# 概念検索についての誤解

六車技術士事務所・所長
六車　正道

PROFILE 約36年間、日立製作所において特許情報の活用企画と実務を担当し、2006年2月に独立して技術士事務所を開設し、特許情報の活用促進に関するコンサルタント業に従事。関係業務の講演や著作が多数。PatentCityの運営者。技術士（情報システム）。

〉http://ipbase.cool.ne.jp/mailmug.htm より　〉050-8012-2416

## 1 はじめに

情報検索において、検索式の代わりに文章で質問でき内容的に一致の可能性の高い順に回答されるシステムは概念検索、または類似文書検索といわれている[1)]。概念検索による日本特許情報の利用は2000年に商用サービスが開始されて以来、多くのシステムが装備するに至っている。

ところが、概念検索はさまざまの誤解により適切に評価されておらず、利用はやや停滞しているように思われる。その誤解の原因として次のようなものが考えられる。

・動作原理が、現在普及している検索式（論理検索）と異なって「重み付け」を利用すること。
・操作が簡単であるためにその習得だけで全体を理解したと誤解しやすいこと。
・過度の期待と上手に使えなかった場合の反動。

本稿では、概念検索の妥当な機能、性能、利用法に関する理解がなされ上手に使われることを期待して、これまで見聞した誤解と思われる事項とその解説を紹介する。

## 2 何をしているのか分からないという誤解

検索式による情報検索に慣れた人は、and,orによる検索式の論理は明快だが概念検索の検索論理は不明瞭という印象を持つことが多い。

検索式と概念検索の最も大きな違いは、検索式の検索論理は指定したキーワードの有無という単純なものであるのに対し、概念検索はワードの出現頻度を利用してワードに自動的に重み付けをおこない、回答特許を順位付けする点である。このため、概念検索は質問文中のワードのandでもorでもない結果になり、検索式は論理が明快であるが概念検索は何をしているのか分からないという評価を受けることになる。

概念検索では、質問文がコンピュータによりワードに分解され、それらの出現頻度（TF）と、データベース全体においてそれらの出てくる特許件数の割合の逆数（IDF）を使って計算される。このテクニックはTF·IDFといわれ[2)] 現在のほとんどの概念検索に共通の基礎技術である。しかし、これ以上の細部、つまりワードの切り出し方の違い、重み付けの計算の細部などはシステムによって大きな違いがあり、結果的に概念検索システムをひとまとめに説明できない状況がある。また、各システムでは特徴を分かり易く説明しなかったり、機能を明確にしないものもあって誤解を大きくしている。

利用者としては概念検索の基本はTF·IDFであるということを理解し、コンピュータ技術の細部の検討は必要最低限にすることも限られた時間の中では必要ではないだろうか。

## 3 役立たないという誤解

適当な文章をポンと入力するだけで有益な結果が得られるはずがないという立場からの誤解とそれに対する一

表1　生ごみ処理の脱臭関連特許の概念検索

※表中の数字は順位。nは上位100位までに無かったもの。網掛けは上位50位まで。

| 検索対象／システム ＼ 関連特許 | A | | B | C | D | E |
|---|---|---|---|---|---|---|
| | 要約 | 請求範囲 | 明細書 | 明細書 | 要約＋請求範囲 | 明細書 |
| 特開平10-290976 | n | n | n | n | n | n |
| 特開2000-042356 | n | n | 49位 | n | n | 28位 |
| 特開2001-070424 | n | n | n | 12位 | n | 10 |
| 特開2001-079521 | 83位 | 34位 | 42 | 15 | n | 47 |
| 特開2001-079522 | 34 | 76 | n | 27 | n | 52 |
| 特開2001-191057 | 5 | 3 | 45 | 20 | n | 44 |
| 特開2001-198557 | n | n | n | n | n | n |
| 特開2001-353419 | 2 | 1 | 32 | 23 | n | 49 |
| 特開2002-186936 | n | n | 37 | 6 | n | 20 |
| 特開2002-204923 | 4 | 35 | 3 | 4 | n | 17 |
| 特開2003-225530 | n | n | 13 | 1 | n | 2 |
| 50位までのヒット件数 | 4件 | 4件 | 7件 | 8件 | 0件 | 8件 |

質問文：家庭から出る生ゴミを微生物で分解処理する際，排気を木材チップに通して脱臭する，木材チップの芳香と微生物で脱臭

応の回答を紹介する。

**(1)概念検索の再現率は低い**

さらに具体的にいうと下記のようなものがある。

*※再現率とはデータベース全体に存在する内容的に正しい回答件数のうち取り出しえた件数の割合。*

・再現率は2、3割程度しか期待できない。

・再現率はシステムによって大きな違いはない。

・漏れを気にしない参考程度の検索に使うものである。

表1のように比較実験を行なうと分かることだが、システムの特徴と利用法の違いによって再現率は大きく異なる[3)]。ところが、それらを考慮しないで再現率を論じるケースを目にすることがある。概念検索システムの作成者やサービスの提供元自体が上手な利用法を十分に説明しない状況もあり、事態は混乱している。

**(2)本格調査の手がかりを得るための予備調査に使うもの**

簡単な概念検索であれば10数分で一定の成果をあげることができるので、予備調査に利用できることは確かである。しかし、筆者の経験では本格的な調査にも十分役立つものである。

**(3)大雑把な検索に適しており絞り込んだ検索には適していない。**

逆である。概念検索は文章中のワードにそって対象技術を絞り込み、回答は順序を付けて出力するものである。目的のテーマに関連する数100件以上もの特許を順位付けしないで広く集めるような調査には検索式の利用が適している。

**(4)技術的調査には使えるが権利的調査には使えない**

技術的調査は目的の技術に関係する特許を広く集めるものであるのに対し、権利的調査とは審査請求前とか無効資料を探すような詳細技術にマトを絞り込んだ調査を指しているものと考える。概念検索は質問文にそって絞り込んで出力するものであり、短時間で非常に好都合の結果を得ることも期待でき権利的調査に適している面がある。

**(5)質問文の文脈やワードの意味を理解していないので使いものにならない**

現在の概念検索は文章の係り受けや否定表現を適切に処理することはできない。ましてやコンピュータがワードの意味を理解しているわけではない。

人が空を飛ぶには鳥のように羽根をばたばたしなくても、固定翼でいいから空を飛んで目的地に早く着くことが大切である。概念検索も実際に使ってみて実用になる程度の回答を得ることができるならばそれを利用するのが正解ではないだろうか。

**(6)同義語、類義語の入力は重要である**

同義語や類義語はあまり気にする必要はなく、むしろ入れ過ぎると弊害の出ることがある。例えば「インターネット競り取引で購入希望価格と最大許容値を入力可能のシステム」という質問文では、1、2のワードが無くてもほとんど同じ結果を得る。例えば「インターネット」というワードの代わりに「電網」などとした場合や全くなかった場合でもほとんど同じ結果になる。これは最近の特許においての「競り取引」は多くの場合インターネットを前提としており、したがって「競り取引」のワードに「インターネット」の意味が含まれていると考えることができる。つまり、概念検索では他のワードによって補ってもらえることが期待できるので同義語はあまり入力しなくてもよいと思われる。

一方、同義語があり過ぎると、それらのために必須ワードの重みが下ったり、時によっては必須ワードが全く無視される場合もある。対策として質問文を分けて別の概念検索を行なうことが有益である。例えば「インターネットのオークションで購入希望価格と最大許容値を入力可能のシステム」などとして別の概念検索を行なうと、新たな特許を見つけることが期待できる。

同義語の問題としてもう一つの観点を紹介する。例えば「モータ」を質問文として概念検索した場合、モータではなく電動機としか書いてない特許は検索できない。これを根拠に概念検索では同義語の入力が必須との議論がある。しかしこれは、概念検索で同義語が必要と理解すべきことではなく、概念検索の特徴を生かした使い方になっていない（つまり、誤った使い方の）問題と考える。概念検索は（通常の特許調査においては）複数のワードで成り立つ一定の狭い技術範囲を指定した検索においてメリットがあるようである。システムの使い方は利用者が決めていいことであるが、特徴を生かしきらない利用法は間違った使い方というべきであろう。

**(7) 上手な質問文の作り方や概念検索のやり方は検索式と同じくらい難しい**

概念検索の利用は検索式の作成やそれによる検索ほどの難しさはない。しかし、それなりの利用法はあるので知っておく必要がある。上手な質問文作成法や利用法を下記に示す。

①質問は概念を明確に特定できる40～80文字程度の文章が最良…長い文章では希望する概念に絞られないために再現率が低くなることが多く、短かすぎると目的とする概念が明瞭に絞りきれず不適当になることが多い。長い文章でもうまくいく場合もあるが、それは目的技術のワードが運よくその長文中で最も頻度高く使われていたケースである。数少ない単語や長文で質問する場合は、高再現率を期待するのではなく簡単であることを生かした利用法と位置づければ意味がある。

②具体的表現にする…特許の検索だからと考えて上位概念の表現や一般化したワードを使うと他の技術との区別がつき難い。

③明細書で良く使われる表現にする。また方法、部品、材料などの技術的な面だけでなく、目的、効果を含む質問文も有益。さらに質問文から切り出したワードを調整したり、重み付けを調整することも有益。

以上をまとめて別の言い方をすると「20秒くらいで社長に説明する文章」とも言える。前提を長々と言うこともないだろうし、他の技術と区別がつくように平易な説明をするだろうからである。

④質問文を変えて数回概念検索を行う…なお、質問文を変えることで少しずつ異なる概念の検索に移っていくことになるので、後の文章が最良とは限らない。各質問文で見つけたヒットを保存しておくことが大切。

⑤必須ワードやIPC、出願人、発行日などにより限定した絞込み検索を併用する…これは非常に有益な方法である。ただし、限定し過ぎにならぬようにorを考慮することが必要になる。

表 2　特開2000-42356（生ごみ処理の脱臭）の関連特許の概念検索

※上位50位までのヒット件数。

①Bシステムによるヒット件数

| 質問 / 対象 | 短文 54字 | 第1請求範囲 216字 | 要約 311字 | 明細書 10,800字 |
|---|---|---|---|---|
| 明細書 | 6件 | 0 | 0 | 0 |
| 請求範囲 | 4 | 0 | 0 | 0 |
| 要約 | 4 | 1 | 1 | 0 |

②Cシステムによるヒット件数

| 質問 / 対象 | 短文 54字 | 第1請求範囲 216字 | 要約 311字 | 明細書 10,800字 |
|---|---|---|---|---|
| 明細書 | 7件 | 0 | 0 | 1 |
| 請求範囲 | 3 | 1 | 0 | 0 |
| 抄録 | 3 | 1 | 1 | 0 |

**(8)検索の対象は明細書と比べて請求範囲や要約でも大差ない**

検索対象は全文明細書の方が再現率は高いことが多い。表2の縦方向は検索対象として明細書、請求範囲、要約を並べ、2つのシステムの実例を示している。いずれも対象は明細書の場合がヒット件数は多い。

**(9)質問文に請求範囲が使えないのでは価値がない**

概念検索を利用しようとする多くの人が、質問文として特許請求範囲を利用したくなるようである。確かに請求範囲は発明の重要点を過不足なくまとめた文章である。しかし概念検索では文章を認識するのにワードの出現頻度を用い、文章の係り受けを重要視する請求範囲とは異質のものであることを忘れてはならない。多くの場合、請求範囲は概念検索の質問文としては冗長であり適していない。

しかし、まれに請求範囲をそのまま質問文としてもうまくいくことがある。そのようなケースとは、次に示すように請求範囲が概念検索の質問文に適したような書き方をされている場合である。例；特開2002-265223の請求範囲；層状チタン酸化物微結晶を剥離して得られる薄片粒子からなるチタニアナノシートの多層構造からなりポリマー介在層を有しないことを特徴とするチタニア超薄膜。

**(10)利用者による工夫の余地は少ない**

通常、概念検索の質問文は作り変えることで適切な質問文になることが多い。これはその間に工夫する余地のあることを示唆している。(7)で説明したような工夫ができる。

**(11)簡単な使い方に特化すべきで利用法の深い研究は間違いだ**

ピアノや太鼓は叩けば音が出るので簡単な利用法に限定すべきと言っているような間違いと思われる。概念検索の操作法は質問文を貼り付けて「検索実行」をクリックするだけという簡単なものだが、妥当な質問文の作成法、絞り込み検索などの利用法など研究する課題は多いし、それにより再現率の高い利用が可能になる。

**(12)注意事項や予備知識が必要というのでは概念検索の存在価値がない**

概念検索システムは通常のコンピュータシステムの一つであり、使いこなしのための注意事項や予備知識は当然必要である。

## 4 なんでもできるという誤解

概念検索は万能のシステムと見る立場の誤解を紹介する。

**(1)思いついた2, 3のワード入力で必要な特許を探せる**

最低限それでも利用できる場合があるということでは間違いではないが、再現率を気にする使い方であればそれだけでは不足である。目的とする技術の内容（概念）を他と区別できるように正確に伝える必要がある。

**(2)質問文は要約や請求範囲、または明細書でよい**

「特許番号指定で類似の特許を検索できる」というのもほぼ同じであり、この場合は種になる文書を要約、請

求範囲、明細書のどの部分にするか指定する必要がある。表2の横の比較で分かるように、質問文を要約や請求範囲、明細書として行なった場合再現率は低い。しかし、簡単に行なえるのでやってみる価値が全くないことではない。

**(3)1つの質問文だけで欲しい特許を全部探せる**

概念検索では質問文を変更して数回行なうことで再現率を上げられることが多い。検索式による検索は関連特許を含めて数100件を出力しそれらを参照して該当のものに絞るため、技術的に広範囲の観点の特許を抽出できる。これに対し、概念検索はピンポイント的に絞って優先順位を付けるために技術的な範囲が狭い。そこで観点を変更した質問文を作成することで広範囲の技術をカバーできるといえる。

**(4)回答リストの第1位の特許が最も関連の高いものである**

最も関連の高い特許が1番目以外に来ることが多いように思われる。これは順位付けする点数は1位と10位ではそれほどの差が無く、些細なワードの違いなどで変わりうる一方で、1位はただ一つの場所であるのに対し、2〜10位は9倍の場所があるという確率的な問題のように思える。

したがって、狭い範囲に限定した通常の概念検索では50位くらいまでは見る必要がある。また、あまり限定しない場合には数100番目まで関連の高いものが出てくることもある。

**(5)目的の技術内容や利用法を知らない人でも簡単に利用できる**

対象の技術内容が分かっていないと妥当な質問文が作成できず、また回答の抄録や代表図面を参照する場合も内容の理解に時間がかかる。技術内容や利用法、つまり上手な質問文作成法を知っていることは重要なことである。

**(6)短時間で高再現率が期待できる**

条件次第ともいえるが一般的には無理である。対象の技術内容を熟知した専門家が概念検索の質問文作成法を知っておれば1時間でもかなりの再現率が期待できる。一方、対象技術の知識が少ない検索専門家（サーチャ）が技術を習得しながら検索式を作って検索する場合は10時間かかっても再現率はあまり高くない。このような比較においては、短時間で高再現率が期待できるといえる。

**(7)概念検索だけですべての検索が行なえる**

「概念検索があれば検索式は使えなくてもよい」とか「近い将来検索式は全く不要になる」などの言い方もある。少々の漏れはあってもよいとするアイデア発想支援的な利用においては概念検索だけで十分といえる。また、特許出願前の簡単な調査では概念検索だけで済ますこと

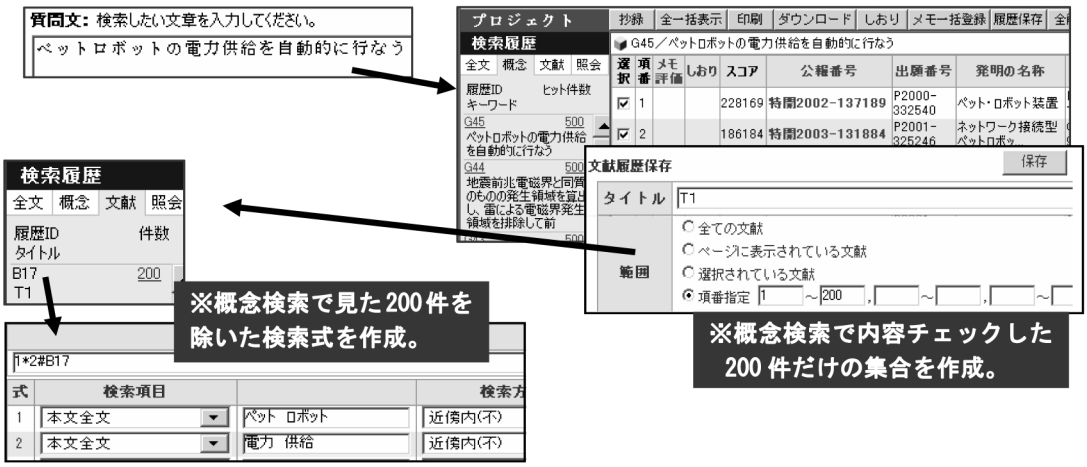


図1　概念検索で参照した特許を検索式で除く利用法

も、方針によっては可能であろう。しかし、高再現率を求める通常の特許調査においては検索式も使える方がよい。まず概念検索を行ない、そこで見た特許を除いて検索式を作るような組合せ利用も効果的である。図1は日立のShareresearch（シェアリサーチ）における組合せ利用の具体例である。

**(8)動向調査が簡単に行なえる**

一般的には概念検索は動向分析には適していない。概念検索は細部まで絞り込んだ対象技術を検索する場合に適しているのに対し、動向調査は多くの関連特許を集めて技術や企業の傾向を分析するものである。したがって、動向分析の対象を狭い技術範囲に絞り込むような工夫をすれば利用可能と思われる。

**(9)現在の概念検索は完成しておりしばらく改良されることはない**

原理的な部分は当分変わらないようであるが、使い勝手などの工夫は少しずつ進んでいる。絞り込み検索の併用や検索式との組合せ利用などは一部で利用可能になっているし、その他にもいろんなアイデアが検討されている。

## 5 終わりに

人が空を飛ぶには、鳥のように羽根をバタつかせるのは間違いであり固定翼にして重い金属の容器に入ることが必須であった。似た文書を探すのに、人間のように文章の係り受けや意味を理解するシステムが正しいのか、そうでないのか長期的に見ていきたい。

これに対し、開発元へのお願いであるが、実務に有益な工夫は早急に検討して実施していただきたい。

概念検索は情報検索を一握りの検索技術の専門家から膨大な数のエンドユーザに開放することは間違いない。それはインターネット検索の隆盛を見れば明らかである。情報検索を職業にしている身であれば、従来のやり方と比較して再現率の違いや多様な出力の利用など主張したい気もする。しかしそれは、昼休みに飲むお茶と茶道のお茶の違いに似ている。役割が異なり、一方の立場から他方を無視するよりも、両者の違いを理解して共に活用するのが前向きな姿勢と考える。

**参考文献；**

1） 特許情報検索の課題と概念検索システムの役割、知財管理、2001.12月、六車正道
2） 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－、発明、Vol.102、No.4、 No.5、2005、六車正道
3） 概念検索の現状－概念検索システムの比較と使いこなし－、JAPIO／創立20周年記念誌、2005.10/12、六車正道