

特許情報検索の課題と概念検索システムの役割

六 車 正 道*

抄 録 コンピュータとネットワークの発展により、昨今は研究開発者や特許技術者など一般ユーザによる特許情報の利用が増加している。しかしながら、現在、多くの情報検索で行われている AND, ORなどを組み合わせた検索式による利用は検索専門家においても簡単ではなく、一般ユーザの利用では大きな壁となっている。

これに対し、質問文を入力するだけで、システム側がその文章を解析して質問に近い特許を探し、優先順位を付けて出力する概念検索システムが実用になりつつある。これにより、自分のアイデアに似た特許を短時間に探すことができ、アイデアのブラッシュアップを迅速に行うことが可能になる。また、特許調査などでも業務の迅速化が可能になる。

目 次

1. はじめに
2. 特許情報の利用法
3. 全文検索の長所と限界
4. 必要とされる発想支援システム
5. 概念検索システムとは何か
6. 概念検索システムの上手な使い方
7. おわりに

1. はじめに

情報とは、見た人の「知識構造を変化するもの」といわれ、情報の利用目的は意思決定に役立てることである。特許情報は出願した各社の研究開発がすべて公開されているものであり、特許情報を適切に利用することで、自社の研究開発費を大きく削減できる可能性がある。

特許情報を利用するには情報検索が不可欠であり、情報検索には AND, OR を組み合わせた検索式の作成が必要である。ところが、上手な検索式の作成にはかなりの熟練が必要である。このため、コンピュータやネットワークなどの環境の改善にもかかわらず、人類の英知を集めた特許情報を、誰もが自由に使いこなせるとはいえない状況である。

一方、インターネットの検索サービスではキーワードを2, 3個入れるだけで、簡単に欲しい情報を手に入れることが実現している。

特許情報においても、欲しい技術の説明文を入力するだけで、関連特許を探してくれるインテリジェントな検索システムがあれば、その本来の価値を発揮できよう。最近、この目的に合うものとして概念検索システムが出現してきた。本稿では、特許情報検索を取り巻く課題を整理し、実際の概念検索システムの機能・性能を参考にしながらその役割を検討する。

2. 特許情報の利用法

特許情報は、研究開発を支援する技術情報と、特許権の獲得・行使に関する権利情報の二つの面がある。いずれの利用においても、できるだけ「簡単、迅速、安価、高再現率、低ノイズ」に検索できることが望まれている。

(1) 簡単

1) 通信とコンピュータの発展により机上のパソコンで誰でも簡単に利用できるようになっ

* (株)日立製作所 知的財産権本部 主任技師
技術士(情報システム) Masamichi MUGURUMA

てきた。しかし、操作が複雑だったり、分かり難いものがいまだに存在する。繰り返し操作を減らし、論理的に明白なことはユーザに行わせるのではなく、コンピュータ（ソフト）で行い、単純化を徹底すべきである。

2) AND, ORを使った検索式の作成は簡単ではない。検索式作成の難しさの一例を示す。「微生物浮遊担体による廃水処理」を探するとき「微生物*水処理*浮遊*担体」の簡単な検索式で検索した場合に比べ、図1のように、キーワードやIPC（国際特許分類）をAND（*）、OR（+）で組み合わせると、約5割増の特許を見つけることができる。

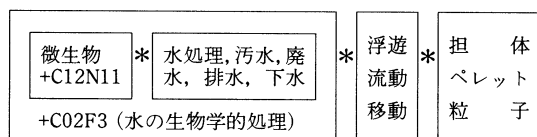


図1 やや複雑な検索式

しかし、このような検索式は、一般ユーザでは作り難いものである。研究開発のキーパーソンは特に多忙であり、このような検索式を考えたり使ったりする時間がない。いつでも、どこでも、誰でも検索できるユビキタスな利用が可能とは言い難い状況である。

(2) 迅速

コンピュータの検索時間だけでなく、特許の抄録や明細書の表示時間も短い方がよい。また、ユーザの理解しやすい抄録出力などの利用で、調査全体の時間を短くすることができる。

例えば、開発会議で出された新しいアイデアについての他社状況がその場で分かれば、詳細検討を次回まわしにせずに、迅速なアイデアの発展・強化が可能になる。

人が記憶から思い出すのは数秒～数十秒である。せめて数分で情報検索の結果を得られれば、思考は中断せずに継続できる。しかし、検索式を作成したのでは数時間より短くするのは困難と思われる。

一方、SDI（情報の選択提供）サービスを受ければ、特許リストをクリックするだけで欲しい特許を迅速に入手できる。しかしこれは、あらかじめ登録しておいた内容に近い最新特許を定期的に受け取るものであり、まったく新しい内容の即時検索・入手とは異なる。

(3) 安価

1) 調査費用はトータルコストで考えるべきである。最近では、データベース利用料金は安くても、人件費が非常に高い。例えば、技術者が1時間かけて無料データベースを使った場合、そのトータルコスト（＝人件費）は数千円～数万円かかっている。

2) データベース利用において、使った分だけ料金を支払う従量料金制は初期経費がないので導入しやすいが、個々の利用で費用が発生するので、利用が専門家中心になる事が多い。これに対し、固定料金制（または社内データベース）はリスクが大きく導入し難いが、導入後は使い放題なので一般ユーザでも利用しやすく、結果的に割安な利用の行えることが多い。

(4) 高再現率

再現率とは、正しい回答全件数の内、取り出し得た件数の割合であり、適合率とはコンピュータの出力や目でみた全件数のうち、欲しい内容と一致した件数の割合である。一般に、高い再現率の調査が良い調査である。

特許調査では権利対策の調査をはじめとして、特に高再現率が求められる。ただし、目的によっては再現率を少々落としてもいいから、費用と時間を限定した高適合率の調査が喜ばれることもある。

ところが、特許調査というと条件反射的に高再現率のための複雑な検索式を作る傾向があり、難しい検索にしまっていることが多い。目的によって使い分けが必要である。

(5) 低ノイズ

特許事件関係の調査の場合は、詳細な点まで

一致していることを求めることが多く、数千件(以上)の特許を見て1,2件のヒットを得るような極端に適合率の低い(ノイズの多い)調査もある。しかし、一般的にはノイズの少ない(適合率の高い)検索が期待される。

開発前や特許出願前の検索では数百件を出力して内容を見ていくことが多い。しかし、これには多くの時間を費やすため、検索でさらに十分の一程度に絞る検索を行なうことがある。ところが、その結果、必要なものが一部漏れてしまうジレンマがある。

最近では、一般ユーザ自ら特許情報の検索を求めており、情報検索を簡単に行なえる体制が望まれている。システムを整備する立場の人は、コンピュータ技術の助けを借りて情報検索の大衆化を進める必要がある。このため、社内の実情を踏まえた特許情報の活用企画を検討していく必要があると考える。

3. 全文検索の長所と限界

最近数年の特許情報のトピックスは、全文検索が手軽に行なえるようになったことである。

特許の重要ポイントの検索は抄録データベースでも一応可能である。しかし、表現がクレームや出願人要約に引きずられた抄録では、キーワードが不足気味のことがある。また、実施例中にしか出てこないワードや詳細説明中で簡単に触れてある程度のワードは抄録に入っていない事が多い。

平成5年の電子公報の発行以来、明細書全文を対象とした検索が可能になっている。これは明細書のどの部分の記述も検索できるために再現率が向上する。ただし、ノイズ対策が必要である。抄録と同じ検索式では10~20倍の回答件数があり、多くのノイズを含んでいる。この対策として、キーワードの近さを指定した近傍検索によりノイズを減らすことが大変効果的である。さらに、明細書を数秒で表示できる機能を

持っていることが多く、これも調査に好都合である。

しかし、全文検索でも検索式の作成は必要であり、近傍検索などは検索を一層複雑にすることになるので、検索の煩わしさは全く軽減しない。これでは一般ユーザ利用の大きな進展は期待できない。

4. 必要とされる発想支援システム

特許情報は、研究開発の効率化のための情報源として期待されている。

一方、製品サイクルの短期間化に伴い、開発工数の短縮化が要求され、洗練されたアイデアが必要とされる。ところが、新分野への参入や技術の細分化・専門化により社内に適当な相談相手のいないことが多い。

しかし、革新的な創造も過去の情報の組み合わせであり、情報の集め方でそれらを補うことが可能と推察される。昔から、三人寄れば文殊の智恵という考え方があり、現在においては、知識を共有化して全体としての効率を高めようとする「ナレッジマネジメント」の手法がある。情報共有の仕掛けはコンピュータを上手に使えると比較的簡単に行える。しかし、メンバーが自分の獲得した情報を吐き出してコンピュータに登録することが、この手法の大きな問題点である。多くの人は、情報は欲しいが自ら情報を提供するの面倒だ、もったいないと考えるからである。

これに対し、特許情報はすでにアイデアをドキュメント化してあり、しかも、社内の限られた人だけでなく、全世界の人がアイデアを蓄積してくれている。「特許情報ナレッジマネジメント」の推進により、世界の研究者の情報を生かす体制を作る必要がある。

ユーザが「○○に関する特許情報が欲しい」と指示するだけで、確度の高い特許をリストアップしてくれるシステムの出現が待たれている

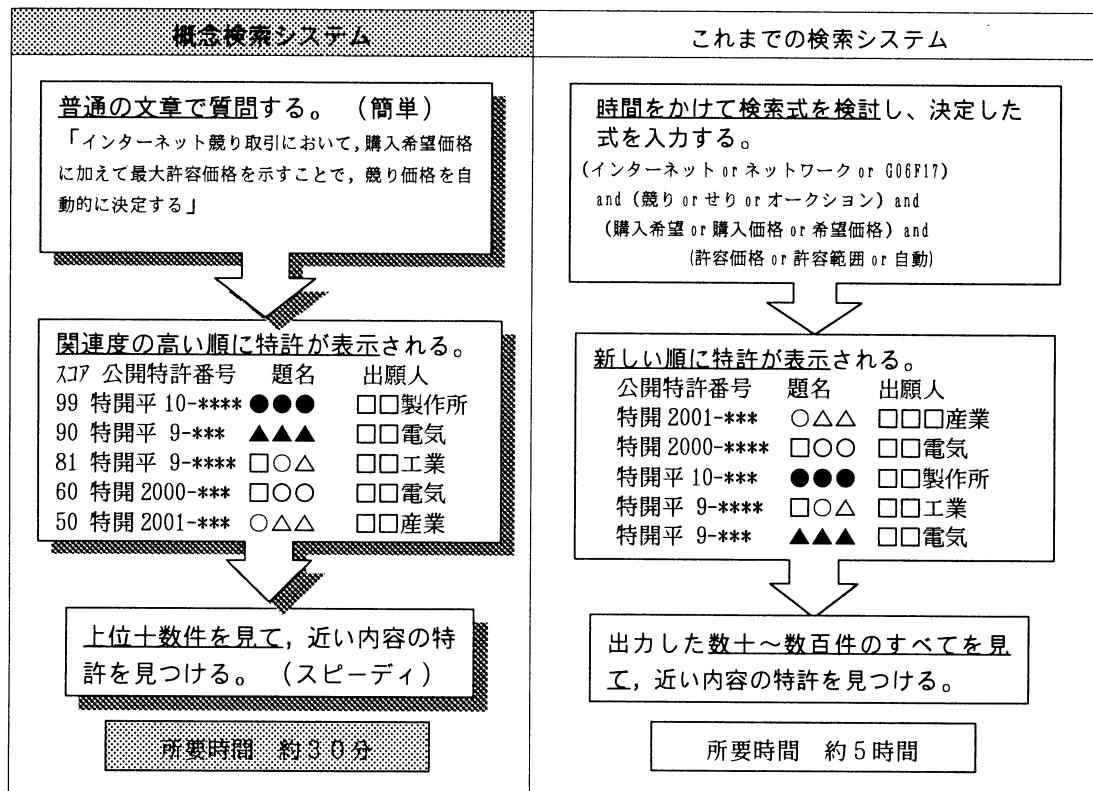


図2 概念検索システムとこれまでの検索システムとの比較

といえる。つまり、現在求められている情報検索システムは、厳密な検索が行えるだけでなく「簡単に使えるシステム」である。

5. 概念検索システムとは何か

検索式を使わないで簡単に情報検索の行える概念検索システムが実用になりつつある。これは、ユーザが技術的説明の文章を質問文として入力すると、システム側で質問文からキーワードと判断されるものを切り出し、それを使って蓄積された多数の文書进行评估し、関連度の高い順に文書を出力してくれるシステムである。したがって、検索専門家でない研究開発者や特許技術者などの一般ユーザでも簡単に利用することが可能になる。

図2は、概念検索システムの利用手順を、これまでの検索システムと比較して示したものである。概念検索システムでは、文章で質問し、

回答の特許リストはスコアの高い順に表示されるので、上位の十数件を見ることで、該当特許を見つけられることが多い。

なお、概念検索においてはコンピュータによるヒット件数とか回答件数は意味がない。つまり概念検索とは、データベースに存在する全てのデータ(特許)に最高の点数(関連度)から最低の点数まで割り振るためである。この結果、利用のたびに先頭(最高得点)から何件まで見るかを決める必要がある。

概念検索システムのキーポイントは、蓄積文書の自動的な関連度評価であるが、これはキーワードの扱いと、関連度の評価ロジックに分けられる。

(1) キーワードの扱い

日本語は英語のようにスペースで区切れていないので、キーワードの扱いが難しく、様々の工夫が行われる。蓄積文書(検索対象データ)

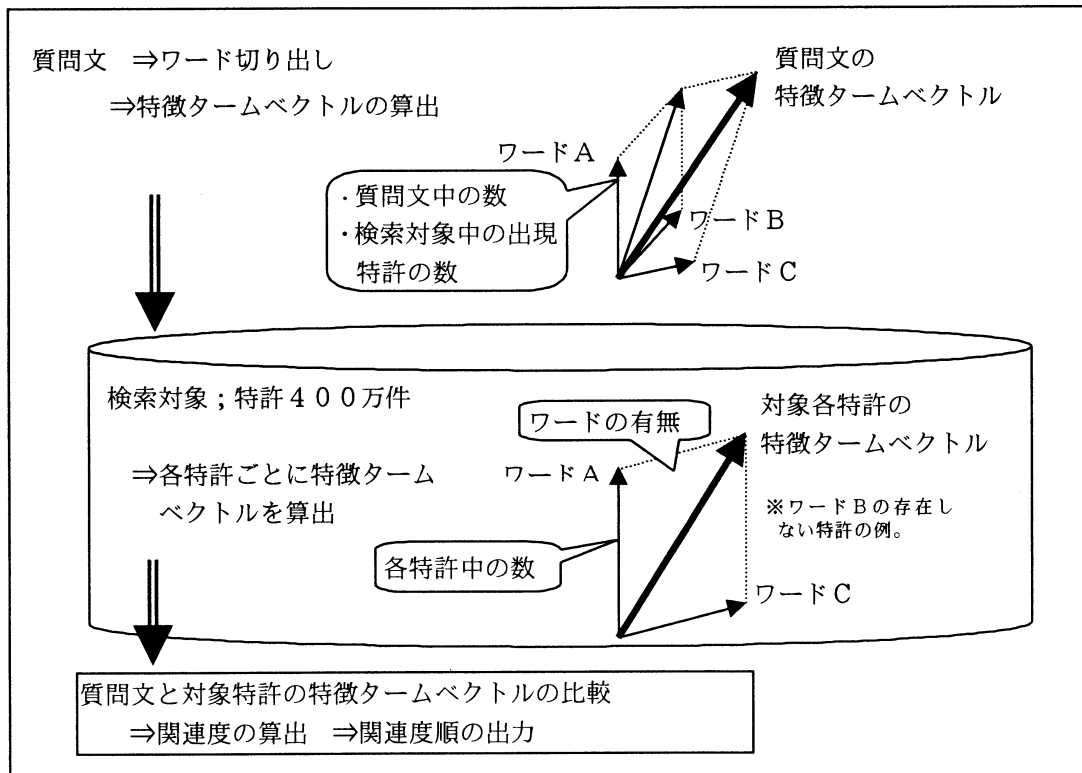


図3 概念検索システムの機能説明図

の処理において、辞書を併用して一定のアルゴリズムでキーワードを切り出す方式と、キーワードとして切り出さずに特殊なインデックスを利用して全文を検索する方式がある。一方、質問文の処理は、文章からキーワードを切り出すシステムが多い。

(2) 関連度の評価ロジック

キーワードの様々の観点での出現頻度などを使って関連度を評価する。この詳細はシステム開発の各社で工夫がなされている。次に具体例で説明する。

図3は、日立製作所で開発した概念検索システムの動きを説明するものである。日立では、平成5年以降の特許情報を対象に、このシステムの利用を開始している。このシステムはまず、質問文として入力された文章からキーワードを切り出し、データベース全体の状況を反映した重み付けを行い、それらの合計値（特徴ターム

ベクトルという）を算出する。次に、nグラムという特殊なインデックスを利用して蓄積されているデータベースを全文検索し、各特許における各ワードの出現頻度などにより質問文に対する関連度算出を行い、関連度の高いものから順に出力する。

この図の特許では、ワードBが存在しないが、ワードAとCの頻度が高いために、この特許の特徴タームベクトルは質問文のそれと似たものになっている事を示している。つまり、質問文中のすべてのワードが検索対象の特許に無くても、他のワードの点数が高いと関連度が高くなるということである。これが、概念検索として機能する理由のひとつである。

関連度の評価基準として主に以下の指標が使われている。

- ① 質問文中に何度も出てくるワードは点数が高い。

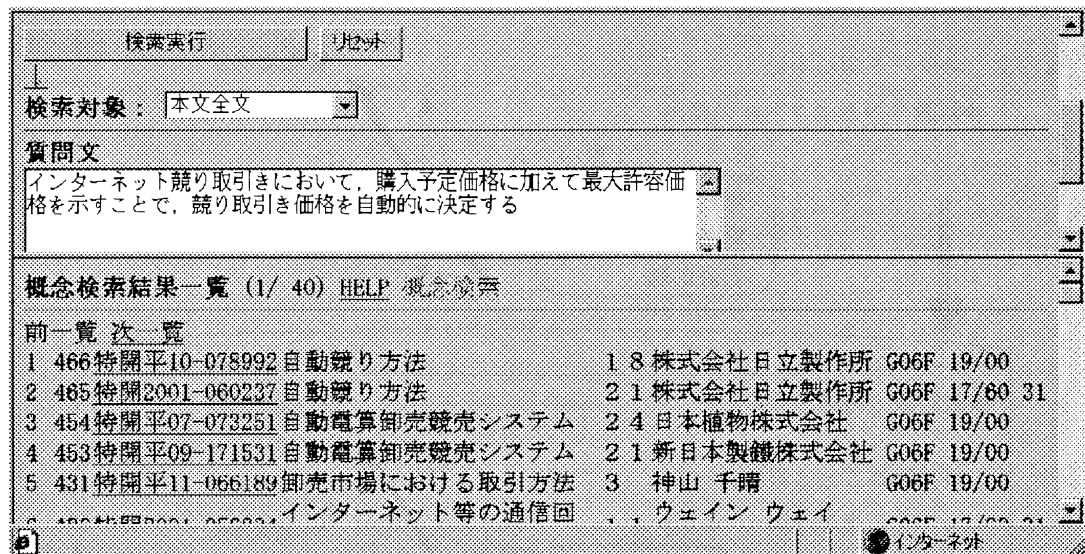


図4 入力された質問文と特許リストの回答画面

- ② データベース全体において、そのワードを含む蓄積文書数の少ないワードは点数が高い。
- ③ 質問文中のワードをたくさん持つ特許は点数が高い。
- ④ 各ワードの出現回数の多い特許は点数が高い。

実際には、各ファクターの重み付けの数値の選定とか、特許文書の部分による重み付けや、特許文書特有の処置なども行われる。これらを総合的に勘案して、特許文書の関連度として算出している。特許文書のように新しい技術や用語が次々と出てくる場合には、検索する時点のデータベースの状況により、ダイナミックに質問文の各ワードを評価する本システムが特に適していると考えられる。

なお、質問文と言っているが、現在の概念検索は文脈を理解する機能は持っておらず、キーワードに分解して処理している。したがって、質問としてキーワードをカンマやスペースで区切って入力した場合とほとんど変わらない結果になる。

図4は、質問文の実際の入力と特許リストが

スコア順に表示されている画面である。図5は、特許内容を表示している画面であり、要約などの部分（ユーザ設定可能）が表示される。「全文」をクリックすることで、図6のように明細書全文を表示させ、詳細を検討できる。実際には、見つけた特許中の新しい表現を追加し、さらに概念検索を行うことにより、芋づる式に多くの特許を見つけることができる。

6. 概念検索システムの上手な使い方

概念検索システムは、ユーザが欲しいと考える技術の説明文を質問文として入力するだけで検索してくれる。このとき、質問文のキーワードにより対象技術の概念を上手に特定することができれば、期待した検索結果を得ることができる。

〈例1〉図4にある「インターネット競り（せり）取引において、購入予定価格に加えて最大許容価格を示すことで、競り取り価格を自動的に決定する」では、欲しい技術の概念をうまく特定しており、期待した結果を得ることができた。さらにこの中で、「インターネット」や「競り」「最大許容価格」など重要と思われるキ

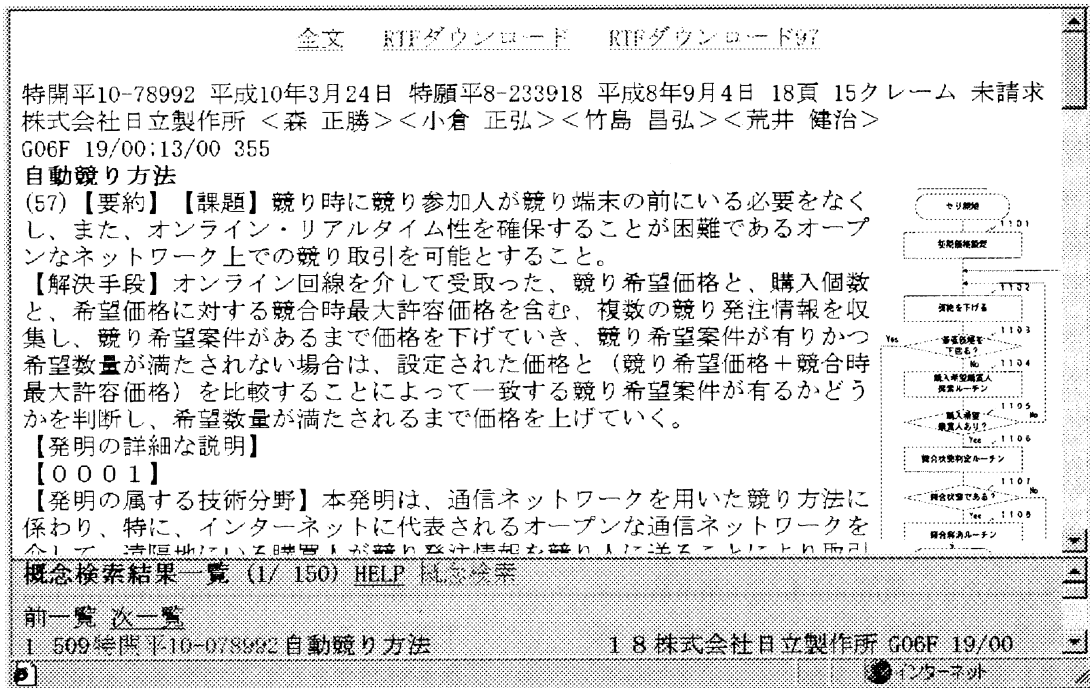


図5 要約などの表示画面

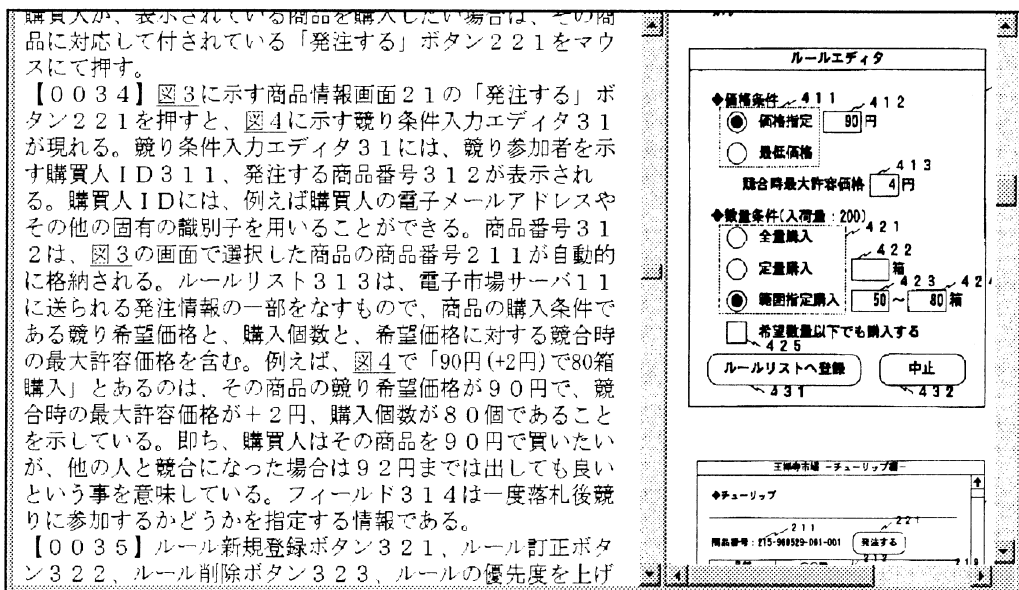


図6 特許全文の表示画面

ワードを1つ省いても、ほとんど同じ結果を得ることができる。これは、1つくらい省いても、現在は類似した特許が少ないのでほとんど同じ概念を指すためといえる。

ただし、2つ省くと、最も近い特許（特開平10-78992）が4、5番目に落ちる。さらに3つ省くと、最も近い特許は10番目に後退する。これは、「取引において、購入予定価格を示すこ

とで自動的に決定する」となることで、最初の文の概念から変っている（概念がボケている）ためといえる。このように、質問文で概念を絞り込む程度によって回答は違って来る。一方、「オークション取引において購入希望価格に加えて容認する価格を示すことで、価格を自動的に決定する」と、かなり違った表現にした場合20番目付近に出てくる。ちなみに目的とした特許には、オークション・容認などは存在しないが、取引・購入希望価格・自動的などにより、比較的類似の概念範囲になっているためと思われる。

〈例2〉似た例であるが「飛行機のエンジン制御」が欲しいとき、質問文として「乗り物のエンジン制御」では質問の概念がずれており、適切な結果を得ることはできない。

〈例3〉質問文を「安価な液晶表示装置」とすると、「安価に作る」と記載されている特許を見つけることはできる。しかし、安価に作るための膨大な数の特許が存在するのに対し、この質問文では欲しい技術の範囲（概念）を特定していないため、安価に作る特定の技術の特許を簡単に見つけることはできない。これに対し例えば、「ゲート絶縁層の接触部分が多結晶シリコンを主成分とする活性層を使う薄膜トランジスタを使った安価な液晶表示装置」などのように具体的な質問文にすると、回答もかなりの確かなものになる。

〈例4〉一方、特許の少ない技術分野では、このように詳細に限定しなくても概念を明確にすることは可能であり、ラフな質問文でも的確な回答が得られる。例えば「ナノサイズのコイル」という簡単な文章でも、期待に近い結果を得ることができる。

以下に、概念検索システムを実際に利用して考察した結果を列記する。

(1) 上手な質問文

- 概念を明確に特定できる短い文章が最良。

- 類似特許が少ない分野なら大まかな特定でよいが、多い分野ならば限定した表現が必要。

- 重要ワードは2,3度繰り返して書く。

(2) 不適切な質問文

キーワードに分解したとき、他の技術概念と区別のつき難い下記のようなものは不適切である。

- 漠然とした概念の指定（短すぎて概念の不明確な文章）

- 関連技術や前提とするワードを数多く、また何度も書いた文章・・・全てのワードを含む概念になるので不適切なことが多い。この理由で、クレームの文章そのままでは不適切なことが多い。また、「要約」の全体を質問文として入力すると不適切な結果を得ることが多いのもこの理由と考えられる。

- 否定表現は使用不可（現在のシステムは肯定否定は判断できない）

(3) 繰り返し検索

最初の検索で似た特許が分かったら、その特許に出てくる新しい表現を質問文として次の検索を行う。これにより、芋づる式に多くの特許を見つけることもできる。この場合も検索式を考える必要がないので、短時間に何度も繰り返して検索を行える。

(4) 検索対象

要約には出てこないが明細書全文には出てくる技術内容があり、このような場合には、検索対象が明細書全文まで行えると好都合である。

例えば「店舗にてICカードの暗号情報で電子財布を使う」という質問文では、要約中にそのような記述のある特許は、対象が要約でも全文でも上位に出てくる。ところが要約では「家庭用ゲーム機で、ICカードを情報センターでの勝敗の決定や参加料徴収に使う」ことを書いて

である特許で、実施例の一部に「店舗で電子財布的な利用にも応用可能」と書いてある特許は、要約だけの検索では順位が低くなる。

(5) 概念検索システムは漏れが多いか

検索専門家の行った手めぐり調査を一部含む検索式方式の結果と、概念検索による結果を比較した。その結果、両方で見つかる極めて近い特許も多いが、一方のみで見つかる特許がそれぞれにある。やや概念検索の方が漏れの多い傾向にあった。

これは、現在の概念検索システムが文脈やワードの意味を人間が理解するように判断していないという本質的な問題に加えて、利用者が質問文の作成法をまだ良く知らないことも原因のようである。より再現率の高い検索を行うには、AND, OR による検索式と併用するのが望ましいと考える。

しかし、これは調査の専門家が長時間をかけて行った調査と、数十分で行った概念検索を比較していることに留意する必要がある。つまり概念検索は、ほとんど準備なしに使えること、誰でも使えること、短時間で使えること、検索式方式と違った観点で検索できることなどの長所がある。したがって、強調されるべきことは、その長所を生かした新たな検索手法の開拓である。

図7は、以上の検討結果を要約して両方式の再現率を象徴的に示したものである。

(6) 一般ユーザでも使えるか

一般ユーザ（研究者、設計者、特許技術者）は検索の対象技術に精通しており、検索専門家は対象技術には詳しくないが検索式の作成ノウハウを熟知している。このため、対象技術を簡明な文章で書く必要のある概念検索システムは、一般ユーザに適していると考えられる。なお、社内で行った実験の結果では、日常的に検索を行っていない特許技術者も十分利用できるものであった。

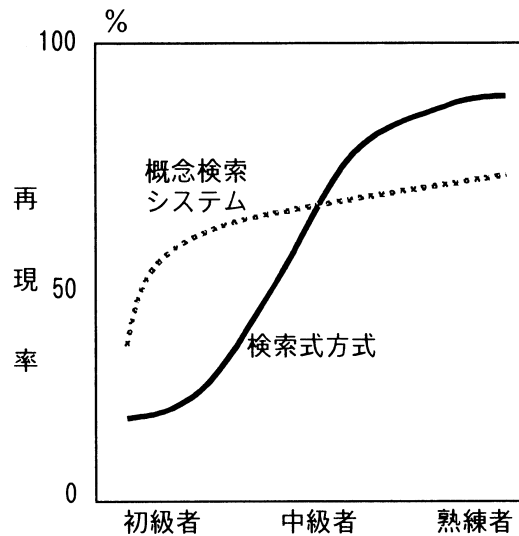


図7 再現率の比較

(7) 概念検索システムは概念を理解しているか

概念や意味を理解するとはどういうことであろうか。例えば、「半導体を冷却する」という場合、半導体に関する様々の技術知識や他の技術との関係、さらに製品・市場の情報など膨大な情報が連携して動くように蓄積され、また冷却についても同様であり、それらが「で」の文脈ではなく「を」の関係で動くように仕組まれている状態といえよう。そういう意味では、システムが技術概念を理解しているとはいえない。

しかし、「飛行機は、鳥のような羽根にしたことで空を飛べたが、翼を固定しない限り飛べなかった」という警句を思い出す必要がある。類似の原理を利用する場合でも、生物と機械では具体的なやり方は大きく異なる場合があり、全く同じやり方を狙っては、長年月を費やしても完成しないことがある。

簡単な文脈を取り扱うシステムは10年くらいで実用になるかもしれないが、意味のわかるシステムの実用時期は不明瞭といえよう。

(8) 料金体制

以上と少々趣の違う観点であるが、概念検索システムを一般ユーザがアイデア育成支援シス

テムとして使うには、固定料金で使えることが重要と考えられる。出力が1件で100円など、個々の利用において費用を気にしていたのでは、情報の内容に意識を集中しきれないように思われる。固定料金の商用オンラインサービスや社内データベースにするのが最良である。

(9) 利用の場

概念検索システムは、簡単に、短時間に、誰でも利用できるのもので、研究開発や特許関係業務において様々の役割を果たすと考えられる。

1) アイデア育成支援

技術開発会議にパソコンを持ち込んで、出されたアイデアをその場で概念検索システムに入力し、全世界の人の似た特許を見つけ、自分のアイデアに生かす使い方に十分利用可能と思われる。以下はその利用法の事例である。

- 自分のアイデアの不足を見つけ、補強する。
- 既に公知のアイデアを習得する。
- 新たな課題を見つけ、新しいテーマに発展。
- 新しい解決法、応用、材料、原理などの発見。
- 他人を納得させるうまい表現を見つける。

2) 特許調査

調査専門家には予備調査的に利用するとか、従来の調査後の再確認として利用できる。また、検索の専門家でない技術者が特許出願前などの調査に利用するのも上手な利用法であろう。

7. おわりに

概念検索システムは、簡単に、短時間に、誰でも、特許情報の検索を行えるので、特許情報を知識として活用する手段となり、また、特許調査にも貢献するものと考えられる。

なお、現在の概念検索システムはインテリジェントな情報検索システムとしては、入門段階のものであろう。技術的ワードはどのように難しいものでも問題はないが、文脈の理解という意味においては頼りにならない。しかし、特許情報を知識として活用したいニーズは、特許情報検索の大衆化を要求し、その結果、概念検索システムは今後さらに発展するものと考えられる。

やがて、コンピュータのインターフェース帽子をかぶるだけで、考えた事柄をデータベースから検索できる時代になるかもしれない。しかし、我々が生きているのは、コンピュータ技術がハードもソフトも数年で様変わりし続ける現代である。究極のシステムを求めすぎると、永遠に次世代のシステムを遠望するだけの、虹を追いかけるようなことになりかねない。完璧さを求めるよりも、すでに一定の完成度に達している概念検索システムを、実務に生かしていくべき段階になっていると考える。

(原稿受領日 2001年7月19日)