

新検索システムとその展望について

調整課 審査企画室

1. はじめに

特許庁の審査官が、日頃、自信を持ち、誇りにしている事とは何でしょう。それはきっと、「どんな文献でも探し出すぞ」というような、心意気のようなものではないでしょうか。発明のポイントを見抜き、その考え方と同じ引用例を探し出すという「調査能力」こそが、審査官の基礎を築いている一つの力であり、また自信や自負に繋がっていることでもあると信じます。

そして、そういった力を根底から静かに支えているものが、庁内にある検索システムです。

2. 検索システムにおける課題について

特許庁では、過去から検索システムを着実に発展させてきました。昭和59年にペーパーレス計画の開発に着手し、Fターム（多観点ターム方式）のシステムを完成させてきました。その後、電子出願の法制整備と実施、ハードウェア処理能力の向上と検索システムの発展等の基礎基盤が整ったことにより、フルテキスト検索も可能になりました。このように、関係者がシステムを常に発展させる努力を続けてきたことや、過去からの着実なインフラ整備による蓄積が、現在の日本国特許庁の検索システムを支えています。

しかし、長期的な視点で特許庁の検索システムを俯瞰すると、様々な問題点が浮かび上がりつつあります。例えば、FタームやFIのインデックス情報は、更新・メンテナンスする総費用や人的給源が限られているため、

良質なインデックス情報を維持し続けていくことは徐々に難しくなりつつあります。

また、近年の出願の内容も、様々な要素技術を織り込んだ技術内容の出願へと複雑化しつつあり、分野横断的な検索が必要になるなど、潜在的に想定される調査対応能力が、量・質ともに広範化・高度化しつつあるという現状があります。

さらに、かつては日米欧が世界の80%の主要特許文献を占めていた時代には、日本語と英語で主要文献が容易に検索出来ましたが、今や、中国・韓国の非英語圏の出願件数は、欧州の出願件数を上回り、更に増加しています。今後は非英語国の文献をどう検索していくべきかという点が大きな課題となるでしょう。

今は、まだ潜在的な問題ですが、これらがやがて顕在化し、大きな問題になる前に手を打たなければなりません。

そこで、こういった問題を一つ一つ解決するために、『特許庁業務・システム最適化計画（検索系システム追補版）』では、新たな技術的手法を導入することとしています。その柱になるものが「知識処理」を用いた新たな支援機能等の導入であり、この成否が次世代の新検索システムに課せられた大きな宿題になっています。

そこで、調整課審査企画室では、これらの新検索技術をしっかり検討するために、各部2名ずつ審査官を推薦して頂き、合計8名の審査企画室の併任審査官¹⁾からなる特別の調査チーム「調整課検証ワーキンググループ」を構成しました。

そして、これらの新検索技術が、困難化する一方の

1) 調整課検証WGのメンバーは、五閑統一郎さん、古川直樹さん、谿花正由輝さん、中田善邦さん、河野一夫さん、清野千秋さん、滝谷亮一さん、脇水佳弘さんの8名の審査官です。

審査官の業務の助けとなることを願い、実証実験においてシステムを構築し、検証WGメンバーを中心に、そのあり方（フィジビリティ）を検証し、検討する作業を行なっているところです。

3. 知識処理を用いた支援技術とは

審査業務においては、審査官はまず本願を理解し、一連の戦略を考え、どのように探すかを思い巡らします。そして先行技術の検索を行い、探した引用例と本願を対比し、判断を行います。これら審査のプロセスでは、人間の頭の中で、非常に高度な知識処理が行われています。この一連の知識処理そのものは、恐らく現行のどのような優れた人工知能技術を使ったとしても、実現することができない程、高度な知識処理レベルに属します。

人間が図面を一瞥して一致性を判断するような処理でさえ、コンピュータ上では実現が極めて困難です。概念の理解などは、どんなにコンピュータ上の意味処理能力や述語論理展開の推論能力を発展させても、人間の知能レベルの実現にはほど遠いのが現状です。ですから、よく飲み会で語られる「(空想の)自動審査マシン」などは、現状のコンピュータの知能処理のレベルからみると、夢のさらにまた夢の技術なのだと思います。

しかし、審査に係る業務を補助するために用いられる、現実的な範囲の「支援機能」であれば、基本的に提供ができるだろう、というのが、現状検討されている「知識処理を用いた支援技術」のレベルの想定です。そこには、もちろん技術的限界があり、決して完全ではありません。留意すべき点は、これら機能はいずれも審査業務を「支援する」ことを目的に構築される機能であり、その機能自体は、決して人を凌駕するような「凄い知識処理」を実現するものではないということです。しかし、コンピュータにも得意な処理部分があるはずで、上手く利用できる局面を探していくことを目標としています。

簡単に例えて言えば、ワープロやメール、表計算ソフトと同じような感覚で、これらの知識処理を用いた支援機能を使えるようにすることが、最終的な目標です。

現段階では、まず、現在製品化されつつある汎用シ

ステム機能を、どれだけ審査業務に特化させることができ、審査業務に実際に活用出来るかということを中心に現実的な検討が進められています。

この審査に係る業務を補助するための「知識処理を用いた支援機能」として3つのテーマを選び、検討ならびに調査（実証実験）を行います。1つが、「データマイニング」と呼ばれる分析支援機能、2つが「概念検索」と呼ばれる検索支援機能、3つが「自動分類付与」と呼ばれる分類支援機能です。

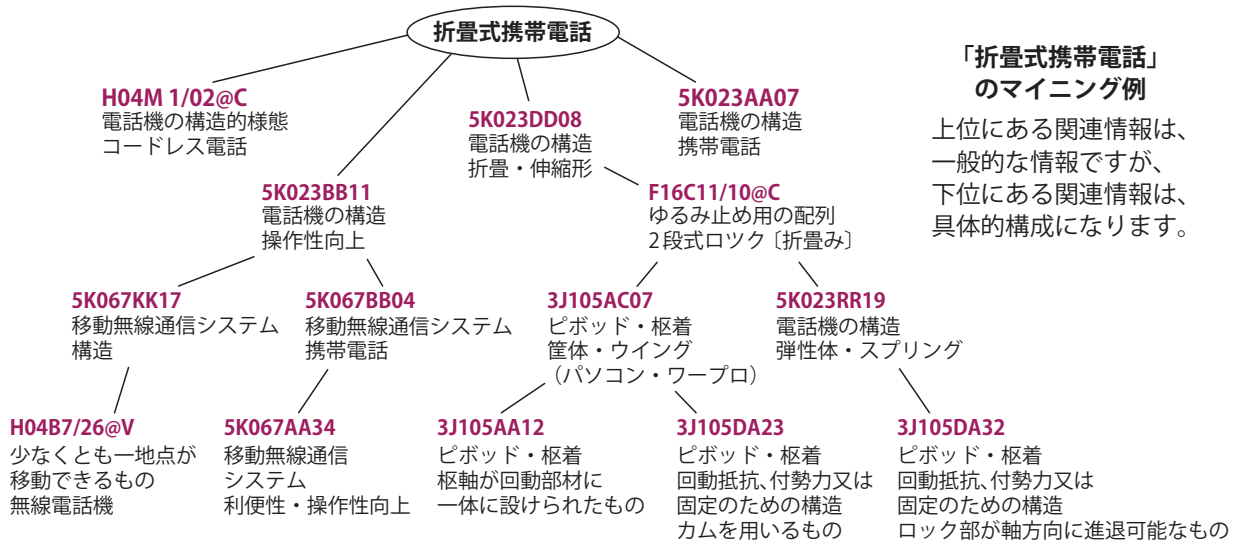
また、非英語国をはじめとする複数の言語の特許文献を一括検索する仕組みを検討するために、「多言語横断検索技術」についても同様に調査（実証実験）を行います（ここでは、機械翻訳技術も知識処理の一つとして捉えます。）。)

(1) データマイニング技術

現行の検索システムでは、検索履歴が自動的に蓄積されます。これらの検索履歴データは膨大な量であり、いずれも検索のノウハウや知識が詰まった情報です。このような検索履歴データ等の情報に対して、統計処理等を行うことにより、有意義な情報を抽出できないだろうかという課題がありました。しかしながら、これまでは十分な検討がなされてきませんでした。

そこで、平成18年に、審査企画室では、「今後の審査システム開発に向けた調査研究」（平成19年3月）を行い、検索履歴データを分析し調査しました。その結果、検索履歴から抽出されたナレッジは、審査官が作成したSSF（サーチ戦略ファイル）等の情報と同等の情報の抽出の可能性があり、このような情報が（データマイニング技術等により）自動的に抽出可能であるならば、積極的に採用すべきであるとの結論となりました。

データマイニングとは、膨大なデータの山から、有用な情報を抽出することです。かつて、米国全土に店舗を有するあるスーパーマーケットにおいて、売上履歴情報から、顧客が何と何を同時に購入するかを分析したところ、「ビール」と「おむつ」の組合せが多いことを発見しました。そこで、このスーパーマーケットでは、「ビール」と「おむつ」の売り場を近くに配置することによって売上を大きく伸ばしました。このように、データマイニングとは、マーケティング等におけるクロスセリングの分析で活用され、一躍脚光を浴び



**「折り畳み携帯電話」
のマイニング例**

上位にある関連情報は、
一般的な情報ですが、
下位にある関連情報は、
具体的構成になります。

図1 検索履歴からのデータマイニングの結果 (例)

た技術です。

先行技術調査でも、検索式を作成する際に、その技術に関連して、どのような検索ワードや検索キー (FIやFターム) を用いて検索を行っているかを知ることが必要です。

そこで、今回の実証実験のモデルシステムでは、過去数年分の検索履歴データや過去10年分の公報データを用いて、その技術に関連して、どのようなキーワードや検索キー (FIやFターム) が同時に用いられているかを分析・表示させ、検索において気付きを得るための「発想支援機能」を提供します。今回のデータマイニングの支援機能の主たる目的の一つです。

今回の実証実験により、このデータマイニングの支援機能が、どのような局面でどう役立つか、どのような効果があるかを定性的に分析しているところです。

(2) 概念検索技術

現在の特許庁のフルテキスト検索においては、審査官がキーワードとして入力した文字列が一致しているか否かのみを判断していますが、そのキーワードが文献中での程度使われているかが考慮されていないため、必ずしも概念が近い文献がヒットするわけではありません。

この点を改善することを目的として検討されている技術が、「概念検索」と呼ばれる技術です。この技術では、文献中のキーワードの出現頻度を考慮して文献間の類似度を判断しています。

「概念検索」の技術とは、一般的に、検索条件を自然文で入力し、システムがその検索条件文を分析し、適合している文献を検索結果として抽出し、検索結果を類似している順にランキング表示させる機能を指します。

現在では、既に複数の会社が、概念検索システムによる特許検索サービスを提供しており、広く普及しつつある技術になっています。

また、国立情報学研究所 (NII) では、NTCIR²⁾ (エンティサイル) というプロジェクトを開催していますが、このワークショップの「特許タスク」には多くの企業や大学の研究者が参加していて、概念検索についても、既に数多くの研究実績があります。

平成20年2月に審査企画室が「特実検索系の将来像検討プロジェクトチーム結果報告書」として報告した検討結果の中で、概念検索は、類似文献の抽出や、スクリーニングの順序変更等、補助ツールとしての利用が可能ではないかとの提言がなされ、今回の実証実験では、その利用方法について評価・検討することになりました。

「概念検索」に関しては、韓国特許庁 (KIPO) では「ラ

2) 「情報検索システム評価用テストコレクション構築プロジェクト」 (NII Test Collection for IR Systems) 略してNTCIR

概念検索エンジンの原理 GETA (GenericEngine for Transposable Association)

その文書に現れる単語の集合情報(頻度)を「特徴」として用いることによって、文書間の類似度の判断を行います。これが、本実験の「概念検索」の原理です。
 ここでは、特開2001-?13254と特開2000-?09324が類似すると判断しています。

	電話回線	インターネット	ウェブ	インク	安価	高速化	メモリ	バッファ	アドレス	データ	パケット	転送	トナー	紙	ネットワーク	印刷	プリンタ
特開2000-?23421	1	4	5	3	8	2	11	8	17	13	9	7			11		
特開2001-?13254		1	2	8	8	3	5			9			12	9	9	21	6
特開1996-?12642		10	8						7		1	13			11		
特開1998-?12341				5										3		11	7
特開2000-?09324			9		6	3	6			7	1		11	5	10	19	3
特開2002-?89312				13	12								15	9		13	8
特開2001-?23241		7	4			3	1	6	9	10	8	6			6		

単語の頻度(数)が格納されているメモリ(WAM)

図2 概念検索エンジンのしくみ

ンキング表示機能」が採用されるなどの先例もあり、実際に庁内で実用化され、システムに導入される可能性が最も高い技術と考えられます。

今回の実証実験では、この「概念検索」に対してチューニングを実施し、審査官の目線から、どの程度審査実務に用いることができるかについて定性的・定量的に評価を行っています。そして、どのような局面で活用が可能かを検討しているところです。

(3) 自動分類付与技術

FI、IPCやFターム等のインデックス情報は、効率的なサーチを行うための有効なサーチツールですが、更新・メンテナンスする総費用や人的給源が限られているため、良質なインデックス情報を維持し続けていくことは徐々に困難になりつつあります。過去の文献に遡って分類付与を行うことは、膨大な手間がかかるためです。

そこで、これらの分類付与作業を自動化、あるいは半自動化し、効率的に分類付与を行うための支援機能として役立てられないかというのが、今回の実証実験における大きな課題です。近年、分類改正に伴う作業負担は増加する傾向にあるため、施策的な観点からも重要なテーマであり、ある程度の実現性への見通しを付けることが望まれています。

今回の実証実験の自動分類付与システムにおける技

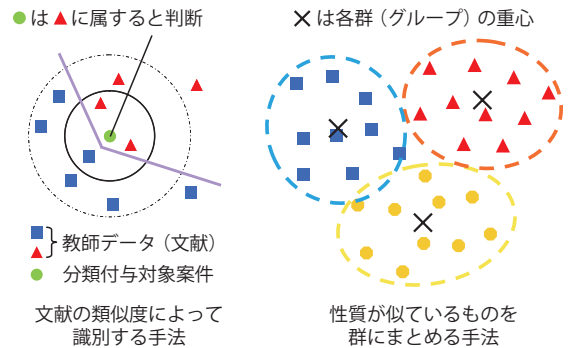


図3 自動分類付与における色々な手法

術的手法には2つあります。

1つは、文献の特徴を類似度によって識別する手法を使うものです。分類付与済みの文献集合を利用して、類似度の高い文献に付与された分類を統計解析して、未分類の文献に対して適切な分類を付与したり、旧分類が付与された文献に対して、新分類で再分類を行うことを実現するためのものです。この手法では、なるべく精度が向上するように工夫を凝らして技術的手段の実現に取り組んでいるところです。

もう1つは、統計的に性質が似ているものを群にまとめるという手法を使って、文献集合を類似する主題を持つ集合(クラスター)にグループ化し、新分類案とともに、新分類作成後の文献数の分布を提示するものです。これについては、インデックス情報のメンテナン

スを支援することへの適用実現性について、検証しているところです。

この2手法は、それぞれ用いられる場面や性質が異なりますので、この2つの手法について、システムを構築して、実証実験を行っています。

今回の実証実験では、これらの自動分類付与技術を、庁内の実際の分類関連業務に適用した場合、どの程度利用可能性があり、どう役立てていくかを審査官の目線から定性的・定量的に評価しているところです。

(4) 多言語横断検索技術

冒頭でも触れましたが、かつては日米欧の三極が世界の80%の主要特許文献を検索出来ましたが、近年では、日米中韓欧の五極が世界の77%の出願を占めるまでになっています。このように、中国・韓国等の非英語圏の出願が徐々に増加しつつあります。

しかしながら、非英語文献（中国語・韓国語）の場合、その文献を理解出来る審査官の数は、英語文献を理解する審査官の数に較べるとごく僅かであり、何らかの形で、これらの非英語文献について検索および理解を容易にするための支援機能を設ける必要があります。

なお、欧州特許庁（EPO）では、欧州機械翻訳プログラム（EMTP：European Machine Translation Programme）を構築し、欧州各国間の翻訳を実現しつつあります。

そこで、今回の実証実験では、概念検索技術と機械翻訳技術を融合した多言語横断検索技術を採用し、実際に中国文献と韓国文献を日本語により検索可能にするシステムを構築しました。

韓国語と日本語では、互いの言語の文法構造が似ているため、単語の置換でほぼ翻訳処理が実現出来ますが、中国語の場合、日本語と言語の文法構造が異なるため、技術的難易度の高い「文法を理解する処理」が必要です。このため、中国語のように、言語文法構造が異なる言語で、どの程度翻訳精度を向上出来るかが大きな課題です。

今回の実証実験では、代表的な分野を中心に、審査官が多言語横断検索システムを実際に用いて、どの程度利用可能性があるか、そして改善を図るためにはどうしたらよいかを審査官の目線から、多面的に評価しているところです。

4. おわりに

実証実験のシステムで、フリーオペレーションを試してみた方は、ことのほか性能が良くなく、すこしがっかりしたと思われた方もいらっしゃるかもしれません。

現在、実証実験中の「知識処理を用いた新検索システム（検証モデル）」は、まだ荒削りのシステムであり、実験レベルの発展途上のシステムです。いわば、やっと歩き出した子供のようなものです。ですから、この実証実験では、どうか一緒に新検索システムを育てるという温かい視点を持って見守って頂けると幸いです。

また、現在の取組みは、やや野心的に見えるかも知れません。「本当に実現出来るのか？」と。しかし、決して夢でも野望でもありません。現実を踏まえ、冷静に客観的にシステム技術の本質を見極め、実現可能性（フィージビリティ）を判断していくことが必要になるのだと思います。また、多くの試みが初めてのことばかりなので、すぐに諦めないで、前向きに検討し続ける視点も大切なことだと思います。

今後10～20年後を展望し、これら知識処理を用いた新検索システムの性能限界を理解した上で、しっかり手入れをし、うまく使いこなすことができる「道具」として着実に発展させる事ができれば、それは、とても役に立つ手段になると思います。これらの情報技術には、そのような強力な潜在能力があるからです。

調整課・審査企画室では、審査官の調査能力を支える検索システムを、世界でも最も先進的なものに発展させ、審査官の皆様の大きな力になることを願い、あらゆる可能性を含めてどのようにしたら良いシステムを構築出来るかを検討していきます。

平成22年秋頃を目途に新検索システムの基本要件を確定させ、開発に着手する予定です。そして、平成27年の知識処理システムの完成を目指していくことになります。

本稿が、新検索システムへの理解を深め、皆様の関心を少しでもお寄せ頂けるきっかけとなればと願います。

また、本稿を執筆するにあたり、多くの皆様からのご協力と、貴重なご助言を頂きました。この場をお借りいたしまして、厚く御礼申し上げます。