

特許調査に役だつ特許情報分類技術

Patent Information Classification System for Patent Research

平 博司 松本 茂

■ TAIRA Hiroshi

■ MATSUMOTO Shigeru

近年のプロパテント化(特許重視)の流れを受けて、各企業においても、特許出願などのいわゆる知的財産関連活動(以下、知財活動と略記)が、これまで以上に盛んになってきている。この知財活動において、重要なものの一つに“特許調査活動”がある。特許調査は、自社及び他社の特許を検索し、分析し、可視化する活動であり、技術トレンドの把握や製品開発の方向性の検討、他社特許の侵害防止(パテントクリアランス)など、企業活動を行ううえでなくてはならないものとなっている。しかし、一般に特許調査は多大な時間と労力を要するものであり、その負荷をいかに軽減するかが大きな課題となっていた。また、調査だけで終わらせず、いかに新しい技術及び特許の創造に結びつけていくかも重要なポイントである。

東芝ソリューション(株)は、このような課題に対応するために、特許文書を自動的に分類する技術を研究開発している。この技術によって、特許調査の効率化が図られ、企業の知財活動が活性化されることが期待できる。

In line with the increasing emphasis placed on patents, companies are promoting intellectual property (IP)-related activities typified by patent applications. Patent research activities, comprised of research, analysis, and visualization of patent movements, are one of the most important IP-related activities because they reveal technological trends and product development orientations while helping to prevent potential infringements by other companies (i.e., patent clearance).

On the other hand, IP-related activities require a great deal of time and effort, and the reduction of human resources devoted to such activities is a significant management issue. A further issue is that IP-related activities should not be confined to research activities but should serve as an impetus for the creation of new technologies and patents.

In response to these issues, Toshiba Solutions Corp. is developing a technology to classify patent documents. This technology is expected to speed up patent research work by companies and enhance their IP-related activities.

1 まえがき

15世紀にベネチアで生まれた特許法は、ガリレオ ガリレイなどの偉大な発明者をその歴史に刻みながら、今日まで発展してきた⁽¹⁾。2004年の全世界の出願件数の総数は約160万件であり、近年の世界的なプロパテント化の流れを受けて、過去20年間で約2倍に増加している⁽²⁾。特にわが国は、世界の特許出願の約1/4を占める“知財大国”であり、政府が進める“知財立国”戦略ともあいまって、企業の内外における特許の価値はこれまでに高く高まってきている。

企業における特許に関する活動は様々あるが、大別すると、以下の二つに分けられる。

- (1) 特許管理 新しいアイデアを考え、特許として出願し、特許権を取得し、権利を維持管理し、必要に応じて権利行使していく活動
- (2) 特許調査 自社や他社の特許を調査し、技術トレンドの把握や製品開発の方向性の検討、他社特許の侵害防止や新しい特許や技術の発案に役だてる活動

ここでは、このうち特許調査に焦点を当て、東芝ソリューション(株)における研究開発の取組みを紹介する。まず一

般的な特許調査プロセスを俯瞰(ふかん)し、内在する課題を整理した後、それらの課題解決に向けた特許クラスタリング及びマトリックスマップ生成の技術開発について述べる。

2 一般的な特許調査プロセスとその課題

特許調査とひと言で言っても、その目的や活動内容は様々である。具体的な例としては、表1のような活動がある。

ここで示した特許調査のうち、現在企業内で高頻度に行われているものは侵害防止調査、特許動向調査、そして新規性調査であり、これらは一般に以下のようなプロセスで行われている。

- (1) 検索 大量の特許公報群から、目的としている調査に必要な特許をある程度絞り込むプロセス
- (2) 分類 検索された特許群に対して、類似した公報ごとにグルーピングして、データのいっそうの絞込みや注目すべき部分の発見を行うプロセス
- (3) 可視化 絞られた特許を対象に、抄録作成やグラフ生成など、目に見える形でのまとめを行うプロセス
- (4) 分析 可視化されたデータを閲覧し、該当するデー

表1. 様々な特許調査活動
Types of patent research

活動名	目的	実施される時期
侵害防止調査 (パテントクリアランス)	他社の特許を侵害するのを防ぐための調査	研究開発時、 事業化時
特許動向調査	他社の特許動向を把握するための調査	研究企画時、 事業企画時
新規性調査 (先行技術調査)	他社の特許を把握し、提案しようとする特許の差異化を図るための調査	特許提案時
無効化調査	他社の特許を無効化するための調査	要注意特許発見時
侵害発見調査	他社が自社の特許を侵害していないかを把握するための調査	有力特許取得時、 関連他社製品発見時
知財経営調査	自社や他社の特許動向と、企業の研究開発費や資本金などの関係を分析するための調査	経営判断時

タからどのようなことが言えるかを判断するプロセス

従来、上記の(1)～(4)のプロセスは、すべて手作業によって行われていた。しかし、1990年代に起こったパソコンの爆発的な普及や高性能化、記憶装置(HDDなど)の低価格化、そしてインターネット技術の急速な発達により、主に検索と可視化に関しては、IT(情報技術)による支援が急速に進むこととなった。

例えば、検索に関しては、従来の“手めぐり検索”に代わり、IPDL(Industrial Property Digital Library)^(注1)や企業内の特許検索システム、そのほかの民間検索会社が提供する“機械検索”がほぼ主流となっている。

また、可視化に関しても、企業の技術者や特許調査を行うユーザー(以下、ユーザーと略記)の端末の高性能化により、市販の表計算ソフトや特許専用ソフトを用いて、リスト化やグラフ化が容易に行えるようになってきている。

しかし、分類に関しては、多くの場合いまだに手作業によって行われているのが実情である。ここで、一般的な特許の分類作業のプロセスを以下に示す。

- (1) 分類軸の検討 どのような観点で特許群を分類するかを検討する(例:課題, 技術, 効果)。
- (2) 特許の分類 (1)で決めた観点に沿って、互いに類似した特許をひとかたまりとして、特許群を分類する。
- (3) 分類名の検討 (2)で分けられたひとかたまりの特許群について、互いに識別可能なわかりやすい名前をつける(例:速度の向上, 精度の向上)。

上記(1)～(3)の分類プロセスを手で行う際には、主に以下のような課題がある。

- (1) 分類基準の非一貫性 分類を進めていくにつれて、初期に考えた基準からずれていってしまい、結果的に、全体を通した一貫した指標で再分類せざるを得なくなる。
- (2) 分類基準の非中立性 人は、それぞれの思いや考

(注1) 独立行政法人 工業所有権情報・研修館が運営する特許電子図書館。

えを内在しながら分類作業を行うため、その基準が個人の得意技術や価値観によるバイアスを受けてしまう。

- (3) 分類基準創出の困難性 何も指標がない状況から、手探りで特許の分類基準を考えていく作業そのものが困難である。
- (4) 分類作業の高コスト性 (1)～(3)の要因を含め、分類作業全体として、多くの時間と人員がかかる。

当社では、この特許分類に関する諸課題を解決するための研究開発を行っている。この分野に関してはNTCIR(NII Test Collection for IR Systems)^(注2)プロジェクトなどで先駆的な研究が続いているが、ここでは、当社のナレッジマネジメント製品であるKnowledgeMeisterTM(ナレッジマイスター)のクラスタリング技術を活用した研究内容について述べる。

3 コスト削減に効果のある特許クラスタリング

クラスタリングとは、類似するデータを“クラスタ”と呼ぶ部分集合に分ける、データ分析の一手法であり、このデータが文書である場合を一般に文書クラスタリングと呼ぶ。文書クラスタリングには様々な手法やアルゴリズムがある⁽³⁾が、その機能を端的に表現すると、“文書群を、あらかじめ分類体系や正解データを与えず(教師なし)に、内容的に均質なくつかのグループに自動的に分類する技術^(注3)である”,といえることができる。

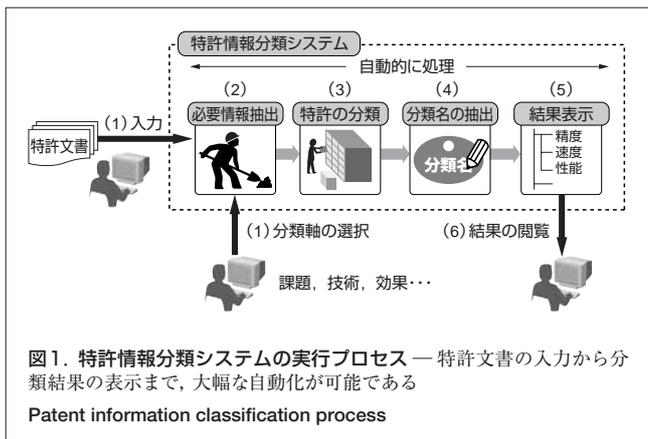
この文書クラスタリングに関する機能を持つシステムの一つとして、当社の“KnowledgeMeisterTM知識分析”(以下、KM知識分析と略記)がある。KM知識分析は、KnowledgeMeisterTMのサービス群を形成する五つの製品ラインアップの一つであり、テキスト情報の分析作業を自動化することで、短時間で傾向や要因を分析し、意思決定を支援できることを目指したものである。

当社では、このKM知識分析が持つ文書クラスタリング技術を、特許文書に対して適用した“特許情報分類システム(仮称)(以下、このシステムと言う)”のプロトタイプを構築した。このシステムを用いた特許分類プロセスのイメージを図1に示す。おおまかな流れを説明すると、以下のとおりとなる。

- (1) ユーザーは、分類したい特許文書群をこのシステムに入力すると同時に、それをどのような観点で分類したいかも併せて指定する。
- (2) 指定された分類軸に応じて、特許文書の中から必要な情報を抽出する。

(注2) 国立情報学研究所などの支援により開催されている、情報検索とテキスト要約や情報抽出などのテキスト処理技術の研究のいっそうの発展を図るワークショップ型共同研究。

(注3) 対になるものとして、ユーザーが分類体系や正解データをあらかじめ定義する“テキスト分類(教師ありの文書分類)”がある。



- (3) 抽出した情報を元に、特許文書を分類（クラスタリング）する。
- (4) 分類結果を元に、各分類グループ（クラスタ）に名称（分類名）や特徴的なキーワード群（特徴キーワード）を付ける。
- (5) 処理結果を表示する。
- (6) ユーザーは、結果を閲覧し、全体の把握や絞込みなどを行う。

上記の流れのうち(2)～(5)に関しては、このシステムによって自動的に処理される。

“洗濯機”関連の特許データを、“技術”の分類軸で分類したクラスタリング結果の例（一部）を図2に示す。

図2の閲覧結果を見ることで、ユーザーは“ドラム”や“ポンプ”、“モータ”など、洗濯機にかかわる主要な技術の構成をまず理解し、次に、興味を持ったグループ（クラスタ）を適宜読み進めることができる。

このシステムを用いることで、従来人手で分類を行う際に

分類名	特徴キーワード	特許件数
ドラム	ドラム 回転 中心 水平 軸 物	236件
横軸	横軸ドラム 式 形 各工程 回転	9件
連続 回転	連続ドラム 回転 中心 物 水平	6件
ネット	ネット 洗濯 袋 収納 でき 入れ	72件
ハンガ	ハンガ ネット 掛け 衣類 セー	6件
収納	収納 ネット でき 物 袋	11件
入れ	入れ ネット でき デリケート	10件
崩れ	崩れ ネット 不用 今 乾燥 10	6件
ポンプ	ポンプ 給水 洗濯 風呂 機 吸水	59件
吸水	吸水 ポンプ 風呂 機 残り 洗濯	13件
供給	ポンプ 供給 風呂 槽 残り 具備	8件
風呂	風呂 ポンプ 自吸 吸水 機 給水	16件
マイクロコンピュータ	マイクロコンピュータ 制御 子	12件
モータ	モータ 洗濯 駆動 機 回転	142件
インバータ	インバータ モータ 駆動 レス	25件

図2. クラスタリング結果の例 — “洗濯機”を構成する技術要素を自動的に抽出できる。

Example of conceptual clustering

あったいくつかの課題を軽減することができる。具体的には次のとおりである。

- (1) 分類基準が一貫して、中立である 機械的な分類を行うため、分類基準が全特許に対して一貫しており、また、非属人的である。なお、もしユーザーが恣意（しい）的に指定したい基準があれば、“重要語”として登録することで、結果に反映させることもできる。
- (2) 分類基準創出が容易になる 機械的に分類されたデータが既に存在するため、全体を眺めて雰囲気をつかんでから、適宜人手により微修正するというアプローチを取ることができ、一から人間が分類基準を考えるよりも簡単になる。
- (3) 分類作業のコストが下がる 上記(1),(2)及び、ある基準に沿った文書分類そのものをも機械的に行うため、分類の手間が削減できる。

更に、当社の特許クラスタリングは、一般的な文書クラスタリングと比較して、次のような特長がある。

- (1) 高機能 階層型クラスタリング、重要語や不要語の指定、及び一つの文書の複数クラスタへの分類が容易にできる。また、クラスタが多数になった際に、選択の判断材料とするために、各クラスタの特徴を抽出できる。
- (2) 高精度 特許用にカスタマイズした重要語や不要語の辞書を用いることで、特許文書に対してノイズの少ない高精度なクラスタリングを実現している。

このシステムを用いることで、分類にかかわるユーザーのコストを引き下げることができると同時に、ある特許分野の包括的な理解や学習が、これまで以上に容易に行えるようになる。

4 技術的な知見の発見につながるマトリクスマップ

更に当社では、このクラスタリング結果を組み合わせたマトリクスマップに関する機能開発に取り組んでいる。

マトリクスマップとは、縦軸と横軸に特許の分類名を、また、軸の交点に該当する特許の数や特許の詳細情報を、それぞれ配置した特許マップである。表形式を採用することが多いが、わかりやすくするために、バブルチャートなど別の形式で表されることもある。ある分野の公開特許を、特許が解決しようとする“課題”と、その課題を解決するための“技術”の2軸に人手で分類し、件数を集計したマトリクスマップの例を表2に示す。

マトリクスマップは、特許を複数の視点から同時に観測できることに大きな特長がある。例えば、表2のマップであれば、特許を解決しようとする課題と、課題を解決するための技術の二つの側面から見ることができる。この結果、例えば、“業務処理の効率化”という課題を、“データベースの統

表2. マトリックスマップの例

Example of conventionally prepared patent matrix map

(単位：件)

		課題				
		システムの改善	創造の支援	安全性の向上	検索の高度化	業務処理の効率化
技術	キー情報の活用	データベース			2	
	システムの構成	CD-ROMからMO	1			
		システム全体				1
		スキャナ読取り	1			
	データベースの活用	ディスクキャッシュ	2			
		データベースの統合化	1	1		5
	インターネット情報		1		1	
データ変換	外国語翻訳				1	
	暗号化			1		

MO：Magneto Optical disk (光磁気ディスク)

*特許流通支援チャート⁽⁴⁾などを元に作成。

合化”という技術で解決する特許は多いが，“外国語翻訳”や“暗号化”といったデータ変換技術で解決する特許はない、といったことがわかり、新商品の企画や新技術の検討、新しい特許の提案などに結び付けることができる。

当社は、このマトリックスマップの作成について、3章で述べた特許クラスタリングを活用した機能開発を行っている。具体的には、マトリックスマップの二つの分類軸について、特許クラスタリングの結果から自分が分析したい分類名をそれぞれ選ぶだけで、自動的にマップを生成することが可能で

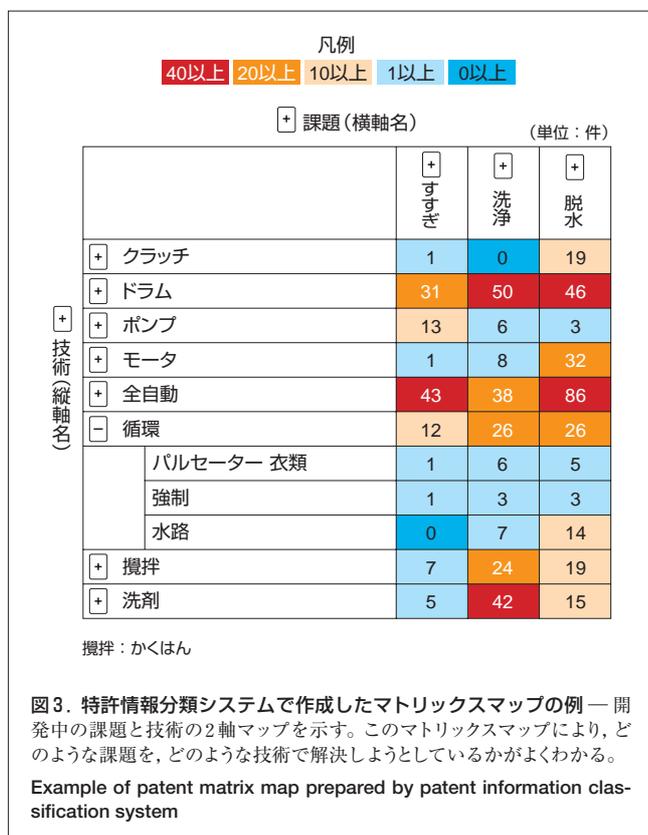
ある。例えば、先の洗濯機に関する特許について、技術と課題の観点でクラスタリングした結果を元にマトリックスマップを作成した例を図3に示す。ユーザーは、このマトリックスマップ(図3)を見ることで、“すすぎ”、“洗浄”、“脱水”の分野の課題に対し、“ドラム”や“モータ”などどのような技術で対応した特許が多いのか、全体の傾向をひと目で把握することができる。そのうえで、自分たちが開発している技術がどのような課題の解決に適用可能かを見つけたり、青や薄青で塗られた特許出願件数が0あるいは少数の、いわゆるブルーオーシャン(未知の市場空間)となり得る領域を発見し、新しい事業や研究の候補として、検討の対象とすることなどができるようになる。

5 あとがき

ここで述べた技術については、既にプロトタイプを用いて当社での実験的運用を始めており、将来的には顧客への適用を実現させたいと考えている。ユーザーからのフィードバックを元に、機能や性能、精度、処理速度などを更に高度化させ、特許調査の負荷軽減や、新たな知見の発見に役立つサービスとして、広く活用されることを目指していく。

文献

- 竹田和彦. 特許の知識【第7版】. ダイアモンド社, 2004, 543p.
- The World Intellectual Property Organization (WIPO). WIPO Patent Report: Statistics on Worldwide Patent Activity (2006 Edition). < http://www.wipo.int/ipstats/en/statistics/patents/patent_report_2006.html >, (参照 2006-10-20).
- 岸田和明. 文書クラスタリングの手法: 文献レビュー. Library and Information Science. 49, 2003, p.33-75.
- 独立行政法人工業所有権情報・研修館. “電気10CRM・知的財産管理システム”. 特許流通支援チャート. < <http://www.ryutu.ncipi.go.jp/chart/H14/denki10/frame.htm> >, (参照 2006-12-22).



平 博司 TAIRA Hiroshi

東芝ソリューション(株) IT技術研究所 研究開発担当。
ビジネスドキュメント処理技術及び戦略的特許活動支援技術の研究・開発に従事。情報処理学会, 人工知能学会会員。
Toshiba Solutions Corp.



松本 茂 MATSUMOTO Shigeru

東芝ソリューション(株) IT技術研究所 研究開発担当主務。
ビジネスドキュメント処理技術及び戦略的特許活動支援技術の研究・開発に従事。
Toshiba Solutions Corp.