

韓国特許調査手法の検討:

ハングルキーワード検索を併用した網羅性向上手法の検討

○田畑文也¹⁾, 前田佳治²⁾, 伊藤徹男³⁾, 中西昌弘⁴⁾

富士フイルム(株)¹⁾, 東亜合成(株)²⁾, アジア特許情報研究会³⁾,
オリンパスメディカルシステムズ(株)⁴⁾

〒421-0396 静岡県榛原郡吉田町川尻 4000 富士フイルム株式会社¹⁾

Tel: 0548-34-5401 FAX: 0548-32-8286

E-mail: fumiya.tabata@fujifilm.com

Study of the Korean Patent Search : Study of Completeness Improvement using Korean Language Keyword Search.

TABATA Fumiya¹⁾, MAEDA Keij²⁾, ITO Tetsuo³⁾, NAKANISHI Yasuhiro⁴⁾

Fujifilm Corporation¹⁾, TOAGOSEI CO., LTD²⁾, Asia Patent Information Society.³⁾,
OLYMPUS Medical Systems Corp.⁴⁾

Fujifilm Corporation¹⁾, 4000, Kawashiri, Yoshidacho, Hibara-gun, Shizuoka, Japan

Phone: +81-548-34-5401 Fax: +81-548-32-8286

E-mail: fumiya.tabata@fujifilm.com

【発表概要】

韓国特許を網羅的かつ効率的に調査する手法を検討した。マルチカントリーデータベースでの、英語キーワードのみを用いた韓国特許検索では、特にクレーム以下の情報がない、または機械翻訳情報のため、漏れが発生しやすい。これらを韓国特許庁系のデータベースである KIPRIS などを用い、ハングルキーワード検索を併用して検索することにより、網羅性を向上させる手法について検討した。また、特にハングルキーワード検索の検索ロジック、キーワード抽出方法、注意点、問題点についても解析した。また、ハングルキーワードの解析手法として、正規表現を利用したハングル対応の検索ツールを独自に作成し、効率的にハングルキーワードを抽出する方法も検討した。

【キーワード】

韓国特許, 網羅性, ハングル, 韓国語, キーワード検索, ワードサーチ, 検索ロジック, 完全一致, 分かち書き, 正規表現

1. はじめに

近年、電気・電子・半導体分野を始めとする先端技術分野での韓国の技術的向上には目覚ましいものがあり、韓国特許を詳細に調べる必要性が増している。しかし、韓国特許庁からの英語情報は、要約までの書誌情報しか提供されていない。よってクレーム以下の情報を含め詳細に調べるには、マルチカントリーDB（データベース）などを用いた英語検索だけでなく、ハングルキーワードを用いた原語検索補完などにより、網羅性の向上が必要である。

英語キーワード検索とハングルキーワード検索を併用した韓国特許調査手法については、2年前のINFOPRO2012にて報告¹⁾した。しかし、その後のハングルキーワード検索の検証で、ハングル表記の分かち書きを含む表記割れの問題と、韓国特許庁傘下のKIPRIS (<http://engpat.kipris.or.kr/engpat/>)などのDBにおいて、ハングルキーワード検索の検索ロジックが複雑で、我々の想定以上に注意が必要なことがわかった。

また、ハングルキーワードを、独自作成の正規表現²⁾検索ツール(RExツール)を用いて、効率的に解析することができ、今まで主に手作業であったハングルキーワード抽出が、より効率的にできることがわかり、ここに報告するものである。

2. 韓国特許の動向とデータの流れ

韓国特許庁公表の出願統計データをまとめたものを図1に示す。ここ10年では日本での出願数は2割強以上の減少に対し、韓国は約5割の増加をしている。

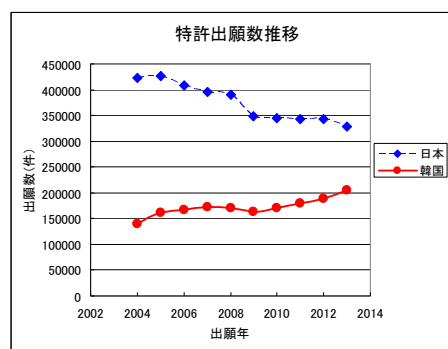


図1. 韓国特許出願推移

韓国特許庁の統計データを基に、特許の内国人出願率の推移をグラフ化し、図2に示す。8割近くが内国人出願である。

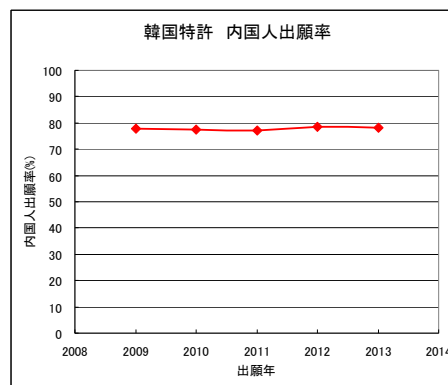


図2. 韓国特許 内国人出願率

さらにWPIを用いて、韓国特許の韓国1ヶ国出願率を調べると、図3のようになり、約6割が対応する外国特許のファミリーがない特許であるため、クレーム以下の精度ある英語情報が十分でないことが推定できる。

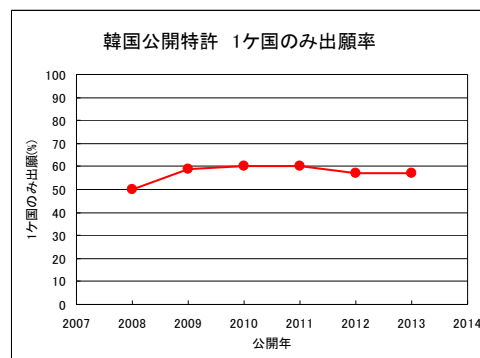


図3. 韓国公開特許 1ヶ国のみ出願率

また、IPC 付与数について、KIPRIS を用いて、付与されている IPC を統計処理(隔年ごと各 n=2000)したものを図 4 に示す。2006 年以前公開のものは IPC が 1 個しか付与されず、検索時には、キーワード検索補完が必須である。

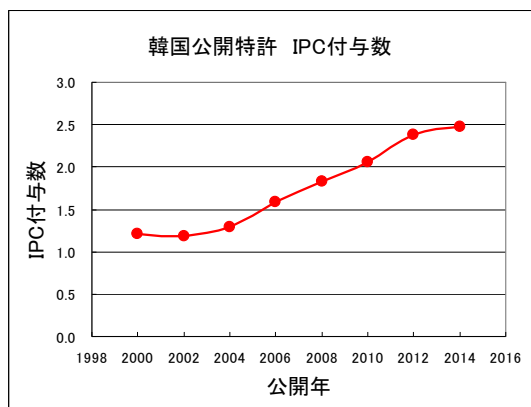


図 4. 韓国公開特許 IPC 付与数

なお、韓国特許については、韓国特許庁(韓国特許情報院)から英文で供給されるのは、要約までであり、そのデータ入手ルートを図 5 に示す。韓国特許庁などから直接データを入手した場合でも、クレーム以下はハングルのみで、それを DB ベンダー等が独自に機械翻訳した英文ものしか存在しないのが現状である。

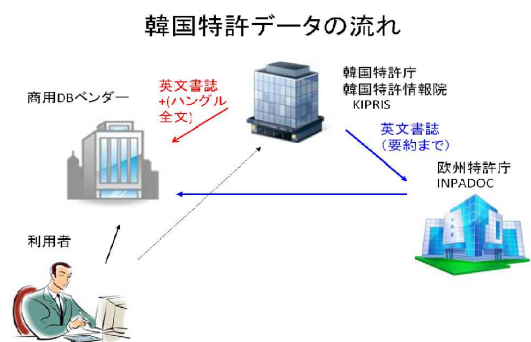


図 5. 韓国特許データの流れ

3. 韓国特許庁系 DB KIPRIS

KIPRIS は韓国特許庁傘下の韓国特許情報院が提供する DB で、無料で使用可能な DB であり、英語メニュー版とハングルメニュー版があり、今回の検証では英語版を用いた(図 6, 図 7)。

(無料) ユーザー登録すると、ダウンロード機能が向上したり、マイフォルダ機能なども使えるようになる。



図 6. KIPRIS DB 英語メニュー版検索画面例



図 7. KIPRIS DB 検索結果画面例

4. 検討内容

検証に使用した DB を以下に示す。韓国特許庁系 DB の KIPRIS (英語メニュー版を使用)、及び韓国 WIPS 社のハングル DB である WIPS ON、Shareresearch((株)日立製作所), PatBase (RWS グループ社)を用いて検証した。主な韓国特許 DB について、機能例を表 1 にまとめた。

表 1. 主な韓国特許 DB と機能例

DB名	検索原語	機械翻訳検索	クレーム	備考
KIPRIS	ハングル	○	○	
KPA	英語	×	×	要約まで
WIPS ON	ハングル	×	○	ハングルのみ
PatBridge	日本語 & ハングル	○	○	日本語検索 & 表示可
Shareresearch	英語	○	○	
Total Patent	英語	○	○	
PatBase	英語 & ハングル	○	○	
Orbit.com	英語 & ハングル	○	○	

4-1) 英語クレーム検索の精度検証

KIPRISにて、化学(的)機械(的)研磨の最も頻繁に使用されるハングル表記のトップ4表記(表2)を用いて、クレーム検索を実施し、対応ファミリーの無いものを178件を抽出し、機械翻訳クレームの収録されたDB-Aで、英語クレーム検索した結果と突き合わせると(表3)、該当する英語表記のヒットは約53%のみであった。

表2. 化学(的)機械(的)研磨のハングル表記トップ4

ハングル表記	日本語表記
화학적 기계적 연마	化学的_機械的_研磨
화학 기계적 연마	化学_機械的_研磨
화학기계적 연마	化学機械的_研磨
화학 기계 연마	化学_機械_研磨

(注) - はスペース(分かち書き)を示す

表3. DB-A 機械翻訳英語クレーム検索結果

クレーム検索式	ヒット	ヒットせず
(chemical* or electrochemi*) *(mechanical*) *(polish*)	94件 (53%)	84件 (47%)

正しく訳されていないものを確認すると、化学(的)の37%が英語にはない単語のchemicに訳されていたり、機械(的)の部分がinstrumentやmachineなどに訳されているものが多かった。

4-2) KIPRIS のインデキシング確認

分かち書きの無いハングル表記の例で、層間絶縁膜(층간절연막)を取り上げ、2000年～2013年公開特許のタイトルを母集団として、インデキシングロジックを確認し、表4にその結果を示す。

表4. 層間絶縁膜のインデキシング

タイトル検索式	意味	ヒット
TL=[층간절연막]	層間絶縁膜	166
TL=[층간절연막!절연막]	層間絶縁膜 not 絶縁膜	165
TL=[층간절연막!절연]	層間絶縁膜 not 絶縁	164
TL=[층간절연막!층간]	層間絶縁膜 not 層間	163

層間絶縁膜は、絶縁膜や、絶縁または層間でもヒットせず、これを漏れ無く検索するためには、必ず“層間絶縁膜”という完全一致のキーワードを含める必要がある。なお、KIPRISには前方一致のワイルドカードの機能はない。

4-3) KIPRIS での分かち書き

ハングル表記での特徴に分かち書きがある。例えば層間絶縁膜(층간절연막)の表記を例に考えると、層間_絶縁_膜と2ヶ所の分かち書きされる可能性のある部分が存在する。図7に2000年～2013年公開特許のタイトルを母集団として、分かち書きのない層間絶縁膜(층간절연막)と、分かち書きされた層間_絶縁膜(층간_절연막)との検索結果を表5に示す。

表5. 分かち書きの有無での検索結果の違い

タイトル検索式	意味	ヒット
TL=[층간절연막]	層間絶縁膜(分かち書きなし)	166
TL=["층간 절연막"]	層間_絶縁膜(分かち書き有り)	257
TL=["층간 절연막" !층간절연막]	層間_絶縁膜(分かち書き有り) not 層間絶縁膜(分かち書きなし)	256

KIPRISのハングルキーワード検索では、分かち書きされたものと、分かち書きされていないものは別のワードとして検索されるため、分かち書きを含め、完全一致させ検索する必要があることが分かった。

4-4) KIPRIS ハングル同音異表記

電池については、日本語でもバッテリーとも表記されるが、ハングルでも外来語をそのまま用いることが多々ある。ノイズ表記を入れないようにするため、電池及び太陽電池のIPC分類であるH01M、H01L31の中で、KIPRISにて以下の検索式にて母集団を作成し解析した。(表6)

・KIPRIS 検索式

IPC=[H01M+H01L31]*OPD=[20050101~20131231]*TL=[battery+batteries] (10907 件ヒット)

表 6. バッテリー タイトル表記率(%)

バッテリー ハンゲル表記	電池IPC タイトル	限定なし タイトル
배터리	99.3	84.9
밧데리	0.5	13.7
베터리	0.1	0.6
바테리	0.1	0.8

ノイズをさけるため、電池関連の IPC でキーワードの抽出を行ったが、2 位以下の頻度は目視で捉えるには困難なほど少なかったが、分野を限定しないタイトル検索では、表記割れの頻度は拡大し、2 位以下の表記も無視できないレベルになった。

4-5) 正規表現検索ツールによる解析

サンプル検索によるハンゲルキーワード抽出を効率化するため、ワーキングメンバーの中西氏に正規表現検索ツール(以下、REx ツール)を作成してもらい、ハンゲルの文字列を分かち書きのスペースを含め解析できるようにした。

解析用の母集団として、KIPRIS で以下の検索式で、カラーフィルターのタイトル母集団 915 件を作成し、解析した結果を表 7 に示す。

・KIPRIS 検索式

IPC=[G02B5/20]*OPD=[20000101~20131231]*TL=[(color+colour)*filter] (機械翻訳有り 915 件ヒット)

・REx ツールによる解析

REx ツール解析用検索式

[색컬갈].{2,5}[터층용]

(色 or カ(ハンゲルまたは英語表記))

2 文字~5 文字以内

(ター or 層 or 用)

表 7. カラーフィルター REx ツール解析結果

ハンゲル表記	意味	頻度(%)
컬러 필터	カラー フィルター	39.7
컬러필터	カラーフィルター	36.2
컬러필터용	カラーフィルター用	8.7
컬러 필터용	カラー フィルター用	6.8
칼라 필터	カラーフィルター	2.7
칼라필터	カラーフィルター	2.5
칼라 필터용	カラー フィルター用	0.8
칼라필터용	カラーフィルター用	0.5
색변환 필터	色変換 フィルター	0.4
색 칼라 필터	色 カラー フィルター	0.2
컬러필터층	カラーフィルター層	0.2

REx ツールにより、ハンゲルキーワードが出現頻度とともに効率的に抽出された。なお検索時には、これらの表記を分かち書き、接尾語的な“用”や“層”を含め完全一致的な形を考慮したキーワード検索が必要である。

4-6) WIPS ON による検索

商用 DB の WIPS ON(図 8)では、ハンゲルキーワード検索時にワイルドカード検索ができ前方一致検索可能である。実際にシルセスキオキサンの例で確認した結果を表 8 に示す。(2000 年~2013 年公開特許にて確認)



図 8. WIPS ON ハンゲル検索例

表 8. WIPS ON ワイルドカード(*)機能検証

WIPS ON タイトル検索式	意味	ヒット数
(실세스퀴옥산).TL	シルセスキオキサン	62
(실세스퀴옥산*).TL	シルセスキオキサン*	67
(실세스퀴옥산계).TL	シルセスキオキサン系	3
(실세스퀴옥산계 not 실세스퀴옥산*).TL	シルセスキオキサン系 notシルセスキオキサン*	0

4-7) PatBase による部分一致検索

PatBase においては、ハングル検索時には部分一致検索が可能で、実際にハングル要約のシルセスキオキサンを検索しても、接頭語がついたポリシルセスキオキサンや、接尾語がついたシルセスキオキサン系のハングル表記も部分一致で検索できている。(表 9)

表 9. PatBase ハングル要約検索 (9.24 確認)

ハングル要約検索式	意味	ヒット数
[KR]: AB=(실세스퀴옥산)	シルセスキオキサン	385
[KR]: AB=(실세스퀴옥산계)	シルセスキオキサン系	22
[KR]: AB=(폴리실세스퀴옥산)	ポリシルセスキオキサン	48
[KR]: AB=(말향족실세스퀴옥산)	芳香族シルセスキオキサン	2
1 or 2 or 3 or 4	上記総和(増えず)	385

5. 考察とまとめ

韓国特許の場合、クレーム以下は英語情報があっても、ファミリー情報がない場合は機械翻訳のみで、その翻訳精度は高くない場合が多い。最も網羅性の高い手法はハングルキーワードを併用した検索で、マルチカントリーDBによる英語検索を補完する手法と考える。

だが実際に、KIPRISなどで実際にハングルキーワード検索してみると、

- ・分かち書き表記の問題
- ・完全一致のワード検索の問題
- ・表音系の場合の異表記の多さ

など、日本のサーチャーが日頃行っている検索では遭遇しないような問題に直面する。辞書ツール等(表 10)もあるが、ハングルに対応したものは少なく、

表 10. ハングル対応辞書ツール例

ツール名	URL
NAVER辞書(英語)	http://dic.naver.com/
日韓化学用語辞書	http://gigadict.com/cgi-bin/pro_che/J/dic-cheJK.cgi?
weblio	http://kijk.weblio.jp/
グーグル翻訳	http://translate.google.co.jp/
WIPO CLIR	http://patentscope.wipo.int/search/en/clir/clir.jsf?new=true

むしろ KIPRIS ハングルメニュー版にあるような拡張クエリツール(図 9)の方が、異表記の表記例が多い場合もある。



図 9. KIPRIS ハングルメニュー版 拡張クエリ

ハングルキーワードを効率的に抽出する手法として、REx ツールによる手法を紹介したが、このツールの使いこなしにもハングルの知識もある程度必要になるなど、現時点では、万人向けとは言い難い。

WIPS 社の PatBridge も本年 9/22 よりハングル検索に対応したとの情報もあり、12 月の発表時には、他のハングルキーワード抽出方法も含めて紹介し、より多くの手法を紹介したいと考えている。

「謝辞」

最後に、本報告は 2014 年度の「アジア特許情報研究会」のワーキングの一環として報告するものであり、報告者として名前を挙げさせていただいた他に、他のワーキングメンバー、テマリーダーの皆様には様々な協力をしていただきました。ここに改めて感謝申し上げます。

6. 参考文献

- [1] 田畑文也 他 : 網羅性のある韓国特許調査(第 9 回情報プロフェッショナルシンポジウム、2012/10)
- [2] 正規表現
<http://www.mnet.ne.jp/~nakama>
http://www4.ocn.ne.jp/~kaerume/k2e/regex_1.html