

特許情報検索の現状と課題

住友電工知財テクノセンター株式会社
日本知的財産協会
知的財産情報検索委員会委員長
臼井 裕一

PROFILE

住友電工入社後、研究開発部門にて研究開発に従事、その後研究開発部門テクノリサーチセンターに異動し、情報調査・管理業務。その後現職に。

✉ Usui-yuichi@sei.co.jp

☎ 045-853-7160

1 はじめに

知的財産権は重要な経営資源であり、事業戦略、研究開発戦略と連携した知財戦略を立案・推進するにあたって特許情報の果たす役割は、非常に大きい。昨年、JAPIO 2006 YEAR BOOKに特許情報検索の現状と課題について述べさせて頂いたが、今回は前回の内容を概括するとともに、その後の状況について述べる。

2 特許情報政策

2.1 昨年までの動向

「知的財産立国」を国家戦略として目指し、2002年7月に知的財産戦略会議において「知的財産戦略大綱」が取りまとめられ、ついで11月に「知的財産基本法」が制定され、翌2003年に内閣に「知的財産戦略本部」が設置され、「知的財産立国」を推進する体制が整った。「知的財産戦略本部」は2003年から毎年、数百項目の施策からなる知的財産推進計画を制定している。この推進計画は、具体的目標と達成時期が明示されている点で、非常に重要であり、かつ影響度の大きいものである。また、特許法改正により2001年10月以降の出願について審査請求可能期間が7年から3年に短縮され、それに

伴って審査請求件数が急増し、審査順番待ち期間も長期化が避けられない状況となった。このため知的財産推進計画2004において、2004年現在で26ヶ月ある審査順番待ち期間を最終的には0とすることを目指し、中期目標としては順番待ち期間がピークを迎える5年後の2008年に20ヶ月台に留め、長期目標としては10年後の2013年に11ヶ月とした。この目標達成のために2004年からは、「特許審査迅速化・効率化の中・長期目標を達成するための実施計画」を知的財産推進計画の作成時期に合わせて策定し前年度の目標および達成状況とともに知的財産戦略本部に報告し、公表している。2005年12月には、この特許審査迅速化の目標達成を最重要課題として、特許庁を中心として経済産業省全省挙げての総合対策を実施するために、経済産業大臣を本部長とする「特許審査迅速化・効率化推進本部」が設置され、2006年1月に官民を挙げて取り組むべき「特許審査迅速化・効率化のための行動計画」が策定されている。一方特許庁では、「知的創造サイクルの活性化（外部ユーザーの利便性向上）」「世界最高レベルの迅速かつ確かな審査」「業務の根本的見直しとレガシーシステム脱却」の3点を目標に掲げて2005年8月に「特許庁業務・システム最適化計画（改訂版）」⁽¹⁾を策定した。この最適化計画はおおよそ10年の長期的スパンで取り組んでいるもので、その前半で新事務処理システムを構築し、後半にて検索システムを構築するというものである。前

半の新事務処理システムの稼働は2011年1月、後半の検索システムの稼働は2014年1月を予定している。なお、この計画において外部ユーザーの利便性向上に関連する記述は以下の通り。

産業財産権情報の対外提供

- ・ 24時間、365日のインターネット出願
- ・ インタラクティブ申請の導入
- ・ インターネット公報の推進
- ・ データ提供のリアルタイム化
- ・ 包袋情報の無料提供の推進
- ・ 審査・審理関連情報の提供
- ・ 検索機能の提供の拡大
- ・ 快適なレスポンスの維持

2.2 その後の動向

昨年の執筆時以降の特許情報検索に関わる主なものは以下の通り。

- ・ AMARIプラン2007⁽²⁾ (2007年1月)
- ・ 知的財産推進計画2007 (2007年5月)
- ・ 特許審査迅速化・効率化中・長期目標を達成するための実施計画 (2007年5月)
- ・ IPC第8版初めての改正 (2007年1月)

2.2.1 イノベーション促進のための特許審査改革加速プラン2007 (AMARIプラン2007)⁽²⁾

安倍内閣の発足をうけて経済産業省では2006年10月に新たな特許行政の基本方針である「イノベーション促進のための特許審査改革加速プラン (AMARIプラン)」を4分野20項目についてとりまとめた。その後2007年1月に先に述べた行動計画において設定した数値目標を改訂するとともにプランの重点施策を4分野26項目に拡充し、これらを一体化して「イノベーション促進のための特許審査改革加速プラン2007 (AMARIプラン2007)」⁽²⁾として策定、公表した。ここで設定された2007年度の数値目標は以下の通り。

- ・ 一次審査件数 31万件
- ・ 審査順番待ち期間 28ヶ月台
- ・ 審査官一人あたりの年間処理件数 (請求項ベース) 1300項以上
- ・ 先行技術調査の外注件数 22.6万件
- ・ 審査に係る直接コスト (1請求項あたり) 2.3万円以下

また重点施策4分野は以下の通り。

- ・ グローバルな権利取得の促進と知財保護の強化
- ・ 特許庁による審査迅速化・効率化に向けた更なる取組
- ・ 企業における戦略的な知財管理の促進
- ・ 地域・中小企業の知財活用に対する支援の強化

なお、このプランは毎年策定され、従来の特許審査迅速化・効率化中・長期目標を達成するための実施計画に替えて運用される。

2.2.2 知的財産推進計画2007⁽³⁾

ちょうど昨年、2006 YEAR BOOKの原稿執筆の直前に発表された知的財産推進計画2006において、特許情報の検索に関連した記述は以下の通りであった。

- ・ 「特許・論文情報統合検索システム」を整備する。
- ・ 特許電子図書館等を通じた産業財産権情報の利用環境を整備する。

これらの具体的内容のうち、大学等における固定URLによる特許公報データへの直接アクセスとJSTによる「特許・論文情報統合検索システム」の試用、またIPDLにおける審査経過から分割出願に関する情報へのリンクやテキスト検索等の検索結果からの審査経過情報へのリンク等の機能の充実、使いやすさの向上等は既に実現されており、検索ガイドブックによる審査官のサーチノウハウも前年度に引き続いて公表され、更に審査官のサーチノウハウを活用した研修 (エキスパート研修・上級) や (独) 工業所有権情報・研修館における審査官

と同等のサーチ端末を用いた検索・閲覧サービスも開始された。ただし、全文テキスト検索機能の追加、国内公報と外国公報とを同時に検索する機能の追加および特許庁内で利用されているFIやFタームなどの検索キーとテキスト検索との組合せによる検索機能の追加については、必要な措置を講ずるとのことであったが現段階では実現していない。

なお、「特許・論文情報統合検索システム」は、2007年9月3日よりJSTPatM（ジェイエスティパットマルチ）として有料で一般に公開されている。

知的財産戦略本部の設置から3年が経過した2006年度からは、第2期として目標を「世界最先端の知的財産立国を目指す」として第2期の初年度の知的財産推進計画2006が策定され、引き続き第2期の2年目として知的財産推進計画2007⁽³⁾が策定された。この推進計画で特許情報検索に関連するのは以下の通り。

- ・ 大学等での特許情報利用促進のため特許情報検索ソフトを2007年度中に工業所有権情報・研修館より無料で提供する。
- ・ 2007年度中にパテントマップ作成ソフトを開発し、大学等に提供する。
- ・ 2014年1月稼働予定の特許庁業務・システム最適化計画に基づく新システム構築完了前においてサービス可能なものから先行して提供する。2007年度中にロードマップを策定。
- ・ 2007年度中にIPDLに全文テキスト検索機能を追加する。
- ・ 現在大学等に限定している特許情報の固定URLサービスを2007年度以降順次一般に拡大する。
- ・ 2007年度から利用許諾の得られた意匠公知資料の公開を促進する。
- ・ 2007年度中に各企業が自社の出願件数や審査実績等の情報を加工、抽出、経年比較できる特許戦略ポ

ータルサイト（仮称）の試行を開始する。

2.2.3 IPC第8版の発効と改正

第7版までのIPCは、特許発行数が多い大規模な特許庁にとっては先行技術の調査には粗すぎ、その一方で小規模な特許庁にとっては細かすぎて、分類付与の負担が大であるという問題があった。このため第8版ではアドバンスレベルとコアレベルとに二分化することになった。各分類項目のレベル分けは、コアレベルがアドバンスレベルの約30%となるように、分類項目毎に個別に設定されている。また改正頻度は、アドバンスレベルでは定期的ではなく、日米欧の三極特許庁及びWIPOで構成される特別小委員会の決定により随時改正される。またコアレベルは3年毎にまとめて新しい版として発効される。そして各国特許庁が付与したIPCデータは、MCD（Master Classification Database）に蓄積されて一元管理される。また、既発行の過去分特許公報についても共同して再分類し、最新のIPCが付与されてMCDに格納される。既発行分の再分類されたデータが全てMCDに搭載された段階で改正IPCが発効されるため、改正された段階で過去分も含めて改正後のIPCで検索されることになる。このMCDの運営は欧州特許庁（EPO）が行い、データは有償で販売される。日本においては、このIPC第8版の既発行分の再分類された過去分データは、2006年9月8日から整理標準化データとして（独）工業所有権情報・研修館から提供されている。

このIPC第8版における変更点としては、上記の二分化、改正頻度の他に分類項目の桁数変更、分類記号の統一、公報表記の変更、標準配列と分類付与ルール等があるがここでは詳細説明は割愛する。

ところで日本においては、改正分の過去分の再分類データは、整理標準化データとしては提供されない。このため、改正分の過去分データについて商用データベースのプロデューサーは、MCDから入手する必要がある。

ところが現段階では各プロデューサーは様子見で明言を避けている。この過去分のデータを搭載するのか、しないのか、また搭載するとしてIPC第7版以前のデータと共存するのか等、そのデータの持ち方によって検索の方法のみならずその結果も異なってくる事が予想されるので、われわれユーザーはこの辺りの動向をしっかり把握しておく必要がある。2007年1月にIPC第8版の初めての改正が行われた。具体的にはA62DとH02Mの2つのサブクラスについて改正が行われた。ところで本来IPC第8版においては、改正分が発効された時点では既発行分の再分類は終了して全てMCDに搭載されているため、発効時には改正IPCにて全期間検索が可能のはずであった。ところが、IPDLでこれら改正IPCで検索すると、2007年以降の特許しかヒットしてこない。これはMCDのデータが入っていないためである。一方Esp@cenetでは既発行分についてもヒットする。先に述べたように日本においては改正時に既発行分の再分類はされても整理標準化データとして出されない。このため、現時点でも日本の商用データベースでは、IPDLと同様に既発行分がヒットしない状態となっている。一方、DWPIのような海外のデータベースにおいては、MCDのデータを入れているため既発行分もヒットする。今後、改正が進んでいくと海外の商用データベースと日本の商用データベースとの乖離がますます増大していくため大きな問題となろう。各データベースのベンダーの動向を見定める必要がある。

3

特許情報データベースの変遷

3.1 公報の電子化と特許情報データベース

現在、日本の特許情報商用データベースは主立ったものだけでも、すぐに10数種類をあげることができる。

筆者が特許情報に関わるようになった頃には、PATOLISしか無かった事を思い起こせば隔世の感を禁じ得ない。この頃は、まだ紙公報の時代であり、特許情報のデータベースは書誌事項と抄録程度であるが人手で作成し搭載するため、時間とコストがかかっていた。今になってみれば、特許情報のデータベースにとって最も影響が大であったのは、平成5年（1993年）の公報の電子化であった。従来の紙公報からCD-ROM（現在はDVD）化され全文が機械可読形式となり、更にSGML（現在はXML）によって構造化されたことにより、このデータをそのまま蓄積していけば容易に、かつ安価に特許情報のデータベースが構築できるようになった。公報の電子化以降にサービスを開始した特許情報データベースはいずれも、この電子データを取り込んで利用している。ただし、実際のサービスとして利用されるようになるには、データの蓄積量がある程度増えるまで待つ必要があった。この蓄積されたデータをもとに、検索エンジンを搭載して、インターネットを介して接続すれば、容易に特許情報データベースサービスを開始することが可能であった。インターネットは、日本においては公報電子化の翌年の1994年に民間に開放されている。このインターネットが情報インフラとして整備され、かつIT技術の進展と相まって新たな全文検索系の商用データベースが続々と誕生することになった。これらの電子化公報のデータを利用したデータベースと同様に、特許庁ではこれらデータを蓄積して一般の利用に供する特許電子図書館（IPDL：Industrial Property Digital Library）サービスを平成11年（1999年）に開始した。また同年より、特許庁は平成10年（1998年）以降の保有する「出願マスタ」、「審判マスタ」、「登録マスタ」、「サーチマスタ」、「国内引用文献マスタ」、その他のマスタの情報をSGMLまたはXMLにて標準化した整理標準化データとして提供を開始した。（現在は、独立行政法人 工業所有権情

報・研修館が提供)

現在の全文検索システムのほとんどは、先の電子化公報のデータと、この整理標準化データを用いたものである。IPDLのサービスは、平成5年以降のデータに限られるが、無料で検索が出来てまた公報を1件ずつではあるがダウンロードする事が可能である。インターネット経由で24時間接続が可能であり、何時でも、どこからでも、誰でもが安価に特許情報にアクセスできる環境が整ったと言えるだろう。しかしながら、当初は接続性が確保できなかったり、反応が遅い等の問題があったが、順次改善されてきた。一方、公報のダウンロードが可能であることからロボットによる大量データのダウンロードにより、ますますトラフィックを混雑させて、一般の使用に差し障りが出る等の問題も顕在化した。また無料で検索が可能である事から、民業圧迫との声も聞かれ、官と民とのベストミックスということが言われた。

平成15年(2003年)の産業財産権情報利用推進委員会の最終報告⁽⁴⁾によれば、国の役割は「正確で基本的な1次情報をマージナルコストで、整理標準化データとIPDLで提供」する事とされている。また更にIPDLで提供するために、基本的な検索サービスを提供するとも述べられている。一方民間の役割は、「国が提供する正確で基本的な1次情報に高い付加価値をつけた情報を提供」する事とされている。具体的には、高度な検索や動向分析等の高い付加価値のついた情報をユーザーに提供していくと述べられている。

このような役割分担のもと、各民間業者は次々と高度な検索手法や高い付加価値のついた情報が得られるような機能を付加したサービスを提供し始めた。また、サービスの提供方法や課金方法についても従来とは異なるような方式を採用するケースが増えてきた。以下にこれらについて説明を加える。

3.2 全文検索

現在の特許情報検索システムは、ほとんど全てがデータとして電子化公報データを利用している。もともとのテキストデータは、出願人作成の電子データであり、どのシステムも同じデータを利用している事になるために、どのシステムで検索を行っても同じ結果になると思いがちである。しかしながら、実際には若干、検索結果は異なってくる。この理由はそれぞれの検索システムでデータの取り込み時の工夫や検索時に異表記展開を行ったり半角/全角の区別をしなかったり、それぞれ工夫をしているからである。従って使用するシステムのそのような事項をよく知らないと、検索漏れを生じたりノイズが増えたりする結果となるので注意が必要である。また全文検索では、キーワード検索の場合には基本的に中間一致での検索(文字列検索、ストリングサーチともいう)であるためにどうしてもノイズが多くなる傾向にある。また自然語による検索となるために従来の抄録ベース/統制語のデータベースの検索に比べてキーワード検索時には、同義語を全て入れなければ漏れが生じることになる。このように全文検索は、システムごとにある程度、特徴があるためにこれらについてよく確認しておくとともに、全文検索のノウハウを確立する必要がある。またデータとして公報データのみで整理標準化データによるアップデートをしていない検索システムもある。このような場合には、更新FIや更新Fタームが使用出来ないだけでなく、2000年より前のデータでは公報にFタームが付与されていないのでFターム検索ができないことになるので、特に注意が必要である。また、最近これら全文検索システムでは公報電子化の平成5年よりも以前の公報をOCRにて読み取ってテキスト化してデータの蓄積範囲を広げている。このテキスト遡及データにおいても単純にOCRで読み込んだだけのものや、特許請求の範囲までを手でチェックしたものの2種類が存在するようで

ある。いずれにしてもこのOCRで読んだだけの遡及データは当然SGMLのタグはついていないため、要約や請求の範囲に限定した検索はできないことに注意する必要がある。

3.3 概念検索

概念検索あるいは類似文書検索は、検索式によるのではなく、検索条件を自然語による文章（種文書）で与え、この種文書と蓄積文書とを形態素解析して特徴文字列を抽出し、例えばベクトル演算をして類似度を計算して類似度順にランキングするといった検索方法で、インターネット上の検索エンジンと同じような感覚で使用するためにエンドユーザーや検索に精通していないユーザーには取り付きやすい検索方法である。基本的にはランキングするため、漏れという概念は無い（つまり全ての蓄積文書に順番をつけるので）が、実際問題としては上位100件とか300件とかのうちに必要なものがほとんど入ってこなければ実際の使用は難しいと思われる。ただし概念検索が世の中に出てきた当初に比べれば、かなり適切な結果が出るようになってきているシステムもあるので、今後どのように発展・進化していくのか期待を込めて見守っていく必要がある。現状での問題としては、検索のアルゴリズムが公開されていないため、思ったような結果が出ない場合に種文書をどのように変更すればいいのかがよくわからない点である。この事はとりもなおさずユーザー教育において、システムの使い方は教育できてもどのように検索すれば最適な解が得られるのかというような事は説明が出来ないことになる。しかしながら、簡便に結果が出る点や、分類による検索でヒットしなかったような結果が出る場合もあるので、従来の分類やキーワード検索を補完するような形で使用するケースもある。

3.4 引用解析

論文の世界では引用情報は、その論文を評価する一つ

のファクターとなっている。すなわち多く引用される論文は価値が高いと一般に言われている。同様のことが、特許でも言える可能性がある。主として米国特許において、引用解析が可能なデータベースが数種類ある。例えばFOCUST、Aureka、Delphion等である。引用情報を使用して何が出来るかについては、知財情報検索委員会でも継続して検討中であるが、例えば被引用回数（ある特許が後の特許から引用された回数）は、ある分野における技術的に重要な特許がどれであるか判断するための一つの材料となりそうである。また特許の評価と言った意味でも、やはり判断材料の一つとなりそうである。図1に引用関係のツリー表示の例を示す。また、日本特許においても引用解析ができるツールが出てきている。FOCUST-J、WIPS、PATENT・NET、StraVisionなどであるが、これらはいずれも審査官引用を利用している。

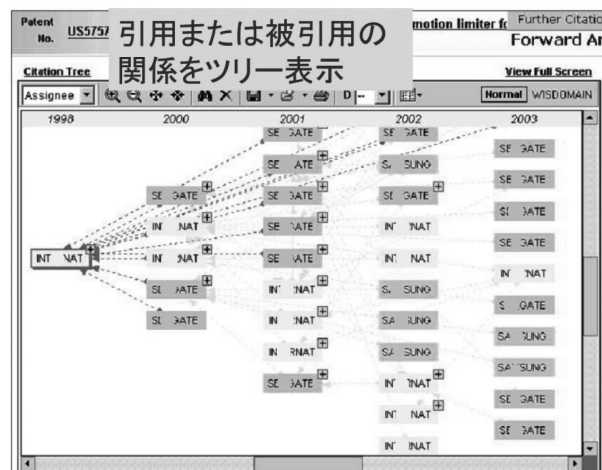


図1 引用関係ツリー表示の例（FOCUSTの例）

3.5 テキストマイニング+ビジュアル化

テキストマイニングは、大量の特許文献の文書データを様々な観点から分析し、役に立つ知識や情報を取り出し、利用する技術であるが、この検索結果の傾向をビジュアルに表現する技術を付加することで、頻出度の高い

ワード間の関連性をマップやグラフとして表示することができる。このようなものとしては、Aurekaのテーマスケープ等が有名である。この技術も肝心なマップ作成のアルゴリズムが公開されていないため、思ったようなマップにならなかった場合に何が原因なのか、またどうすればいいのかが分からないという問題がある。ただし、うまく使えば特許ポートフォリオの比較等に利用できそうにも思われる。この辺の技術は、今後まだまだ発展するものと思われるので、常に情報収集しておく必要があらう。なお、図2にAurekaのテーマスケープの例を示す。

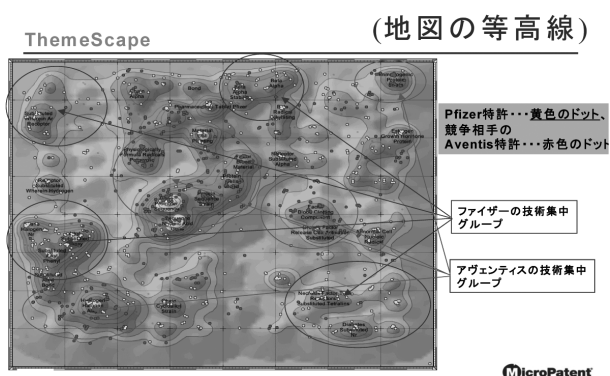


図2 Aurekaのテーマスケープの例

3.6 インハウスからASPサービスへ

高度な検索とか高度な解析のような具体的なサービスではないが、新しいサービスの提供形態としてASP (Application Service Provider) サービスが始まっている。このサービスは従来のインハウスデータベースのように自社分類やメモをつけたりした独自データを自社のサーバー (社内、あるいは社外) に持つことができ、その一方で検索用の特許情報データは社外の共有サーバーのものを利用することで、データサーバーのメンテナンスやデータの投入作業、あるいはシステムのバージョンアップ等の労力を省くことができるサービス提供の形態であり、従来のインハウスデータベースと商用オンラ

インデータベースの良いところ取りしたものといえる。また、これら最近のサービスでは、課金方法についても従量制から固定制に変わってきている。この事は非常に重要な内容を含んでいる。すなわち、固定料金制の故に大量の検索結果を例えばCSV形式等でダウンロードして、このデータを利用して特許マップを作成する等の解析作業をデータ入手のコストを気にせずに行うことができる点である。これまで、従量制でデータ入手のコストを気にしたり、躊躇していたことを思えば、このような解析をするための環境が整ってきたともいえよう。

3.7 エンドユーザーサーチ

従量制の商用オンラインデータベースから固定制のデータベースに変わってきて、エンドユーザーにデータベースを開放する企業が増えてきている。これはIPDLや各国特許庁が無料で特許情報データベースを提供している事と無関係ではない。研究者や技術者がこれら無料のデータベースを使用し、自然より利用しやすいツールを求めようになってきた。また課金が固定制になって管理しやすくなった点もこの理由の一つにあげられるだろう。全体的には検索が高機能化すると同時に低価格化がはかられ、またその一方で自然語で検索できる全文検索や概念検索が出てきた事によって、エンドユーザーが感じる検索の敷居が低くなってきているのかもしれない。しかしながら、既に述べたように全文検索の難しさや概念検索の現状については、よく理解した上で使用させる事が必要である。エンドユーザーサーチを導入した場合の教育や講習をどうすべきか、きちんと計画を立てて実行することは非常に重要である。

4 おわりに

「知的財産戦略大綱」に述べられているように、知的創造サイクルを早く大きく回すために、企業活動の各段階で、それぞれの段階に応じた特許情報がタイムリーに提供される必要がある。このように知的創造サイクルを回すのに役立つ特許情報こそが、価値ある特許情報といえる。また2章で述べたように、審査官のサーチツールが今後、無償で提供される等により民間の商用データベースは、今以上に高機能化やあるいは分析・解析機能が付加されていくであろう。情報を扱う人は、これらの新機能や解析機能等を使うことによって、本当に価値ある特許情報が得られるのかを確認し、使いこなす必要がある。また解析ツールを使いこなして何らかの結果が得られたとしても、最終的にその結果をどう読み取るか判断するのは結局人間であり、読み取り判断する能力を養う必要がある。またこのようにツールを使いこなす判断ができる人材を育成することも必要である。このように今後、特許情報を扱う人は単にサーチするだけでなく、人材育成にも力を注ぎ、その一方で、特許情報を元にした高度な解析を行える実力を養う必要があるであろう。

参考文献

- (1) 特許庁業務・システム最適化計画（改訂版）
http://www.jpo.go.jp/torikumi/system/system_optimize_re.htm
- (2) イノベーション促進のための特許審査改革加速プラン2007（AMARIプラン2007）
http://www.jpo.go.jp/cgi/link.cgi?url=/torikumi/hiroba/sinsa_kaosku.htm
- (3) 知的財産推進計画2007
<http://www.kantei.go.jp/jp/singi/titeki2/kettei/070531keikaku.pdf>
- (4) 産業財産権情報利用推進委員会最終報告
http://www.jpo.go.jp/shiryoutoushin/kenkyukai/pdf/riyousuisin_houkoku.pdf