

中国特許調査における CNIPRデータベースの役割

アジア特許情報研究会 **伊藤 徹男**

PROFILE

2004～現在：日本知的財産協会セミナー講師、2006～現在：日本特許データベースセミナー講師、2009～現在：発明推進協会セミナー講師、2007～2009：検索競技大会委員、2008年：アジア特許情報研究会設立

✉ patentsearch2006@yahoo.co.jp

1 はじめに

日本特許庁はじめ各国特許庁データベースサービスの充実ぶりには目を見張るものがある。無料でワールドワイドな特許検索が可能なヨーロッパ特許庁の espacenet が商用英語データベースのような検索キーワードのハイライト機能や検索結果一覧からのダウンロードもエクセル形式だけでなく csv 形式を追加したり、複雑なクレームでは独立項と従属項を一瞥できるクレームツリー機能を追加したのは、つい2、3年前である。WIPOのデータベースである PATENTSCOPE が、例えば日本語で検索すると英語だけでなく、ドイツ語やフランス語、ポルトガル語、中国語などに翻訳して検索できるクロス言語検索機能を採用し、検索結果も各国語に機械翻訳してくれるサービスを追加したのは記憶に新しい。

アジアにおいては、韓国特許ツール KIPRIS のハングルデータベースが英語でも検索できるようになったり、審査経過情報に英語のサポートが付き、我々外国人にとって気軽にアクセスできるようになった。台湾特許庁データベースにおいては、2008年に無料データベースとしては珍しく履歴演算機能を付加したり、検索結果を一度に300件ダウンロードできたり、審査経過情報や年金納付状況までダウンロードできることになった。中国においては、中国特許庁データベースとは別の「CNIPR」が2010年4月に概念検索機能、クロス言語検索機能や簡易解析機能などを付加して登場したと思ったら、2011年11月には、検索結果一覧で生

死情報が確認できるなどいくつかの機能を追加し、さらに、これまで無料であった英語データベース C-Pat が2012年3月に、突然有料データベースとなったりするなど目まぐるしい動きがあった。また、中国特許庁は2012年4月には包袋情報をインターネットからアクセスできるサービスも追加した。

さらに、インド特許庁、ロシア特許庁、ブラジル特許庁など新興国の特許データベースが、やはりこの2、3年で矢継ぎ早にデータベースを更新した。特に、アジア・新興国特許庁データベースの特許情報については、使い勝手が悪いものの商用英語データベースの収録が不十分な点を補完するデータベースとして今でも役に立っている。

このような各国特許庁データベースの動向のうち、本稿では、中国特許調査に際して多くの有用な機能を有する中国語データベース CNIPR の実務でも使える最新機能について紹介する。

2 中国特許調査と調査ツール

外国特許調査に用いられている商用英語データベースが欧米特許調査と共に中国特許調査ツールとして使われているのが一般的である。抄録やクレームを人手翻訳しているデータベースもあるが、多くは機械翻訳された情報を収録しているデータベースである。商用英語データベースの台湾や韓国などアジア・新興国のデータ収録は極めて不十分で満足な調査ができないのも現状であるが、中国に関しては2、3年の間にデータ収録が大幅に改善され、直近の1～数か月のタイムラグを除けばほと

んどの商用英語データベースの書誌・抄録・全クレームの収録には問題がない。問題があるとすれば、多くは機械翻訳であるため、誤訳やスペルミスも欧米系のデータに比べて多いことである。

このような商用英語データベースを補完するデータベースとして使われているのが、中国語データベースである。中国語データベースは、中国特許庁（SIPO）だけでなく、国家知識産権局の直属機関である知識産権出版社（IPPH）、中国專利情報センター（CPIC）、上海や各省などいくつかのデータベースが存在するが、最も信頼されて実務で使われているのが知識産権出版社のCNIPRである。

CNIPRの各種機能の紹介は次項に譲るが、ここでは何故、中国語データベースによる補完が必要かについて少し触れることにする。1つは、中国語から英語への翻訳に際して同一用語、同一出願人名（発明者名等）が多様な表現で英訳されることである。中国出願人の多様な異表記に関しては赤壁氏が紹介しているので¹⁾、ここでは用語の異表記について1例を紹介する。

中国特許で「太陽電池」は、「太阳能电池」または「太阳电池」などと表現されるが、これに対応する商用英語データベースの翻訳は、そのほとんどが「solar cell」や「solar battery」（語尾変化も含む）であるが、「photovoltaic cell」、「solar energy cell」、「solar power generation」の他、「solar module」や「solar assembly」といった多少広義の概念にまで翻訳されることがある。

英語データベースでこのような幅広い用語まで網羅して検索すれば中国語データベースを利用する必要もないが、異表記のすべてを網羅することは困難であるし、ノイズを拾ってしまう場合も多い。そこで英語での検索とは別に中国語で検索することによって漏れない、かつノイズの少ない検索ができるのである。常に英語データベースの方が異表記が多いという訳ではなく、同一概念の用語であっても英語用語よりも中国語用語の方が異表記が多い場合もある。

最近、いくつかの商用英語データベースで原語（中国語も含む非ラテン言語）を搭載したデータベースが現れ、英語でも中国語でも検索でき、英語検索での漏れをカバーしようとする試みがなされているが（ハイブリッド検索システム²⁾）、現状では原語の収録率が100%ではないので、このハイブリッド検索システムだけに頼ることもできない。原語収録率が向上すれば極めて有望な調査ツールとなる。

したがって、現時点での中国特許調査は、やはり商用英語データベースとそれを補完する中国語データベースが必要となり、CNIPRが多用される理由ともなっている。中国語データベースと言うと中国語など全く理解できない自分には無用のもの、と思われがちであるが、英語の技術用語をしっかりと把握できれば中国語の技術用語を集めるのは、図1に示すような無料のWeb辞書³⁾などを利用することによって比較的容易に、中国語検索式を立てられる。



【図1】 英中科学用語辞書

3 中国語データベース CNIPR の基本機能と最近追加された機能

そこで中国語データベースとして実務で使われている「CNIPR」の概要について基本的な機能と共に最近追加された新機能などについて紹介する。CNIPR 中国語版には有料版と無料版があるが、SDI 機能、検索履歴の保存など一部機能のみ利用できないだけであり、実務では無料版で充分である。後ほど紹介する簡易統計解析機能が旧バージョンの有料版では 10 万件、無料版では 5 万件処理できたものが、2011 年の改訂でそれぞれ 5 万件、1 万件と制限されてしまったのは少々残念である。1 万件もの解析をすれば充分と思われるかもしれないが、動向調査や予備検索などの集合をざっくり見てみたい場合には数万件の情報を解析したいときもある。

3.1 CNIPR の基本機能

出願番号など各種の番号や日付、出願人、発明者、代理人、代理人事務所、発明の名称などの書誌事項のほか、抄録、請求の範囲、全文中の用語などが各フィールドから検索できるし、各フィールドのコマンドを利用したコマンド検索も可能となっている（図2）。無料のデータベースにはめずらしく数種の近接演算も利用できるのはノイズカットの点ではうれしい機能である。コマンド検索では、フィールド検索にはない国際出願番号、国際公開番号の他、分割出願番号からの検索もできる。

検索ステップごとの検索集合を組合せた履歴演算が

できないのが最大の欠点である。これは他の各国特許庁のデータベースいずれも同様であるが、少なくとも有料版だけでも履歴演算が可能になればかなり使いやすくなるのはまちがいない。履歴演算ができない点をカバーするのがコマンド検索であるが、ここでは各種のコマンドを使ったかなり複雑な検索式を立てられる。2011 年の改訂前までは膨大な検索式を入力できたが、改訂後は 4000 バイト以内に制限されてしまった。履歴演算ができないのであれば、この入力文字制限は解除して欲しいものである。

その他の検索・表示機能として以下のものがある。

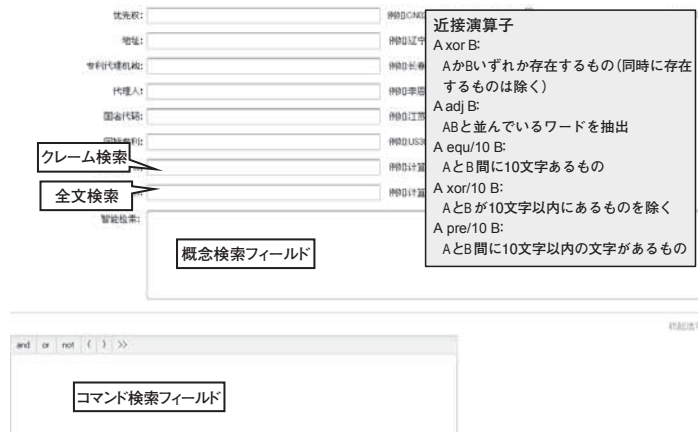
①失効検索

拒絶や取り下げ、年金未納などで失効した特許、実用新案、意匠に関するデータベースが別途用意されており（ボタンひとつで切り替えられる）、検索結果が大量になったときなどは、この機能を利用して生きている登録特許や公開特許などを抽出し、生きているものから先に査読（解析）していく、などということができる。

②法律状態検索

フィールド検索やコマンド検索など通常の見出し画面とは別に審査経過情報や年金未納による失効などが検索できる。SDI などに関連重要特許を見つけ、その後の動向をフォローする、いわゆるウォッチングとして活用できる。

特許と実用新案の二重出願について「重複授權放棄」を調べることにより、競合他社や分野別の二重出願状況なども把握できるなど活用範囲は広い。



【図2】 コマンド検索画面

③検索結果のダウンロード

表示・出力できる機能には制限がなく、何万件でも可能であるが（日本の特許電子図書館 IPDL やヨーロッパ特許庁の espacenet などは表示できる件数が 1000 件や 500 件に制限されている）、1 回に書誌・抄録をダウンロードできる件数が 10 件ごとであることが実務上悩ましい。この点は有料版でも同様である。今年中にも日本に専用サーバーが設置され、ダウンロード件数が有料版で一括 500～800 件程度まで可能になることが知識産権出版社から予告されているが、まだその動きはない。

3.2 CNIPR の新機能

2010 年 4 月の大改訂以降に追加された新機能について紹介する。各項目ごとに追加された年月も示した。

①出願人辞書機能（2010 年 4 月）

特定の出願人名を入力すると関連会社を含めた出願人一覧を表示してくれる。ここで得られた出願人名を元に検索ができる。グループ企業が充分網羅されているとは言えないし、社名変更などにも対応できていない。出願人検索時の参考程度と考えた方がよい。

②概念検索（2010 年 4 月）

抄録や請求の範囲の文章などを入力して入力文章に近い該当特許を抽出できることになっている。精度は不十分ではあるが、通常の演算検索では抽出できない類似特許を抽出できることもある。

③類似性、新規性、侵害性検索（2010 年 4 月および 2011 年 11 月）

特定の特許（実用新案なども）について、全文を対象に類似特許を抽出したり、特定特許の出願日以前の類似特許や生死情報データも加味して侵害性のある特許などを抽出する機能である。この機能の抽出精度については十分な検証データが手元にないので何とも言えないが、参考情報を得る程度と考えておいた方がよい。

④ワールドワイド検索（2010 年 4 月）

日本、米国、ドイツ、台湾、韓国など 85 の国・地域の特許を中国特許と共に串刺し検索できるようになっているが、ほとんどの国のデータ収録が 20%未満であり、現状では役に立たない。

⑤クロス言語検索（2010 年 4 月）

英語を入力すると中国語に機械翻訳されて検索できることになっているが、異表記、同義語のカバーが不十分でノイズ用語を削除できないなど、検索用語を取捨選択できない。したがって、この機能も現状では役に立たない。

⑥検索結果一覧における生死情報識別マーク（2011 年 11 月）

検索結果一覧にグリーン（権利として生きているもの）、イエロー（審査請求前後の公開特許（但し、審査請求期限を徒過して失効しているもの、放棄したものは失効としてグレー表示）、グレー（失効しているもの）の 3 種の表示が加わった（図 3）。

検索結果が大量の場合には、権利化されている登録から、次いで審査請求されているものおよび審査未請求の公開、最後に失効しているものと、別々に査読（解析）す



【図 3】 生死情報識別マーク



【図4】 統計解析機能

ることができる。場合によっては失効特許は査読しなくて済む。検索結果のスクリーニングには強い味方である。

⑦簡易統計解析機能（2010年4月および2011年11月）

検索結果一覧画面上部にある分析ボタンをクリックすると検索でヒットした該当のものについて出願推移やIPCなどの特許分類分析、出願人ランキングなど簡易な解析ができる機能が2010年4月に加えられたが、さらに2011年11月に全文を解析するクラスター分析が加わった（図4）。

この統計解析機能では、出願人や発明者およびワードで解析する際にはそれぞれの異表記を統合する機能がないことに留意が必要である。図5に示すように出願人の法人格が異なるだけで別出願人として扱われてしまう。ワードの異表記も同様に統合できないのでクラスター分析などに影響する。

中国特許庁の統計情報などを解析したレポートとして知られる「専利統計簡報 2011年07期」でも出願

三星SDI株式会社[396]	宇部兴产株式会社[42]
比亚迪股份有限公司[275]	三菱化学株式会社[39]
松下电器产业株式会社[263]	索尼公司[38]
三洋电机株式会社[226]	东莞新能源电子科技有限公司[35]
索尼株式会社[219]	日产汽车株式会社[33]
深圳市比克电池有限公司[129]	东莞新能源科技有限公司[30]
天津力神电池股份有限公司[125]	麦广树[29]
丰田汽车株式会社[108]	中国电子科技集团公司 [28]
株式会社LG化学[104]	上海比亚迪有限公司[28]
株式会社东芝[53]	深圳市赢合科技有限公司[28]

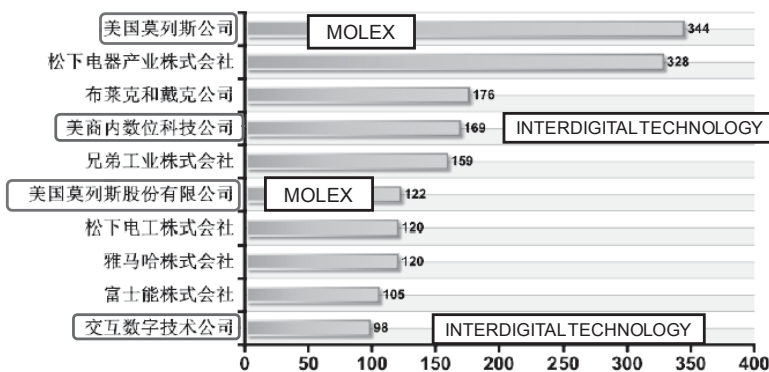
【図5】 出願人の異表記

人の名寄せがされないまま報告されている⁴⁾。おそらくCNIPRの統計解析機能のようなツールで処理されたものではないかと思われる（図6）。

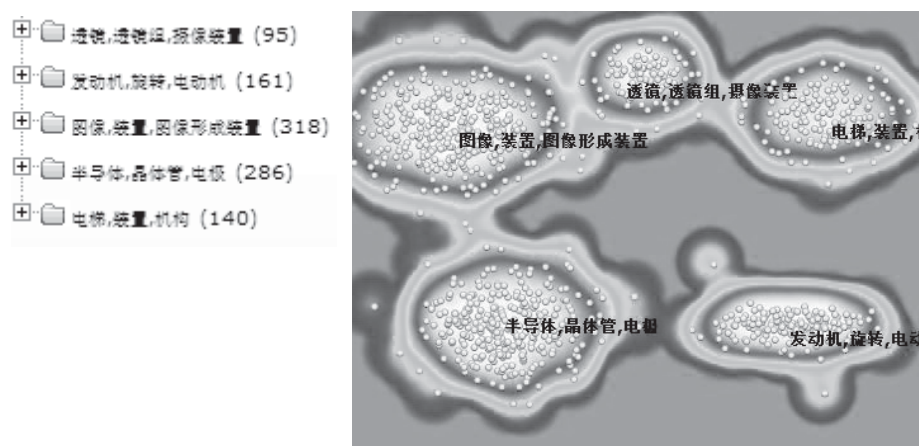
したがって、統計解析機能で得られたランキング情報などをデジタルデータでダウンロードし、エクセルや汎用の解析ツールなどにインポートして名寄せしたり、用語の統一をして解析することが好ましい。

クラスター分析では、検索結果の集合を指定した特定数のクラスターに自動分割し、判別してくれる機能であり、クラスターごとの集合を取り出して専用の解析ツールにデータを流し込んで詳細な解析をすることも可能で

图 9 有效实用新型专利量前十位的国外专利权人



【図6】 専利統計簡報 2011年07期の有効実用新案ランキング



【図7】 クラスター分析

ある（図7）。解析方法の詳細を紹介する余裕はないが、現状では、このクラスター分析は2000件までのデータしか処理できない。

4 おわりに

CNIPR 英語データベースも中国語データベースと同様にコマンド検索できる機能などが追加され、フィールド検索にも全請求項、全文からの検索フィールドが設けられた。しかし、汎用の安価な固定料金制の商用外国特許検索システムと比べると、やはり履歴演算ができない点は拭うべくもなく、また、最近の商用システムでは検索結果一覧から全文表示まで機械翻訳ではあるが簡単に日本語表示できるなど表示機能の点でもかなわないなど実務での有用性を考えると大きなメリットもないことから紹介を割愛した。

本稿では、商用英語データベースのみでは欠落を生じる中国特許調査の補完的データベースとして各種の機能を有する中国語データベース CNIPR の最近追加された機能について、実務でも使えるか、という観点から紹介した。検索結果一覧での生死情報識別マークや名寄せなど若干の問題点はあるものの簡易統計解析機能などが実務上有用な機能として追加された。その他、商用データベースには見られないユニークな機能も各種保有してい

るが実務的な観点からは不十分な点も多い。CNIPR データベースにワールドワイドな検索ができることまで求めてはならない。

データベースとして最も基本的な検索式集合の履歴演算機能および検索結果のダウンロード機能が有料版のみの機能であっても強化されればさらに強力な中国特許調査ツールとなることは間違いない。今後に期待したい。

また、2012年8月から中国も日本、台湾、韓国と同様に優先審査制度が施行されるに至り、従来以上に出版願から1年以内に登録となる特許が増えることが予想されることから、商用英語データベースだけに頼ってはいけません。重要特許も逃すことになる。商用英語データベースの補完ツールとして今後ますます中国語版 CNIPR データベースの活用が求められる。

参考文献

- 1) 赤壁：知財管理 VOL.61 NO.4 P521(2011)
- 2) 田畑他：第8回情報プロフェッショナルシンポジウム(2011)
- 3) 郑州大学・英中科学用語辞書 (<http://www3.zzu.edu.cn/zjdict/>)
- 4) 中国特許庁・専利統計簡報 2011年07期 (http://www.sipo.gov.cn/ghfzs/zltjtb/201104/t20110422_600236.html)