

海外特許を用いた パテントマップ分析の主な注意点

鈴木 伸一郎*

抄 録 海外特許によるパテントマップ分析のためには、言語上の問題をはじめ、利用できる情報検索システムの選定、知的財産権制度とその運用、産業活動に関する情報等、国内特許の分析にはない新たな能力とアプローチが必要となります。本稿では米国、中国、欧州（EPO及びドイツ）を中心として、海外特許のパテントマップ分析に必要な母集団の構築、データの洗浄等について注意すべき事項と対応法について説明します。

目 次

1. はじめに
 2. パテントマップ分析の流れ
 3. 分析における主な注意点と対応
 3. 1 母集団の構築
 3. 2 データの洗浄
 3. 3 日本語翻訳
 3. 4 詳細分析
 4. おわりに
- ① グランドデザイン（基本計画の策定）
 - ② 母集団の構築（特許情報検索による特許文献データの収集と洗浄）
 - ③ 母集団データを用いた統計分析・指数分析
 - ④ 母集団データの複数観点による詳細分析
 - ⑤ テーブル・チャート等による可視化
 - ⑥ 総合評価と報告書の作成

1. はじめに

1968年に日本国特許庁が「明日をひらく特許」を発表してから今年で50年となります。特許と技術開発の結びつきを国際的側面から分析したこの報告書は、グローバル化が更に進んだ今日の経済活動においても不可欠なものとなっています。本稿は、海外特許、特に米国、中国、欧州（EPO及びドイツ）の特許情報を用いたパテントマップ分析やパテントランドスケープ分析（以下特許情報分析と呼びます）を行う際の留意点とその手順について説明するものです。

2. パテントマップ分析の流れ

特許情報分析は以下の手順で進められます。

3. 分析における主な注意点と対応

3. 1 母集団の構築

(1) 母集団構築の手順

特許情報分析においてはその基礎となる「特許情報の塊」は「母集団」と呼ばれます。母集団は、基本計画に定めた利用目的に合致したものでなければなりません。同時に重要な情報の漏れや、ノイズ情報を多数含めば、分析結果の信頼性と分析者の評価を著しく低下させることとなります。

18世紀に確立された特許分類システムは、特許出願の性質上避けることができない用語法的不安定性に対応するため、当初から発明の主題

* 弁理士 Shin-Ichiro SUZUKI

事項に注目し、審査資料の整備や審査官の指定の基本的ツールとしてきました。更に全く新しい発明の出現に対応するため、早い時期から「他に分類されないもの」や「雑」というクラスを用意しました。この考え方は、特許情報分析における「漏れのない情報収集」というサーチ戦略においても基本となりました。一方、分類システムには「分野横断的な検索」ができないという弱点があります。これを補うために新たに開発されたのがファセット分類方式・ハイブリッド方式でしたが、個別の「文献の分析」や「分類付与」には膨大な労力を必要としたことから、その普及は限定的なものに留まりました。

情報処理技術が急速に進展し、テキストベースのデータベースが構築され、自然語によるテキスト検索が実用化すると、分類システムとテキスト検索のかけ合わせによる絞り込みが可能となりました。この方式は先行技術調査では一定の成果をあげていますが、特許情報分析においては漏れの可能性が懸念されています。

この結果、現在の特許情報分析では、国内外を問わず、分類システム検索とテキスト検索を並行して実施し、この検索結果を合体し、重複を整理した後に目視によりノイズを除去するという手法が一般的とされています。

(2) 利用する分類システム

特許分類システムはその国の文化を反映するものです。極寒の地域では食品が凍ることを防ぐ冷蔵庫が必要となりますが、その他の地域ではこのような分類は必要としません。このため「海外特許の調査は、それぞれの国で基本分類システムを使用すべし」という原則が生まれ、適用されてきました。この原則は国際特許分類(IPC)が世界共通の特許分類となった現在も、専門家の常識となっています。

2014年に米国知的財産権者協会(IPO)が特許情報ユーザグループ(PIUG)の協力を得て

発表したパテントランドスケープ実態調査報告書¹⁾は、特許情報分析のための検索ツールとして多くの支持を得ているのがIPCであったことを明らかにしています(表1参照)。

表1 特許情報分析に用いる検索ツール

	IPC	ECLA or CPC	USPC	FI
強く同意	37.3%	28.4%	11.9%	1.5%
同意	38.8	35.8	40.3	14.9
その他	23.5	35.9	47.8	83.6

出典：「IPO Patent Landscape Survey Report」(2014)

この流れを変えるかもしれないとみられているのが、2013年に欧州特許庁(EPO)と米国特許商標庁(USPTO)の協力の下で誕生したCPC(Cooperative Patent Classification)です。CPCは、EPOの特許分類システム(ECLA)をベースに、EPO審査官専用分類ICOと、米国特許分類(USPC)のXRACsやdigestsを組み入れて開発されたもので、2018年1月現在、USPTO及びEPOの発行する特許文献データの99.9%、世界知的所有権機関(WIPO)の発行するPCT文献データの99.7%で利用可能とされています²⁾。更に、欧州特許機構加盟国、中国、韓国、イスラエル、ブラジル、チリ、カナダ、メキシコ、ロシア、オーストラリア等、その利用国は急速に拡大しています。

しかしながら、特許情報分析では最新の情報だけでなく、過去に発行された情報も重要な意味を持ちます。2015年にCPCが基本的特許分類システムとなったUSPTOは、既発行文献の再分類もほぼ完了しています。中国では2014年以降に発行された特許、実用新案データへのCPC付与が開始されましたが、既発行文献の再分類は手つかずのままといわれています。ドイツではEPOの発行した特許文献のほとんどにCPCが付与されていますが、ドイツ特許商標庁(DPMA)の発行する特許、実用新案文献への

適用は進んでいません。DPMAのホームページは、ドイツの特許情報検索の基本的特許分類システムは、依然としてIPCであるとしています。欧州各国の特許情報分析を行う場合には、CPCだけでなく、IPCを利用することが必要となる場合もあることに留意する必要があります。

(3) 利用する検索システム

USPTO, 中国国家知識産権局(SIPO), EPO, DPMA, WIPO等が提供する世界の特許情報サービスのほとんどがIPCによる検索が可能となっています。しかしながら、検索結果を母集団とするために、そのまま表計算ソフトやデータベースソフトにエクスポートできるものは限られています。

PATENTSCOPEはWIPOが提供する無料の特許情報検索サービスで、PCT国際公開と米国、中国、ドイツをはじめ、世界51国・地域の約7億件の特許文献をカバーしています³⁾。検索メニューとして、簡易検索、詳細検索、構造化検索、多言語検索 (CLIR) 及び化学構造式検索が用意されており、検索結果は図面入抄録リスト等フォーマットを選んで表示することができます。主要な書誌事項については、そのまま表計算ソフトフォーマットでダウンロードすることができます。ドイツや中国の実用新案出願についても検索・出力が可能です。このサービスの特徴のひとつに、出願人、発明者等のランキング表示機能があります。この機能は、ヒットした文献の上位出願人や発明者を表示するものですが、後述の「名寄せ」は行われていません。また翻訳機能を備えており、ノイズ文献の除去や分析の際に利用することができます。

Espacenetは、EPOが提供する無料の特許情報サービスで、EPOが発行する特許文献だけでなく、世界中の1億を超える特許文献情報をカバーしています。検索メニューとして、Smart Search, 高度検索, 分類検索が用意されていま

すが、IPC検索は分類検索ではなく、高度検索のページで行う点に留意が必要です。検索結果は書誌情報を含む特許文献リストとして表示され、表計算ソフトに対応したフォーマットを選択し、エクスポートすることができます。検索結果画面に表示される「Download covers」ボタンをクリックして、ヒット文献の第1ページ集をダウンロードすることができます。言語上の問題がなければ、これを用いて目視によるノイズ除去にも利用することができます。

Espacenet利用の最大のメリットは、CPCが利用できることにあります。検索メニューから「分類検索」を選択して表示されるCPC分類表を用い、最適なグループを選択すれば、そのまま検索が可能です。更に高度な検索を行うためには、分類記号を「高度検索」のページに貼り付け、発行特許庁や公開年等を特定した検索を行うことも可能となっています。

(4) テキスト等による補完検索

特許情報検索サービスの多くは、分類検索機能だけでなく、出願人名や明細書文のテキスト検索機能も備えています。特許文献は原則として発行庁の公用語により発行されますから、翻訳ミスを避けるためにテキスト検索は公用語による検索が中心となります。また、世界共通の用語と考えられていたものが、一部地域では別の意味に用いられることもあります。このため、テキスト検索を行う場合には、専門家等の助言も有効です。

WIPOのPATENTSCOPEは無料で多言語検索サービス (CLIR) を提供しています。CLIRは統計的解析システムを利用した拡張検索機能で、特定の言語で単語やフレーズを入力することにより、多言語の類義語拡張検索を行うことができます。このシステムでは、多言語検索用辞書として、WIPOの特許文献対訳コーパスが用いられ、日本語のほか、中国語、韓国語、英

語、ドイツ語等12カ国語に対応しています。

(5) 引用頻度データの取り込み

特許情報分析では、発明或いは特許の評価のために、その後の出願に引用された回数（引用頻度）が広く用いられます。引用頻度は時間の経過とともに変化するものですから、一般的に公開時期が早いものほどその高い値を示し、新しいものは低い値を示すものとなります。引用された出願人との関係も注目されます。同じ出願人のもの（自己引用）か、別の出願人のもの（他者引用）であるのかを識別する必要があります。

USPTOのFull-Text DatabaseやEPOのEspacenetでは、ヒット文献ごとに用意された「Referenced By」又は「Citing Documents」をクリックして表示することができます。

(6) パテントファミリの利用

特許制度の原則のひとつに、権利行使はそれぞれの国（地域）で登録されたものでなければならないという「属地主義」があります。このため古くからグローバルビジネスを目指す企業や研究機関は、多数国への出願により、全世界をカバーする権利化を進めてきました。このような同一の発明（優先権）に基づく出願群はパテントファミリと呼ばれます。一般に海外への特許出願は国内出願以上の費用を必要とするため、米国では、早い時期からパテントファミリ数を出願の経済的価値を評価する手段として注目してきました。

これとは別に、ひとつの出願からも様々な特許文献情報が誕生します。これらの情報群は拡張パテントファミリと呼ばれ、審査経過の評価により発明評価の価値を示すものとなるだけでなく、重複を回避し、文献量を圧縮するためのツールとしても利用されます。

一方で、優先権主張に基づくパテントファミ

リを重複回避の目的で使用しますと、出願人の国際戦略の分析ができなくなりますので、留意が必要です。

3. 2 データの洗浄

(1) ノイズ文献の除去

多くの場合、分類システムを用いた特許情報検索には、関連のない特許文献（ノイズ）が多数含まれることとなります。これを排除する作業がノイズ除去と呼ばれるもので、その基本は、専門家の目視によるノイズ除去です。多くの場合、この作業はディスプレイ上に表示される個別の検索結果を用いて行われます。

言語上の問題が少ない場合には、海外特許についても同様にノイズ除去をすることができます。この際、代表図面や、機械翻訳を利用することも有効です。このためには、前述のPATENTSCOPEなどの検索システム付属の翻訳機能を用いることが効果的ですが、ブラウザの翻訳機能を用いることもできます。

ノイズ除去作業は通常1回で完了するものではなく、分析の各段階でノイズが発見される度に、その都度行わなければなりません。

目視によるノイズ除去は、別の意味で特許情報分析の重要な役割を持っています。この作業は解析者が母集団の全データを時系列的に把握できる数少ない機会となりますから、これを利用して、母集団に含まれる発明、時間的変化、主導する出願人等を把握し、詳細分析の切り口を知る絶好の機会となります。

(2) 出願人名等の統一（名寄せ）

特許情報は最新の技術情報、権利情報であり、かつ経営情報であると高く評価される一方、表現上の揺らぎや統一されていない用語法・翻訳の乱れなどにより「生のままでは使えない情報」とも言われます。これは、合併・分社・M&A等による出願人名等の変更、フルネーム表記と

省略表記、原語表記と翻訳表記の混在、グループ化すべき出願人の分散、異なる標準（地域コード等）の利用、単語の前後或いは単語中に含まれる空白、点、制御記号の存在、ひとつのセルに記載された複数の出願人・発明者などによるものと考えられています。

「洗浄された出願人」はIPOの実態調査報告書においても特許情報分析における最も重要なカテゴリとされるものですから慎重に扱うことが求められます。特許情報プロバイダの中には、すべての出願人にコードを付し、データベースを常に最新となるようメンテナンスしているものや、サービスにデータの洗浄機能を備えているものもあります。

これらの機能が利用できない場合には、自ら洗浄することが必要となります。

セル中の制御記号や空白については、表計算ソフトの関数を用いて除去することができます。名称変更等への対応については、ピボットテーブルによりアルファベット順の一覧表とし、知識と経験を有する専門家の目により修正することになります。

また、都市名を冠した企業・大学・研究機関では意図しないグループ化が行われることもありますので、このような場合には分離・整理することが必要となります。

目視による洗浄は、表計算ソフトを利用して出願人名のアルファベット順にソートして行われます。この際、海外特許については、地域の専門家や出願人等情報が十分でない場合も少なくなく、予め必要な情報を用意することも必要です。

(3) 洗浄ソフトウェアの利用

WIPOが2016年に発表した「WIPO Manual on Open Source Tools for Patent Analytics」⁴⁾は特許情報分析にも利用できる無料の洗浄ツール「Open Refine」を紹介しています。このソ

フトウエアは、表計算ソフトのデータに蓄積された出願人名等データを数種のアルゴリズムを用いてチェックし、同一出願人である可能性のあるエントリを提示するもので、利用者がふたつのエントリ（出願人等）が同一のものと判断し、マージすることを指示した場合に限り、複数のエントリをひとつのエントリに統合することができものです。

3. 3 日本語翻訳

(1) 翻訳が求められるタイミング

いうまでもなく、言語の違いは海外特許の分析において、最も大きな障壁となるものです。特許文献については、発行機関の公用語が用いられるのが原則ですから、必要な言語は英語に限らず、中国語、ドイツ語など多様なものとなります。翻訳文の有無が致命的なものとなるのが、母集団からのノイズを落とす作業、特許情報全文を読み込んで解析する作業です。また分析結果への添付が不可欠といわれる主要(重要)特許の要旨リストについても日本語化されていることが好ましいと考えられています。これらの作業前には、易読言語による翻訳文の作成が必須のものとなります。

(2) 機械翻訳の利用

特許情報の翻訳には多くの費用と有能な翻訳者を必要とします。そこで、ノイズ文献除去等の作業では無料の機械翻訳サービスを利用することが効率的と考えられます。

前述のとおり、WIPOのPATENTSCOPEでは、抄録文や書誌情報の翻訳ができ、代表図面とあわせて一覧表として利用することができます。更に、OCRで読み取ったデータを用いて明細書全文のテキストについても機械翻訳による日本語として利用することができます。翻訳は技術分野毎の辞書を備えたWIPO独自の翻訳システムと、Google翻訳システムをはじめ各種

の翻訳システムを利用することが可能で、各分野や原語に最適なシステムを選択することができます。

EPOのEspacenetも翻訳機能を備えています。英語以外の言語については直接日本語に翻訳することができず、また代表図面と抄録文の同時表示などの問題も残されています。

機械翻訳結果が、そのまま利用できない場合であっても、その後のマニュアル翻訳の箇所を特定するためにも利用することは可能です。

これらの機能を活用することにより、効率的な分析と報告書の作成が可能となります。

3. 4 詳細分析

最初に紹介したIPOの実態調査報告書は、特許情報分析で用いるカテゴリとして、「利用者のマニュアルによるカテゴリ」が7割近くの支持を得ています。このことは、海外においても詳細分析の主流が、特許文献を読み込み、独自分類を付与し、体系化して行われることを意味するとも考えられます。

海外特許の分析の場合、この作業のために多数の有能な解析者を確保し、解析の時間を確保することは大きな負担となります。この作業を効率化する方法として利用されるのが、詳細分析の対象文献を絞り込む方法です。母集団には様々な特許が含まれていますが、注目される出願に絞り込むことにより、その作業負担を大幅に軽減することが可能となります。

いまひとつは、文献の読み込む範囲を特定する方法です。USPTOでは、分類の付与にあたって、明細書に記載された発明の目的、発明の解決すべき課題、課題の解決手段、発明の効果等を特定することから行うことを指示しています⁵⁾。これらの項目は必ずしも明細書に明示的に示されているものではありません。例えば、

発明の解決すべき課題については、「従来技術との関係」や「発明の効果」から把握することができる場合があります。

前述のIPOパテントランドスケープ実態調査報告書は、現在注目されている「アルゴリズムによる特許評価」についての利用者の見解が大きく分かれたことを指摘しています。そして、現在のところ特許情報分析におけるその有効性は不明と結論づけています。

4. おわりに

海外特許の分析には更なる課題も残されています。そのひとつが新興国を中心として急増する実用新案の問題です。特に2016年の出願が150万件に迫った中国については、その知的財産権としての重要性と作業負担の面から、新たな対応が求められることとなります。

注 記

- 1) Daniel Aleksynas, IPO Patent Landscape Survey Report pp.56-60 (2014) IPO
- 2) EPO・USPTO, 5th CPC Annual Meeting pp.6-11 (2018) WIPO
- 3) WIPO PATENTSCOPE The User's Guide (updated July 25th, 2017) WIPO
- 4) Paul Oldham, WIPO Manual on Open Source Tools for Patent Analytics (2016) WIPO
- 5) Richard Lehman, CPC Block B
<https://www.uspto.gov/video/cbt/cpc-blockB/index.htm> (参照日2018.5.5)

参考文献

- ・ Anthony Trippe, Guidelines for Preparing Patent Landscape Reports (2015) WIPO
- ・ USPTO, CPC for USPTO
<https://www.uspto.gov/video/cbt/cpc-overview/index.htm> (参照日2018.5.2)

(原稿受領日 2018年5月28日)