

特許調査支援システムの試作と評価

Trial manufacture and evaluation of patent investigation support system

平 博司*¹
Hiroshi Taira

松本 茂*¹
Shigeru Matsumoto

*¹ 東芝ソリューション株式会社
Toshiba Solutions Corporation

Patent research activities 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. In response to these issues, we are 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. はじめに

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、自社および他社の特許を検索、分類、可視化して、技術トレンドの把握や製品開発の方向性の検討、さらには他社特許の侵害防止(パテントクリアランス)につなげる、いわゆる「特許調査活動」が盛んにおこなわれるようになってきている。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る。

このようにコストがかかる理由の一つは、パテントクリアランスや特許マップ作成といった多くの特許調査タスクにおいて、数千～数万件という大量の難解な特許文書を読みきらなければならないという点にある。これは、わずかな調査漏れが莫大な損害賠償金につながってしまうことなどから、「きちんとした」調査を行う必要があるという特許調査の性質に起因している。

そこで筆者らは、特許を自動分類する機能をコアとした、特許調査支援システム(以下、本システム)を試作した。大量の特許群を自動分類することで、特許群の全体傾向を把握することができ、重点的に調査すべき特許を見つけたり、特許マップ作成時の分類項目の参考としたりすることが可能となる。

本論文では、本システムの概要と、実利用の観点から行った評価結果について概説する。

2. 特許調査支援システムの概要

本システムの機能構成概要を図1に示す。



図1 特許調査支援システムの概要

本システムは特許検索システムで粗く絞り込んだ特許群を対象に、公報中に記載されている項目に基づいて、特許文書を自動分類する。また、自動分類結果を組み合わせて、2軸マップを自動生成する。

2.1 自動分類機能の概要

本システムの自動分類機能は、「属性分類」、「ルールベース分類」、「単語クラスタリング」の3つからなる。これらは以下のような特性を持つ。

・属性分類:

特許公報中に含まれる「出願人」や「IPC」といった書誌情報を、「東芝」や「G06F」といった同一属性ごとにカテゴリ化する分類手法である。

・ルールベース分類:

ユーザ自身がどのような分類構造とするかを検討し、分類のためのルールをカテゴリごとに設定し、そのルールに従ってカテゴリに対して特許を分類する分類手法である。システムはユーザが決めたルールに従って、特許を自動分類する。なおルールは、各カテゴリに分類したい特許群の特定のパート(例:「要約の課題」など)に共通して含まれるキーワードを指定する形式で設定する。

・単語クラスタリング:

ルールベース分類と同じく公報中の「要約の課題」などに記載されているテキストを用いる分類であるが、分類構造をユーザが考える必要なくテキスト中の特徴語間の関係に基づき自動的に構築できる点に特徴がある分類手法である。

このうち、単語クラスタリングについて、2.2節で詳述する。

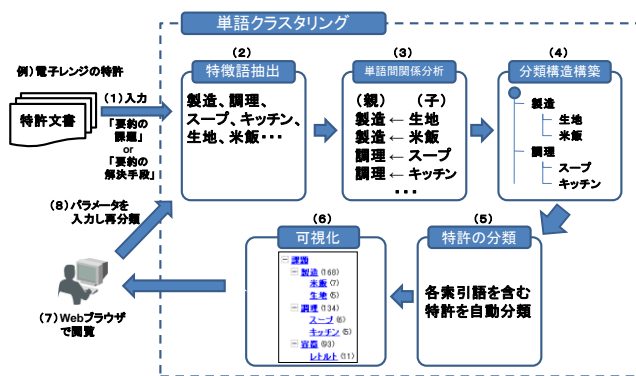
2.2 単語クラスタリングの詳細

文書群の全体傾向の把握の支援としては、文書相互の表現の類似性に基づきクラスタを生成する文書クラスタリング[岸田03]が用いられるケースが一般的である。この際、カテゴリラベル(クラスタラベル)は、一般に各カテゴリに割りつけられた文書群の重心ベクトルが用いられる。例えば、「受像機、電波、カラー」

をカテゴリラベルとして提示し、それにより「テレビ関連の特許群」がカテゴリに多く含まれていることを示す、というように、カテゴリ名とカテゴリ内の文書は比較的緩やかな関係で結ばれている。

一方、特許調査は調査漏れに厳しく、調査の確実性が強く要求される。このため多くの場合、「テレビ」というカテゴリ名には「『テレビ』という語を含む特許群」がすべて収集されていることが求められる。

このような特許調査の特性に対応した分類構造を自動生成するためには、特許群から有用な特徴語を抽出し、その特徴語ごとにカテゴリを生成し、生成したカテゴリに対し特徴語を含む特許を分類するアプローチが適切である。しかし、数千～数万の特許群に対しては、特徴語の個数は膨大なものになりがちである。そこで、特徴語間の関係に基づき、階層的な分類構造を生成するようにする方法を選択する。処理フローを図 2 に示す。



以下、図 2 の各プロセスを概説する。

- (1) **特許公報の入力**: 分類対象の特許公報群をシステムに入力する。分類の対象テキストは、特許公報中のテキスト記載部分であればどこでも可能であるが、試作システムでは「要約の課題」と「要約の解決手段」をそれぞれ対象とする。これは、特許文書の本質は「どのような課題をどのような技術で解決するか」というところにあるからである。
- (2) **特徴語の抽出**: 特許公報群から特徴語を抽出する。特徴語としては、TF(Term Frequency) や DF(Document Frequency)が中頻度の語が文書の高い識別能力を持つ性質を利用し[Salton 75]、これらの語を中心として採用する。ただし、「装置」「当該」といった不要語を削除しているほか、「要約の課題」については、「精度」や「性能」といったいわゆる「課題語」が出現しやすくなるようチューニングを行っている。
- (3) **単語間関係分析**: 抽出した特徴語間の親子関係を分析する。親子関係については、特徴語それぞれの頻度と、2 つの特徴語がともに出現する頻度(共起頻度)などを元に算出する。
- (4) **分類構造構築**: 分析した単語間関係を元に、各特徴語の階層的な構造化を行う。ただし、子階層の枝狩りを行うなど、妥当な分類階層となるように調整する。その上で、各特徴語をそれぞれカテゴリとした分類構造を構築する。

- (5) **特許の分類**: 最後に、特許の分類を行う。分類構造の各単語を含む特許(ただし、子階層については親階層の言葉も含むもの)を分類する。

ユーザは Web 画面上で分類結果を閲覧できる。

なお、クラスタリングはシステムが分類構造を自動生成するため、分類結果がユーザの意図と合わない場合がある。このような場合は、複数のパラメータを用いて再分類をおこなうことで、分類結果をチューニングすることができる。パラメータとしては、不要語や重要語、関連語などのほか、分類項目数や階層化の有無などを指定できる。

2.3 各自動分類手法の特徴と、これらを組み合わせた分類構造の構築

本システムの 3 つの自動分類の特徴を以下の表 1 に示す。

表 1 自動分類の特徴

	属性分類	ルールベース分類	単語クラスタリング
明瞭性	◎	◎	△
コスト	◎	△	◎
柔軟性	×	◎	○

ここで「明瞭性」とは、分類結果のわかりやすさ・説明しやすさを、「コスト」とは、結果が得られるまでのユーザの作業負荷を、柔軟性とは分類をユーザの思い通りにカスタマイズする際のしやすさをそれぞれ表したものである。

属性分類は、特許調査において最も明瞭で確実に活用可能な分類である。ラベル名と中に入っている特許が完全に合致しており(例:「東芝」ラベルには必ず東芝の特許が入っている)、分類構造全体も必要十分なものが構築される。また、ユーザが特に何か設定する必要はなくコストもかからない。一方、書誌情報を超えた分類はできないため柔軟性はほとんどない。

ルールベース分類は、ユーザが自分自身で分類ルールを設定するため、明瞭で柔軟性の高い分類である。ただし、妥当なルールをユーザが考える際に大きなコストがかかる。

単語クラスタリングは、システムが自動的に分類構造を構築するため、「どうしてこのような分類構造になったか」を説明するのは難しい。一方、システムが分類構造を自動生成するためコストはかからない。また、チューニングにより、自分が分類したい分類構造へ修正を行うことができるため、多少柔軟性のある分類である。

もし、これら 3 分類の長所を活かした分類ができれば、特許調査に大いに役立つ。そこで、本システムでは、各分類手法をシームレスに利用することができるようになっていく。図 3 にその利用例を模式的に示す。

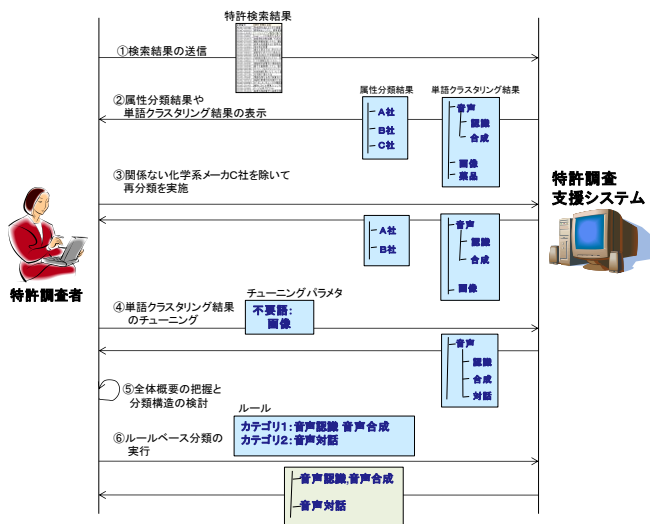


図3 自動分類を用いた分類構造作成例

ユーザが特許検索結果を送信すると(①), 属性分類と単語クラスタリングの自動分類結果がシステムより提示される(②). ユーザはまず属性分類の結果を見て, 明らかに今回の分類に関係のない特許群(図3の例ではC社の特許)を取り除いて再分類を行う(③). 次に得られた単語クラスタリング結果を見て, 不要語等を設定して(図3では「画像」を不要語に設定)チューニングを行う(④). こうして得られたチューニング後の単語クラスタリング結果で全体概要を把握するほか, 自分が構築したい分類構造を検討する(⑤). そして最後にルールを決めてルールベース分類を実行し自分の思い通りの分類を構築する(⑥).

このように3つの分類を相互に利用することで, いきなりから分類構造を考えるよりはるかに効率的に分類を構築することが可能となる.

各分類結果は, 図4のように分類構造と各分類に含まれる特許のリスト, および特許内容を同時に閲覧することができる. また, 各特許と同じ分類に所属する別の特許群を逆引きすることも可能である.



図4 特許調査支援システム 分類結果画面

*1 チームとは1人以上のユーザにより構成される特許調査のグループのこと. 多くの場合, システム開発を行うプロジェクトと同じである. なお本システムは2007年5月末から東芝ソリューション(株)内で, 技術者や研究開発者を対象に試行運用を開始し, 複数回のバージョンアップを挟みながら現在まで継続して運用されている.

2.4 分類結果を組み合わせた2軸マップ機能

また, 本システムではさらに, 「要約の課題」と「要約の解決手段」といった複数の観点で分類された結果を組み合わせて2軸マップとして閲覧することができる機能を設けている. 図5に電子レンジに関する課題と解決手段に関する2軸マップ(一部)を示す. 画面上部に並んでいる語が課題, 左部に並んでいる語が技術である.

	縦軸: 技術(発見型)										横軸: 課題(発見型)										凡例					
	製造	工程	米飯	生米	調理	調理	入フ	キチン	防止	定内	湯気	音	トレット	充電	断熱	本形	透過	低コト	安全	省スペース	58以上	39以上	20以上	1以上	0	
加熱	45	3	3	1	68	2	2	3	25	2	2	28	3	3	3	1	2	8	15	1						
蒸気	2				13			1	7	2	2	3	3	1			2	1	5							
電子レンジ	5				11	1	1		5		7		1				1	5								
高周波					10				6	1	2							1	1							
設定 操作					5			1	2	1									3							
樹脂	34	1			11	1	1		20	4		28	3	3	3	3	2	14	5	1						
可塑	7				1				1		4							1								
積層	3				5	1			7	3		9	1	1	1	1	1	2								
組成 以下	5								1																	
平均 粒子 径	4																									
容器	19	1	1	1	28	2	1	1	20	1	26	6	7	8	7	6	12	7								
開口	5	1	1		9				10		1	28	2	5	5	5	3	6	1							
所定 形成												4														

図5 特許調査支援システム 2軸マップ画面(一部)

この2軸マップを見ることでどのような課題をどのような技術で解決しているケースが多いの概要をつかむことができるほか, 「自分が持っている技術をどのような課題に適用できるか」, 「自分が抱えている課題をどのような技術で取り組んでいる技術者が多いか」といった, 新しい知見を発見することが可能である.

たとえば, 図5の赤い矩形で囲まれたエリアは, 「調理」関連の課題に対し, 「加熱」関連技術と「樹脂」関連技術でそれぞれ解決している特許群を示している. 加熱技術は主に電子レンジの加熱方法に関する特許が, 樹脂技術は主に(電子レンジの中に入れる)容器やラッピングに関する特許が, それぞれ多数含まれている. このような, 同一の課題に対し, 分野の大きく異なる別々の技術で解決しているのを知ることは, 技術者にとって視野狭窄を防ぐうえで大いに有用と言える.

3. 評価

本システムの評価については, 精度や性能といったオーソドックスな自然言語処理的観点からの評価と, 特許調査で実際に「使いものになった」「どのように使われたか」か, という実用的な観点からの評価の2種類が可能であるが, 本論文では, AI応用というセッションの趣旨に鑑み, 後者の「特許調査業務に対して実用的なレベルであったか」という実証的観点からの評価結果について説明する.

3.1 評価の概要

評価は以下の2つの観点から実施した.

- (1) 特許調査作業の省力化をどの程度達成できたか
- (2) 作業品質をどの程度高めることができたか

(1)は本システムの主目的である作業時間の削減をどの程度達成しているかを評価するものである. (2)は一般に作業時間とトレードオフの関係にある作業品質をどの程度維持(あるいは逆に向上)できたかを評価するものである.

評価は, 本システムを利用したチーム*1を対象にして, アンケート形式(約30~50項目)で行った(ただし, より詳細な情報を

把握するため、一部ヒアリングを行ったチームもある。対象チームは全 29 チームであり、特許マップ作成が 18 チーム、技術動向調査(特許マップは作成しない)が 8 チーム、新規性調査が 3 チームである。各作業は質的な違いがあり、これによりシステムの使い方などにどのような違いが生じるのかについても適宜調査した。

3.2 評価結果

評価結果について以下に示す。

・作業の省力化

まず本システムを活用することでどの程度作業を省力化できたかについての評価結果を図 6 に示す。なお、ここでいう省力化とは、本システムを利用した部分をすべて手作業で行った場合を想定して、どの程度作業時間が削減できたかを作業工数(人日)で示すというものである。

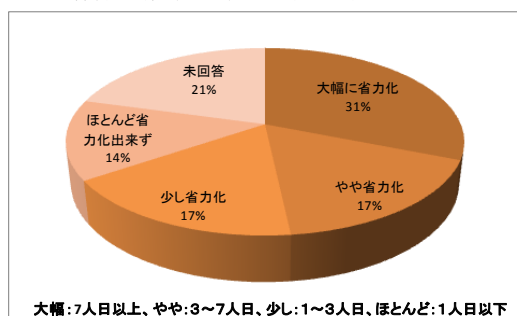


図 6 「省力化」の観点での評価結果

7 人日を超える大幅な省力化につながったと回答したチームが 9 チーム(31%)で、少なくとも 1 人日以上の省力化を達成したチーム全体では 19 チーム(63%)となった。特に全体作業工数の多い特許マップ作成チームの省力化への貢献が大きく 18 チーム中 7 チームが 7 人日以上の削減につながったと回答した。

省力化の理由については、新規性調査チームは「単語クラスタリングを用いて、関係のある特許に簡単にアクセスできた」など、キーワードを提示することによる関連特許へのアクセス性の向上をポイントとして挙げた。また技術動向調査チームは、「書誌情報とキーワードを組み合わせた 2 軸マップ(例:「出願人」と「要約の解決手段の単語クラスタリング結果」)が短時間で技術動向把握に非常に有効だ」との主旨のコメントを述べた。これらのチームは、属性分類と単語クラスタリング分類の出力結果をそのまま利用し、短時間で希望する特許を発見したり、傾向を把握したりする行動が見られた。

一方特許マップ作成チームは、単語クラスタリングで特徴語を発見し、不要語や同義語を設定してチューニングしたあと、そこで得られた分類結果をもとにマップの分類構造を考え、ルールベース分類にかけて自分の使いたい分類を獲得する、といった複数の自動分類を相互に利用して特許マップを作成する行動が多くみられた。これにより、「2 週間程度かけて作成していたマップと同等のものが、単語クラスタリングを中心とした 2 時間程度のチューニング作業で作成できた」と回答したチームもあった。

・作業結果の高品質化

また、各チームに作業品質の維持・向上に本システムが役立ったかに関するアンケート結果を図 7 に示す。

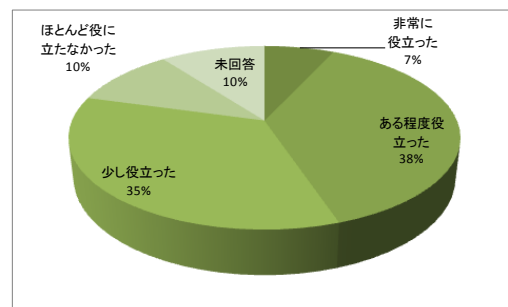


図 7 「高品質化」の観点での評価結果

分類結果の全部、あるいは大部分を最終報告に活用できたことを示す、「非常に役立った」、「ある程度役立った」と回答したチームが 13 チーム(45%)で、何らかの形で品質向上に役立ったチーム全体では 23 チーム(80%)となった。特にルールベース分類を含めた全分類を活用したチームほど評価が高かった。

どのような点で品質向上に役立ったかの理由については、特に特許マップ作成チームから、気付かなかった分類項目を発見できたというコメントが多かった。中でも、「課題」の観点に関するキーワード(例:時間短縮, 低コスト, など)については、「技術」その他の観点からの分類以上によくわからないケースが多く、特許マップを作成する上で大いに利用できたとの回答が得られた。

一方、本システムでは自動分類で可能な範囲の分類しかできないため、例えば文書の移動やカテゴリの統合などが柔軟にできるシステムにしてほしいとの要望もあった。

4. まとめ

本論文では、特許を自動分類する機能をコアとした、特許調査支援システムについて紹介した。また、実際の特許調査業務に適用した実証評価を行い、特許調査の省力化と高品質化の観点から効果的であることが確認された。一方、本システムを発展させて分類作業を完遂するまでサポートしてほしいというニーズがあることも明らかになった。このようなユーザの声に対応するため、筆者らの研究グループでは、ユーザとシステムの協調作業により、分類構造構築作業全般を支援できる「対話型分類技術」の研究開発に取り組んでいる[宮部 09]。誰でも簡単に大量の情報を整理・分析できるシステムを目指して、分類アルゴリズムの改善や機能の拡張、ユーザインタフェースの改良など、さらに研究を進めていきたい。

参考文献

- [Salton 75] G. Salton, etc.: A theory of term importance in automatic text analysis, Journal of the American Society for Information Science Vo.26, No.1, ASIS&T, 1975.
- [岸田 03] 岸田和明: 文書クラスタリングの技報: 文献レビュー, Library and Information Science No.49, 三田図書館・情報学会, 2003.
- [宮部 09] 宮部泰成: ユーザの意図を反映した対話型文書分類技術, 東芝レビュー Vol.64, No.2, 東芝 技術企画室, 2009.