

概念検索システムの現状と使いこなしの検討

・・・知財力強化に貢献する概念検索・・・

技術士（情報工学）

六車正道

日立公共システムエンジニアリング株式会社

シニア・コンサルタント

5．上手な概念検索のためのその他の考察

5．1 同義語の設定

筆者の経験によると，概念検索においては質問文に同義語や類似語の設定はそれほど気にしなくてよい。ただし，重要なワードの場合は，同義語や類似語はある程度必要だし，同じワードを複数回繰り返しておくことも重要である。

例えば「モータ」で概念検索すると「電動機」としか書かれていない特許は検索できない。このような数少ないワードで概念検索する場合，高い再現率を望むのであれば，利用者が入力するかシステムの助けを借りて同義語を補う必要がある。しかし，これは概念検索の普通の使い方としては，妥当な使い方であろうか。

一般的に，システムの利用には操作法と上手な使いこなしの二つの段階がある。数少ないワードで概念検索するのは操作法としては正しくても，通常の利用では上手な使い方ではないと思われる。

概念検索の特徴は，質問文中のワードとその出現頻度で成り立つ技術概念に類似した順に，データベース中の特許をリストアップするものである。もし，「モータ」など数少ないワードで概念検索をすると，数万件以上の関連特許を対象に単純にそのワードが使われている回数の多い順に特許を並べることになる。そして，それらの内のわずかに上位数十件を見るわけであり，通常の利用においては，意味のある結果を得ることはないように考えられる。つまり，何か具

体的な詳細技術や典型的なものを入手したいのであればそれに限定した文章にすべきである。概念検索の特徴から，通常の利用としてはもう少し技術の細部まで限定し，それに添っているかどうかの観点で回答の優先順を決めさせるのが上手な使い方のように考える。

例えば，「モータの回生運転時に回生エネルギーを消費する電力形抵抗器を本体に接して外部に設置した可変速制御装置」などのように詳細技術に限定するのが上手な使い方と考える。さらに，最も特徴的な部分を表す「電力形抵抗器を本体に接して外部に設置」などを繰り返して質問文に入れておくことは，いっそう上手な利用法である。そして，このような利用においては，モータの同義語である電動機を質問文に追加するかどうかにかかわらず，ほとんど違いの無い結果を得る。つまり，モータの代わりに「電動機」としか書いてない特許も抽出することができる。それは，モータというワードが無くても，他のワードの存在によりほとんど同じ意味の広がり＝概念になるためと考えられる。

5．2 文章とワード入力は同じか？

現在の概念検索システムは，文章をワードに切り離してその出現頻度を利用するので，基本的にはワードを入力した場合と同じとみて良いと思われる。（入力したワードがさらに再分割されることもある。）

ところが，実際の利用に際しては，全てワードで入力するよりも，名文である必要はないが

文章を入力しこれに重要ワードを追加するのが上手な利用と思われる。その理由は、ワードだけを入力すると、知っている同義語を無意識のうちに加えてしまう恐れがあり、それが検索に悪影響を及ぼすことがあるからである。

実例として、「携帯電話で自動販売機の支払を行うとき本人確認する」ことに関する特許の検索をおこなった。一つの質問はこのままの短文とし、もう一つはワード入力とするために次のキーワードを入力した。

「携帯電話 移動通信 自動車電話 自動販売機 支払 購入 料金 入金 対価 本人確認」
これには携帯電話と支払いに関する同義語、類義語がいくつも書かれているが、自動販売機と本人確認は一語ずつになっている。

表5-1は、両質問による上位20位までのヒット件数とその実例である。短文入力ではほとんどヒットだが、ワード入力では本人確認に関するものが少なく抽出件数は半減している。

表5-1 文章とワード入力の比較

	短文入力	ワード入力
上位20位までのヒット件数	18件	9件
短文入力のヒットの実例；		
1位 特開 2003-337911 携帯電話を利用した料金収納法		
2" 特開 2002-183641 携帯端末を利用した本人確認		
3" 特開 2003-006549 指紋認証を搭載した携帯電話		
4" 特開 2001-351155 携帯電話を利用した自動販売機		
5" 特開 2003-091687 料金支払の切替が可能な携帯		

これから分かるように、ワード入力する場合は、そのワードが重要かどうか、他のワードと意味が重ならないかなど考えつつ同義語を補うような気遣いが必要であり、面倒なだけでメリットは少ない。それよりも文章を書き、重要ワードを追加して、合計で数十～100文字程度の質問にする方が、簡単で効果的と考える。

5.3 分割による上手な検索

上手に検索できない場合、次のような観点で分割して概念検索を行なうとうまくいく。

- ・テーマを分割する
- ・出願人で分割する
- ・発行時期で分割する

(1) テーマを分割する

複合的な技術テーマの場合には、いくつかに分けて検索することで上手な検索の行なえるこ

とがある。分割においては、意味的に and の部分では分けずに、or の部分で分けるのが良い。例；セキュリティ対策の本人確認システムで指紋や静脈パターンや音声パターンのいずれかを使う。

上記の文章で検索すると、指紋、静脈パターン、音声パターンの3つを共に使うものが上位に来ることになる。しかし、この文章の目的は指紋、静脈パターン、音声パターンのいずれかを使うものである。つまり「セキュリティ and 本人確認 and (指紋 or 静脈パターン or 音声パターン)」となっている。したがって or 部分で分けた3つの文でそれぞれ検索するとうまくいくことがある。

一例；セキュリティ対策の本人確認システムで指紋を使う。

なお、この分割検索が特に役立つのは次の場合である。

- ・3つを共に使うシステムが大量にある場合…個別に使うものが上位にならない。
- ・3つの細分技術の製品や特許に大きな差がある場合…適用の多い特定の細分技術が上位を独占し、他のものが上位にならない。

(2) 出願人で分割する

技術的な説明文章だけではノイズに埋没してしまい、必要なものが上位に見られないことがある。このような場合に、あらかじめ分かっている先行企業や専門メーカーなどで絞り込んで概念検索することで、劇的にうまくいくことがある。

IPC、または特定のキーワードによる絞り込みも同様に大きな効果がある。しかし、これらの場合には、絞り過ぎによる漏れに注意する必要がある。

(3) 発行期間の限定による検索

対象特許を発行期間で分割する検索は、例えば出願日以前だけに限定するような利用において役立つが、次のような使い方でも非常に有益である。

図5-1は、ある概念検索をしたときの特許とそのスコアの分布のイメージ図である。2000年前後にスコアの高い特許がたくさん発行されている。このような状態では、その後いつ検索しても2000年前後の特許が上位に出てきて、それ以外の状況が分からないことになる。そこで、1993～1997年とか、2003年以降など期間を区切

って概念検索すると、その期間内でスコアの高い特許が上位に出てくることになる。つまり、虫眼鏡のように部分だけを拡大してみることで、全体でのハイスコア以外の特許の状況を知ることができる。

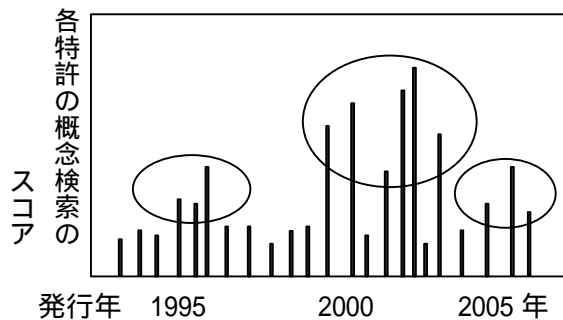


図5 - 1 概念検索結果の分布

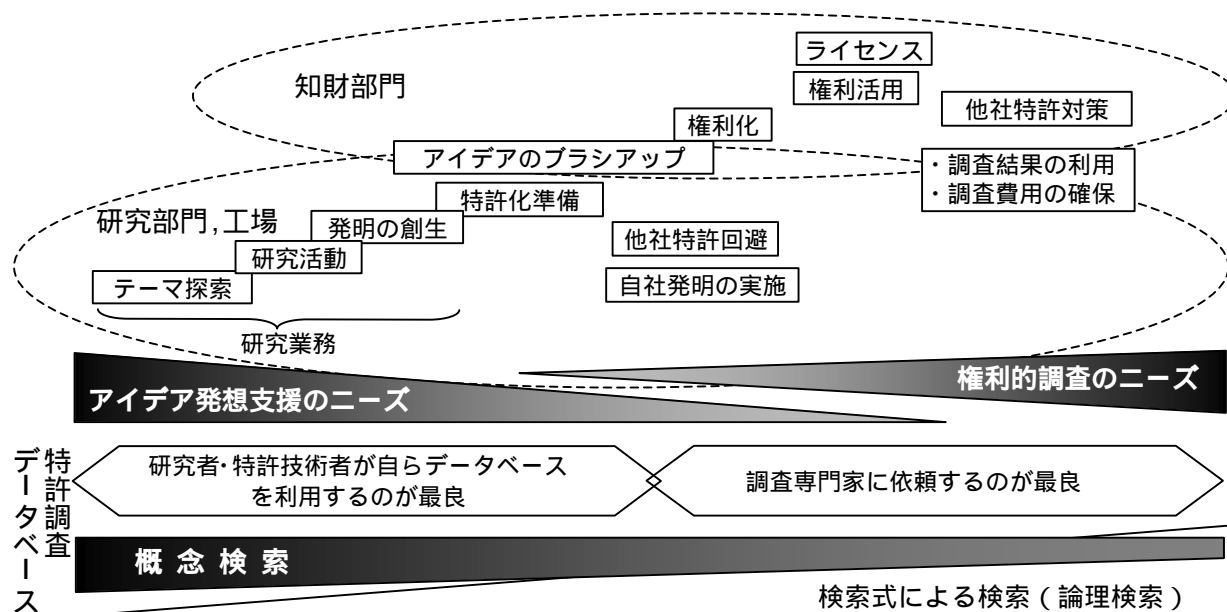


図6 - 1 強化すべき知財力とデータベースによる支援

6. 概念検索の適用の拡大

概念検索の利用は人間の行動パターンに近いものであるため、現在主流の論理検索を取り込みながら、将来大きく発展するものと考えられる。概念検索自体の改善はシステムの研究者によって行われるだろうが、利用法は特許情報の利用者、またはその取り纏めを行っている部署で行う必要がある。

そこで、簡単に考えると概念検索には不向きと思われるいくつかの使い方について検討する。

6.1 概念検索のアイデア発想支援への利用

(1) 知財力強化の高まり

知財力強化の機運が高まっており、概念検索による支援を考察する。図6 - 1は、研究のテーマ探索から始まる知財関連業務を知財部門と研究所・工場に分けてみたものと、データベースによる支援を示したものである。

発明活動や研究活動の段階で、概念検索はアイデア発想支援ツールとして大きく貢献しうる。そのためには、研究者自らデータベースを利用することにより、最も効果的な利用が期待できる。このとき、検索式の作成は容易ではないが、概念検索の利用は比較的容易である。つまり、発明の創生、育成のための知財力強化として、研究者や特許技術者が概念検索によりデータベースを利用するのが効果的である。

また概念検索は、特許の権利化や権利活用においても、重要な調査手法の一つとして論理検索と共に利用することが可能である。詳細は6.4でも検討する。

(2) 特許マップの役割

特許情報を概念検索で利用することによるアイデア発想支援への活用とは「概念検索を使うことで自分の課題やアイデアと類似の特許情報を簡単に探すことができる、これを利用して、先行知識を学んだり、自らのアイデアの不足部

分を発見することができ、アイデアの創生やアイデアの完成につなげることが期待できる」といえる。

その方法についてはすでに述べられている¹⁾が要約すると、見つけた明細書をよく読んで知識を増やしておくことで、課題を一生懸命に考えるときアイデアが閃くという直感的な利用を期待するものである。

これに対して、知識を整理する方法としてマップを作成することも有益といわれる。図6-2は直感的利用とマップによる分析的利用の役割を示したものである。明細書を読み込むことで、必ずしも明瞭ではないが全体的な姿を直感的に把握することが期待できる。一方、それらを通じて知った重要と思われる観点にそって分析マップを作成することで、観点は限定的だが明瞭な知識を得ることができる。これらは補完的な関係にあるが、ベースとなるのは抄録や明細書の読みこなしである。

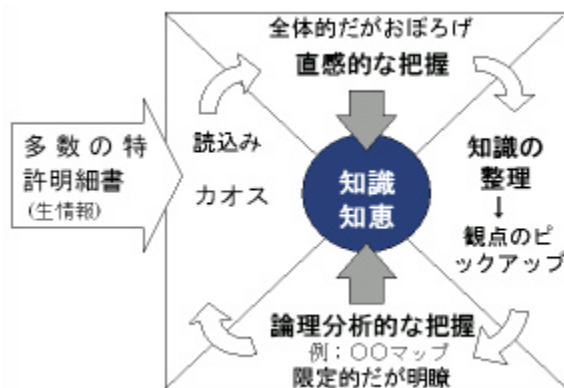


図6-2 直感的利用と分析的利用

6.2 特許マップや動向分析への利用

概念検索はテーマの細部を限定した検索に適用している。したがって、関連特許の全体を出力しておいて、人手により詳細技術や発行年ごとに仕分けする特許マップや動向分析に利用することは難しそうに見える。しかし、概念検索による分析が優れている点もある。

例えば、検索式で数百件(～数千件)を取り出してその内容を逐次見ていく方式に比べて、分析観点の設定やその変更が容易である。つまり、観点を決めて大量の特許を見て仕分け作業をしている途中で、観点の変更・新設が必要になると最初からまたやり直しになり、労力も多くなる。これに対し、概念検索を利用すると、

観点の変更・新設をする場合、そのつどその観点到合わせた概念検索を行えばよい。つまり、概念検索による特許マップ作成や動向分析は、気軽に観点を設定できるのが大きな特長である。

したがって、概念検索による特許マップ作成は、小規模、短期的、技術的マップに適しているように思われる。

6.3 固定費・ASP利用方式

アイデア発想支援などデータベースに不慣れたエンドユーザによる利用においては固定費で利用できることが重要である。少額であってもデータベースの利用量に応じて課金される体制では初心者の活用マインドは抑制される。

図書館が、利用するたびに入場料をわずかでも徴収すれば、利用者は減少するであろう。税金とか会社の費用で運営し個々の利用では料金をとらないからこそ、効果の推定が不明瞭であっても出かけていき、費用を気にせず調べたい内容に集中して利用できる。

アイデア発想支援などにデータベースを利用する場合も同様で、コストと成果の関係は分かり難い。アイデア発想支援に利用するには、データベース費用を固定費としてまとめて支払っておき、個々の利用においては費用の発生しない体制が重要である。

社内データベースを構築すると上記の目的は達せられるが初期投資が大きくなる。特に、特許庁の審査経過まで収録するなどデータベースが複雑化になると社内データベース化はいつそう困難になる。しかし、ユーザID数や利用量ごとに費用の発生する商用データベースでは、年に数回しか利用しない人までユーザIDを持たせることは困難である。

そこで、社外の情報センタでサーバを運営してもらい、ユーザ企業側には社内ユーザIDを管理するコーポレート・サーバ(以下、CSと略称する)を置くという一種のASP(アプリケーション・サービス・プロバイダ)サービスが期待される。図6-3はこのシステムの典型的な一例である。CS側にはキャッシュディスクを付けることで、一度センタ・サーバから取り込んだデータを保存して以後の表示を速めるとか、社内での機密メモを保存するなどの便利な運営もできる。

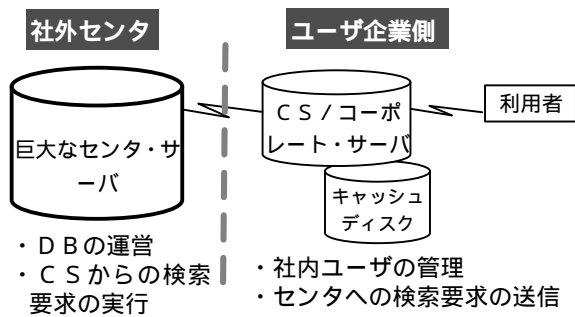


図6-3 コーポレート・サーバ利用のASP

このような方式では、契約で決めた数以内(例えば数個)の検索式などの実行をCSからセンタ・サーバに同時要求できる。一方、CSは社内に発行している多数(例えば数千)のユーザIDを管理し、このうちの一部のユーザID(例えば数十)の接続を受け、リアルタイムに管理する。CSが管理している、また実際に接続しているユーザID数はセンタ・サーバには影響しないので、ユーザ企業側でいくらでも増やせる。多くの場合、大勢の利用者がCSにつないでいても、センタ・サーバに大きな負荷を与える検索の実行は低い割合でしか行なわれないので、このような利用が可能になる。もし、利用者から契約で決められた数を上回る検索実行が

要求された場合は、CSで待ち行列に入り先の検索が終わってからセンタ・サーバに送られる。もしそのための処理の遅れが顕在化する事態になれば、契約を変更して同時接続数を増やす必要がある。

このような構成にすることで、社内の出費を押さえる一方で利用のたびに料金が請求されるような体制を避けることができ、アイデア発想支援などのようなエンドユーザによる利用も促進させられる。

6.4 概念検索を利用した特許調査フロー

これまでの手めくり調査とか検索式による検索などに加えて概念検索が実用化されるにしたがって、受託による特許調査の方法も変わっていく。図6-4は概念検索が実用になった時代の特許調査の実施フローである。だけ行うのが従来の調査のやり方である。これに加えて、概念検索が利用可能な場合には、まず調査依頼を受けたとき、依頼書の主要な文章で概念検索を行なう。これは予備検索の位置付けになる。文章を変えて簡単に数回実施して近い特許を知る。それを利用して本格的な調査にとりかかる。

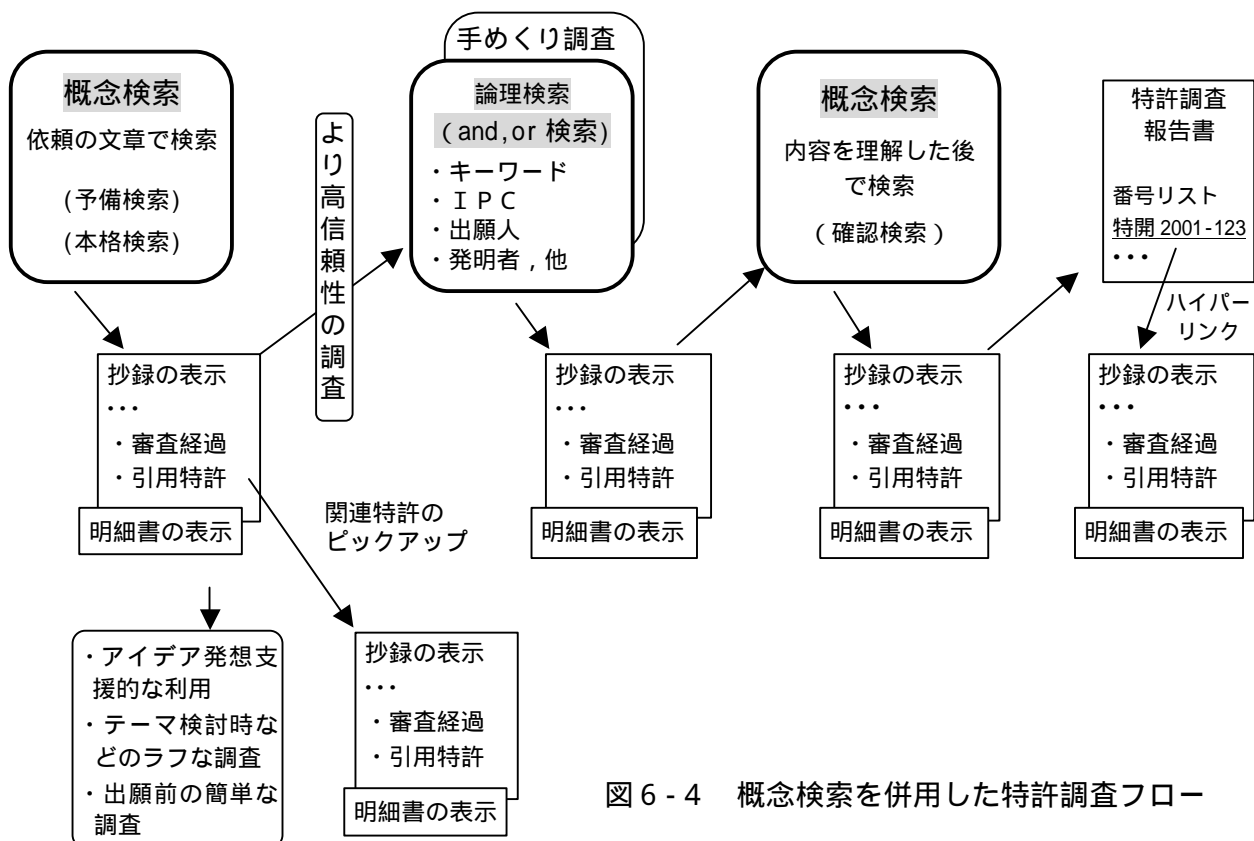


図6-4 概念検索を併用した特許調査フロー

もし、審査引用データが収録されているデータベースでは、近い特許が分かればそのデータにより芽づる式に関連特許を増やすこともできる。

なお、アイデア発想支援などある程度簡単な調査でよい場合には、だけで終わりになることもあり、その場合は予備調査ではなく本格調査の位置付けになる。

さらに、調査対象の技術をよく理解した後で簡単に概念検索を実施することも再現率向上に有益である。これは確認検索と言える。

7. 将来の概念検索

複数の概念検索システムがすでに実用に耐えるものになっており、今後さらに改善が進み、特許情報の検索も将来は概念検索が中心となっていくものと考えられる。

1970年頃までは、特許調査は特許分類のみを頼りに、特許部門の書架に並べられた紙の明細書を使って行なわれていた。その後最初に見る資料として明細書に代わって抄録が使われるようになった。1980年前後からキーワードによる検索式でコンピュータ検索を行うことが始まり、まず調査専門家が補助的に利用を開始した。1993年の電子公報の発行以来、明細書を対象とした検索が可能になり、ワード間の近さを指定する近接演算が日常化され、また、操作法や表示法も改良された。そして、パソコンや通信の発展と共にエンドユーザによる利用が増加し、紙による調査はほぼ過去のものになった。

この延長として概念検索が実用化され、情報の利用者自身により、情報が利用されるその場で行なわれようとしている。

今後は、いっそう人間の行動、つまり利用者のニーズにそった技術が研究され、実用化されていくものと思われる。たとえば、文章の係り受けを理解できるもの、特許1件の中でどの文章が重要か分かるもの、特許の要点が一致するものを探すか質問文に沿った部分があれば一致とみなすか選択できるもの、などは利用者として必要なものではないだろうか。このような発展と共に、将来は概念検索が中心となり、希望すれば論理検索も行える方向に進むと考えられる。

一方、将来の検索システム、つまり概念検索を考えると、展望できることと実際に使えることを区別して取り組むことが重要と思われる。あまりに短い時間でほとんど漏れの無い調査を求めると困難である。また、特許請求範囲をそのまま質問文として使って妥当な結果を得ることは、一部を除き、かなり難しい状態が続くであろう。

しかしそれらの事情を承知しておけば、概念検索は今後ますます便利な道具として、知的活動に貢献するものと考えられる。

参考文献

- 1) 概念検索による特許情報の活用法, 六車正道, 発明, 2003.4月, 5月, 6月