

中国・台湾および韓国特許庁データベースの全文検索機能とその活用:

○伊藤徹男¹⁾, 沖 祥嘉²⁾
日本パテントデータサービス¹⁾, 東ソー²⁾
〒105-0003 東京都港区西新橋 2-8-6
Tel: 03-3580-8021 FAX: 03-5512-7810
E-mail: ito@jpds.co.jp

Full-text search capability and the use in China, Taiwan, and South Korea Patent Office data base.:

ITO Tetsuo¹⁾, OKI Yoshitaka²⁾
Japan Patent Data Service(JPDS)¹⁾, Tosoh Corporation²⁾
2-8-6, Nishi-shinbashi, Minato-ku, Tokyo 105-0003 Japan
Phone: +81-3-3580-8021 Fax: +81-3-5512-7810
E-mail: ito@jpds.co.jp

【発表概要】

韓国特許データベース(KIPRIS)、中国特許データベース(CNIPR)に次いで2008年7月には台湾特許庁の特許データベースでも明細書全文からの検索(要約だけでなく特許請求の範囲(クレーム)や発明の詳細な説明(明細書全文)も)が可能となった。

これら3特許庁関係特許データベースの書誌の収録については、既にいくつかの報告があるが、抄録やクレーム、明細書全文の収録状況については未だ報告されていない。

本報告では、これらのデータベースにおける抄録・クレーム・明細書全文の収録状況について報告するとともに、収録が遅れており、収録のタイムラグも大きい商用データベースのアジア3特許庁における最新特許情報収集の問題点と各特許庁関係特許データベースの活用について提案する。

【キーワード】

特許データベース, アジア特許, 中国特許, 台湾特許, 韓国特許, 全文検索,

1. はじめに

ここ数年、中国、韓国、台湾の特許庁関係のデータベースが相次いで改善され、使いやすいものとなってきた。中国特許庁データベースを管理・運営している知識産権出版社による特許データベース「CNIPR」の中国語データベースでは数年前にクレームや明細書全文からの検索が可能となったが、昨年になって英語データベースのクレームや明細書全文を機械翻訳ではあるが英語で閲覧可能とした(但し、現時点では翻訳データに欠落あり)。

韓国特許データベース(KIPRIS)においては、2007年にハングルデータベースでクレームや明細書全文からの検索がハングルだけでなく英語でも可能となり、さらに昨年には、検索結果の「発明の名称」および「審査経過情報」部分に英語を併記するなどの改善があった。

そして昨年7月、台湾特許庁中国語データベースにおいても、クレームや明細書全文からの検索が可能となった。

他方、従前からワールドワイドな特許情報を提供してきた英語商用データベースでは、アジアの特許情報の欠落、収録のタイムラグの問題について、日本企業のアジアに対する関心の高さに押され、特に中国に関して収録のタイムラグの改善が見られるものの、韓国や台湾のデータについては収録のタイムラグが大きい(台湾においてはタイムラグ1年以上、という状況である)。

これら3特許庁関係特許データベースの書誌の収録については、既にいくつかの報告があるが1),2),3)、抄録やクレーム、明細書全文の収録状況については未だ報告されていない。

本報告では、これらデータベースの抄録やクレーム、明細書全文の収録状況を検証するとともに、英語商用データベ

ースでは得られない直近の最新情報を各特許庁データベースから入手し、社内に配信する、いわゆるSDIとして活用する方法についても提案する。

2. 検証方法

商用の英語データベースには、「AB/FA」や「AB=YES」などの抄録収録の有無を確認するコマンドが用意されているが、各特許庁のデータベースにはそのようなコマンドはない。したがって、極めて初歩的ではあるが、抄録、クレームおよび明細書全文に最も頻出するワード(英語データベース)および原語(原語データベース)を予め調べ、それらを基に各データベースにおける各年度の収録状況を調べた。

2-1. 収録データの検証

まず、各特許庁データベース(表2-1)の抄録を対象に出現頻度の高そうなワード(英語データベース)および原語(原語データベース)を試行錯誤的に抽出し、それぞれ2007年度の収録状況を調べ、書誌収録数を100として収録率を求めた。収録率の高いワードおよび原語の集合を本検索用ワードおよび原語として用いた。

次いで、本検索として予備検索で得られた高頻出度のワード集合を用いて各年度の抄録、クレーム、明細書全文の収録数を調べ、書誌収録数を100として収録率を求めた。

ワード集合の文字数が多いほど書誌収録数に近くなり収録率も100%に近づくことになるが、クレーム中や明細書全文からの検索では、文字数が多くなると検索に時間がかかり、場合によってはタイムアウトとなって検索できないこととなるため、抄録検索用のワード集合より若干少ないワードでクレームや明細書全文の収録を調

べることとした。したがって、収録率が97～98%以上であれば、ほぼ100%収録されているものと見做すことができる。

表2-1に各特許庁の検証対象データベースと検証フィールドを示した。

表 2-1. 検証対象データベース

	AB	CL	FULL
SIPO (CN)	○		
CNIPR (CN)	○	○	○
TIPO (TW)	○	○	○
KIPRIS (KR)	○	○	○

また、英語商用データベースとしてDWPIについて、中国、台湾、韓国の書誌および抄録の収録(AB/FA)を調べ、各特許庁のデータと比較した。

2-2. キーワード検索による検証

出願前の新規性調査（公知例調査）や無効化資料調査では、具体的な化合物名や物質番号（CAS No.、カラーインデックス番号など）、あるいは特殊な技術用語や化学式、単位記号などを検索することもある。これらが抄録に顕現するのは、発明のポイントがそれらにある場合であり、通常はクレームや明細書全文中の記載にとどまることが多い。

本検討の対象である中国、台湾、韓国などアジアの特許データベースにおいては、商用データベースのクレームや明細書全文からの検索は不十分である（US,EPなどにファミリーがあれば、それらファミリーのレコードを対象に検索は可能であるが、ファミリーがない場合にはクレーム等からの検索は不十分となる）。

そこで、一般的な技術用語として「ポリカーボネート」について、化学式として珪素系の化合物であるシル

セスキオキサンを表わす「RSiO1.5」および「RSiO3/2」など数種のキーワード、化学式についてアジア各特許庁の英語および原語データベースと商用データベースの抄録、クレーム、明細書全文の収録を比較し、調査時の有用性について判断した。

3. 検証結果

3-1. 収録データの検証

商用英語データベース(DWPI)の抄録および各特許庁データベースの抄録(AB)、クレーム(CL)、明細書全文(FULL)の収録タイムラグの概要を、表3-1(英語データベース)、表3-2(原語データベース)に示した(2009/08 現在)。

表 3-1. 英語データベースタイムラグ

	AB	CL	FULL
DWPI(CN)	1M	-	-
(TW)	12M	-	-
(KR)	2M	-	-
SIPO(CN)	4M	-	-
CNIPR(CN)	1M	-	6M
TIPO(TW)告 開	8M 8M	- -	- -
KIPRIS(KR)	3M	-	-

表 3-2. 原語データベースタイムラグ

	AB	CL	FULL
SIPO(CN)	1W	-	-
CNIPR(CN)	1W	2W	2W
TIPO(TW)告 開	1W 2W	1W 9M	1W 9M
KIPRIS(KR)	1W	1W	1W

TIPO は公開と公告、その他は公開
(M:Month W:Week) 2009/08 現在

英語データベースは商用、各特許庁とも収録タイムラグが大きく、収録はほとんど抄録までである。これに対し、原語データベースの収録タイムラグは各特許庁のデータ更新と同時であり、中国および韓国においては毎週、台湾において

は公告:10日、公開:15日となっており、抄録だけでなく、クレームや明細書全文からも同じタイムラグで検索できた(台湾の公開特許のみクレーム、全文は9ヶ月のタイムラグがあった)。

3-2. キーワード検索による検証

「polycarbonate」や添字付きの化学式(「RSiO1.5」「RSiO3/2」)なども抄録と同様のタイムラグでの収録が確認できたが、原語データベースのクレームや全文から得られた情報がDWPI および各特許庁英語データベースでは得られないことも確認できた(詳細なデータについて発表することとしたい)。

また、特に台湾の特許情報については、DWPIの書誌・抄録とも公告特許の収録は2008年の途中まで、公開特許は2005年および2009年は未収録であり、台湾特許庁の英語データベースでも公告特許は2008年以降、公開特許は2009年2月以降抄録の収録がなかったため英語データベースで最新情報を配信することはできない。

DWPIのこれら3国・地域における抄録収録は、原語抄録の収録数よりもいずれも多くなったが、「1/CYC」と1国出願のみに限定すると原語収録数よりも少ないかゼロとなったことから、US、EP、PCTなどのファミリー特許抄録からの抽出によるものと考えられる。

4. 考察

US、EP、PCT特許と異なり、アジアの特許情報に関しては、商用データベースおよび各特許庁の英語データベースでは、データ収録のタイムラグが大きく、最新特許情報の配信としてのSDIとしては問題があり、日本特許と同様のタイムラグでSDI配信するには各国・地域特許

庁の原語データベースに頼らざるを得ないことがわかった。

原語検索および原語検索結果で得られた情報は、中国語やハングルであり、原語入力および翻訳が必要となるというハンディが存在するが、少し工夫することで解決することができる。2004年以降の台湾原語データベースおよび韓国特許データベースにおいては、英語でも検索できるようになり、両特許原語データベースの検索結果の「発明の名称」は、原語と共に英語表記されており、情報の配信先において関係ある情報かどうかを容易に判断できるようになった。

しかし、商用データベースにおいても2009年に入って、人手翻訳による中国、韓国特許データベース(Thomson Innovation)、機械翻訳によるフルテキストデータベース(QPAT)などが提供されるようになり、タイムラグも大幅に改善され、クレームや全文からの検索も可能となってきたことは喜ばしい。

したがって、今後のアジアの特許情報は、遡及検索は商用データベースで、SDIなどの最新特許情報は原語データベースを利用して補完する、というような利用方法が有用ではないかと考えられる。原語データベースの簡易検索法、得られた原語検索結果の簡易翻訳法も含めて提案したい。

5. 参考文献

[1] 伊藤徹男, 外国特許調査(日本知的財産協会研修C9Aコース), (2008/09)

[2] 邊玲美, 船戸奈美子, 有料データベースにおけるアジア特許の収録の状況と速報性, 第5回情報プロフェッショナルシンポジウム, p.87-91 (2008/11)

[3] アジア主要国の特許収録状況, (化学情報協会), <http://www.jaici.or.jp/stn/pdf/asiapat.pdf>, (accessed 2009-07-15)