

次世代検索システム体験記

特技懇編集委員



去る11月4日特許庁近くの某ビル内、ついに私達の目の前に、審査関連次世代検索システムがその姿を現しつつありました。今回、次世代検索システムについての調査研究を担当していただいている日立製作所の方々のご厚意により、なんと次世代検索システムの検証用プロトタイプを一足先に体験する機会を得ることができたのです。

今回体験させていただいたのは、次世代検索システムの新機能のうち、日立製作所の方々が開発をご担当されている、特に『データマイニング』、『概念検索』、『新分類体系作成支援』の三機能。最前線の開発者から、マニュアルからは得難い、新検索システムのコア部分(実現したいこと、思想)についてわかり易く手ほどきを受けつつ、それぞれのプロトタイプを体験することができました。三機能とも、現在特実審査官が用いている検索システムには無い、多くの魅力が詰まっています。この記事の掲載号発行の頃には、庁内でも検証用モデルがリリースされ、特実審査官の方々は既にこのプロトタイプの機能を体験済みかもしれませんが、実際に最前線で次世代検索システムの開発に携わっている方々から直接説明していただいた貴重な機会を体験記として報告させていただきます。

なお、データマイニングや概念検索といった各検索技術自体については、本特集内で別途ご紹介させていただいております。是非、そちらの記事もご一読ください。

1. データマイニング

1.1. 新機能、ついに起動

次世代検索システムへのログイン後、ついに待望の新機能の一つ『データマイニング』が画面に姿を現しました。膨大な審査関連情報から、関連の高い分類やキーワード(データ)を掘り出す(マイニング)機能とのことなので、早速『自転車』というワードをマイニング条件として関連するワードを掘り出させていただきました。数秒後に画面が更新されると、なんとといっても私達の目を惹いたのは、同時に表示された画面右下のツリー表示でした(図1)。

従来の検索システムでは見たこともないそのウインド

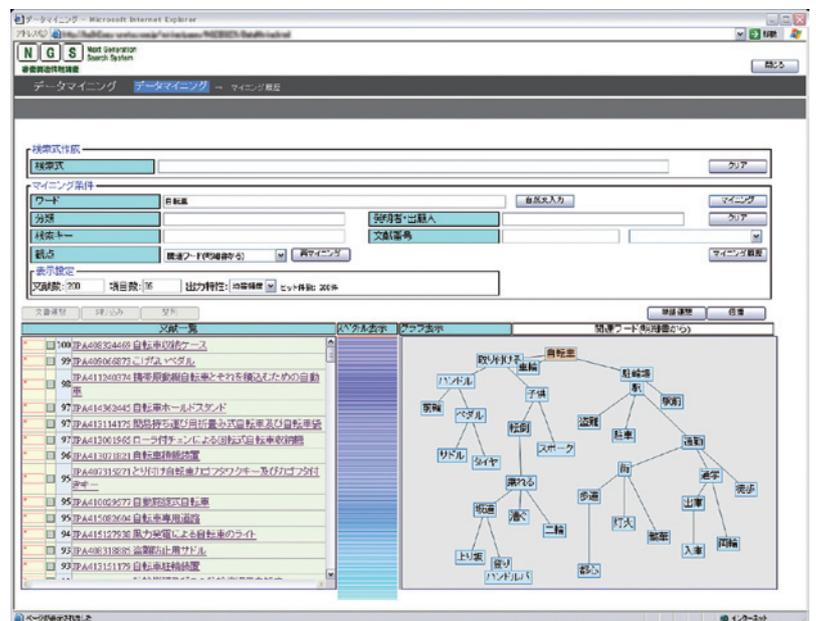


図1 データマイニングシステム起動トップ画面

ウに、私達の目は釘付けです。一番上部に『自転車』というボックス、それに連なって『通勤』、『駅』、『盗難』など、関連ありそうな言葉のボックスが、ツリー上に配置されています。『自転車』というたった一語の条件入力だけで、本当にいろいろな情報が掘り出されてきました。

1.2. 『自転車』という一語から……

左下欄の文献は、ワード『自転車』を含むという条件で検索された結果の文献集団だそうで、確かに自転車関連の文献が多く表示されています。右下のツリー表示は、これらの文献の中から頻度の高い情報（今回の場合はワード）を取り出して頻度順に上から並べた図とのことで、各ボックスは最も関連の高い他の要素と線で結ばれていました（図2）。この場合だと、『自転車』に関連して『駐輪場』・『盗難』・『駅』といった単

語が特許文献でも関連深く用いられているということですか。なるほど、どのような語が関連して用いられているかが視覚的にわかりやすく、この二次元配置図を見れば関連語が直感的に把握できそうです。

また、スペクトル表示部にマウスカーソルを当てると、ツリー表示では表示しきれなかった関連語も見ることができました。審査官はこのツリー表示等から、検索キーワードに関連しそうな情報を抽出することによって、検索式の拡張・補充に利用できそうですね。

しかも、ツリー表示中で選択したボックスの情報は瞬時に左下欄の文献群に反映され、その情報を含む文献に★マークが付与されます。この★の数で整列させることも可能で、それによって、審査官の関心対象の情報を多く含む文献にアクセスしやすい（スクリーニングもできました。）という点は、非常に便利であると感じました（図3）。

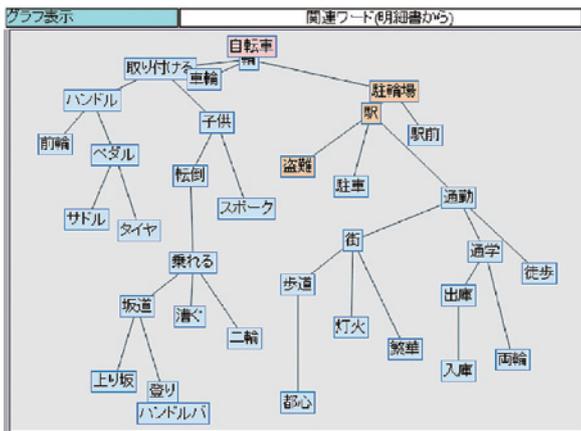


図2 ツリー表示部

1.3. 検索式の発想をアシストする

マイニング条件は、ワードに限らず分類や自然文によっても行うことができるようです。日々の先行技術調査において、「こういった思想のものをサーチしたいが、テキスト・分類がいまいちよくわからない。どのように検索対象を絞ってよいかわからない……」という悩みは尽きないと思います。そんなときに、例えば「駅前での自転車の盗難を防止する。」、そんな抽象的な条件でも関連する分類や検索語を集めてくれる機能があれば助かりますよね。これは担当分野以外の発

図3 文献整列

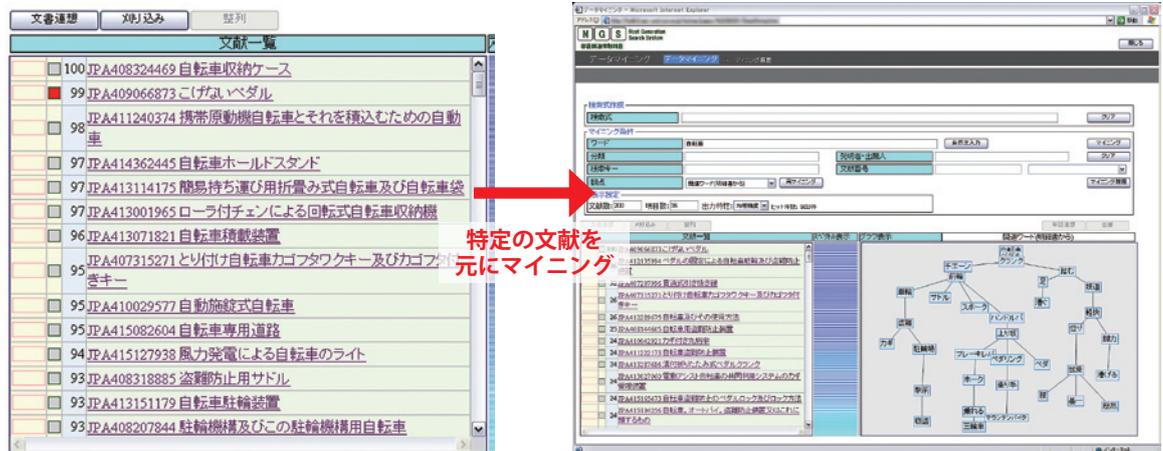


図4 文献からの連想

明を審査するときには当該分野で良く使われている分類やテキストを知ることができるなど、かなり期待が持てる機能なのではないでしょうか。

特実審査官は、ワードや分類の掛け合わせによる検索を通じて先行技術調査を行っています。一つの検索式で調査を終えることはまずありません。調査中に派生してきた関連ワードや関連分類をも検索式に組み込み、試行錯誤しながら常に調査範囲を変化させていきます。こういったリアルタイムの思考変化での発想を支援するためにも、ツリーで表示されたボックスの関連情報や文献自体をマイニング条件にして、一つの結果からさらに派生した関連情報を表示させるような機能もしっかり用意されていました(図4)。

以上をまとめてみると、データマイニングを使えば、自力で関連ワード・分類等を集めて検索式に用いるという従来型の検索から、以下のように検索式構築のバリエーションが増えていくのではないのでしょうか。

- ① まずデータマイニングを使ってワード・分類・自然文等を使って関連する情報を掘り出してみる。
- ②-1 関連ありそうなワードが掘り出されるので、それらを検索式に組み込んでサーチしてみる(→従来型検索式構築の負担軽減)。
- ②-2 掘り出された結果中に表示された、本願に近そうな文献・ワードが気になったので、これらをマイニング条件として設定し、より類似した文献や関連ワードの漏れを掘り出す。掘り出された関連ワード・分類等を検索語として検索式に補充する、も

しくは関連文献を拾って技術背景の理解に役立てる(→検索式発想に幅が増える)。

検索式作成の発想支援とはまさにこのこと。本機能の威力が十分に発揮され、審査官の頭で行われている検索式発想と同様の思考過程がデータマイニングによるアシストでも達成されれば、当機能は先行技術調査における審査官の負担を大いに軽減してくれる、頼れる存在となると感じました。

1.4. 次世代新検索システムへの導入に向けて

今回入力したワードのような明細書に元々ある情報によるマイニングだけではなく、検索履歴等の審査関連情報などを元にしたマイニングも可能とのこと。例えば、本願に関連する出願の審査の際に、どのようなワードや分類が検索式として用いられる傾向があるか等、特実審査官の審査ノウハウの発掘・共有においてもデータマイニングは威力を発揮するのではないのでしょうか。

検証用モデルでの難を言えば、マイニング結果表示までに時間がかかる、ブラウザの「戻る」機能があると良い(履歴を使って戻るのは若干手間)、「単語連想」の精度が若干不透明、といった点はありましたが、検証用モデルリリースによる試行段階を経て、それらの点も改善されていくと考えられます。

次世代新検索システムへの導入の際には、マイニング対象とする文献の拡充(検証段階では平成6~15年の国内の特許公報(実用は除く)が対象)の他、例えばマ

インテグレーション結果の表示の仕方等について学習機能を付ける可能性など、より扱いやすいインターフェースを目指しているとのことで、導入が非常に待ち遠しい機能です。

2. 概念検索

2.1. 審査に使う概念検索とは

次に私達が目にすることができたのは『概念検索』でした。巷でも、一般的な概念検索については最近耳にしますし、その言葉通り、概念で検索する、と説明されれば何となく理解できそうなものです。とすると、導入予定の概念検索と、巷の概念検索や従来型の検索との違いは何でしょうか。そう考えを巡らせていると、早速概念検索の画面が立ち上がりました(図5)。

画面を見ると何やら本願の書誌情報と共に『パラメータ設定』なる欄があり、どうやら、『本願』と概念的に近い文献を、この『パラメータ設定』によって拾ってくる、という点にこの概念検索の特徴があるようです。

2.2. 本願に近いものを、ランキング表示

今回は、検索範囲を『請求項』と設定したときの検

索結果を見せていただきました。右上欄には『順位』や『スコア』なる欄と共に、本願に近そうな文献一覧の表示(図6)。なるほど、この表示からみて、検索サイトの結果表示のように、検索したいものに近い順に上から表示されるわけですか。結構便利ですね。……が、何を元に検索が実行されたのか、気になるところです。

説明を聞いてみると、①本願からデータマイニングと同様に特徴語を抽出、②特徴語に対して自動的に重みが設定される(図7)、③重みの設定された特徴語が、どの程度『請求項』に含まれるか、という基準で検索している、とのこと。つまり、本願の内容に概念的に近いものが上位に表示される、そういった検索のようですね。

従来の特実検索システムだと、このような検索結果のランキング表示自体が不可能なので、「本願発明の思想により近いものからスクリーニングしたい」、というときに非常に便利かと思います。従来にも確かに類似文献表示機能はあったのですが、どのような概念に基づいて『類似』と判断されているのかわからず、精度も若干不透明でしたし……。

今回の概念検索プロトタイプでは、クエリ等を変更することでどの部分の情報に重点を置くかを審査官が

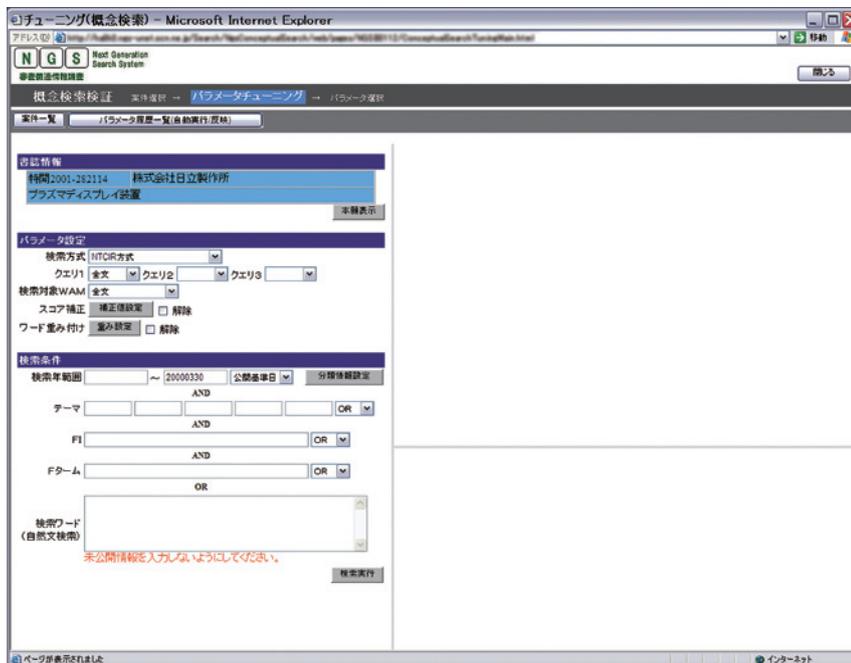


図5 概念検索起動トップ画面



図6 文献一覧表示画面 (ウインドウ)



図7 重み設定画面 (ウインドウ)

設定することができる上、単語や自然文を検索条件として追加する機能や、本願の特徴語の重み（どれだけ重要視するか）を補正する機能も付けられていました（※検証用モデルでリリースされるバージョンでは付けられるか否か不明とのことでした）。

今回の概念検索の導入により、本願発明の概念と近いものを高精度で抽出できるのでは、と期待が高まります。

2.3. 概念検索から見てくること～技術分野毎にも特徴が～

興味深いことに、概念検索を行ってみると、技術分野毎に望ましいパラメータ設定に傾向があるかもしれない、というお話でした。例えば、技術分野によっては、実施例に重点を置いてランキングするとか、FタームやFIを限定した上で検索を実行する、等の条件を設定した方が高精度の検索結果が得られる場合があるということです。

今回の検証用モデル段階では、これらのパラメータのチューニング機能が付いた状態でリリースされる予定との話でしたが、ある程度の情報が蓄積されれば、次世代新検索システムへの導入の際には、それこそパラメータ設定すら不要となるかもしれません。つまり、特実審査官は、本願番号を入力するだけで本願発明と思想が近い文献が（高精度で？）自動的に得られる、という将来像もあり得るといえます。

2.4. 検索結果フィードバックによって精度を向上

本願番号の入力だけで概念的に近い先行技術文献が自動的に表示される、このようなシステムは確かに便利ではあります。しかし、検索結果の精度は本当に高いのか、パラメータ設定は本当に最適か、などの信頼性が高くなければ、概念検索が実用レベルであるとはいえないでしょう。そのためにも、本願（又は他の出願）の審査で既に用いられた引用文献について、概念検索を行った文献ランキング中で何位にあったのか、といった情報表示の機能が有ります（図8の右下欄 ※検引用モデルでのリリースにおいて本機能はありません）。この機能により、得られた結果を元にパラメータ設定を調整し、（技術分野毎の）精度向上を図って検索に用いる、ということが可能となります。

2.5. クラスタ検索との連携

今回体験することはできませんでしたが、特実審査官が使うクラスタ検索との連携も非常に魅力的であると考えられます。従来のクラスタ検索では、ランキング表示のような機能はありませんでしたが、概念検索

を組み合わせることにより、①クラスタ検索で文献集団作成→②該集団に概念検索を掛け、条件に見合うものをランキング表示、という検索フローが実現します。「検索式で絞ったもののノイズが多く、本願に近いものから表示させたい……」、そのような悩みを抱える審査官は多いと思います。この連携が実現すればそんな悩みはすぐに解消するのではないのでしょうか。

3. 自動分類付与

3.1. 整備された分類情報は検索の大黒柱

審査官向けに先日リリースされた検引用モデルでは『データマイニング』、『多言語横断検索』、『概念検索』の機能が体験できたものと思いますが、それらの様々な検索においても、特許文献の先行技術調査の際には、発明の分類情報が非常に大きなウェイトを占めるといっても過言ではありません。一方、技術の進歩に応じて発明の分類は細分化・変化しており、従来の分類情報だと絞りきれない、といったケースが徐々に顕在化しています。

最後に『分類付与支援』と『新分類体系作成支援』

順位	公開番号	発明の名称	出願人	スコア	補正前順位	スコア
1	特開平10-116936	プラスマディスプレイ装置	株式会社富士通ゼロラル	0.271	1	0.271
2	特開2001-022282	プラスマディスプレイ装置	松下電器産業株式会社	0.205	2	0.205
3	特開平08-162002	プラスマディスプレイパネル	株式会社パナソニック	0.205	3	0.205
4	特開平11-024523	プラスマディスプレイパネル装置	松下電器産業株式会社	0.203	4	0.203
5	特開平11-381319	プラスマディスプレイ装置	株式会社富士通ゼロラル	0.2	5	0.2
6	特開2001-117644	プラスマディスプレイ装置	パイオニア株式会社	0.199	6	0.199

引用文献	順位	補正前順位	スコア	補正後スコア
特開平10-116936	1	1	0.271	0.271
特開平08-320996	-	-	-	-
特開平11-251777	9	9	0.195	0.195
特開平08-241172	-	-	-	-
特開平11-327449	11	11	0.194	0.194
特開平11-272307	20	20	0.182	0.182
特開2000-010493	105	105	0.164	0.164
特開平10-321498	-	-	-	-

図8 引用文献ランキング図（ウインドウ）

を紹介いただきました。分類付与支援は、概念検索技術を分類付与に適用した技術とのことで、上記問題点を解消する有力なツールであるとのことでした。『分類付与支援』とは、分類を付与したい対象文献と類似する文献を概念検索で抽出し、類似文献に既に付与されている分類を付与候補として表示できる、という機能です。この自動分類付与が可能となるだけでも、特許出願等の分類付けの際に、非常に処理負担を軽減できるものと思われます。

さて、『新分類体系作成支援』に話を戻しますと、これは、内容の似た特許文献同士を集めた塊（クラスタ）を作成する、クラスタリングという技術を応用した技術であり、分類改正の支援を目的としています。以下のように、既存の分類を統合したり、分類1, 2, ……を分類a, b, ……という観点で分類し直したり、といった分類改正のシミュレーションが可能となります（図9）。

3.2. バックログ解析・分類再付与の期待の星か

現状でも新しく分類を作成・変更したい等の要望は多くあるようですが、過去の文献まで遡ってすべての

文献を再分類するのは、非常に多くの負担を伴うものであり、実際のところあまり頻繁に行うことはできません。導入予定の新分類体系作成支援及び分類付与支援については、精度がどの程度であるのかが大きなポイントとはなりますが、この技術により過去文献に対しても半自動的に再分類することができれば、これまでの負担は非常に軽減され、技術進歩に伴った発明分類が随時可能となるのではないかと期待が持てました。

4. インタビュー

今回の次世代検索システム見学では、実際に開発に携わられた日立製作所の方々や研究所の方々に直接インタビューさせていただく機会も得ました。

○企業の検索システムにおいても概念検索やデータマイニングなどは行われていますが、特許庁の検索システムはどのような点が違うのでしょうか。

今回のシステムは、企業で用いられるシステム（例えば特許マップ作成の手法）等を応用して開発しています。データマイニングを利用して検索式を作る、という点は企業も特許庁も同じですが、発明の内容に応

The screenshot shows a web application interface for 'New Classification System Confirmation'. It features a table with columns for 'No.', 'Current Classification (Cluster Name)', 'Description', 'Count', 'New Classification (Cluster Name)', 'Description', and 'Count'. Below the table is a 'History of Classification System Creation (Classification Creation History)' section with columns for 'Initial State', '1st Step', and '2nd Step'.

例：偏りのある既存分類の再分類

履歴表示
 1回目：分類の統合
 2回目：新たな観点で再分類

図9 分類統合・再分割フロー画面図

じた分類体系を作る、という点は、企業ではあまり必要とされていません。また、企業においては、検索に使えるような単語がざっと一覧等で見られれば足りる一方、特許庁では、より広いキーワードを対象に、単語同士の関係や頻度・重要度等について多く指標を設定して検索・マイニングするというニーズがある点でも違うかもしれません。審査官の審査履歴・検索履歴を検索対象とするという点も非常に有用で、共有化された審査ノウハウを活かした検索が可能、という点も特許庁内の環境を活かした機能であるといえます。

一般的に、概念検索は、こうすれば類似文献がヒットできるだろう、と思って検索してもそう簡単にうまくいくものではありません。特許庁での概念検索の場合、特許出願に特有の文章の構造や、重要語が文章のどこにあるかを考慮すれば検索精度が（飛躍的に）向上する可能性もあるので、検証用モデルでの試用段階を通じて精度向上を図っていく予定です。

○将来的に概念検索等は従来のクラスタ検索を代替するものとなる、ということでしょうか。

概念検索等は、これによって従来の審査（テキスト、FI、Fタームを用いたクラスタ検索）が置き換わるのではなく、従来の検索を補完するもの、という位置付けとなると思います。適切な場面で概念検索等を利用することによって、従来の審査で時間をかけて一つ一つ

チェックしていた労力の軽減や効率アップを図るというイメージでしょうか。例えばクラスタ検索による検索結果を本願との類似度順にランキング表示する例が良い例です。また、概念検索は、（自分の審査担当分野でない）別分野にある文献をサーチする場合や馴染みのない分野でも、入力文章と検索対象文献のそれぞれにて使用されているワードの分布の類似性をもとに文献や関連情報を抽出してくれるので、それをクラスタ検索へ展開することによって、サーチ範囲を拡充していくという目的にも有用かと思います。

○特許庁の検索システムを検討・開発して、初めて気づいた点がありますか。

審査官の考え方や検索における思考過程について理解を深めることができたので、どのような検索システムが有用なのか、より深いレベルで考えることができたことも印象的です。

5. おわりに

以上、検証用モデルについて体験記を報告させていただきましたが、最後に、この場をお借りして、今回の取材にご協力いただきました日立製作所の須藤毅様を始めプロジェクトメンバーの皆様、及び庁内関係者の皆様に深く御礼を申し上げます。



日立製作所の開発チームの方々

左より、データマイニングを研究されている相園様、自動分類付与支援の研究をされている佐藤様、プロジェクトマネージャの須藤様、プロジェクトリーダーの八木様、概念検索及び自動分類付与支援を研究されている間瀬様。