Stream 2: Paradigm change in Searching Asian documentation

New Search Tech. applied to KIPRIS

4. 20. 2018  |  Myung Hun, Moon
Contents

I Introduction of KIPRIS
II KIPRIS Search Function
III KIPRIS - Cross lingual search
Introduction of KIPRIS

KIPRIS is the most comprehensive free intellectual property information search service in Korea on the Internet

<table>
<thead>
<tr>
<th>IP Information Search Service</th>
<th>Non-Patent Information Search Service</th>
<th>Customized Service</th>
<th>IP Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domestic</strong></td>
<td><strong>Prize-Winning ideas</strong></td>
<td><strong>IP Contents</strong></td>
<td></td>
</tr>
<tr>
<td>Patent &amp; Utility Model</td>
<td>Standard Technology Information</td>
<td><strong>TODAY KIPRIS</strong></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td>Webzine-Newsletter</td>
<td></td>
</tr>
<tr>
<td>Trademark</td>
<td></td>
<td>Today's IP Keyword</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td></td>
<td>(Patent / Design / Trademark)</td>
<td></td>
</tr>
<tr>
<td>KPA</td>
<td></td>
<td>Recent patent Issue</td>
<td></td>
</tr>
<tr>
<td><strong>Overseas</strong></td>
<td></td>
<td>GUIDE</td>
<td></td>
</tr>
<tr>
<td>Foreign Patent</td>
<td></td>
<td>Tutorial service</td>
<td></td>
</tr>
<tr>
<td>- 23 countries PCT, EPO</td>
<td></td>
<td>Video manual</td>
<td></td>
</tr>
<tr>
<td>Foreign Trademark</td>
<td></td>
<td>Easy-to-follow KIPRIS GUIDE</td>
<td></td>
</tr>
<tr>
<td>- JP, AU, CA US, OHIM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign Design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- JP, US, WIPO</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction of KIPRIS

Korean website: http://www.kipris.or.kr
- access to Korean IPR data and foreign patent data
- trial data search (IP Trial and Appeal Board, Patent court, Supreme court)
- sentence search service, similarity search function
- cross lingual retrieval for search foreign data
  (Keyword translation, Korean → English)

English website: http://eng.kipris.or.kr
- access to Korean IPR data only
- linkage between a publication and corresponding trial record (not searchable)
- cross lingual retrieval for searching Korean data
  (Keyword translation, English → Korean)
## KIPRIS Search Function - Sentence search

<table>
<thead>
<tr>
<th>Extracting keywords</th>
<th>Comparing similarity</th>
<th>Extracting stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyzing patents and extracting keywords with TEXT MINING SOLUTION</td>
<td>Index design, operating similarity of keywords, applying ranking algorithm</td>
<td>Extracting stop words with patent format analysis</td>
</tr>
<tr>
<td>- extracting from a typical text data considering frequency of specific words or phrases, sentence length, space between keywords, statistic model</td>
<td>- Designing index with basic meta information (e.g., number, technique classification, date, personal information)</td>
<td>- Stop word from analyzing KIPO publication format (XML, SGML)</td>
</tr>
<tr>
<td></td>
<td>- Indexing keywords for improving speed of reviewing process</td>
<td>- Extracting stop word from index data with stop word recommending solution (XML Tag, repetitive idiom, suffix, prefix)</td>
</tr>
<tr>
<td>Extracting keywords for the feature index representing atypical text document</td>
<td>Improving accuracy by using meta information and similarity operation algorithm</td>
<td>Improving quality of key keywords extraction by eliminating words for document form</td>
</tr>
</tbody>
</table>
KIPRIS Search Function - Keyword extraction

1. Keyword Extraction
2. Topic Analysis
3. Custom Field Analysis

Information Radius
It refers to similarity scale based on Korean divergence. Dissimilarity is calculated from each keywords amount of information between two documents.

\[ IRAD(P, Q) = \sum_i p_i \log \frac{p_i}{(p_i + q_i)/2} + \sum_i q_i \log \frac{q_i}{(p_i + q_i)/2} \]

Index search
It refers to search technique improving search function by extracting word positioning value. It is exclusive to types of original data (e.g., DB, file, webpage) and a location-based search appropriate in real-time search.

analysis result
- relational analysis
- analysis on co-occurrence
- space between keywords

interwork search engine
- similarity ranking
- search result clustering
- exclusion, inclusion filtering

A typical text data (publication)
KIPRIS Search Function - Algorithm for similarity comparison

Algorithm for detecting duplication

Text matching algorithm

- Information
  - Invention title
  - Abstract
  - Claim
  - Specification
  - Personal info
  - IPC

- Comparing meta info
- Similarity measurement

- Extract keyword
- Weight by field
- Detect similarity & duplications

Keyword extraction algorithm

- Requesting review
- Comparing similar publication
- Similarity measurement

- Extract key technical term
- Compare similarity

1. Use function of extracting stop word
   - Extract stopword from publication indexing data
     - Extract stopword from high-frequency word in publication

2. Analyze meta data to use
   - Analyze meta data to use
     - Analyze each of related patent publication
     - Reset weight of technical term

3. Design search algorithm using meta information
   - Design integrated similarity analysis algorithm
     - Design index for using raw patent publication but meta data
KIPRIS Search Function - Stopword dictionary

1. Patent publication
2. Select stopwords
3. List of stopwords
4. Stopwords refinery
5. Stopwords dictionary

- 20,000 records
- Indexing similarity/duplication
- Solution of recommending stopwords
- Select stopwords
- List of stopwords
- De-duplication

<table>
<thead>
<tr>
<th>Division</th>
<th>Count</th>
<th>Korean</th>
<th>Total</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format to extract</td>
<td>18</td>
<td>976</td>
<td>994</td>
<td>Partitioning compound noun</td>
</tr>
<tr>
<td>Index to extract</td>
<td>71</td>
<td>1,780</td>
<td>1,851</td>
<td>Select mechanical extraction list (remove English medical term, Korean technical term)</td>
</tr>
<tr>
<td>Total</td>
<td>89(0)</td>
<td>2,752(4)</td>
<td>2,841(4)</td>
<td>Review format, index</td>
</tr>
</tbody>
</table>
KIPRIS - Sentence search

Sentence search for ① KR patent, utility model, ② JP patent (translated in Korean) ③ Contest prize ideas
Search Field

» **Number** Select one of search interfaces, Publication number(OPN), Registration number(GN) and Application number(AN)

» **Text** Copy and Input abstracts or the first claim of patent document (Maximum 20,000 characters)

Copy the abstract from document and paste in text search interface. Then click SEARCH button.
KIPRIS - Sentence search

Search with abstract ⇒ Patents listed in order of descending similarity score

Smart TV control system capable of controlling TV with simple movement and an operation method thereof (Similarity: 82.1)

Smart TV control system capable of controlling TV with simple movement and an operation method thereof (Similarity: 76.8)

Apparatus, method and computer readable recording medium for controlling smart TV. (Similarity: 69.0)
KIPRIS - Cross-lingual search

How it works?

- Use English data as much as possible in Database
  - Title of invention in English submitted by applicant
  - Name of applicant, Inventor, Agent
  - Address of applicant, Inventor, Agent

- Match Data between English and Korean patents
  - There is no dictionary matching personal information between KR ← ENG

- Cross-Lingual Retrieval is useful for Keyword Search
  - Search Full-text, Abstract and Claims
  - More reliable for Nouns, terminologies describing certain things
  - One adjective, adverb can have various forms in Korean depending on noun
    - Ex.) Red: 레드, 붉은, 빨간, 붉다
KIPRIS - Cross-lingual search

Only “Phone” in Eng.

Cross-lingual retrieval

“Phone+전화”