

概念検索による特許情報の活用法 その2

六車正道 (むぐるま まさみち：㈱日立製作所 知的財産権本部 特許主幹，技術士)

前号で概念検索の下痢や特徴を紹介した。今回は、さらに詳細な特徴や発展の方向、上手な利用法などを紹介する。

4. 質問文の長さによる再現率の比較

概念検索では、まだ質問文と回答の関係が十分に分かっていない面がある。部分的に分かっている事は、例えば、クレームは検討されつくした論理的な文章であるが、それを質問文としてシステムに与えても、上手な概念検索はできないことである。特許番号をシステムに与えて、それと似た特許を検索する類似文書検索といわれるものも、結局は指定した特許のクレームや要約を使うものであり、似ており、やはり再現率は低い。平均的にみて2, 3割り程度の再現率はあるようだが、それでは実用にならない。

図5は、質問文として長い明細書の「全文」と「要約」「クレーム」や短文を比較⁴⁾したものである。短文以外での再現率は低い。

対象	A社システム					B社システム			
	全文	全文	要約	全文	全文	要約	クレーム	要約	
質問	全文	クレーム	クレーム	短文	短文 and 必須ワード	要約	クレーム	短文	
特許									
回答									
No.1	数件	×	×	×	○	◎	×	×	×
No.2	"	△	△	×	○	◎	×	△	×
No.3	"	△	△	△	◎	◎	×	×	◎
No.4	"	×	×	×	△	○	×	×	×
No.5	"	×	×	×	△	◎	×	×	△

図5 質問文の長さによる違い (2002.2月実施)

- ・結果は上位50件までの評価。
- ・短文は技術上のポイントを表す1～2行の文
例；圧縮ガスにガスタービン排気での熱交換で得た飽和蒸気や過熱蒸気を混合して供給ガスとする (必須ワード；混合)

(1) 短い質問文が上手に検索できる理由

質問文の長短は概念検索システムを利用する要件ではないものが多い。しかし、実際には長い文章ではうまくいかないことが多い。

数行の短い文章のワードでは、探そうとしている概念が明瞭になるので、概念検索はうまく

いくことが多く、長い文章には多くのワードが含まれ、多くの概念を併せ持つことになり、その文章のどの部分が重要なのか分からないため、うまく行かないことが多いようである。

(2) 短かすぎる質問文

例えば、「レンズ付きフィルムの不正詰め替え防止」で概念検索すると、上位にビンの詰め替えなどが上位にきて、十分検討された検索式による検索結果よりも再現率の低いことがある。その原因として、この文章はやや短かすぎると思われる。「レンズ付きフィルムの不正詰め替え、詰換えを無許可で行い、流用や再利用、再使用、リサイクルを防止するもの」と言う程度の文章が良い。このように、**質問文は数回作り変えることが必要**である。また、不正詰め替え防止の具体的な方法を特定できる場合は、それらの説明文を追加することは大変有益である。

(3) クレームでなぜ上手な検索ができないか？

現在の概念検索システムは、質問文をワードに分解し、ワードの出現頻度で質問文と対象文献の類似度を判定するものである。つまり文章としての良し悪しは考慮されない。これに対し、クレームは文脈による意味が重要である。また、クレームにおけるワードは、上位概念の表現にすることが多く、日常表現的な技術ワードではカバーできないことが多い。例；ポリ塩化ビフェニルを有機塩素化合物という。

またクレームでは、前提とするようなワードを「前記○○の・・・」というように何度も記述することがあり、このときは何度も書かれたワードが重要なワードと判断される間違いもおきる。このような理由で質問文としてクレームをそのまま利用するのは不適当なことが多い。

(4) 特許番号指定の類似特許の検索

特許番号を指定するだけで類似特許を探せ

れば大変好都合である。番号指定の検索とは、明細書全文や要約などを質問文とすることである。しかし、図5からも分かりますとおり、上手な検索は難しい。それは、質問文から中心課題をピックアップするときの問題がある。つまり、質問文としての長い明細書などからシステムの基準でワードが選択される時、利用者が重要と考えた概念やワードが必ずしも選択されないということである。

取引	(□)	100	競	(□)	100	価格	(□)	100
ネット競	(□)	63	購入	(□)	63	希望	(□)	63
ネット	(□)	63	許容	(□)	63	インター	(□)	63
動的	(□)	63		(□)			(□)	

図6 質問文から切り出したワードと重み付けの例

(注. この画像の著作権は株式会社日立製作所に帰属します。)

質問文 インターネット競り取引で、購入希望価格と最大許容価格を入力し、自動的に競り取引を行う

つまり、利用者にとって重要な事項を含む特許があり、その事項がその特許の重要ポイントでないとき、その特許の明細書全文や要約を使って概念検索をしても、うまくいかないということである。

5. 概念検索の発展

概念検索はいろんなやり方があり、その有益性から考えて、今後次々と改良がされるであろう。改良の実例のいくつかを紹介する。

(1) 配点の変更, ワードの変更

概念検索ではシステムがワードを選択し、対象特許を自動的に評価して点数を与える。一般的な利用では、内部処理の詳細はブラックボックスとして十分利用可能である。しかし、うまく検索できない場合など、質問文からどのようなワードを切り出しているのか知りたい、また切り出したワードを変更したい、さらに重み付けを変更したいという要望がある。また、検索式方式の検索に熟練した人も同様の希望をもつことがある。

図6は、「インターネット競り取引で、購入希望価格と最大許容価格を入力し、自動的に

競り取引を行う」という質問文を与えたときに切り出されたワードとその一次的な重み付けである。概ね人間が考えるワードに切り出されているが、「インター」と「ネット」が二つに分かれていたり、「ネット競」など馴染みのない切り出され方もある。また「取引」などは2回使っているので点数が100点であり、1回ずつのワードは63点となっている。もし、ここで「許容」は重要ワードだから点数を100点

に変更すれば、概念検索の結果は若干違ったものが期待できる。また、ワードの変更や追加もできる。例えば「インター」と「ネ

ット」と分かれているのを「インターネット」にしたり、「オークション」などを追加することもできる。

(2) 必須条件での絞り込み

概念検索システムは蓄積された全体を対象に行うので、利用者は文章を与えるだけで検索できて、便利である。しかし、全く異なる技術分野の特許が上位に紛れ込んで上手に利用できない場合がある。このようなときは、必須条件で絞り込んだ結果を対象に概念検索を行えると非常に効果的である。図7は、キーワード、またIPC(国際特許分類)や出願人などで、orやandにすることができるシステムの例である。

図8は、「ポリ塩化ビフェニルをアルカリを加えた超臨界水で分解処理する無害化法」の検索をしたものである。同図の(1)は上記の文章で検索したもの、(2)は必須キーワードとして「超臨界」を設定した概念検索である。この限定の結果、全体を対象とした場合7番目であった特許が2番目に上がっている。また数十番目にあったものが3、4番目に来ており、超臨界を含んでいるのでヒットの可能性が大きい。と

ころが、この例で必須キーワードを「超臨界水」にすると「超臨界条件」などと書かれた特許が漏れることになるので、長いキーワードとか、詳細な IPC で絞るのは慎重にする必要がある。さらに、この絞込みをやり過ぎるとそのときに漏れが発生するので、あまり多重の and は避けた方がよい。つまり、必須条件の設定とは検索式方式との組合せ方式であり、上手に利用すれば効果は大きいですが、絞りすぎると漏れの恐れが

図7 補助機能として必須ワードを設定した概念検索
(注. この画像の著作権は株式会社日立製作所に帰属します。)

大きくなる。

この絞込み機能は、あくまで補助手段と考えるのがよい。この絞込みが常時必要であると、簡単に利用できる概念検索のメリットが全くなり、簡単、短時間にエンドユーザが利用するという新しい道は開けない。

このように補助的な機能を付けることによって、概念検索がブラックボックスで一方向的なものではなく、システムと利用者がインタラクティブ（双方向）な関係となる。これらは再現率の向上だけでなく、親しみやすさの面からも意味のあることである。当然であるが、利用者の工夫を反映させられるので、熟練サーチャはいっそう効果的な利用が可能になる。

6. 上手に概念検索を行なうには？

短時間で再現率の良い概念検索を行う方法をまとめる。

(1) 質問文は短文が良い

質問文としては、概念が特定できる程度の短文が良い。要約やクレームも長すぎ、質問事項のポイントからなる数行程度の短文がよい。ただし、概念が特定できない短すぎるのも良くない。

(2) 結果をみて、数回やり直すのが良い。

最初から最適の質問文を作ろうとすると、難しい。まず思いついた文章で検索し、その結果を見て、修正するのが実際的である。そのとき、

見つかった類似特許の表現を利用するか、不要なワードの削除などをおこなう。
(3) 特許明細書に使われているワードを使う

概念検索はワードを頼りに検索するものであるから、対象である特許明細書で使われている用語をほとんど使っていない質問文では検索できない。しかし、同義語の説明でのべたように、全てのワードが明細書に使

われていないと検索できないかということそうではない。

例 「写真機やカメラにおいて、レンズ付きフィルム不正詰め替え防止」という質問文はあまり良くない。明細書に写真機とかカメラは書かないものがかなり存在するため、質問文にそれらを入れると漏れが多くなる。

(4) 重要なワードは複数回使う

重要ワードは、同義語、類似語も使う。

(5) 特徴タームの設定、点数の再設定も有益

補助的に利用すると効果的である。

(6) 必須ワードの設定は有益

必須ワードの絞りが狭いと、漏れが多くなるので、あまり狭くしないような注意が必要。補助的に利用するのがよい。

(7) 検索対象は明細書がベスト

検索対象としては、要約やクレームよりも明細書全文の方が都合が良い。要約やクレームで

はその発明の代表的なポイントは書かれるが、複数の重要なポイントが全部記述されるとは限らない。これは、概念検索に限らず一般的な検索や調査でもいえることである。

例 「パソコンの冷却用の磁気式液体振動ポンプ」の質問文で要約やクレームを対象として概念検索すると、下記特許はパソコンなどの記述がないために上位に出てこない。しかし、明細書全文を対象に検索すると、上位に出てくる。

1	4080	特開平 10-225632	超臨界水による
2	3764	特開平 09-103672	ポリ塩化ビフェ
3	3659	特開平 10-118618	ポリ塩化ビフェ
4	3595	特開平 05-253318	塩化合物の無
5	3468	特開 2001-300299	有害ハロゲン化
6	3440	特開平 10-314695	有機塩素化合物
7	3358	特開 2000-116814	有機ハロゲン化

(1) 質問文だけの概念検索

図8 「ポリ塩化ビフェニルを、アルカリを加えた超臨界水で分解処理する無害化法」の概念検索

特開平 7-***・・・【発明の属する技術分野】本発明は例えばノート型のパーソナルコンピュータ(以下ノートパソコンと称す)などの持ち運び可能な電子機器に用いる冷却装置に関するもので・・・

※質問文作成のテンプレートはないか？

作成は難しいように思われる。質問文は、類似特許が少ない間は簡単でよいが、多い場合には詳細にする必要がある。例えば、「家庭電化製品の不法投棄の監視システム」に関する特許は2001年あたりまでは少なく、このままの質問文で概念検索を行っても十分実用になった。しかし、2003年になると関連特許が多くなり、もう少し詳細な技術内容を書かないと自分の欲しい特許が上位になることが難しい。このように、探す内容をどの程度詳細に書くかという問題は、データベース全体の状況で変るものである。そのようなことまで承知したテンプレートは難しそうである。一方、質問文は、主語、動詞などの整った文である必要はないのであるから、その面からもテンプレートは不要

と思われる。

7. おわりに

概念検索のメリットをまとめてみる。

- (1) 技術的な説明文で検索できる・・・検索式を作らなくてよい。
- (2) 優先順位の回答・・・全件見なくても良い。
- (3) 同義語の心配がほとんど不要
- (4) 件数を適当に絞ることが不要。

1	4080	特開平 10-225632	超臨界水による
2	3358	特開 2000-116814	有機ハロゲン化
3	3229	特開 2001-218866	ハロゲン化合物
4	3024	特開 2001-047026	有害物に汚染さ
5	2842	特開平 08-173572	ポリ塩化ビフェ
6	2491	特開 2002-360728	P C B無害化処
7	2311	特開 2002-001089	超臨界水反応装

(2) 「超臨界」を必須ワードとした概念検索

- (5) 技術者が短時間で利用できる・・・本来の情報のユーザであるエンドユーザの直接利用が可能になる。

ところで、概念検索は原理がいまひとつ分かり難く使う気がしないとの意見が、サーチの一部にある。(エンドユーザにはほとんどない。)

この疑問に対する答えは自動車の例で考えてみると分かりやすい。自動車の走る原理はブラックボックスではないものの細部は良く分からないが日常的に利用されている。これは、練習を通じて、使い方(運転)と結果の関係が運転者に理解されているためである。しかし、自動車出現の初期においては、馬車の親方は馬車の方がシンプルだと言い、また、走行原理の細かな開示を迫る人がいたかもしれない。それは、馬の機嫌をとるのに苦労した人の意見であったり、また、自動車の使い方に慣れていない人の意見であったろう。なお、現在でも自動車より馬車が良いという人はいるが、馬

に乗るために自動車で出かけることもある。

概念検索もよく似た状況にある。なぜ機能するかそのロジックは、本稿で紹介したようになり明らかにされており、ブラックボックスではない。上手な利用を行うには原理の細部を理解するだけでなく、使い方と結果の関係を理解し、慣れていくのがよい段階にきている。

実際に使ってみて、大勢の利用者の意見を聞いていると、将来間違いなく概念検索は発展すると思われる。しかし、ユーザやそれを代表する人が、新システムが必要であることを意識し、

改善を続けなければ利用できるのは遅くなるように思える。I T社会の急激な進展は情報洪水をもたらし、情報麻痺を起こしているように思う。

ほしいものは触覚を伸ばして探し、改善を働き続けていくことが必要ではなかろうか。

参考文献；

4) 日立製作所における特許情報の活用，六車正道，INFOSTA シンポジウム 2002（2002.6）