

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・・・

技術士(情報工学)

六車正道

日立公共システムエンジニアリング株式会社

シニア・コンサルタント

- 1. はじめに
- 2. 概念検索の特徴
- 3. 概念検索システムの機能比較
- 4. 概念検索システムの性能比較
- (以上は4月号, 以下は5月号)
- 5. 上手な概念検索のためのその他の考察
- 6. 概念検索の適用の拡大
- 7. 将来の概念検索

1. はじめに

多くの方が、インターネットを利用するとき意識しないで「情報検索」をする時代になっている。その検索エンジンの多くは一種の概念検索であり、利用者は可能性の高い方から表示されたリストの数件を見て自分のほしい情報を入手している。

特許情報データベースも増えているが、簡単・便利に利用でき多くの人に利用されることで、ますます有益なものになる。その方法の一つが概念検索であり、この1, 2年でいくつものシステムが利用可能になってきている。

本稿では、特許情報を対象とする概念検索システムの機能・性能比較を行ない、様々のものがあることを明らかにする。また、上手な利用法を探すため、実例データを明らかにしてやや詳細に検討する。さらに、知財関連業務への概念検索の適用を広げるためのいくつかの手法を検討した。特に、概念検索をアイデア発想支援に利用すると大きな効果が期待できるが、その

際、個々の利用では料金請求の無い方式が望まれる。その解決法としてセンタ・サーバとコーポレート・サーバを組み合わせたASPサービスについても検討する。

2. 概念検索の特徴

2.1 概念検索とは何か?

概念検索システムとは、データベースに質問として文章や複数のワードを与え、それらとデータベース中の対象文献において、ワードとその出現頻度で構成される概念の類似度を算出し、回答は類似度の高い順に出力されるものを行っている。なお、ここで言うワードとは日本語の全文検索の環境においてコンピュータの処理により切出される文字綴りであり、利用者が考えるキーワードとは必ずしも一致しない。

類似検索とか類似文書検索という表現もされているが、機能的には概念検索と同じと考えてよい。特許番号を指定して概念検索を行なうことを類似文書検索と限定しているサービスもあるが、一般的な区分けではない。

and,orなどを組み合わせた検索式を作成して検索する論理検索モデル(ブール検索モデル)に対して、概念検索はベクトル空間モデルと称される手法を用いる。ベクトル空間モデルとは、質問文やデータベース中の特許を対象に、関連ワードのいくつかの観点の出現頻度を利用¹⁾して類似度を算出する方式である。これを「ベクトル空間モデル」というのは、質問文と対象特許のそれぞれにおいて特徴を数値で表すとき、

関連する n 個のワードで構成される n 次元の座標（空間）を想定し、各ワードの出現頻度をベクトル合成して一つの値を算出するからである。さらに概念検索は、対象特許が質問文の概念を満たしている確率を考慮した確率モデルとか、その他様々に分けられる。しかし、実現しているシステムはそれらの典型的なものとは限らず、部分的に利用しているために明確な区別の難しいものもある。

検索式方式と概念検索を次のように分けることも実務的には有益である。

- (1) 検索式方式・・・and, or を使う検索式とか論理式、ブール検索モデルといわれる方式。
- (2) 検索式方式 + 優先順回答・・・検索式により数千件程度まで絞り込んだものを対象に、概念検索または類似の方法で該当技術との類似性を計算し、優先順をつけて出力する。しかし検索式の利用が必須であり、その段階で漏れた情報は絶対に取り出すことができないので、利用に当たっては概念検索の気軽さはない。
- (3) 概念検索・・・質問や対象特許のワードの出現頻度を巧妙に利用するベクトル空間モデルにより類似度を計算し、優先順に回答をリストアップする。
- (4) 概念検索 + 絞り込み検索・・・概念検索と共に、キーワードや IPC（国際特許分類）、出願人、発行期間などで限定して出力するもの。絞り込みは補助的に行なうものであり、絞り込む件数の制限などはない。利用者からみれば概念検索のさらに発展した使い方といえる。
- (5) 係り受け概念検索・・・文章の係り受けの関係を認識した概念検索。例えば「半導体を冷却する」が「半導体で冷却する」を区別できるもの。2005.2 月の時点ではまだ一般利用できるものはないが、大学とメーカーで研究が進んでおり、将来さらに再現率（ ）の良い概念検索が出てくると考えられる。

再現率・・・正しい回答全体に対し、ある検索で取り出しえた割合。

2.2 概念検索の比較検討

概念検索を利用しようとする場合、次のような事項を検討することが多い。

- (1) 概念検索の動作原理
- (2) 検索できるもの、できないものの内容の相違
- (3) 正解が何番目に出るか？

- (4) 再現率はどれくらいか？
- (5) どのような質問（文）が良いのか？
 - ・番号指定で類似特許を探せるか？
 - ・請求範囲は質問文に適しているか？
 - ・短文と長文の比較？
 - ・キーワードと文章の違いはあるか？
 - ・同義語は必要か？
 - ・対象技術により再現率の相違はあるか？
 - ・システムによる相違

(6) 絞り込み検索や検索式の併用の利用法、効果

(7) どのような知財関連業務に使えるか？

- ・技術的調査と権利的調査のどれに使えるか？
- ・所要時間と再現率はどの程度か？
- ・アイデア発想支援に利用できるか？
- ・動向分析に使うには？
- ・SDI サービスに使えるか？
- ・公知例調査に使えるか？
- ・特許性調査に使えるか？

上記において、(1)～(4)は基本的で重要なことであるが、これらに拘泥されて、実際の利用に進めないのでは利用者のメリットにならないので注意が必要である。これに対し、(5)以降は実務上重要なことであり、利用を進めると共に研究を続けるべきものとする。

かつて、データベース検索が台頭してきたとき「紙の調査資料は絶対に無くならない」と言う意見があった。ある見方ではそうも言えようが、大局としては、現在の特許調査はほぼ紙から脱却している。データベース検索も概念検索が充実しつつあることを認識し、前向きな姿勢が必要と考えられる。

概念検索の動作原理はすでに紹介¹⁾されているので、以下では、現時点における概念検索の機能と性能の比較をおこなう。

3. 概念検索システムの機能比較

2003 年夏頃までは、特許を対象とした概念検索システムは 1, 2 であったが、2005 年になると以下のように多くのものが使える状態になった。

- ・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
- ・PATOLIS
- ・PATOLIS - J

・ A T M S / G - Search
 ・ R I P W A Y / リコーテクノシステムズ
 この他にも、Bluesilk / 三菱総合研究所で 2, 3 年分の特許を対象に概念検索の実証実験が行なわれている。また、独立行政法人工業所有権情報・研修館から財団法人日本特許情報機構 (Japio) に委託されている特許流通データベースにおいてはライセンス希望として登録された特許を対象に概念検索が利用可能である。

表 3 - 1 は、現在利用可能な概念検索システムの機能を比較したものである。これから分かるように、一口に概念検索システムと言っても利用できる機能は大きく異なる。以下に、特徴的な部分を解説する。
 収録対象・・・古くから収録されている方が利用者にとっては好ましい。

検索対象・・・多くの場合、検索対象は要約や請求範囲よりも明細書の方が再現率が高い。(4章の性能比較で検討する。)

番号指定の概念検索・・・関心を持つ人が多いと思われるので項目として取り上げたが、実際の利用ではあまり再現率の高い利用は期待できない。

質問文の切出し・・・切出しワードの個数は多い方が良いと思われそうだが、実証結果はない。ワードの追加や重み付けの変更は、補助的役割であるが、行なえる方がよい。

必須項目の設定 (絞込み検索)・・・これも補助機能であるが、重要である。必須のキーワードやIPCを制限した概念検索は、再現率を大きく上げられる。出願人や発行期間の制限も貴重である。

表 3 - 1 概念検索の機能比較

システム		A	B	C	D	E	F
収録対象	公開公報	請求範囲 S61 ~ H5 ~	S58 ~	S58 ~	S58 ~	H5 ~	S58 ~
	公告公報	S61 ~	S58 ~	S58 ~	S58 ~	H6 ~	S58 ~
検索対象フィールド	明細書	×	○	○	×	×	
	全文フィールドの任意指定	×	○	×	×	×	×
	要約	○	○	独自抄録	×	(まとめで指定する)	(まとめで指定する)
	請求範囲	○	○	○	○		
特許番号入力の場合概念検索		×	○	○		×	
検索結果の番号指定の場合概念検索				○	○	×	
質問文のワード	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)	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	×	○	○	○		
(絞込み検索)	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	×	○	×	×		

4 . 概念検索システムの性能比較

概念検索はシステムによって再現率が大きく異なる。多くの場合、検索対象は長文、質問文は短文の方が抽出件数は多い。以下、実例で検討する。

4 . 1 比較検索 1 (生ごみ分解脱臭)

(1) 短文による検索

テーマ; 家庭などの生ごみを微生物で分解処理する際に発生する悪臭を、木材チップを通して微生物などで分解処理するもの。

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」

関連特許；概念検索や検索式による5，6回の検索でみつけた11件の特許とした。

表4-1は、5つの概念検索システムの結果であり、上位100位に出てきたものの順位と、網掛けは上位50位以内のものである。ヒット件数は、大きな差がある。この差は、システムにより検索対象が明細書と請求範囲などと異なることに加えて、類似度の評価法が異なるためと考えられる。論理検索においてはシステムによるこのような大きな差はないが、概念検索では

珍しいことではない。概念検索が、ある用途に適・不適を検討する場合は、このような状況を知った上で行なうことが必要と思われる。

なお、本記事の検索は2004年10月～2005年2月におこなったものだが、Fシステムのように検索時期が異なると結果に大きな違いの出る場合がある。その原因は、蓄積データが増えることによる数値の微妙な変化に加え、システムの細部の変更(チューニング)がある。さらに、システム本体の改善が行われると、大きな改善のなされる場合もある。

表4-1 生ごみ処理の脱臭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。網掛けは上位50位までの件数。

関連特許 検索対象	システム					
	B	C	D	E	F	
	明細書	明細書	請求範囲	要約+ 請求範囲	明細書	
					2004.11月	2005.2月
特開平 10-290976	n	n	n	n	n	n
特開 2000-042356	49位	n	n	n	n	24
特開 2001-070424	n	12	n	n	84	8
特開 2001-079521	42	15	55	n	14	41
特開 2001-079522	n	27	90	n	100	47
特開 2001-191057	45	20	16	n	24	40
特開 2001-198557	n	n	n	n	n	n
特開 2001-353419	32	23	32	n	5	42
特開 2002-186936	37	6	n	n	n	19
特開 2002-204923	3	4	n	n	36	15
特開 2003-225530	13	1	n	n	34	3
50位までのヒット件数	7件	8件	2件	0件	5件	9件

質問の短文； 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

表4-2 特開 2000-42356 (生ごみ処理の脱臭) の関連特許の概念検索

抽出件数；上位50位までの件数。指定特許を除く。

(1) Bシステムによる抽出件数

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システムによる抽出件数

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・・・(省略)

要約；【要約】【課題】脱臭のために使用する材料が安全で容易に入手・・・(省略)

(2) 請求範囲などによる概念検索
すでに指摘されている²⁾ことであるが、多く

の場合において、質問文としては請求範囲や明細書よりも目的に合致した短文の方が再現率は

高い。(1)と同一テーマで特開 2000-42356 をターゲットとして検討する。

表 4 - 2 は、二つのシステムにおいて、質問文を短文、特開 2000-42356 の第 1 請求範囲、出願人作成の要約、明細書とし、検索対象を明細書、全請求範囲、要約として、概念検索をおこなったものである。両システムは、質問文は 50 文字程度の短文が最も良く、その中でも検索対象は明細書が最も良いという類似の傾向がみられる。

質問文として、なぜ長文よりも短文が良いのか検討するために、以下、短文と第 1 請求範囲の文章を比較検討する。

短文 (54 字); 「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 (216 字); 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みガス・・・(以下省略)

図 4 - 1 は両質問文の主なワードと、その出現回数を概念検索における影響の大きさをイメージできるように示したものである。これから分かるように短文の方は「脱臭、微生物、木材、チップ」と「家庭、生、ゴミ、分解、排気、芳香」は 2 回と 1 回の差しかなく、重点の差の小さい検索を行うことが分かる。これに対し、請求範囲の方は「ガス、脱臭、導入」に強く偏っており、「仕切、充填」などにもやや偏った検索を行うことが伺える。調査目的からみると、短文の方が明らかに妥当なワードに重点化されていることが分かる。

<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>(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:短文</p>	<p>質問文;第1請求範囲</p>

図 4 - 1 概念検索における主なワードとその重み付けのイメージ図

これから伺えることは、短文に比べて長文または請求範囲は概念検索に適した文章ではないことがありうる、ということである。特に、請求範囲の文章をみると、この例のように技術的な中心概念のワードが必ずしも頻度高く使われないことがあるということが分かる。つまり、ワードの出現頻度を用いる現在の概念検索においては、請求範囲の多くは概念検索の質問文としては適していないということが伺える。

4.2 比較検索 2 (ナノシート)

(1) 3つの短文による検索

テーマ;チタニア超薄膜およびその製造方法
ターゲット特許;特開 2002-265223

質問文;この発明は3つの観点があるので次の3つの短文を用いた。

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。

関連特許;情報科学技術協会のオンラインユーザグループ(通称OUG),特許分科会の主査である鈴木利之弁理士(鈴木利之特許事務所)による調査でピックアップされた21件とした。これには特許庁の審査による引例3件も含む。

表 4 - 3 は、5システムで3つの質問文を使って概念検索した結果である。検索対象としては、請求範囲など(D, Eシステム)よりも明細書(B, C, Fシステム)の方が抽出件数は多い。なお、いずれのシステムでも検索できない数件があるが、これは準備した3つの質問文がこの技術の専門家でない者が作成したことも影響していると考えられる。専門家が作成すれば、さらに若干の再現率向上が期待できるはずである。

(2) 請求範囲による検索

この特許の請求範囲は、前記4.1の比較検索の請求範囲の文章と趣が異なり、本稿で説明している概念検索用の短文に似た形をしている。つまり、下記に示すように、技術説明に必須のワードが無駄な繰返し少なく使われており、概念検索の質問文として使える可能性がある。

【請求項1】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。(74字)

表4-4は、4システムで質問文を第1請求項として、明細書と全請求範囲などの対象を、概念検索した結果である。B、Fシステムは傾向が似ており、対象が明細書、請求範囲にかかわらず抽出件数は多い。

表4-3 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。

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。

関連特許	システム 検索対象 質問文	B			C			D			E			F		
		明細書			明細書			請求範囲			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	n	n	n	n	n	
2	特開平 1-41121	n	n	n	n	n	n	n	n	n	n	n	n	n	n	
3	4-305018 引例3	23位	n	n	n	n	n	n	n	n	n	n	n	n	n	
4	8-245815	n	84	n	n	21	n	n	11	99	n	31	n	n	11	
5	9-21100	n	n	n	n	n	n	n	n	n	n	n	n	n	n	
6	9-25123	3	n	n	5	n	n	n	n	n	95	n	n	6	n	
7	9-67124	40	n	n	n	n	n	n	n	n	n	n	n	n	n	
8	9-227122	32	n	n	48	n	n	n	n	n	n	n	n	n	n	
9	9-227123	31	n	n	42	n	n	n	n	n	n	n	n	n	n	
10	10-100306	n	45	n	n	1	n	n	n	n	n	50	n	n	8	
11	10-167707 引例2	n	1	84	n	2	n	n	n	19	n	n	n	n	1	
12	10-259023	35	n	n	n	n	n	n	n	n	n	n	n	n	n	
13	11-90213	n	n	n	n	n	n	n	91	39	n	n	n	n	13	
14	11-140656	n	n	n	n	n	n	n	n	n	n	n	n	n	n	
15	11-319542	n	n	n	n	n	n	n	n	n	n	n	n	n	76	
16	11-329001	n	n	n	n	n	n	n	n	n	n	n	n	49	n	
17	特開 2000-86241	13	8	n	13	68	n	n	n	n	n	n	n	n	n	
18	2001-270022 引例1	4	27	2	1	33	2	1	10	2	19	n	n	4	27	
19	2002-104825	n	n	n	n	n	n	n	n	n	n	n	n	26	n	
20	2002-225172	30	n		n	n	n	n	n	n	n	n	n	15	n	
21	2002-241130	73	n	14	n	n	n	n	n	n	n	n	n	53	n	
50位までのヒット件数		9件	4	2	5	4	1	1	2	3	1	2	0	5	5	0
" 合計		12件			8件			4件			3件			9件		

注1. 表中の関連特許 21件は情報科学技術協会・オンラインユーザグループ(OUG)・特許分科会主催の鈴木弁理士による調査結果³⁾

注2. 表中の質問文イ、ロ、ハは下記の通り。

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、静電相互作用に基づく「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。

表 4 - 4 特開 2002-265223 関連特許の請求範囲による概念検索 (指定特許を除く)

システム 検索対象	B		C		E	F	
	明細書	請求範囲	明細書	請求範囲	要約+請求範囲	明細書	要約+請求範囲
上位 100 位までの件数	12 件	11	8	2	5	9	10
上位 50 位までの件数	11 件	10	8	1	4	7	10

注 1. 質問文は第 1 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。

注 2. 上記の請求項は, 技術説明に必須のワードのみが無駄な繰返し少なく利用されている文章なので概念検索の質問文として適しており, 抽出件数が高くなっている。

表 4 - 5 特開 2002-265223 関連特許の概念検索

抽出件数; 上位 50 位までに出てきた件数。指定特許を除く。

(1) B システムによる抽出件数

質問 対象	短文 64 字	第 1 請 求範囲 74 字	要約の 構成のみ 225 字	全要約 387 字	明細書 9,900 字
明細書	9 件	11	7	4	1
請求範囲	8	10	6	5	1
要約	5	5	6	6	1

(2) C システムによる抽出件数

質問 対象	短文 64 字	第 1 請 求範囲 74 字	全要約 387 字	明細書 9,900 字
明細書	5 件	8	3	(省略)
請求範囲	1	1	1	(省略)
抄録	1	2	1	(省略)

短文; 層状のチタン酸化物を溶液中で振動攪拌 (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

第 1 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。

要約; 【要約】【課題】 本発明者が先に開発したチタニアナノシートと・・・(省略)

表 4 - 5 は, 質問文が, 短文, 第 1 請求範囲, 要約/構成のみ, 全要約, 明細書において, 抽出件数がどのように異なるかみたものである。(C システムは一部を省略。) 質問文が 200 文字を超えると漏れが多くなり, 検索対象は明細書のように長い文の方が抽出件数は多いことが分かる。

4.3 質問文は短文が良いか?

質問文は数十文字 ~ 100 文字程度が良い理由は, 利用者が調査したい技術内容を表すワードが質問ワードとして使われる可能性が大きいためと考えられる。

なお, 上手に概念検索のできる短文の作成法が難しいとの意見もある。しかし, これまでの実例からも分かるように, 該当技術の専門家なら容易に作れる程度のものである。(とはいえ, 「適当な文章を入力して欲しい特許を探せる」というほど簡単ではないが。)

筆者は説明会で「欲しい情報を社長に 2, 30 秒で説明するつもりで作ればよい」, 「一般化したり, 同義語を補ったりする必要はない」と言

っている。

また, 表 4 - 5 によると, 請求範囲が短文と同等以上に抽出件数の多い場合もあることが分かる。上記の特許は化学材料であるので, 同分野の特許ならいつでも請求範囲が質問文として適しているかと考えるとそうではない。

例えば, 特開平 9-118663 は「光照射で硬化するスルホニウム塩化合物、重合開始剤を含有する硬化性組成物」に関する特許である。この請求範囲は化学構造式を含んでいるが, それを除いた下記の請求範囲(281 文字)を質問文として概念検索すると自分自身以外ほとんど検索できない。

【請求項 1】一般式〔 I 〕【化 1 〕〔式中、R1, R2 はそれぞれ C1-18 アルキル基、ヒドロキシ基、C1-18 アルコキシ基、C1-18 アルキルカルボニル基、芳香族カルボニル基、芳香族チオ基またはハロゲン原子を表す。R3 は・・・(中略)・・・X は非求核性のアニオン残基を示す。〕で表されるスルホニウム塩化合物。

これに対し, 下記のような質問文にすること

で、抽出件数を大幅に増やせる。表4 - 6はその上位の件数である。

質問文；加熱または光照射により、極めて薄い膜から厚手の膜まで短時間で硬化するスルホニウム塩化合物、カチオン重合開始剤、硬化性組成物。

表4 - 6 光硬化スルホニウム塩化合物

順位	特開#	内容
2	08-325225	スルホニウム塩化合物、重合開始剤・・・光照射で硬化
3	07-300504	〃
4	07-300505	〃
5	07-126313	短時間の光照射で硬化するカチオン硬化性組成物
6	2003-277352	スルホニウム塩、カチオン重合開始・・・光照射で硬化
7	2003-277353	〃

このように、化学材料分野であれば請求範囲で概念検索が行なえるわけではなく、請求範囲が概念検索の質問文として利用しやすい形であればそのまま利用できるということである。その形とは「技術説明に必須のワードが無駄な繰返し少なく使われている」場合と考えられる。

4.4 複数の番号指定の概念検索

特許番号を指定して類似特許を探せるならば大変便利である。表4 - 7は、4.2比較検索2（ナノシート）と同じ課題を対象に特許番号を使って概念検索を試みたものである。質問ワードを抽出する項目を要約、特許請求範囲、明細書の3通りとし、指定する特許を特開2002-265223の1件だけでなく、ほぼ同じ内容の特許2, 3, 4件の4通りで比較した。その結果は、表4 - 5に示す質問文を短文などの妥当なものにした場合より、いずれも抽出件数は少ないことが分かる。

また、抽出件数が比較的多いのは、質問ワード抽出は明細書よりも要約や請求範囲から行なう方がわかる。これは先に述べた数十文字程度の短文からワードを選択する方が再現率が高い事実とも一致する。

また、指定特許は1件の場合よりも複数件の場合が抽出件数は多い。これは、図4 - 2に示すように類似内容の特許を複数件指定すること

で、目的の技術が強調されることによるものと考えられる。

しかし、表4 - 7で4件を指定した要約の場合は、逆に抽出件数が5件に減っている。この原因を探るために、4件目だけを加えた2件で検索してみると、抽出項目が要約の場合、3件しか出てこない。これは4件目の特許が、人が明細書を読むと類似内容と思えるが、要約は類似内容と言えないような書き方になっているためと推察される。このように、類似内容特許を複数件指定しても必ずしも抽出件数が増えるとは限らない場合がある。

表4 - 7 複数の特許番号指定による概念検索
抽出件数；上位50位までに出てきた件数（指定特許を含む）
対象；Bシステム、検索対象；明細書

指定特許件数	質問ワード抽出項目 指定特許	要約	請求範囲	明細書
1件	特開 2002-265223	5件	4件	2件
2件	+ 特開 2002-104825	7	8	5
3件	+ 特開 2002-225172	8	8	5
4件	+ 特開 2002-241130	5	8	6
2件	特開 2002-265223 + 特開 2002-241130	3	4	3

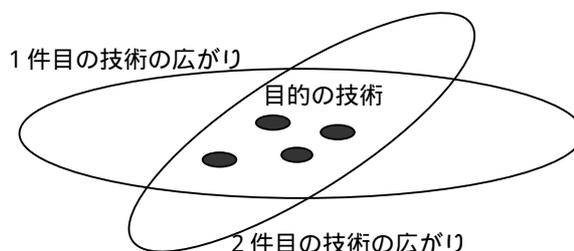


図4 - 2 複数指定の概念検索のイメージ

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，2001.12月
- 2) 概念検索による特許情報の活用法，六車正道，発明，2003.4月，5月，6月
- 3) 情報科学技術協会，OUG，特許分科会研修会資料，鈴木利之，2004.6月

