

特許庁の次世代検索システム 開発に向けた調査について

特許庁 総務部総務課
システム開発室課長補佐

戸島 弘詩

PROFILE

平成11年特許庁入庁、平成19年より現職



1 はじめに

特許庁では、政府の電子政府構築計画に基づき2004年10月に「特許庁業務・システム最適化計画」を策定し、2005年8月には、計画内容の更なる明確化のための改定を行っている。この最適化計画では、特許庁のシステムである「特許庁総合基盤システム」を、「特許庁運営基盤システム」と「特許庁新検索システム」の2つに分けて、段階的に開発していくこととしている。「特許庁運営基盤システム」とは、出願や中間手続の受付から、方式審査、実体審査、権利の登録、公報発行、審判に至るまでの全ての手続と申請人登録や料金等に関係する事務処理を行うシステムである。「検索システム」とは、申請された出願の内容と同一又は類似の文献を検索するもので、例えば、特許における先行技術文献調査に利用されるシステムである。

先行開発とされていた特許庁運営基盤システムは既に開発段階に入っているが、後発開発とされていた特許庁新検索システムについては、平成20年10月に、「特許庁業務・システム最適化計画（検索系システム追補版）」が策定されている。

本計画では、特許庁新検索システムにおいて、先行技術文献等の情報量の増大や外国文献等の検索対象の多様化に追従できる、最新のIT環境による世界最先端の検索システムを構築することとしており、そのために、概念検索、多言語横断検索、イメージ検索等の新たな技術の導入や、過去の検索履歴や引用文献情報等の審査関連

情報を審査ナレッジとして活用する仕組みを検討することとしている。そこで平成20年度には、これらの新たな技術・仕組みについて、次世代の検索システムへの適用の可能性と、それに必要な機能要件等を明らかにするための調査が実施された。

本稿では、上記調査について紹介する。

2 次世代検索システム 開発に向けた調査

本調査においては、審査関連情報を活用した技術（データマイニング技術、概念検索技術、自動分類付与技術）、多言語横断検索技術、画像イメージ検索技術について、実際にプロトタイプを開発し、さらにユーザ評価を行うことによって、検索の精度や効率の向上の可能性や、次世代検索システムの実現に必要な機能要件を検証した。

2.1 審査関連情報を活用した技術

本調査では、次世代検索システム実現に向けて、検索インデックス情報（FI、Fターム、フリーワードなど）、検索履歴、引用文献情報、書誌情報、公報テキスト情報等の審査関連情報の活用における問題点を分析し、審査関連情報のあるべき姿を検討し、「データマイニング技術」「概念検索技術」「自動分類付与技術」を審査業務に適用した場合の有効性について検証した。調査を通じて、これらの技術を審査業務に適用することの有効性が示唆された。

2.1.1 データマイニング技術

(a) 審査ナレッジの共有と有効活用

審査官の知の情報（検索履歴、引用文献等）にデータマイニング技術や概念検索技術を適用し、審査業務の様々な場面で役立つ知識が得られる体系的なプロセスを確立することを試みた。

審査官の業務の流れを整理し、現行システムが抱える問題点について分析を行った結果、審査業務で必要とされる 20 種類の審査関連情報（本願の関連文献、請求項の従属関係、本願の関連文献に付与されている分類、本願の関連文献に高頻出のワード等）と、それらを活用する機能（対話型マイニング、辞書機能、関連文献の可視化機能等）を、体系的に整理することができた。

(b) シソーラス辞書の構築と活用

検索履歴に含まれる分類・検索キーの論理積（AND 条件）、論理和（OR 条件）の関連性から生成した関連語／関連分類辞書の利用性の検証を行った。

この結果、特許文献で用いる用語等を含む、技術分野別のシソーラス辞書まで含めた辞書全般の課題について、長期的な取り組みを要するものであると整理された。

また、検索履歴に含まれる分類・検索キーの論理積（AND 条件）、論理和（OR 条件）の関連性から関連語／関連分類を抽出し、関連語／関連分類辞書を整備することの可能性が示唆された。

(c) 検索結果表現の高度化

本調査では、検索結果に含まれる各文献間の技術的関連の近さや、引用・被引用文献やファミリー文献の芽づる式な関係等を可視する機能、また絞り込み検索の操作等を支援する機能の実現性について検証を行った。

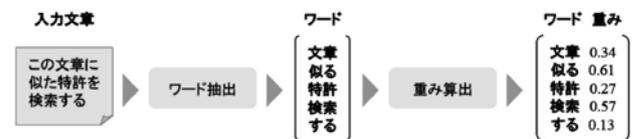
この結果、本願の関連文献や、引用文献候補の関連文献で使われている分類やワード、検索履歴から抽出された関連性の高い分類やワードを分析して、重要度の高い分類やワードを提示するなどの、効率的な検索式作成支援が可能であることが示唆された。

2.1.2 概念検索技術

従来の分類・検索キーによる検索に加えて、本願に

含まれる自然文等と類似度の高い文書を検索する概念検索を用いることで、増え続ける情報量に対して高い検索効率を維持する支援環境の構築が可能かどうかの検討を行った。

本調査の概念検索技術では、入力された文章に対し形態素解析を行うことで、文章からワード（形態素）の集合を抽出し、入力文章及び蓄積文献全件について、切り出した各ワードについて出現頻度から重みを算出した。



さらに各ワードと重みとを対としたベクトルをもとに類似度を算出して、類似する文献を検索した。

	出現頻度					
	文章	似た	特許	検索	する	...
文献A	15	7	3	23	50	
文献B	4	8	2	3	48	
文献C	6	5	0	0	34	
文献D	14	9	2	32	53	

調査の結果、技術分野に応じて分類・検索キー、特徴語、重視する文書構造（全文、要約、請求項、実施例）などを指定することで、技術分野によっては審査業務に適用可能なレベルまで、概念検索の精度が向上するケースが確認された。

機械分野や化学分野では検索結果の上位 50 件で審査官が引用した文献の半数が発見される可能性や、電気分野でも検索結果の上位 50 件で引用文献の 4 分の 1 が発見される可能性とともに、高い業務効率を維持するための支援環境構築の可能性が示唆された。

2.1.3 自動分類付与技術

本調査では、分類改正作業やメンテナンス（文献に対する分類の再付与）作業の負担を軽減し、分類・検索キーの品質、さらには検索精度を維持・向上させることを目



的として、分類・検索キーの分析や新分類の付与を支援する機能について検討を行った。この結果、以下の可能性があることがわかった。

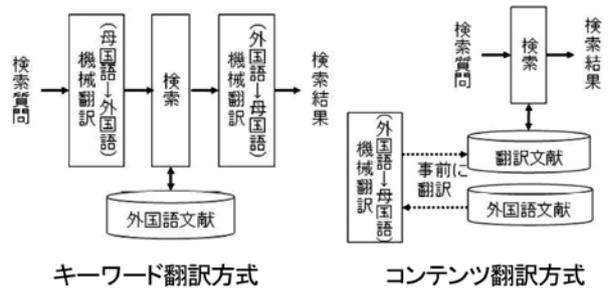
- ・自動分類付与技術の一つである文書クラスタリングや特徴語のグラフ表示を用いることで、FIの分類改正における新分類体系の発想支援や、改正後分類の付与文献の分布シミュレーションが実現でき、分類改正作業の業務負担を軽減できる可能性がある。
- ・FIの再分類作業において、業務負担を軽減できる可能性がある。
- ・適切な分類付与ルールを適用することで、技術分野によっては比較的高い精度で自動再分類付与精度を実現できる可能性がある。
- ・自動分類付与結果の確からしさと分類スコアの間には一定の相関があるため、分類スコアが高い付与結果に対しては人手での確認作業を省略できる可能性がある。

2.2 多言語横断検索

多言語横断検索技術は、母国語で検索条件を与えることによって、複数の言語で記載された文書を、言語を意識することなく検索可能にする技術である。本技術は例えば、日本語のキーワード、文章などの検索条件（検索質問）を入力して、中国語、韓国語などの外国語で記載されている特許文献を一括して検索する技術である。

中国や韓国の特許文献の検索では、英文抄録も利用可能であるが、抄録では先行技術調査に必要な実施例レベルの検索をすることができない。このため、現在重要性が高まりつつある中国特許文献や韓国特許文献に対して必要な先行技術調査を効率的に行うには、実用的な多言語横断検索技術が必要になると考えられる。

多言語横断検索技術は、翻訳技術の実現形態の観点で分類すると、検索質問を検索時に機械翻訳する「キーワード翻訳方式」と、検索対象の全文書をあらかじめ機械翻訳しておきデータベースに登録しておく「コンテンツ翻訳方式」の2つの形態に大別できる。



調査の結果、韓国特許文献については実用可能であることが報告された。しかし中国特許文献については、まずは注力分野を絞って、完成度の高いシステム提供を目指すべきであると報告された。

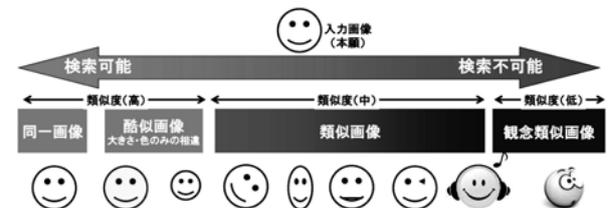
すなわち、平成26年の特許庁新検索システムの開発完了までに、全ての技術分野にわたって辞書を整備して中国語翻訳の精度を高めることは、極めて困難であると思われる。そこで、中国の技術優位性が高い「伝統的医薬品（A61K等）」など、分野を絞って辞書整備を行い、完成度の高いシステムを特定分野から順次提供していくことが重要であると考えられる。

また、ファミリー文献が存在しない中国や韓国の国内文献については、網羅的にサーチする手段が乏しい。そこで、このような中国・韓国の国内文献から蓄積を開始し、ファミリー文献が存在する文献については段階的に蓄積し検索可能とすることも検討すべきと考えられる。

このほか、翻訳方式としてはコンテンツ翻訳方式の方がやや優れていたが、翻訳精度向上の成果を即時に反映できる点で、キーワード翻訳方式にも利点があった。

2.3 画像イメージ検索

画像イメージ検索技術は、画像そのものを入力して、類似する画像を検索する技術である。



調査の結果、現在の技術レベルは、全体形状が同一・酷似の画像が画像イメージ検索では検索結果の上位 20 件で正解の 90%以上を網羅可能であるが、それ以外の類似画像の検索には対応できないものであった。

今後、以下のような課題が解決できれば、検索精度が向上するものと考えられる。

- ・画像の一部（要部）を対象とした検索手法において、多数の検索対象に対して、画像の特徴となる要部を効率的に抽出する必要がある。
- ・「大局的な類似画像」「部分的な変形」「部分的な位置変動」といった形状の相違がある画像の類似性を判断するために、有効な画像特徴を抽出する必要がある。

形状の相違 (チューニング項目)	入力画像	正解画像
大局的な類似画像		
部分的な変形		
部分的な位置変動		

- ・商標審査業務においては、画像的な類似だけでなく、観念的な類似も対象となる。このため、画像検索技術に、ウィーン図形分類やキーワード等を併用する必要がある。

3

おわりに

産業財産権の審査において検索対象となる過去の文献は、年々累積されその数が増加するとともに、英語以外の言語の文献の重要性が増大するなど、その種類も多様化してきている。したがって、世界最高水準の迅速かつ的確な審査を引き続き推進していくためには、このよう

な状況の変化にも対応していくことが必要不可欠となっている。

今回調査を行った次世代検索技術は、今後さらなる発展を遂げることが期待される。実用段階に達したのから活用していくことで、このような状況の変化に対応できる次世代検索システムが実現すると期待される。