

どの程度の長さが質問文に最適か？

六車技術士事務所、六車正道

## 要約

概念検索においては、目標技術を狭く限定した使い方が一般的と思われる。そのためには文字数が 40～80 文字程度の質問文が多くの場合適している。これに対して、10～30 文字程度以下では目標を絞り込まない広い技術範囲の概念検索が行なえ、関連技術まで知りうるなどの特徴がある。質問文が 200 文字程度以上の長い質問文では特許の要約や明細書をそのまま質問文として使えると言うことでもあり、簡単に利用できる大きなメリットがある。

## 1. はじめに

検索式は指定したキーワードの有無でヒットかどうかを判断する単純なシステムである。それに対して概念検索は以下のような特徴があり、単なるキーワードの有無ではないインテリジェントなシステムといえる。

- ①質問文から複数のキーワードを切り出す
- ②それらのキーワードに重み付けする
- ③上記のことがデータベースの現状を反映して自動的に行なわれる
- ④それらの重みに基づいてデータベース中の全文献に点数を付ける
- ⑤点数の高い順に出力する

概念検索においては目視チェックする件数は利用者がそのつど決める必要があり、その件数によって再現率、適合率は変化する。そのため、本稿では上手な検索や漏れが多い・少ないという場合、特に断らない限り上位 50 位（または 100 位）程度までの一致とし、目視チェック件数の細かな違いの議論をするものではない。

要するに、利用者が特定の狭い技術範囲の特許を探している場合に、（それらを含んでいたとしても）広すぎる技術範囲の特許を出力した場合には上位 50 件程度に目標特許が集中しないので上手な検索とは言えない。逆に、利用者がある製品の全体に関する特許を探している場合に、そのごく一部の技術範囲の特許を出力しても上手な検索とはいえない。上手な検索のためには目標とする技術範囲と検索の技術範囲が同じ程度に絞り込まれていることが重要である。

筆者は 2001 年に概念検索の質問文は「概念を明確に特定できる短い文章が最良」<sup>1)</sup>であることが多いと指摘し、質問文を変えて数回繰り返して検索することの重要性を指摘した。さらに 2005 年には 40～80 文字程度の技術的な説明文で作成するとうまく検索できることが多い<sup>2)</sup>と述べた。2008 年には、より本質的にいえば質問文の長さが問題ではなく含まれるキーワードにより検索目標とした事項をいかに限定しうるかが問題であり、長い質問文でもその中のキーワード分布が検索目標と合っていれば妥当な検索を行なうことができる<sup>3)</sup>と指摘した。

一方、40 文字以下の短すぎる質問文でもある程度検索できる。利用者が細かく絞り込んだ技術を目標としている場合、例えば「インクジェットプリンタ」とか「インク目詰まり防止」程度のアバウトな質問文では類似特許が多すぎて満足な絞り込みができず役に立たない。ところが、類似特許が少ない技術分野では 10 文字以下の短すぎる質問文でも必要なものを見つけられることがある。また、類似特許が多くても利用目的を変えることで有益な場合もある。さらに、逆に 200 文字程度以上の長い質問文でも似たことがいえる。

本稿では、探す目標技術と概念検索の絞り込みの問題や質問文の長短の使い分けについて考察す

る。

## 2. 概念検索における絞り込み度

情報検索においては探そうとする目標の技術範囲が狭い場合と広い場合がある。「目標の絞り込み度」は利用者が最終目標としている技術範囲であり、利用者の思考の中に存在するが文章になっているとは限らない。例えば丁寧に書かれた特許調査報告書の「調査の技術的範囲」などに書き出されることは多い。

一方、検索条件の技術範囲も狭い場合と広い場合がある。「検索の絞り込み度」は概念検索においては質問文により概念が絞られている度合いであるが、概ね質問文の長さやキーワードの数、およびそれらのキーワードが核心を突いたものであるかどうか、により決まってくる。

図1は、目標と検索の絞り込みを軸として検索の特徴を表したものである。例えば、目標の絞り込みを狭い技術範囲に想定しておきながら、検索の絞り込みをアバウトにして広く行なったのでは上手な検索にはならない。概念検索では「思いついたキーワードだけで検索できると聞いてやってみたらうまく検索できない」とか「請求範囲で検索したがさっぱりダメだ」という場合はこの絞り込みの不一致が生じていることが多い。

なお、概念検索をアイデア発想支援的に使う場合には、利用の初期においては利用者自身が目標を明確に絞り込んでいないが、何回か繰り返すうちに徐々に絞り込まれていくような使い方もある。

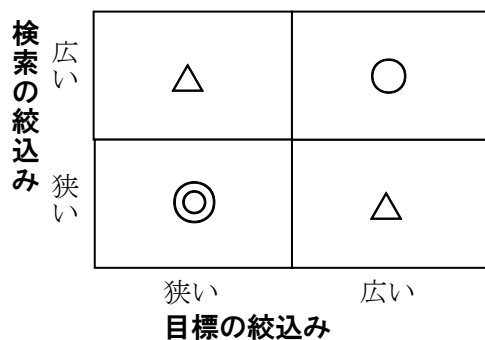


図1 情報検索における絞り込み

### 2. 1 目標の絞り込み度

多くの場合、目標技術が絞り込まれている場合には目標の特許件数は少ない。例えば、「プリンタの目詰まり防止」の技術範囲は「プリンタの目詰まり防止としてノズルを振動させる」よりも広く、前者では1,000件以上の特許があり、後者は10数件になる。

概念検索では、目標とする特許が1,000件以上もある場合と10数件の場合では利用法が異なるように思われる。

概念検索は質問文と類似している程度を計算し、点数の高い順から出力するものである。一方、多くの利用者は50件程度であれば意識を集中して抄録や明細書の内容を目視チェックすることができる。したがって目標特許が10数件の場合、ノイズと入り混じって順序が乱れていても上位50位や100位程度までに多くの特許が出てくることが期待できるので数回繰り返すことでかなり再現率の高い検索を行いうる。ところが目標特許が1,000件にもなると、ノイズも混じることで数千件を見ないと目標特許を拾い出せないことになる。これでは効率的な検索は期待できない。また多く

のシステムでは上位 500 件とか 1,000 件までしか参照できないようにしているの、その面からも全体の把握は困難になる。したがって、目標の絞り込み度が低い場合は利用に際して特別の工夫が必要になる。

利用者が目標の絞り込み度を判断する際には次のような方法が参考になる。

- ①準備した質問文に該当する特許が 100 件以上あると思える場合。
- ②概念検索の結果を上位 100 件程度まで見ても該当特許が頻繁に出てくる場合、結果的に目標の絞り込み度が低いといえる。

## 2. 2 検索の絞り込み度

検索の絞り込み度の高い場合は、技術的に狭く限定した検索を行なうことができる。概念検索では、キーワードがある程度核心を突いたものであった場合、質問文の適度な長さ（≒適度な数のキーワード）が絞り込み度の高い検索となる。概念検索は質問文中のキーワード分布と一致している特許が上位にランクされるものであるから、質問文の長短とは関係ないように思われるかもしれない。しかし、実際に検索してみると、質問文は短すぎても長すぎても上手に検索できないことが多い。

表 1 は、質問文が短文（54 文字）の場合最もうまく検索でき、短すぎる文や、要約や明細書のように長い文ではほとんど検索できない実例である。

**表 1 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。質問特許を除く。

### ①Bシステムによるヒット件数

質問 対象	短すぎる文		短文 54 字	請求 範囲 216 字	要約 311 字	明細書 10,800 字
	6 字	9 字				
明細書	2 件	4	6	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	短すぎる文		短文 54 字	請求 範囲 216 字	要約 311 字	明細書 10,800 字
	6 字	9 字				
明細書	0	2	7	0	0	1
請求範囲	—	—	3	1	0	0
要約	—	—	3	1	1	0

短すぎる文；ゴミと微生物（6 字）、および、ゴミと微生物で脱臭（9 字）

短文；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  
第 1 請求範囲；タンク状の脱臭機本体と、上記脱臭機本体の内部を上方の脱臭室と下方の要処理ガス導入室とに仕切る仕切材であ  
って、後記充填材の通過を許容せず、ガス及び水分の通過のみを許容する仕切材と、前記要処理ガス導入室に要処理ガスを導入  
する導入手段と、前記脱臭室からの処理済ガスを排気する排気手段と、前記要処理ガス…（以下、省略）

図 2 は質問文の長さ（またはキーワード数）と検索の絞り込み度のイメージを示したものである。つまり、質問文が 10 文字程度以下（キーワードが 4 個程度以下）ではポイントが絞りきれず、絞り込んだ検索は困難である。一方、特許の要約や明細書などのように長すぎると目標技術と関係の少ないワードが多くなり絞り込んだ検索が難しくなる。適度に絞り込まれた質問文は多くの場合、一定の範囲の文字数（またはキーワード数）であるといえる。

ところが、短すぎたり、または長い質問文でも類似特許の見つかる場合がある。次にそれらの利用法を検討する。

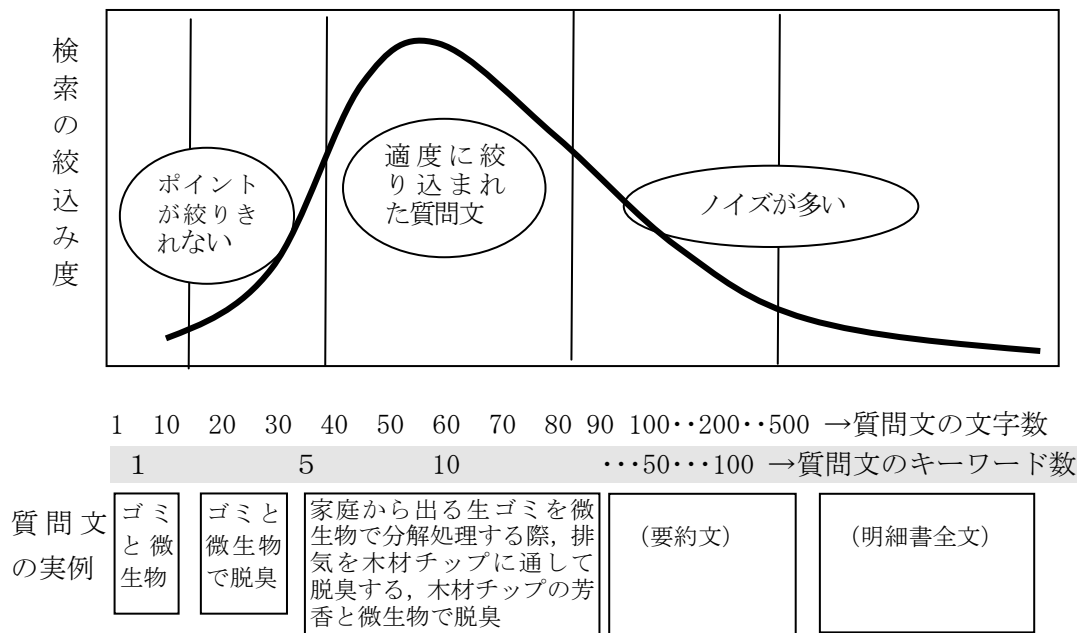


図2 質問文の長さ概念の絞り込み度

### 3. 短すぎる質問文

概念検索の質問文は、多くのシステムにおいて、数文字から数千文字以上の長いものまで可能である。ここでは、短すぎる質問文とは文字数では10文字程度以下、キーワードでは4個程度以下のものとする。

なお、本題とは関係が少ないが知っておきたいことがある。それは、キーワードとは人が見て理解できる技術用語であり、形態素解析を使うシステムでは概ねそのような文字列が切り出される。しかし、n-gram インデックスを利用するシステムでは人間が考えるキーワードとは無関係にn文字ごとに切り出される。さらに発展したインクリメンタル n-gram ではデータベースにおいてn文字以上の適切な文字列も自動的に切り出す機能がある。例えば、「インターネット競り取引」では、通常のキーワードとしてはインターネット、競り、取引などであるが、システムによっては「インター」と「ネット」を切り離すとか「ネット競」を切り出すこともある。概念検索はこのようにして切り出した文字列を特徴語ということが多々ある。

ある程度複雑な技術の場合、2、3個のキーワードでポイントが絞り込まれることは、多くの場合期待できない、つまり検索の絞り込み度は低い。多くの場合、他の技術と区別するために5個前後～10数個のキーワードが必要である。

ただし、新しい技術の出始めの段階で該当特許が少ない場合はそうとはいえない。例えば1970年代は「安価な液晶表示装置」などはそれだけで近い内容の特許を取り出しえたであろう。しかし、2010年の現在においては、目標とする技術をさらに詳しく記述しなければ欲しい技術内容を特定することはできない。その結果、多くの場合5個前後～10数個のキーワードが必要になる。

では、短すぎる質問文は全く使えないかと考えるとそうではない。表1にもあるようにそれなりの回答を得ることができる。コンピュータによる情報検索は指定した検索条件の限りにおいて絞り込まれた検索である。問題はその絞り込みが利用者が目標とした技術内容や絞り込みになっているかどうかである。このように考えれば、短すぎる質問文であっても役立つ利用法がある。

### 3. 1 数ワードの質問文

多くの場合、2～3ワードの短すぎる質問文による概念検索の回答は技術的に絞り込まれていない。例えば「ゴミを分解して脱臭」という質問文で概念検索した結果は、「ゴミ分解装置で、悪臭を木材チップの芳香と微生物で分解し脱臭する」というところまでは限定できない。

ところが、「発電所の燃焼排ガスから炭酸ガスを分離して回収する」という調査の場合には、類似特許は多くないので、「炭酸ガスの回収」のように一般的には短すぎる質問文でもかなり絞り込んだ概念検索を行なえる。

図3は、概念検索の絞り込み度を見たものである。(一定以上の品質の質問文による概念検索では) 出力順位が1位近辺では単位件数あたりの一致(例えば回答10件中の一致件数)が多いが、順位が下がるにしたがって徐々に少なくなる。その際、絞り込まれた検索であれば一致は数10番目以前に集中しており、一方、絞り込まれていない検索であれば1,000番目であってもかなりの割合で一致する。

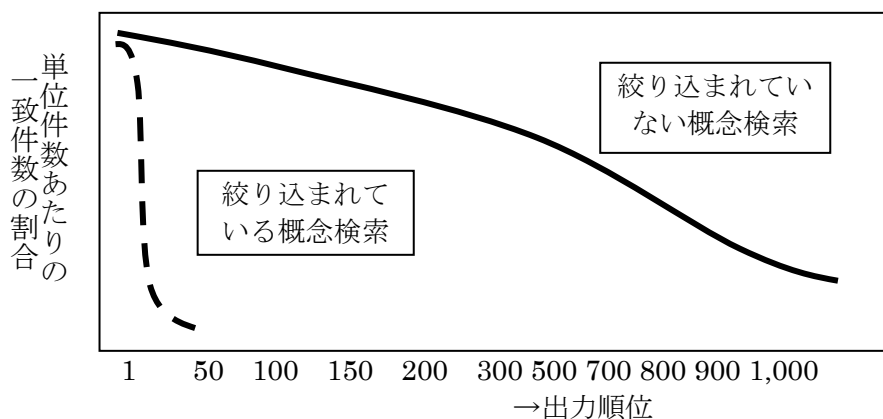


図3 検索の絞り込み度

### 3. 2 短すぎる質問文の利用法

先に述べたように、多くの調査においては、短すぎる質問文では技術的に絞り込まれていない広い検索になるのが一般的である。このような技術的に絞り込まれていない概念検索はどう使えば役にたつであろうか。その方法を以下に事例的に紹介する。

#### (1) うまくいく場合もある

データベース全体の中に類似特許が少ない技術の場合には、数ワードでも絞り込んだ検索が行なえる。

#### (2) 技術範囲の広い技術の検索

技術範囲を絞り込まないで広い範囲の全体を検索したい場合は、短い質問文にせざるを得ない場合がある。ただしこの場合、目標特許が数100件以上ある場合には全体をピックアップできるとは限らない。

#### (3) 概略を知れば良い場合

全件を見る必要はなく、類似特許の一部を見て、技術や出願人の概略の動向を知ればよい、というような場合には利用できる。

#### (4) 関連技術を参照できる

詳細な事項を限定していないので、関連特許が広く検索されることになる。そのため、目標技術

以外の技術も広く知ることが期待できる。

#### (5) 予備検索として使う

少々漏れがあったとしても予備検索と位置づければ下記のような知見を得る機会になり貴重である。

- ・類似特許が多い場合、目標や検索条件を絞る必要があるとわかる。
- ・類似特許が少ない場合、条件を緩くするか別の観点の検索が必要とわかる。
- ・類似特許のいくつかを知ることができる。

#### (6) 別の検索と組み合わせる

目視チェックできる最大の件数を 500 件とか 1,000 件に制限しているシステムではそれ以上を見ることができない。その対策として、検索式と併用する方法がある。つまり、概念検索で知ったことを利用して検索式で検索し、その回答から概念検索で目視チェックした特許を除いて目視チェックする。ただし、この組み合わせ利用のできないシステムもある。

## 4. 長い質問文

質問文が長くなればその文中のキーワードが増えるが、目標技術に対して不要なワードも増える。したがって、一定以上長い質問文では概念検索を上手に行なえないことが多い。

ここでは、長い質問文とは文字数が 200 文字程度以上、キーワードでは 30 個程度以上のものとする。なお、上手な概念検索の行なえる文字数は 40~80 文字程度としておきながら、長すぎる文字数は 200 文字程度以上と言っている。つまり、80 から 200 文字程度まではグレーゾーンであり、長くなるにしたがって徐々に上手な検索のできないことが多くなるということである。公開特許の「要約」は 300 文字程度のものが多く概念検索の質問文としては多くの場合長すぎるといえる。

### 4. 1 長い質問文

要約や明細書などを使った長い質問文は、検索の目標技術と質問文のワード分布の不一致が大きくなるため上手に検索できないことが多い。

表 2 短文と要約、明細書の主なワードの出現回数 ※対象；特開 2000-42356

(1) 作成した短文	(2) 要約	(3) 明細書
(2 回)脱臭、微生物、木材、チップ (1 回)家庭、生、ゴミ、分解、排気、芳香  ※作成した短文；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	(7 回)処理、脱臭、 (6 回)導入、(5 回)ガス (3 回)排気、水抜 (2 回)充填、仕切、運転 (1 回)生、ゴミ、木片、活性、汚泥、バルブ、タンク・・・ (0)悪臭、微生物、家庭、芳香、木材、チップ、分解、通過、許容	(162 回)処理、(135 回)脱臭、 (120 回)生、(107 回)ガス、 (55 回)ゴミ、(47 回)導入、 (39 回)排気、活性、(37 回)汚泥、 悪臭、(35 回)木片、(30 回)微生物、 (29 回)充填、(24 回)仕切、(23 回) 水抜、(12 回)家庭、通過、(11 回) 運転、(10 回)分解、(6 回)タンク、 (5 回)バルブ、(5 回)芳香、(4 回) 許容・・・ (0 回)木材、チップ

表 2 は、表 1 で紹介した短文と要約、明細書の中の主なキーワードとその出現回数である。要約や明細書では検索目標と異なるキーワード分布であり、それらに偏った結果になることが予想される。さらに言えば、要約や明細書では、家庭、分解、芳香、木材、チップという表現は非常に少な

いか全く使われておらず、これらのウェイトの高い特許は上位に位置しないことが予想される。つまり、この例の場合は、作成した短文で概ね表現されている目標技術と、要約や明細書のキーワード分布が離れているといえる。目標技術のワード分布と一致しない質問文は検索に不相当であり、長文の場合はそうなる可能性が高いといえる。

特許明細書などが目標技術のワード分布と一致しないことが多いのは、詳細説明の際に、その技術を構成する部分構成要素についてまで仔細に説明するためと思われる。例えば、パソコンのマウスにカメラを付けた発明の場合、特許明細書ではパソコン側の信号処理やカメラのピント合わせの詳細まで説明するようなことが多く、場合によってはマウスの材質についても説明するようなこともある。そのような特許明細書を質問文にすると目標技術以外のキーワードの頻度が高くなり、焦点のぼけた質問文になってしまうということである。

## 4. 2 長い質問文の利用法

多くの場合、長い質問文では目標とする技術範囲に絞り込まれないために、漏れの多い検索になる。しかしながら、長い文章は設計や製造の説明資料、特許の出願提案書、特許の要約や明細書など様々な場面で入手できる。そのため、これらを利用した概念検索もできる限り上手に使いこなす工夫をすべきであろう。その方法を以下に事例的に紹介する。

### (1) うまくいく場合もある

質問文が長文でもキーワード分布が質問内容と似ている場合はうまく検索できる。例えば、他の技術に言及が少ない場合や、細部の説明を過度に行なわない場合である。

### (2) 利用が簡単

ある特許番号の要約や明細書を指定して概念検索できるということは何と言っても簡単である。図4は特開 2000-42356 の要約を種文書として概念検索をやろうとしている画面の実例である。したがって、予備検索としてやってみることは無益ではない。ただし、類似特許が見つからない場合に50件（または100件）以上を時間かけて目視チェックしたり、上位に類似特許が見つからない場合に概念検索全体を無益と考えるのは早計である。

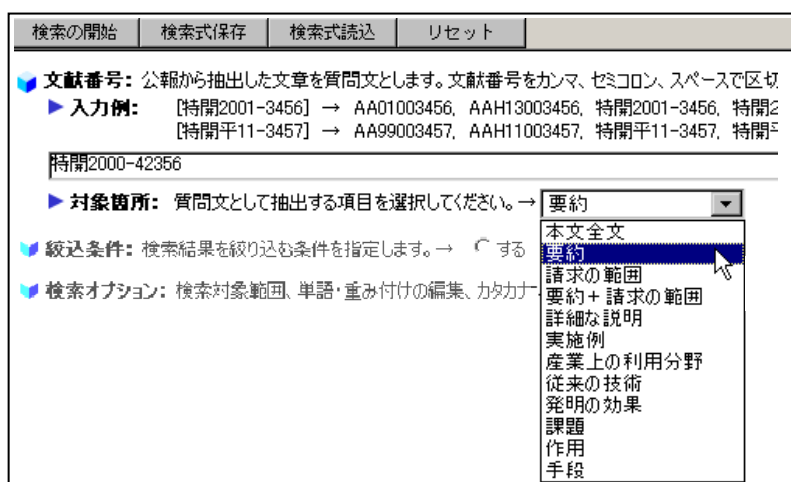


図4 ある特許番号の要約を質問文に指定した概念検索

出展；日立/Shareresearch（シェアリサーチ）、画面を一部加工

### (3) 問題点に分かる

類似特許が見つからないとか異なるものが出てくるとき、なぜそうなるのか考える手がかりを得ることが期待できる。

#### (4) 絞り込み検索を併用する

必須の IPC やキーワードを設定して絞り込んだ結果を対象に特許番号指定の概念検索を行うことは有益である。ただし、短文の質問文でもそうであるが、絞り込み検索では絞り過ぎにならないような配慮が必要である。

#### (5) 検索したあとで質問文を作り直す

長文での概念検索の結果を見た後で、知りえた知識を利用して質問文を短文に作り変えて再度やり直すならば大変有益といえる。

## 5. 文字数が合えば良いのか

### 5.1 妥当な文字数の質問文

質問文の妥当な長さは経験的に 40～80 文字程度であり、本質的にはキーワードの数が 5 個程度から 10 数個程度といえる。しかし目標特許が少ない技術分野の場合は、さらに少ない文字数で可能である。

当然のことであるが、質問文の文字数やキーワード数がこの範囲内でもマト外れの語句の文章では上手に検索できるわけではない。検索したい技術を簡単明瞭に表現するキーワードを含んだ文章が必要である。質問文としての短文の作成法はすでにいくつも<sup>2),3)</sup>書かれている。

### 5.2 探し難いケース

明細書のごく一部分で簡単に言及しただけの特許を、その記述を指定して概念検索で上位に取り出すことは難しい。それは、その特許が類似特許とは言い難いためといえることができる。

例えば、「人間が散歩する際に同行するペット型ロボット」に関する特許の公知例を探す場合、明細書本文に「高齢者の散歩相手としてのヒューマンフレンドリーロボット」の記述がただ一回のみ書かれている特開 2000-9426 を概念検索で探すことは非常に困難である。その理由は、同特許が「ロボットや自律走行車等の視覚装置で物体の三次元位置、物体面の向き、傾き、法線方向を計測するのに好適な面補間距離取得装置」に関する事項が繰り返し書かれているのに対して、目標技術に必要な「散歩」は 1 回のみ、「人間」や「ペット」は 0 回であり、技術内容が近い特許ではないと見ることができる。

しかし、このような場合も、「散歩」と「ロボット」を 30 文字以内という絞り込み条件を付けて主題の短文で概念検索を行なうと、20 番目付近に取り出すことができる。

## 6. 終わりに

概念検索は検索目標と質問文の技術範囲が一致した場合に上手に検索できるといえる。その一致はキーワードの内容と数の問題があり、本稿は主に数の面から考察したものである。

ところで、いかなる点でも改善の必要のないシステムはほとんど存在しない。ある程度の完成度であれば利用者が工夫して使いこなしているシステムは多い。概念検索も一定の完成をしているものであり、「使えるか使えないか？」ではなく「どうすれば上手に使えるか？」について検討されるべきであろう。

筆者はこれまで概念検索の上手な利用に関して質問文のあり方が重要であることを繰り返し書いてきた。ところが、他のいくつかの論文では質問文の特徴に無頓着なものとか、特許請求の範囲



で検索するとうまくいくというものが見られる。2009年になって概念検索の開発者が、今後の課題として「本願全体を入力文章とするのではなく、本願の中でどこが重要であるか・・・をユーザが指定し、・・・段階的に検索を繰り返す」ことを書いた論文<sup>4)</sup>が出ている。

質問文とは概念検索を行う場合の技術内容の定義書であり、ワードの頻度しか理解しない検索システムへの唯一の意思伝達手段である。したがって、システムに適した形で伝える必要があり高度な文法を理解する人間に適している特許請求の範囲の文章が適しているとは限らないはずである。そのようなことで引用文献の多くが筆者のものに偏っていることをご了承いただきたい。

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 六車正道、知財管理、Vol. 51, No. 12、2001. 12月
- 2) 概念検索の現状 ---概念検索システムの比較と使いこなし--- 六車正道、JAPIO／創立20周年記念誌、2005. 10/12
- 3) 概念検索はなぜ上手に検索できるのか？ 六車正道、JAPIO 2008YearBook、2008. 11月
- 4) 概念検索技術および特許検索への適用可能性について、八木敬宏、間瀬久雄、岩山真、特技懇、2009. 1. 3、No. 252