

# 特許を対象とした 概念検索の技術課題

## 概念検索精度の向上に向けて

株式会社日立製作所 システム開発研究所 間瀬 久雄

### PROFILE

平成2年(株)日立製作所入社、システム開発研究所に配属。以来、特許や新聞記事、Webページ等を対象とした、分類自動付与、検索、文章要約、テキストマイニング等の日本語処理の研究に従事。2007年度から特許版産業日本語委員会委員。

✉ hisao.mase.qw@hitachi.com



## 1 はじめに

任意の自然言語文章を入力して、これに類似する文書を検索する概念検索(類似文書検索、自然言語検索)が普及してきている。

特許庁では、業務・システム最適化計画<sup>1)</sup>推進の一環として、平成20年度調査研究「審査関連情報を活用した次世代検索システム開発に向けた調査」<sup>2)</sup>において、概念検索の特許審査業務への適用可能性について調査している。学界では、国立情報学研究所が主催するタスク型ワークショップNTCIR<sup>3) 4)</sup>において、概念検索によって審査官引用文献を検索する精度を競う特許検索タスク<sup>5)</sup>が開催されるなど、特許を対象とした概念検索技術の研究が加速している。

概念検索は、全文検索や書誌検索と並び、検索の一手段として定着しつつある。しかし、精度的にはまだまだ改善の余地が大きい。そこで本稿では、特許を対象とした概念検索の現状の技術課題と、今後の研究の方向性について考察する。以下、2章では、概念検索の概要を説明する。3章では、特許を対象とした概念検索精度の現状について述べる。4章では、検索精度を低下させている要因について考察し、概念検索技術研究の今後の方向性について提言する。

## 2 概念検索とは

本章では、概念検索の概要について説明する。

### 2.1 概念検索の概要

概念検索は、任意の文章を入力として、その文章の内容に類似する文書を、類似する度合い(類似度)の高い順に出力する検索方式である。

図1に示すように、まず、入力された文章からその内容を特徴付ける語(特徴語)を抽出し、語の出現傾向(出現頻度や出現位置など)からその特徴語の重要度を表す重みを算出する。次に、検索対象となる文書集合から予め抽出しておいた重み付き特徴語を格納したインデクスと照合し、共通する特徴語がどのくらいあるかに基づいて、入力文章と各文書との間の類似度を算出する。最後に、類似度の高い順に文書を並べ替えて、検索結果を出力する。概念検索では、入力文章と共通する特徴語をより多く含む文書や、重みの高い特徴語を含む文書が、検索結果の上位に出力されやすい性質がある。

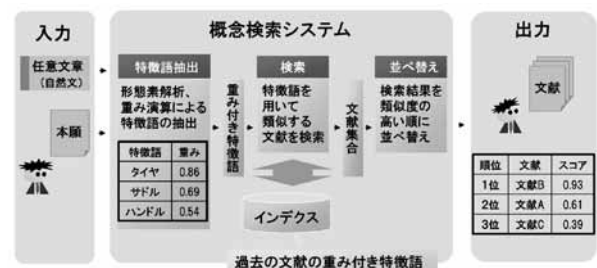


図1 概念検索の処理の流れ

## 2.2 概念検索の用途

特許を対象とした概念検索は、主に以下の場面での適用を想定できる。

### (1) 先行技術文献調査

特許庁や外部調査機関でのサーチ業務、大学・企業での公知例調査などに適用することで、先行技術文献を効率良く探す。

### (2) 技術動向調査（特許マップ作成）

企業などでの特許マップ作成時に適用することで、膨大な特許集合から、作成対象となる特許集合を絞り込み抽出する。

### (3) 特許分類付与

出願特許に内容の類似する特許集合を概念検索によって抽出し、これらの特許の多くに付与されている分類をその出願特許に付与することで、特許に分類を自動付与する。

## 2.2 文章入力の状態

特許を対象とした概念検索では、以下の2種類の文章入力の形態がある。

### (1) 既存の特許を指定する

作成済みの特許（文章全体／一部または出願番号）を入力とする形態である。この形態では、入力文章が長く、文章構造（タグ構成）が検索対象と同じであるため、より精緻な概念検索が実現可能である。

### (2) 任意の文章を入力するもの

論文や Web ページからコピーした文章や、頭の中に思い浮かんだ文章など、任意の文章を入力とする形態である。この形態では、入力された文章がどんな観点（発明の課題、手段、効果、発明の対象など）で記述されているのかを解析する必要がある。また、入力文章が比較的短いため、記述内容の良し悪しが検索精度に大きな影響を与える。

# 3

## 特許を対象とした概念検索の現状精度

本章では、既存の特許文章を入力とした場合の概念検

索精度の評価結果について述べる。

## 3.1 評価環境と評価方法

ここでは、著者らが NTCIR 特許検索タスク（以下、検索タスク）と、上述の特許庁調査研究（以下、調査研究）で行った精度評価結果を紹介する<sup>2) 6)</sup>。

検索タスクでは、公開特許公報 10 年分（93 年～02 年、350 万件）、調査研究では公開／公表特許公報 10 年分（94 年～03 年）を検索範囲としている。ただしどちらも、入力特許の公知基準日より古い特許のみを検索対象としているため、検索範囲が必ずしも 10 年分であるとは限らない。

ここでは精度を評価する指標として、直感的に分かりやすい N 位再現率（概念検索結果上位 N 位までに含まれる正解特許件数の割合、N = 50, 100）を用いる。ここで正解特許とは、審査官が拒絶通知の中で引用した特許（引文）である。

## 3.2 評価結果

検索タスクの精度評価結果を表 1 に、調査研究の結果を表 2 にそれぞれ示す。ここでは概念検索方式として、検索タスクでは 4 種類<sup>6)</sup>、調査研究では 2 種類<sup>2)</sup>の方式を採り上げている。

50 位再現率は、検索タスクで最高 47.1%、調査研究で 25.9% であり、どちらも精度的にはまだ改善の余地が大きい。検索タスクの方がかなり高い数値となっているのは、検索タスクでは公知基準日が比較的古いものが多く、実質の検索範囲が 2、3 年という入力特許が多いためであると考えられる。

検索方式としては、請求項など特許の一部分の文章を入力とするよりも、特許全文を入力として分類などの書誌情報も使う方式が最も精度が高い。

また、詳細は割愛するが、技術分野別の精度にもばらつきが見られる。素材、生命・環境、応用科学の分野で精度が高い反面、情報通信や画像処理の分野では検索精度が低い。このばらつきの原因としては、使用する語句の分野固有性（一般的な語句が多く使われる分野は精度

が低い) が挙げられる。

表1 特許概念検索精度 (NTCIR-5 課題 1189 件)

方式	50 位再現率	100 位再現率
方式 1	36.4%	43.6%
方式 2	41.2%	49.9%
方式 3	43.4%	52.7%
方式 4	47.1%	57.7%

方式 1 : 請求項 1 のみを入力として特徴語を抽出  
方式 2 : 全文を入力として重みの高い特徴語 70 語を抽出  
方式 3 : 方式 2 で文章構造を考慮して特徴語の重みを補正  
方式 4 : 方式 3 で分類の共通性を考慮し検索スコアを補正  
※公知基準日が古い課題が多く、実質の検索範囲が狭いため、  
正解特許が検索結果の上位に集まりやすい

表2 特許概念検索精度 (100 テーマ 13,820 件)

方式	50 位再現率	100 位再現率
方式 2	23.8%	31.1%
方式 3	25.9%	33.6%

方式 2 : 全文を入力として重みの高い特徴語 70 語を抽出  
方式 3 : 方式 2 で文章構造を考慮して特徴語の重みを補正  
※方式 1 と方式 4 は未実施  
※公知基準日が新しい課題が多く、実質の検索範囲が広い  
ため、検索タスク (表 1) に比べ、検索漏れが多くなっている

## 4 概念検索精度を低下させている要因

本章では、概念検索精度の低下要因について考察し、今後の研究の方向性について提言する。

### 4.1 精度評価の難しさ - 類似するとは何か -

要因を列挙する前に、概念検索精度評価の難しさと、精度数値の扱いについて触れておきたい。

概念検索に関する技術論文では、「本システムの検索精度は 70%」とか、「検索精度を 15% 向上できた」など、検索精度を表す数値が記述される。しかし、実際に精度を測定したり、複数の検索方式の良し悪しを比較したりする場合には、精度に対する以下の配慮が不可欠である。

#### (1) 正解特許の定義 (「類似」の基準の設定)

実際の業務への適用を踏まえ、「類似」の基準を明確

にした上で、正解特許を定義すべきである。

検索タスクをはじめ、特許を対象とした概念検索の精度評価では、引文を正解とし、これが概念検索結果の何位に出力されるかを評価することが多い。

しかし、入力特許の発明を拒絶できる特許が複数件あった場合、必ずしもそのすべてが引文となるわけではなく、拒絶できるほど類似しているにもかかわらず、正解とみなされない特許も存在する。

また、先行技術文献調査では、発明内容が完全一致する特許のほか、部分一致する特許も重要視される。したがって、部分一致する特許をどこまで正解特許に含めるか、完全一致する正解特許とどう区別して精度を評価するかを決める必要がある。

一方、技術動向調査において、分析対象とする特許集合の抽出に概念検索を適用する場合、引文を正解として精度評価するのはナンセンスである。

実際の評価では、正解特許を定義することが作業コスト的に難しいため、引文のような既存データを安易に利用しがちである。しかし、概念検索の精度評価では、「類似するとはどういうことか」をきちんと定義した上で、正解特許を決めることが本質的に重要である。

#### (2) 評価指標の多様性

適用業務の特性や、「類似」の基準を考慮した上で、精度評価に用いる指標を選定すべきである。

学術界では、文書検索の精度を表す評価指標<sup>7)</sup>として、平均精度 (average precision) が使われることが多い。また、再現率 (recall、検索漏れの少なさを表す) と精度 (適合率、precision、ノイズ文書の少なさを表す) もよく使われる。

一般に平均精度は、1 件の入力特許に対する正解特許数が比較的多い場合に有効な評価指標である。したがって、引文のように 1 特許あたりの正解特許数が少ない場合に適用するのは不適切である。また、平均精度は上位 10 位程度までの検索結果に大きく依存する「適合率重視型」の評価指標なので、検索結果上位 100 位程度までに正解が入っていれば良いという「再現率重視型」の

用途で概念検索を適用する場合には、不適切な指標となる。さらに、これらの評価指標は、正解特許数によって数値が大きく変わる。1特許あたりの正解特許数が多いと、再現率は低下するが、平均精度や適合率は向上する。

### (3) 入力として使うデータ

概念検索精度を比較する場合、入力として使う文章データの範囲を統一すべきである。

表1、表2で示したように、入力文章の情報量が多いほど、検索精度が高くなる。また、特許文章を入力とするか、任意の文章を入力とするかによっても精度は変わる。一般に、任意の文章を入力とする場合、文章が比較的短いために精度が悪くなることが多いが、検索したい内容を端的に記述すれば、ピンポイントで検索できる場合もしばしばある。

### (4) 検索範囲（検索対象、検索年範囲）

検索対象や検索年範囲などの検索範囲についても考慮すべきである。入力文章が同じでも、検索対象（特徴語を抽出する範囲）が違えば概念検索精度は異なる。特許検索では、特許全文を検索対象とする場合に、検索精度が最も高くなる傾向にある。

また、先行技術文献調査では、入力特許よりも古い特許のみが検索対象となるため、公知基準日によって検索年範囲が狭まる。したがって、公知基準日が古い特許の場合、検索精度が高くなりやすい。一方、技術動向調査では、調査分析対象の特許を絞り込むことが目的であるので、公知基準日を考慮しない場合が多い。この場合、全特許が検索対象となるので、検索精度が比較的低くなる。

### (5) ベースラインの検索精度

概念検索精度の改善方式の有効性をチェックする際には、比較対象となる方式（ベースライン）の精度の高さをチェックすべきである。

ベースラインの精度が低い場合、どんな改善方式を施

しても精度は向上するが、ベースラインの精度が高い場合には、精度はなかなか向上しない。

上記5項目を鑑みるに、概念検索の精度評価では、精度の値だけでなく、評価環境や評価方法を併せてチェックし、概念検索の目的や用途と照らし合わせた上で、システムや方式の有用性を判定すべきである。とかく一人歩きしがちな検索精度数値を鵜呑みにしないことが重要なのである。

## 4.2 精度の低下要因と解決アプローチ

ここでは、概念検索精度を低下させる要因と今後の研究の方向性について考察する。

検索精度を低下させている要因を一言で言うと、「入力文章に書かれた発明内容と、正解特許に書かれた発明内容を、システムが同一視できない」という当たり前のことになる。では、なぜ同じ発明であると特定できないのか。

### 4.2.1 要因1：発明概念を特定できない

現在ある多くの概念検索システムでは、特徴語を抽出する際に、入力文章を「均一な文章」として扱っている。特許の場合も然りで、請求項も実施例も要約も同じアルゴリズムで均一に解析している。

言い換えれば、特許に書かれている発明概念（要するにこの発明のポイントは何なのかを端的に記述した部分）は何かについて何も考慮していない。その結果、発明概念を端的に表していないノイズ特徴語が多く混在してしまい、概念検索結果にノイズ文書が多く含まれる一因となっている。

特許の発明概念を特定し、その発明概念を持つ過去の特許を検索できれば、ノイズの少ない検索が実現できる。しかし、発明概念を特定するためには、発明概念が特許のどこにどのような形で記載されているのかをパターン化（オントロジー化）したり、記載文章の意味にも踏み込んだりする必要がある、チャレンジングな研究課題である。



発明概念に基づく概念検索のイメージを図2に示す。

発明概念解析では、請求項の文章構造や発明概念の書かれ方をモデル化した「発明概念オントロジー」を参照して発明概念を特定し、シソーラス辞書を参照して表記揺れの概念レベルに変換する。次に、この発明概念を入力として概念検索を行うことにより、ピンポイントでブレのない検索を行う。

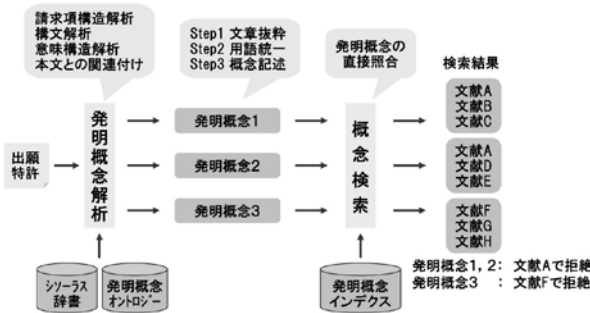


図2 発明概念に基づく概念検索

発明概念に基づく概念検索を具体例で見てみる。

図3は、NTCIR 課題特許 1013 の請求項1を入力文章とした時の正解特許の検索順位を比較したものである。正解特許は8件あり、正解特許Aはこの1件だけで拒絶できる特許であり、残りの7件の正解特許B1～B7は、発明内容が部分的に一致しており、これらの組合せで拒絶できる特許である。

請求項1全文を入力文章とした場合、正解特許Aは1位に出力できているが、他の7件の正解特許は、B5を除いて検索順位が100位以下と低い。

これに対して、請求項1を構成している構成要素1「頂面に燃焼室を凹設するとともに該燃焼室の周縁にスキッシュエリアを設けたピストン」を一つの発明概念と捉え、発明対象「ディーゼルエンジン」を付加したものを入力文章として概念検索すると、正解特許B1とB2の検索順位が飛躍的に向上する。この例では、「スキッシュエリア」「ピストン」に発明のポイントを絞って概念検索することにより、請求項1全文を入力した時に比べて、二つの正解特許を上位に出力できるようになって

いる。

正解特許	入力文章（請求項1）				
	請求項1全文	1構成要素+発明対象（末尾）			
		要素1	要素2	要素3	要素4
A	1	1	11	164	56
B1	132	19	51	1251	500
B2	140	10	57	2077	312
B3	222	123	3169	1146	10001
B4	1755	1991	7220	1841	1311
B5	35	366	274	333	295
B6	2030	1675	7043	5566	3045
B7	1289	6599	3408	4114	3843

- NTCIR 課題 1013 の請求項1の構成要素
  - 構成要素1 頂面に燃焼室を凹設するとともに該燃焼室の周縁にスキッシュエリアを設けたピストンと、
  - 構成要素2 前記燃焼室に対向するドームを凹設したシリンダヘッドと、
  - 構成要素3 前記ドームの中央部に取り付けられて該ドームの側壁面に向かって開口する噴口を備えた噴射ノズルを備え、
  - 構成要素4 前記ドームの外径を燃焼室の口径とほぼ同一に形成したことを特徴とする
- 発明対象 ディーゼルエンジン。

- 正解特許 B1 の請求項1文章
 

シリンダ上端の内周の一部分の外側と内側とにまたがって主燃焼室を設けるか、または前記のシリンダ上端の内周の一部分の外側とシリンダ内空間に面したシリンダヘッドの一部分のどちらかに主燃焼室を設け、ピストン上面とこれに対向する前記の主燃焼室以外のシリンダヘッドの対向面とをほぼ平行にするとともに、上死点に達した時にピストン上面が前記対向面に接近する様にしてスキッシュエリアを形成し、前記の主燃焼室とシリンダ内空間とを連通させ、前記の主燃焼室内の壁面に燃料噴射弁を設け、そして前記の主燃焼室内に衝突面を設けて、燃料噴射弁の噴射方向を前記衝突面に向け、衝突面上の燃料噴流の流れ方向の終端部に角部を形成して、衝突面に衝突後の燃料噴流が前記の角部から飛散する為の空間を主燃焼室内に設ける事を特徴とする燃料の衝突拡散を行うディーゼルエンジンの燃焼室。

図3 請求項の構成要素別の検索順位比較

発明概念を特定するには、文章要約などの高度な文章解析技術が必要である。先の例では、構成要素文字列そのものを発明概念として抽出しており、解析粒度が粗く、用語も統一化されておらず、概念レベルでの記述にもなっていない。発明概念オントロジーを網羅的かつ高精度で構築できるか、発明概念をどこまで定式化できるかが鍵となる。また、そのために特許明細書をどのよう

1 この正解特許は引文以外の特許も含んでいる。

に書いたら良いかという、産業日本語の側面からのアプローチも、解析精度向上の観点からは重要になると考える。

一方、発明概念は一つの特許に複数存在することもあり、各発明概念に対応する先行技術文献を特定する必要がある。この場合、一度の概念検索ですべての先行技術文献を取得するのではなく、発明概念毎に、あるいは発明概念を複数組み合わせる概念検索を行うなどの検索戦略が必要となる。この検索戦略の立案は、審査官やサーチャのノウハウであり、機械処理は難しいため、発明概念に基づく概念検索では、検索戦略に関するユーザとのやりとりを介してインタラクティブに概念検索できるようなユーザインタフェースが重要になると考える。

#### 4.2.2 要因2：用語を同定できない

キーワードを論理演算子（AND/OR/NOT）で結合した論理式による全文検索では、所望の文書をヒットさせるために、同義語・シソーラス展開をサポートしている。このことから、概念検索でも特徴語を同義語・シソーラス展開することにより、精度が向上すると考えるユーザが多い。ごもっともな考え方である。概念検索でも、表記の異なる特徴語を統一することにより、精度向上が期待できる。

しかし、実際の概念検索において、同義語・シソーラス展開が検索精度にどう影響しているかを見てみると、人間の直感とは異なった振る舞いをするところがあることが分かる。

図4を例に考えてみる。入力特許では、特徴語「頂面」が使われているが、正解特許Aでは、「頂部」が使われている。同義語展開をしない限り、この二つの語はマッチしない。したがって、特徴語「頂面」だけから見ると、正解特許Aは入力特許に類似していないことになる。ここで、同義語展開をすると、「頂部」と「頂面」は同一の特徴語と認定される。その結果、正解特許Aは入力特許に類似しているとみなされ、正解特許Aの類似度が少し高くなり、正解特許Aはより上位に出力される。

#### ●入力特許の請求項1文章（NTCIR 課題 1013）

頂面に燃焼室を凹設するとともに該燃焼室の周縁にスキッシュエリアを設けたピストンと、前記燃焼室に対向するドームを凹設したシリンダヘッドと、前記ドームの中央部に取り付けられて該ドームの側壁面に向かって開口する噴口を備えた噴射ノズルを備え、前記ドームの外径を燃焼室の口径とほぼ同一に形成したことを特徴とするディーゼルエンジン。

#### ●正解特許の請求項1文章

ピストン頂部を凹形状にすることにより第一の燃焼室を形成し、且つ前記第一の燃焼室に対向するようにシリンダヘッド下面のシリンダボア軸心上を凹形状にすることにより第二の燃焼室を形成し、前記第一の燃焼室及び前記第二の燃焼室に向けて加圧燃料を噴射する燃料噴射ノズルを設けることを特徴とする直噴式ディーゼルエンジンの燃焼室構造。

図4 入力特許と正解特許の特徴語の比較

この例ではこれでめでたしたが、検索には副作用が必ず存在する。もし、正解特許Aの記載が「頂部」でなく「頂面」だったら、同義語展開する必要はない。ここで同義語展開してしまうと、正解特許Aの類似度は変わらないが、「頂部」を含む他の特許の類似度が高くなるため、正解特許Aの順位は低下する。

また、正解特許Aの記載が「頂部」であり、同義語展開によって「頂面」と「頂部」が同一の語と認定されたとしても、正解特許Aの順位が下がるということも起こりうる。同義語展開による正解特許Aの類似度の増加分が、「頂部」を含む他の特許における類似度の増加分よりも小さい場合、正解特許Aが追い抜いた特許の件数よりも、追い抜かれた特許の件数の方が多くなり、順位を下げることがある。

この種の副作用は、論理演算式による全文検索における同義語展開でも起きている。「頂面」で全文検索した場合に100件ヒットしたとすると、「頂面 or 頂部」と同義語展開して検索すると、当然ヒット件数は増え、ヒット件数は100件を超える。その結果、検索結果を精査するのに多くの時間がかかる、検索結果を100件に絞り込むための新たな検索条件を追加する必要が生じるなどの副作用が起こる。

概念検索では、所望の特許がヒットする／しないではなくて、どのくらいヒットしているか（類似しているか）という程度が類似度として数値化される。したがっ

て、上記の同義語展開をしなくても、他の特徴語が多く照合していれば、検索結果の上位に出力されるし、逆に、同義語展開をしても、他の特徴語の照合状況次第では、上位に出力されないこともあるという複雑な挙動を起こす。

では、概念検索で同義語展開は精度向上に貢献しないのかというと、答えはノーである。では、検索精度向上のための同義語展開はどうあるべきか。

第一に、機械的に一律に同義語展開するのではなく、同義語展開すべき語を厳選した上で同義語展開をすることである。例えば、先に述べた発明概念を表す記載箇所に含まれる特徴語については、重点的に同義語展開を行うべきである。これにより、同義語展開の副作用を最小限に留めることができる。

第二に、技術分野および特許の特性を踏まえた質の高い同義語辞書やシソーラスを用意するとともに、文章を解析して語の多義性を適切に解消することである。市販の同義語辞書をそのまま使用しても、概念検索精度は向上しない。なぜなら、技術分野および語の多義性を考慮した文章解析ができていないからである。例えば、「CD」という語には、「キャッシュディスク」「コンパクトディスク」「カドミウム（元素記号）」など多くの意味がある。技術分野を考慮し、文章におけるその用語の使われ方を解析しないと、「CD」がこれらの語すべてに同義語展開されるため、ノイズ文書が混入してしまう。

質の高い同義語辞書を作成し、逐次それをブラッシュアップしていくのは、コスト的に問題がある。特許は毎年大量に出願され、使われる語も日々変化していくので、特許文書から同義語辞書の雛形を自動抽出する技術や、Wikipediaのように多数の人々が少しずつ知恵を出し合って協動的に辞書を作成・メンテしていくような仕掛けにより、辞書の網羅性と品質を高めていくなどの方策が必要であろう。

#### 4.2.3 要因3：語句の文脈を同定できない

同じ特徴語が同じ意味で使われていても、文脈が違う

ために検索ノイズとなることがある。

図5の例において、特徴語「検索」が使われているが、文章Aは画像の検索、文書Bは辞書見出し語の検索について記述されている。ここで、入力文章における検索対象は画像であり、辞書見出し語ではない。したがって、文書Bは検索結果として不適切である。しかし、文書Bは「検索」を含んでいるので、その分類似度が高くなり、ノイズ文書となる。

入力文章
「インターネットの画像を検索する手段と、・・・」
検索対象文書A
「画像DBから画像を検索し、・・・」
検索対象文書B
「辞書の見出し語を検索して意味を抽出し、・・・」

図5 文脈の異なる特徴語

このようなノイズを除去するために、全文検索では、近傍条件を用いた近傍検索が使われる。すなわち、「画像」と「検索」という二つの語が一定距離（N文字）以内に現れる文書のみを検索する機能である。

この考え方は概念検索にも適用できる。現在、概念検索で使われている特徴語は、一単語毎に独立して抽出している。特徴語をその前後で使われている語と関連付けて対（近傍語対）にして抽出することにより、特徴語の意味を特定でき、単語の多義性を解消できる。構文情報や共起関係を手掛かりに近傍語対を抽出して概念検索するといった研究も見られる。

このような文脈を考慮した近傍語対による概念検索では、近傍語対として扱う条件をどう規定するかが課題となる。すべての近傍語対を特徴語として機械的に抽出すべきか、ある制約条件に基づいて抽出された近傍語対だけを特徴語として使用するかである。また、近傍語対に対する重み付けをどのように行うか（一語からなる特徴語の重みとのバランスをどうするか）も課題となる。さらに、近傍語対を構成する各々の語の表記が統一されていないと、近傍語対が正確に照合されないため、同義

語・シソーラス辞書による表記の統一は不可欠であろう。

## 5 おわりに

特許を対象とした概念検索の精度向上に関する技術課題と研究の方向性について述べた。今後は、「単語から概念へ」と解析を深化させ、人間の直感に適合した概念検索結果を実現するための研究を地道に進めていく必要がある。また、ユーザとのやりとりにより検索戦略を反映させた検索を実現するためのユーザインタフェースも重要となる。

本稿で提唱した発明概念の考え方を突き詰めていくと、究極の論理検索になると考えている。入力文章から発明概念を定式化して、検索対象文書と照合させるには技術的ハードルが高いが、今後は人間が頭の中で行っている思考プロセスを少しずつでも採り入れていくことが必要となると考えている。

一方で、特許版産業日本語委員会で推進しているように、特許を書く側のガイドラインを設け、語彙を効率良く収集する仕掛けを作ることにより、文章解析精度や検索精度が改善されると期待している。

### [参考文献]

- 1) 特許庁業務・システム最適化計画（改訂版）  
[http://www.jpo.go.jp/torikumi/system/system\\_optimize\\_re.htm](http://www.jpo.go.jp/torikumi/system/system_optimize_re.htm)
- 2) 特許庁調査研究報告書  
[http://www.jpo.go.jp/shiryuu/toushin/chousa/pdf/kensaku\\_saitekika/kensaku1\\_1\\_2.pdf](http://www.jpo.go.jp/shiryuu/toushin/chousa/pdf/kensaku_saitekika/kensaku1_1_2.pdf)
- 3) NTCIR ホームページ  
<http://research.nii.ac.jp/ntcir/index-ja.html>
- 4) NTCIR – 6 概要  
N. Kando: Overview of the Sixth NTCIR Workshop, Proceedings of NTCIR-6 Workshop Meeting, 2007.

- 5) A. Fujii and M. Iwayama: Overview of the Patent Retrieval Task at the NTCIR-6 Workshop, Proceedings of NTCIR-6 Workshop Meeting, 2007.
- 6) H. Mase: NTCIR-6 Patent Retrieval Experiment at Hitachi, Proceedings of NTCIR-6 Workshop Meeting, 2007.
- 7) 岸田和明：検索実験における評価指標としての平均精度の性質，情報処理学会論文誌TOD（データベース），Vol.43, No.SIG02, pp.11-26, 2002.