

1 概念検索とは何か？

1975年(昭和50年)頃から実用化されたコンピュータによる特許情報の検索の発展は紙による調査をほぼ過去のものにするに至った。そして今、検索式を必要としない概念検索が実用の時期を迎えつつある。概念検索とは、技術的説明文を質問として与え、文中のワードとその出現回数によりデータベース中の各特許と類似度を算出し、検索回答として類似度の高い順に特許をリストアップするものという。

1. 1 概念検索の原理

現在の概念検索はワードの出現頻度を基礎にしているが、ワードの切り出し方とか出現頻度の取扱い方の細部に関しては、システムによって少しずつ異なる。図1-1は、日立製作所の社内で広く使っているシステムの原理である。質問文が与えられると、システムがデータベース全体の蓄積状況を参照しながら検索に最適のワード(特徴タームともいう)に質問文を切り離す。次に各ワードを重み付けする。そのやり方は、①質問文中のそのワードの出現頻度、②データベース全体でそのワードを含む特許件数の割合の逆数が使われる。次にこれらの数値をベクトル合成し、質問文を表すただ一つの特徴タームベクトルを算出する。ベクトル合成とは、例えばA、B2つのワードの場合、各ワードの数値を二乗し、そ

れらを合計し、平方根を出す。これを質問文の有効なワード全体に繰り返す。ただし、コンピュータの実際の処理は、これと同等であるが、別の効率的な手段で行なわれる。こうして、質問文中の全ワードで成り立つ座標(空間)において1つの特徴タームベクトルが決まる。

次に、蓄積されたデータベース中の各特許において、質問文の各ワードの出現回数を算出し、各特許の特徴タームベクトルを算出する。最後に、質問文と各特許の特徴タームベクトルを比較し、近似度を算出し、その高い方、つまり質問文に似ている順にリストアップする。

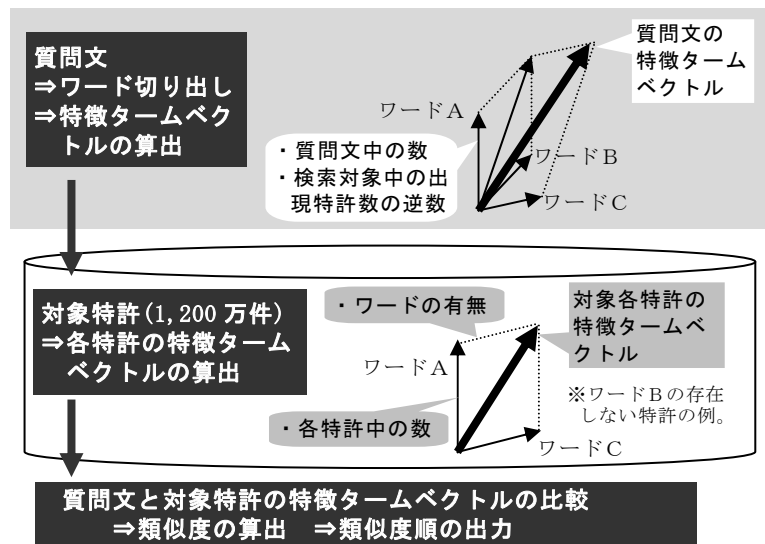


図1-1 概念検索の原理

なお、上記以外にも、データベースを蓄積するときに各特許の特徴的なワードを切り出して質問文とは関係なくその特許の特徴を決めておく方式とか、質問文からのワード切り出しを文法的な特長により行なう方式など、様々のものがある。

2. 概念検索システムの機能比較

現在、特許情報データベースで概念検索が利用できるのは、NRIサイバーパテントデスク、日立総合特許情報検索システムSharereseach, PATOLIS, ATMS/G-Search, RIPPWAY/リコー, HYPAT/発明通信社などである。表2-1は、現

存するシステムの機能を比較したものである。再現率に大きく影響すると思われるのは、後で説明するように検索対象フィールドや絞込み検索などである。なお、各システムの処理方式も重要と思われるが、公表されていないシステムがあるとか、分かり難い表現になっているなどの状況であり、収録しなかった。

表2-1 概念検索の機能比較

システム		A	B	C	D	E
象 収 録 対	公開公報	請求範囲 S61~ 要約 H5~	S58~	S58~	H5~	S58~
	公告公報, 登録公報	S61~	S58~	S58~	H6~	S58~
イ ー ル ド 対 象 フ	明細書	×	○	○	×	○
	全文フィールドの任意指定	×	○	×	×	×
	要約	○	○ }まとめるこ とも可能	○独自抄録	○(まとめて 指定する)	○(まとめて 指定する)
	請求範囲	○	○	○		
特許番号入力 of 概念検索		×	○	○	×	○
検索結果の番号指定 of 概念検索		○	○	○	×	○
ワ 質 問 文 の	ワードの数	不明	10 個(30 ま で追加可能)	20 個(40 ま で追加可能)	10 個(数十ま で追加可能)	15 個(19 ま で追加可能)
	重み付け変更	×	○	○	○	○
	追加・変更・削除	×	○	○	○	○
設 絞 込 み 検 索 の	IPC	△3 桁	○	×	○	○
	キーワード	×	○	×	△検索結果 で指定可能	○
	キーワード; フィールド指定	×	○	×	×	×
	出願人	×	○	×	○	○
	発行年月	×	○	×	○	○

3. 概念検索の性能比較

概念検索の結果はシステムが似ていると判断した順にリストアップされるので、利用者はその上位 50 件程度を見て、内容的に近い数件の特許をピックアップする。1 位に最も近いものがくるとは限らない。また、質問文を変更して数回、検索することが多い。

表3-1は技術を指定した概念検索であり、表3-2は特許番号を指定してその公知例が探せるか検討したものである。この中で、システムB, C, Eは検索対象が明細書であり、

概ね再現率は高く、6~8割の再現率を得ている。また、表3-2では特許庁での審査引例も概ね検索できている。これらの結果から、次のようなことが言える。

- ①再現率は1割以下のものから8割を越すものまで、様々のものがある。
- ②概念検索で、特許審査の引例のような権利的な調査もカバーしうる。
- ③概念検索は実用になるものがすでに存在するとみてよい。

表 3-1 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。

システム 検索対象	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

表 3-2 特開 2002-265223 関連特許の短文による概念検索

※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1~50位のもの。-は未収録範囲。

※引例 1~3 は特許庁の審査引例。ただし引例 3 は、実際は公告公報。

システム 検索対象	A			B			C			D			E		
	要約			明細書			明細書			要約+ 請求範囲			明細書		
	イ	ロ	ハ	イ	ロ	ハ	イ	ロ	ハ	イ	ロ	ハ	イ	ロ	ハ
1 特開昭 63-171671	-	-	-	n	n	n	n	n	n	-	-	-	n	n	n
2 特開平 1-41121	-	-	-	n	n	n	n	n	n	-	-	-	n	n	n
3 特開平 4-305018 引例 3	-	-	-	23	n	n	n	n	n	n	n	n	99	n	n
4 特開平 8-245815	n	n	n	n	84	n	n	21	n	n	31	n	n	11	n
5 特開平 9-21100	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
6 特開平 9-25123	n	n	n	3	n	n	5	n	n	95	n	n	6	n	n
7 特開平 9-67124	n	n	n	40	n	n	n	n	n	n	n	n	n	n	n
8 特開平 9-227122	n	n	n	32	n	n	48	n	n	n	n	n	n	n	n
9 特開平 9-227123	n	n	n	31	n	n	42	n	n	n	n	n	n	n	n
10 特開平 10-100306	n	1	n	n	45	n	n	1	n	n	50	n	n	8	n
11 特開平 10-167707 引例 2	n	18	n	n	1	84	n	2	n	n	n	n	n	1	n
12 特開平 10-259023	n	n	n	35	n	n	n	n	n	n	n	n	n	n	n
13 特開平 11-90213	n	n	n	n	n	n	n	n	n	n	n	n	n	13	n
14 特開平 11-140656	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
15 特開平 11-319542	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
16 特開平 11-329001	n	n	n	n	n	n	n	n	n	n	n	n	27	n	n
17 特開 2000-86241	33位	n	n	13	8	n	13	68	n	n	n	n	n	n	n
18 特開 2001-270022 引例 1	n	n	n	4	27	2	1	33	2	19	n	n	4	30	n
19 特開 2002-104825	n	n	n	n	n	n	n	n	n	n	n	n	22	n	n
20 特開 2002-225172	n	n	n	30	n	n	n	n	n	n	n	n	13	n	n
21 特開 2002-241130	n	n	n	73	n	14	n	n	n	n	n	n	39	n	n
50位までの件数	1件	2	0	9	4	2	5	4	1	1	2	0	6	5	0
〃 合計	3件			12件			8件			3件			10件		

※関連特許 21 件は情報科学技術協会・オンラインユーザグループ (OUG)・特許分科会主査(当時)の鈴木弁理士による調査結果
 ※質問文イ；層状のチタン酸化物を溶液中で振動攪拌することで、層状から単層状に分離させ、チタニアナノシートが分散した状態の溶液を作る。ロ；チタン酸化物のコロイド溶液とカチオン性高分子イオン溶液とに基板を交互に浸漬して、静電交互作用に基づく「溶液からの交互積層法」によってチタン酸化物層とポリマー層が交互に積層した超薄膜を製造する。ハ；チタン酸化物とポリマーが交互に積層した超薄膜について、紫外線照射または加熱処理によってポリマーを分解・除去してポリマー介在層が存在しない超薄膜を得ること。

4. 概念検索の使いこなし

概念検索は 6~8 割の再現率の検索も可能である。ところが、概念検索は「漏れを気にしない参考程度の検索に使うもの」とか、逆に「思いついたワードを 2, 3 個入力するだけで必要な特許を探せる」という言い方をされることがある。このような間違っただけの理解は、概念検索システムが文章を入力して [実行] ボタンをクリックするだけという、操作が簡単であるが故に、数回使っただけで概念検索

を理解したと誤解する状況になっているものと推察される。以下、上手な利用法について考察する。

4. 1 質問文として請求範囲は概ね不适当

表 4-1 は、質問文として短文を使った場合と、ある特許の請求範囲や要約、明細書としたときのヒットした件数である。質問文が短文の場合に比べて、請求範囲、要約、明細書を質問文にした場合のヒット件数は非常に少ない。

表 4-1 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。元の特許を除く。

① B 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

② C 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

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)

表 4-2 ワードとその重み付けの質問文による相違

<p>質問文中の出現回数</p> <p>(2 回) 脱臭, 微生物, 木材, チップ</p> <p>(1 回) 家庭, 生, ゴミ, 分解, 排気, 芳香</p> <p>質問文：短文</p>	<p>(6 回) ガス, 脱臭</p> <p>(5 回) 導入</p> <p>(3 回) 仕切, 充填</p> <p>(2 回) 通過, 排気, 許容</p> <p>(1 回) タンク, 木片, 生, ゴミ, 活性, 汚泥</p> <p>質問文；第 1 請求範囲</p>
---	---

表 4-2 は、質問文が短文と請求範囲の場合を比較検討したものである。探したい技術は表 4-1 下の短文に書いてあるとおりである。この文をワードに分解してシステムが理解する概念は表 4-2 左のようなものである。これに対し、第 1 請求範囲を分解して得られるワードは、同表・右である。請求範囲を質問文にした場合は、ガス、脱臭、導入、仕切、

充填などが重要ワードと判断されて、それらに偏った結果が上位にリストアップされる。探したい技術とだいぶ異なることが理解されよう。なお、この請求範囲の書き方は多くの特許で見られる一般的な書き方である。このように、請求範囲では探したい技術に関するワードが多く使われるとは限らず、概念検索の質問文としては不適當なことが多い。

ただし、請求範囲の文章を概念検索に利用することも出来る。1つは、請求範囲からキーワードを抜き出して、概念検索の質問文に利用する場合である。もう1つは、請求範囲の文章が不要ワードの繰返しが少なく、概念検索に適したものになっているケースである。化学材料分野では、概念検索に適した請求範囲になっているものがときどき見られる。

4. 2 質問文は 40～80 文字の短文が最良

質問文が短文ならヒットが多いが、要約や明細書では少ない。例えば、「ハイブリッドカーにおいて、地図情報やナビゲーションシステムの情報により、バッテリーの充放電を最適化して長寿命化を図る」(57文字)程度の短文の場合がうまくいくようである。しかし、例えば「安価なハイブリッドカー」という短文は、短すぎて技術的課題が分かり難く不適當である。

数百字以上の文章になると、利用者の探したい技術とシステムが重要と判断したワードに食い違いが起きるものと推察される。

4. 3 質問文のその他の工夫

(1)概念検索の質問文は、具体的な表現が適している。特許データベースの利用経験のある人は「特許の検索だから上位概念のワードが良いのでは？」と考える向きも多いと思われるが、そうではない。

(2)質問として、いくつかのキーワードを入力することもシステム的には同じだが、実務的には、文法的には少々不正確でいいので40～80文字くらいの文章を作り、これに、重要なワードを数個追加するのが上手な質問である。

(3)概念検索では、実務上は、同義語について

はそれほど気を使う必要はない。これは、概念検索においては、他のワードが同義語と同じ役割をしてくれることが期待できるためである。つまり概念検索では複数のワードが互いに意味を補いあっており、いくつかのワードがなくても似た概念になるためと推察される。しかし、主要なワードの多くを除くと、違う概念の検索になってしまう。

(4)質問文に、技術的説明文に加えて目的や効果などを入れることも上手な質問文である。

(5)質問において重要なワードは2,3度繰返して書くことで重み付けを大きくできる。

(6)キーワード,IPC,出願人,または発行年などで絞り込んだものを対象に概念検索を行なうことで、非常に的の絞れた結果を得ることができる。

(7)質問文は、新しく知ったワードを追加するなどして数回作り直すことが重要である。

4. 4 検索対象は長い方が良い

表4-1から分かるように、検索の対象は要約とか請求範囲などよりも、明細書全体の方が再現率が高い。この理由は、要約や請求範囲ではその特許の中心課題が簡単に書かれているのに対し、明細書では重要事項は繰返し書かれているためと思われる。さらに、中心課題でなかった場合でも、長い明細書のどこかに質問文のワードが一定の頻度以上存在すればヒットになることも理由の一つと思われる。

4. 5 検索対象を熟知した人は簡単に使える

質問文作成のノウハウは以上のようにいくつかあるが、検索式作成ほど難しいものではない。一方、検索の対象技術を知らない人は、

それを理解し最適の質問の作成に時間がかかる。これに対し、検索対象を熟知した研究者や特許エンジニアは、すでに質問事項が頭に入っており、質問文作成時間はほぼゼロである。したがって、概念検索はエンドユーザに適している。

5. 概念検索の利用

図5-1は十数時間にわたる手めくりや検索式による調査と1~2時間程度で行なった概念検索の再現率を比較したものである。比較的短時間で、実用になる程度の再現率を得ている。

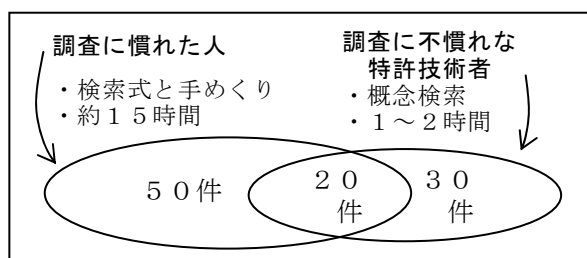


図5-1 検索式方式と概念検索の比較

- ・ 対象48テーマ件数は一定評価以上の合計。
- ・ 明らかに概念検索に不向きな質問文を使い、ヒット0件だった6テーマを含む。
- ・ 出典；発明協会発行「発明」2003年4月号

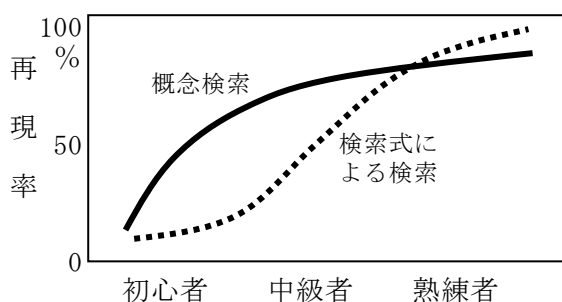


図5-2 習熟度別の再現率のイメージ

図5-2は、検索式による検索と概念検索の再現率のイメージ図である。概念検索は情報検索に不慣れた人もわずかの事項に気がつけば再現率を上げられることを示している。

研究者などのエンドユーザの利用では、簡単に短時間で行なえることから、アイデア発想支援ツールとしての利用とか、特許出願前の簡単な検索などに役立つと考えられる。

概念検索は関連特許を広く集めるよりも、ピンポイント的にマトを絞った検索に適している。また、かなりの再現率が期待できることや審査引例の特許を検索できることから考えると、審査請求前の調査や他者特許の公知例を探すような検索においても役立つことが考えられる。

なお、概念検索は簡単に実用になるレベルになっているが、これだけで全ての調査を行うことは出来ないと考える。概念検索がうまくいかない場合や究極の再現率を求めるような調査においては、検索式の利用や外国特許の検索など様々な手法を組み合わせるべきである。

6. 終わりに

概念検索は既に実用になる時期に来ていると思われ、利用目的や使い方によっては既存の検索法に勝る成果を上げる可能性がある。

一方、文章の品詞や係り受けを利用した概念検索の研究も進められており、今後、長年月をかけて発展するものと思われる。しかしその間、完璧なシステムの出現を待つだけでなく、使い方を工夫して上手に使っていく姿勢も大切と考える。