[Claim(s)]

[Claim 1]

In a photovoltaic power generation apparatus provided with a solar cell array which carried out series parallel connection of two or more solar cell modules, and formed them, and a power conditioner which has a converter which carries out conversion into ac of the direct-current generated output from this solar cell array, an actinometer is attached so that said solar cell array, an installation direction, and installation ***** may become the same, and a temperature sensor is
AIPN (Advanced Industrial Property Network)

- Japanese examination-related information is available for inspection in English by means of Japanese-English machine translation
- Available for foreign IP offices
- Free of charge

Advantageous effect

1. Reducing burden on examiners in foreign IP offices
2. Accelerating quick examination and acquisition of the right
File Wrapper Document

- Examination-related information is available in English through machine translation system.

Application information with legal status

Dossier List

Office actions
Search reports
Written argument
Examiner’s note
etc.

Patent application number 2001-123456

1. 2004.01: Request for a Patent
2. 2004.01: Description
3. 2004.01: Drawings
4. 2004.01: Abstract
5. 02.10.2001: Ex Officio Correction of Defect of Abstract
6. 23.04.2002: Request for Examination (Others’ Request)
   -- -- ----- : -- Certified or Appended Information
7. 28.09.2002: Written Notice (Others’ request)
8. 22.07.2003: Search Report by Designated Searching Authority
10. 02.09.2003: Notification of Reasons for Refusal
11. 04.11.2003: Written Argument
12. 04.11.2003: Written Amendment
13. 21.04.2004: Decision to Grant a Patent

[Translation done]
Application for patent 2001-123456
Filing date: 12/02/2000
Priority date: 12/02/2000
Inventor: NISHIKAWA, Shigeki
Address: KAM, Nishihara 1041 300

(a) The claimed invention is refused for the reason mentioned below:

Reason:

The description and claims do not meet the requirements of the Patent Law Section 28(2) for the reason that the claimed invention(s) is(are) not clearly described and made by persons who have common knowledge in the technical field to which the claimed invention(s) relates.

Claim(s):

1. In the flexible copper-clad sheet with which the copper layer was formed on the flexible polymer base material, the independent minute metal membrane at homogeneity, the part which is not dotted with the metal membrane with the minute surface of a flexible polymer base material has average depth 0.01 - 2.0micrometer impression structure from the surface, and covers a minute metal membrane and impression structure on the surface of (C) flexibility polymer base material. The flexible copper-clad sheet characterized by forming the intermediate metal layer and the copper layer in this order.
Pre-modification for machine translation

Pre-modification is necessary to reduce mistranslation arise from format error.

examples

1. Replacement (elimination) of tag information
   Unnecessary line feed: \n\n\n\n → \n\n<BR>

2. Deleting unnecessary spaces
   「引用文献等一覧」→ 「引用文献等一覧」
   “The list of cited documents”

3. Pre-formatting
   Document number: “yyyyxxxxxx” → “yyyy-xxxxxx”
   Avoiding Consequent quotation: “bcd)(xyz” → “bcd)xyz”
Collection of unknown words
Translation result (IPDL/AIPN)

Refer to the description of "****".
The invention in "****" is not clear.

Unknown word is displayed as "****".

IPDL server
AIPN server
Unknown Words log
Collection of mistranslation logs

Adding the meaning, part of speech etc. to the unknown words based on frequency.
Example:
ニッケロセニル-iso-プロピル基 → nickelocenyl-iso-propyl
重合体カテキン → polymer catechin

Registration of the unknown words
(5,000 words/year)
Provision of dictionary data at marginal cost
(65,000 words : as of March 2009)
JPO has received mistranslation feedbacks from AIPN users. (EPO and USPTO)

Feedback process (once a month)

1. Verification of repeatability
2. Checking up original sentences and adding appropriate meanings
3. Review of added meanings
   by referring PAJ, IPDL and some technical term dictionaries
4. Assessment of side-effect caused by additional words registration
5. Authorization by JPO and INPIT
6. Registration in user dictionary
   (It takes 2 months to complete the process)
Approaches for Improved Translation Result

Research for accurate translation result

**Brief summary of the research**

**Numbering**

Some types of numbering cause mistranslation.

**example**

<table>
<thead>
<tr>
<th>Original sentence</th>
<th>この出願は、特許請求の範囲の記載が下記の点で、特許法第36条第6項第2号に規定する要件を満たしていない。</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation result without numbering</td>
<td>This patent application does not fulfill the requirement of the provision of Patent Law Section 36(6)(ii) in terms of the item(s) listed below regarding the statement of the scope of claims.</td>
</tr>
<tr>
<td>Translation result with “理由1__” (Reason1__)</td>
<td>Reason 1 -- this application is not satisfying the requirements which a description of Claims is the following point and are specified to Patent Law Section 36(6)(2).</td>
</tr>
</tbody>
</table>

**Usage of variable for fixed expressions**

Standardizing numbers and fixed expressions in the documents by using variable.

**Analysis of free description part**

Finding frequent patterns by analyzing last few words, statistically. (extending word by word)

**Extraction of phrases and technical terms**

1,000 frequently-used phrases and technical terms are examined.

**examples**

- Claim 1, Claim 2 → Claim “n”
- July 7, 2009 → “n”

**examples**

- 不明瞭である。が 不明瞭である。
- 技術的範囲 が 不明瞭である。
- その 技術的範囲 が 不明瞭である。

Analyzing by extending word by word
## Approaches for Improved Translation Result

### Replacement of MT engine

#### Technical term & IP dictionaries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Term</td>
<td>1.83 million</td>
<td>2.50 million</td>
</tr>
<tr>
<td>dictionaries</td>
<td>words</td>
<td>words</td>
</tr>
<tr>
<td>Total number of IP</td>
<td>100,000 words</td>
<td>200,000 words</td>
</tr>
<tr>
<td>terms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Translation memories

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial Property</td>
<td>6,000</td>
<td>20,000</td>
</tr>
<tr>
<td>terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic terms / Business terms</td>
<td>150,000</td>
<td></td>
</tr>
<tr>
<td>Technical Terms:</td>
<td>400,000</td>
<td></td>
</tr>
<tr>
<td>(Chemical / Medical, Biotech. / Machine, automobile / electronics)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Excerpt from brochure of the product
Optimization Plan

- Developing the latest search environment for timely and high-quality examinations
- Providing an environment of utilizing patent information for R&D or management strategy in universities and companies
- Streamlining search systems
- Reducing operational expenses
- Improving security and reliability

Sharing information on intellectual property and academic papers, Developing a seamless search environment

**Public User**
- Utilization of the same environment as examiners’

**University/research institution**
- Data held by the university / research institute
- Internal System
- Academic papers

**company**
- Company data
- Internal system
- Manuals and catalogs, etc.

**Shared Innovation Infrastructure**
- Reference under a license

**Access to Patent Information**
- Chinese / Korean Patent Documents (Machine Translation)
- Japanese / Foreign Patent documents

**Non-patent documents**
- Manuals and catalogs, etc.
- Information Dissemination Policy
Research for “Cross-lingual retrieval system”

JPO has a plan to establish “Cross-lingual retrieval system” which enable to search Chinese and Korean documents.

- Evaluation of the effectiveness of the two translation methods (Keyword translation, Contents translation)

**Keyword translation**

Query in Japanese is translated into Chinese or Korean in each transaction for search. Patent documents, written in the original language, are searched by using translated query.

- Evaluation of translation results (Korean-Japanese, Chinese-Japanese)

<table>
<thead>
<tr>
<th>Korean-Japanese</th>
<th>&gt;&gt; Translated sentences are fairly understandable. (practicable level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese-Japanese</td>
<td>&gt;&gt; Further development is needed. (especially, enrichment of the dictionaries)</td>
</tr>
</tbody>
</table>
It’s necessary for the JPO to make an effort to develop MTs of Chinese-JP and Korean-JP.
IP5 Foundation Projects

Mutual Machine Translation (lead: KIPO)

Aim: To help the offices overcome the language barrier of patent information and allow greater access to each other’s patent information.

IP5 Foundation Projects

- Common Documentation Database (lead: EPO)
- Common Approach for a Hybrid Classification (lead: EPO)
- Common Application Format (lead: JPO)
- Common Access to Search and Examination Results (lead: JPO)
- Common Training Policy (lead: KIPO)
- **Mutual Machine Translation (lead: KIPO)**
- Common Rules for Examination Practice and Quality Control (lead: SIPO)
- Common Statistical Parameter System for Examination (lead: SIPO)
- Common Approach to Sharing and Documenting Search Strategies (lead: USPTO)
- Common Search and Examination Support Tools (lead: USPTO)
Thank you