

# 概念検索技術および 特許検索への適用可能性について

(株) 日立製作所 公共システム事業部 八木 敬宏  
システム開発研究所 間瀬 久雄  
中央研究所 岩山 真

## 1. はじめに

特許庁の検索システムとして全文検索が導入されてから約10年が経過した。現在では、民間の特許検索サービスでも全文検索が採用されており、特許を検索する一般的な手法として普及している。全文検索では、発明の内容を端的に表すキーワードや分類などから構成される論理式をユーザが作成し、この論理式を満たす特許を検索する。しかし、的確な検索結果を得るためには、検索範囲の指定や同義語の指定、近傍検索などを駆使した、高度で複雑な検索式を作成する必要があり、ユーザの作業負担が大きくなっているという課題がある。

一方、概念検索と呼ばれる検索手法が広まってきている。概念検索が全文検索に比べて最も異なる点は、論理式の代わりに文章を検索条件として検索できることである。概念検索では、入力された文章と内容的に類似する文献を、その類似度の高い順に出力する。文章が類似しているか否かは、文章の中に現れる単語の出現傾向がどのくらい類似しているかによって判定する。

概念検索を適用することによって、複雑な検索式を作成することなく、入力文章に類似する文献を検索することができる。しかし、一方で概念検索では、検索結果として出力される文献集合を何件まで見れば十分なのかの境界線を定義するのが困難であるという課題がある。また、なぜその文献が類似しているかを、ユーザが直感的に理解するのが難しいという課題もある。このように、概念検索および全文検索には一長一短があり、相互補完的に組み合わせて利用するのが現実的であると考えられる。

本稿では、概念検索技術についての説明とその適用可能性について述べる。2章では概念検索の基本的な考え方と処理方法について説明する。3章では特許検索業務における概念検索の適用イメージについて述べる。4章では概念検索を特許検索に適用した際の検索精度向上に向けたこれまでの取り組みについての事例を紹介する。5章では概念検索精度をさらに向上させるための今後の課題とアプローチについて述べる。

## 2. 概念検索について

### 2-1. 概念検索とは

概念検索は、自然言語で書かれた任意の文章や文献を入力とし、その文章の内容に類似する文献を、検索対象となる文献集合の中から検索し、類似度の高い順に出力する検索方式である。概念検索は、自然言語文を入力として類似する文献を検索することから、自然文検索、自然言語文検索、類似文書検索、連想検索などと呼ばれることもあるが、検索の仕組みはどれもほとんど同じである。ユーザは、入力文章に最も類似する文献から順番に閲覧することができるので、全文検索に比べて効率的に所望の文献にたどり着くことが期待できる。

また、検索結果として出力される文献を選択し、その文献に類似する文献を芋づる式に検索することもできる。さらに、検索された類似文献の中でよく使われる単語群を提示することにより、次の検索のためのキーワードをユーザが発想することを支援することも可能となる。

「概念検索」という言葉を聞くと、計算機が文章の意

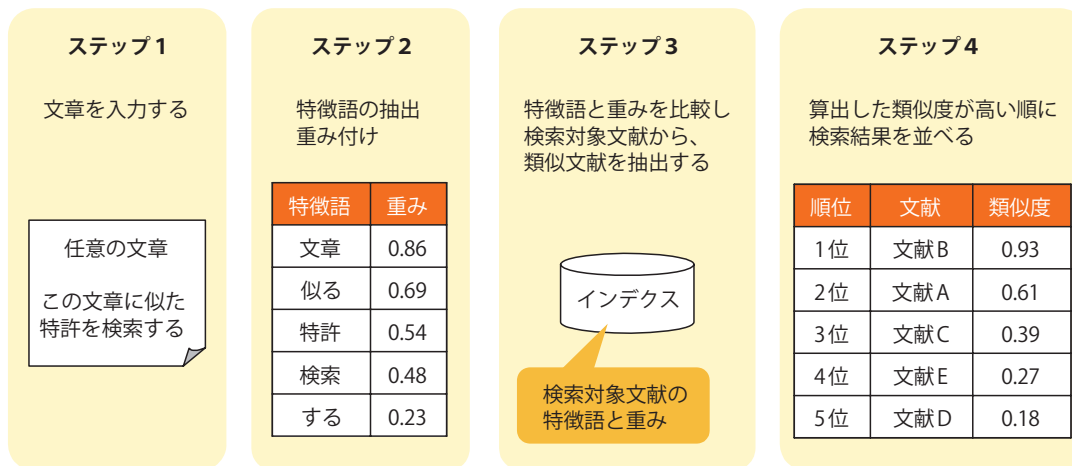


図2-1 概念検索の処理概要

味を理解して検索しているように感じるが、概念検索で言うところの「概念」の定義は、人間が考える「概念」の定義とは異なる。概念検索では、概念を「一つ以上の単語の集まり」として定義する。例えば、「この文章に似た特許を検索する」という文が表す概念を、「文章、似る、特許、検索、する」という五つの単語の集まりとして規定する。

一般の概念検索は、以下のステップで実行する（図2-1）。

#### （ステップ1）検索条件の入力

ユーザはまず、検索するもととなる文章または文献を入力する。

#### （ステップ2）特徴語の抽出と重み付け

ユーザが検索実行を指示すると、システムはまず入力文章を解析し、文章の内容を特徴付ける単語（特徴語）を抽出する。一般に特徴語は、名詞、動詞、形容詞のような自立語で構成される。また、各特徴語に対し、その重要度を表す「重み」を算出し特徴語に付与する。重みに関しては2-2章にて説明する。

#### （ステップ3）類似度の算出

重み付き特徴語の集まりをもとに、類似する文献を検索する。検索対象文献の各々から予め重み付き特徴語を抽出し、インデクスとして格納する。入力文章から得られた重み付き特徴語の集まりを、各検索対象文

献から得られた重み付き特徴語の集まりと照合し、特徴語がどの程度共通しているかを示す「類似度」を算出する。類似度に関しては2-3章にて説明する。

#### （ステップ4）類似度に基づく検索結果並べ替え

検索された文献を、算出された類似度の高い順に並べ替えて出力する。

### 2-2. 特徴語の抽出と重み付け

文章から抽出された特徴語を一律に扱うより、文章の中でその特徴語がどの程度重要な役割を果たしているかを定量化し、特徴語にメリハリを付けることにより、文章がどの程度似ているかをより正確に判定できる。例えば、先に挙げた例では、特徴語「する」は、特徴語「文書」「検索」などと比べると、文章の内容を端的に表す単語とは言えない（一般語に近い）ので、その重みを低くする、といった工夫を施す。

一般の概念検索では、以下の三つの統計的観点から、特徴語の重みを算出する。

#### （1）一つの文献内での出現頻度

##### （TF（Term Frequency））

解析対象となる文献の中で、その特徴語が何回出現しているかという出現頻度を算出する。概念検索では、出現頻度が高いほど重要な単語であると判定し、その重みを高くする。実際には、出現頻度の値そのものを

用いるのではなく、出現頻度の値の対数を用いる場合がほとんどである。

(2) 検索対象文献集合における出現文献数 (DF (Document Frequency))

たとえ一つの文献内での出現頻度が高い特徴語であっても、その特徴語が検索対象文献集合の中の多くの文献に出現する場合、その特徴語は重要であるとは言えない。例えば、特許文献では、「請求項」「発明」などと言った単語はどの文献にも出現するため、その重要度は低くすべきである (図2-2)。そこで、検索対象文献の中でその特徴語が出現する文献数の逆数 (IDF (Inverse Document Frequency)) を算出し、この値が大きい特徴語は重要であると判定する。

(3) 文章長

概念検索では、上述したTFとIDFを掛け合わせて重みを算出するTF-IDF法が重み付け方法の主流である。ただし、TFの値については、文章が長いほど値が大きくなる傾向にあるので、文章の長さに応じてTFの値を補正することが多い。例えば、各特徴語のTFの値の平均値を算出し、各特徴語のTFの値をこの平均値で割ることにより、TFの値が文章長に依存しないように補正する。

	出現頻度			
	請求項	発明	検索	...
文献A	5	7	0	
文献B	4	8	24	
文献C	6	5	0	
...				

「請求項」「発明」は多くの文献に出現する ⇒ 重みを低く  
 「検索」は特定の文献に偏って出現する ⇒ 重みを高く

図2-2 単語に対する重み付けの考え方

このような統計的観点から適切な重みを持つ特徴語を抽出し、これらを検索に使用して類似性を判定することによって、人間の感覚に適合した形で類似性を判定し、概念検索精度を高めている。

2-3. 類似性の判定

概念検索では、「ある文章の内容が別の文章の内容と類似している」とはどういうことかについて定義することが必要である。すなわち、「類似性を定量化する計算式」を定義することが必要である。文章間の類似性の判定については、ベクトル空間モデル、確率モデル<sup>1)</sup>、言語モデル<sup>2)</sup>などといった様々な方法が提案されている。しかし、どれも文章に出現する単語の出現傾向をベースとしたモデルであるという点では大差がない。

ここでは、ベクトル空間モデルを例に取って紹介する。ベクトル空間モデルは、一つの文章を、その文章に出現する特徴語を要素とするベクトルとして定式化する。ベクトルの要素の値としては、その特徴語の重みを割り当てる。そして、文章間の類似度を算出する際には、特徴語の出現傾向がどのくらい似ているか、すなわち、その文章に対応するベクトルがどのくらい同じ方向を向いているかに着目する。具体的には、ベクトル間の内積または余弦を算出し類似性を判定する。

ベクトル空間モデルでは、類似度の大小は以下の二つの観点に左右されることになる。

(1) 共通する特徴語の数

二つの文章に共通して現れる特徴語が多い場合、類似度は高くなる。

(2) 重みの高い特徴語の共通性

二つの文章が重みの高い特徴語を共通に持つ場合、類似度は高くなる。

- 1) 確率モデルでは、文書がクエリ (入力文章) に対して適合している (relevant) かどうかを確率的に推定する。ここでは単語が確率変数となる。ベクトル空間モデルに比べ、確率理論に支えられているという利点がある。ベクトル空間モデルと確率モデルは相互に成果を吸収し合い、双方の最新のモデルではほぼ等しい精度を出している。
- 2) クエリを構築する際は、検索して欲しい文書を思い浮かべ「その文書にはこの単語が含まれているはずだ」という目論見のもとで単語を選ぶ。この過程をそのままモデル化したのが言語モデルである。検索対象の各文書について、文書内の単語の確率分布がクエリ内の単語を生成する確率を計算し、この確率値で文書をソートする。単純なモデルながら、従来のベクトル空間モデル、確率モデルと同等の精度を出したことで一躍脚光を浴び、近年盛んに研究されている。

## 2-4. 汎用連想計算エンジンGETA

概念検索のサービスや製品には様々なものがあるが、本稿で紹介する研究事例では、汎用連想計算エンジンGETA (Generic Engine for Transposable Association) を用いている。GETAは、情報処理振興事業協会 (IPA) が実施した「独創的情報技術育成事業」の研究成果であり無料で配布されている。ソースコードやドキュメントが公開されているため、研究用のエンジンとして用いられることが多いが、最近では、書籍検索サイトや通販サイトなどのWebサイトにGETAを採用する例も出てきた。代表的なサイトは、新刊本や大学図書館の蔵書を連想検索するWebcat Plus<sup>3)</sup>である。

## 3. 概念検索の特許検索への適用

特許検索における各作業フローに対し、概念検索を適用する例を示す (図3-1)。

### (1) 本願理解への適用

本願を理解する上で関連する文献を事前に収集する場合に、概念検索を適用できる可能性がある。例えば、馴染みの薄い技術が含まれているような場合に、概念検索を用いて予備的な検索を実施し、当該技術分野における類似文献を収集することにより、技術水準を効率的に調査できるという効果があると期待できる。

### (2) 検索式の作成支援への適用

概念検索で予備的に検索した結果の特許文献に対し統計解析することで、検索式の作成に有効な分類や単語を抽出できる可能性がある。例えば、概念検索により得られた文献集合によく使われる分類や単語を抽出することで、検索式作成に要する作業負担の軽減が期待できる。

### (3) スクリーニングへの適用

類似度の高い順に文献を閲覧することで、文献を効果的にスクリーニングできる可能性がある。例えば、全文検索で作成した文献集合を概念検索の類似度順に

検索結果を並び替えることで、スクリーニングにメリハリをつけることができる。

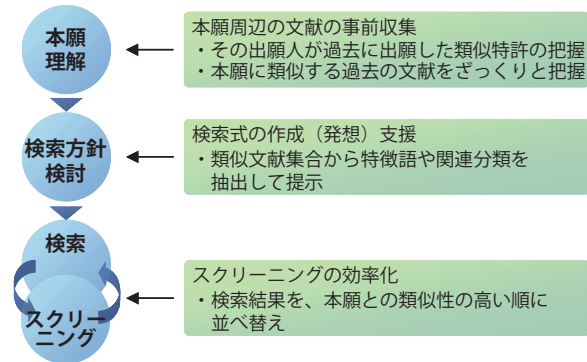


図3-1 概念検索の特許検索への適用イメージ

### (4) 分類付与支援への適用

検索業務だけでなく、分類付与支援へ適用することも考えられる。まだ分類が付与されていない文献を入力とし、既に分類が付与されている文献を対象に概念検索を行い、検索結果上位の文献に多く付与されている分類を付与候補分類として提示することが可能であり、分類付与支援として期待できる (図3-2)。

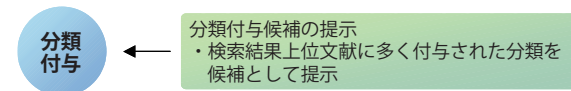


図3-2 概念検索の分類付与業務への適用イメージ

## 4. 概念検索に関する研究事例

### 4-1. 特許検索への適用に向けた課題

概念検索精度の向上に向けた一般的な課題は次の三点である。

- (1) 文献を特徴付ける単語をどのように抽出するか
- (2) 単語の重みをどのように算出するか
- (3) 文献間の類似性をどのように計算するか

さらに、概念検索を特許検索に適用した類似特許検索の観点では、上記三つの課題に加えて次のような課

3) 国立情報学研究所 (NII) が提供するGeNii (ジーニイ): NII学術コンテンツ・ポータルを構成するサービスのひとつである。

題もある。

- (4) 特許文献固有の文書構造をどのように活用するか  
 (5) 特許の属性情報（分類や引用関係など）をどのように活用するか

以下では、上記課題に対してこれまで実施されてきた研究事例について紹介する。また、4-4章で特許を対象とした概念検索技術に関する評価ワークショップを紹介する。

概念検索は、検索対象文献の種類に合わせて工夫を行うことが必要なため、一般的な技術水準を例示することは難しい。これは検索エンジンが計算する類似性の観点が検索対象毎にケースバイケースである、または、検索対象文献毎にみられる特徴がケースバイケースであることが原因である。

#### 4-2. 単語の重み付けに関する研究事例

本章で紹介する事例は、「ある単語の特許文献における出現箇所」に着目したアプローチであり、検索に用いる単語をどのように抽出するか、単語の重みをどのように算出するか、特許文献固有の文書構造どのように活用するか、という三つの観点から検索精度の向上を試みた事例である。

一般に、請求項1または要約から抽出した単語を用いて概念検索を実施すると、他の項目から単語を抽出し概念検索を実施した場合よりも良い検索精度が得られる。そこで、請求項1と要約にはその文献を特徴付ける単語が多く含まれていると考えることができる。しかし、実施例などの詳細な説明に重要な単語が含まれているケースもあり、このようなケースに対しては、やはり本文全体から単語を抽出しないと精度良く検索できない。そこで、本文全体から重みの高い上位30単語を抽出し、請求項1及び要約から抽出された単語とともに検索で使用することで検索漏れを抑制し、ノイズも最小になるように工夫している。

また、単語の重みに対する補正として、上記3箇所（請求項1から抽出した全単語、要約から抽出した全単語、および、本文から抽出した重みの高い上位30単語）のうち何箇所に出現した単語であるか調査する。重要な単語は、文献の全体に亘って出現する傾向にあるため、3箇所全てに出現した単語の重みはそのまま検索に使用し、2箇所しか出現しない単語の重みを80%に削減

表4-1 単語の出現箇所に応じた重み補正

抽出単語	単語出現箇所			単語の重み補正率
	請求項1	要約	全文（重みの高い30語）	
検索	○	○	○	100%
抽出	○	○	×	80%
照合	×	○	×	60%
精度	×	×	○	
文章	○	×	×	

し、1箇所しか出現しない単語の重みは60%に削減し検索する（表4-1）。

このようにして、抽出箇所を限定することにより厳選された単語を抽出し、かつ、それらの単語について出現箇所数で重みを補正することで、概念検索の精度を向上することができる。

#### 4-3. 検索結果表示に関する研究事例

本章で紹介する事例は、「本願とその引用文献の間の関連性」に着目し概念検索結果を並べ替えるアプローチであり、特許の属性情報をどのように活用するかという観点から検索精度の向上を試みた事例である。

一般の概念検索では、発明内容がどんなに類似していても、使われている単語が異なっている場合、その文献は検索結果の上位に出力されない。このような現状の概念検索の限界を補う手掛かりの一つとして、引用文献は有用であると考えられる。そこで本方式では、概念検索結果として出力された上位N件の文献集合を対象とし、引用関係にある文献をクラスタリング（グルーピング）することにより、検索結果上位の文献と引用関係にある文献をまとめて表示する（図4-2）。

図4-2の左側は、クラスタリング前の概念検索結果である。この結果では、個々の文献が類似度の高い順に独立に出力されている。この結果では4件ある正解文献のうちの2件が出力されているが、残りの2件は上位に出力されていない。一方、図4-2の右側は、クラスタリング後の概念検索結果である。検索結果1位にランクされた文献と引用関係にある検索結果中の文献を1つのクラスタにまとめて出力している（図4-2では検索結果に網掛けがしてあるが、1つの網掛け部分が1つのクラスタを形成している。）。この結果、正解文献が4件とも

検索結果（クラスタリング前）

正解	順位	公開番号	発明の名称	出願人
*	1	1994-081652	直噴式ディーゼルエンジンの燃焼室構造	トヨタ自動車株式会社
	2	1996-093480	ディーゼルエンジンの燃焼室	日野自動車工業株式会社
	3	1993-001543	直噴式ディーゼルエンジン	マツダ株式会社
	4	1996-074578	副燃焼室付きディーゼルエンジン	三菱自動車工業株式会社
	5	1995-042562	ディーゼルエンジンの副燃焼室	株式会社いすゞセラムックス研究所
	6	1996-105325	ディーゼルエンジンの燃焼室	日野自動車工業株式会社
	7	1993-098965	直噴式ディーゼルエンジン	日野自動車工業株式会社
	8	1995-317542	筒内直噴式エンジン	トヨタ自動車株式会社
	9	1994-264745	滴流室付きエンジン	三菱自動車工業株式会社
	10	1996-144766	直噴式ディーゼルエンジン	いすゞ自動車株式会社
	11	1993-296045	直噴式ディーゼルエンジン	河野通方
	12	1995-208176	直噴ディーゼルエンジンの燃焼室	トヨタ自動車株式会社
	13	1994-033771	副燃焼室式の火花点火エンジン	株式会社クボタ
	14	1993-010129	新熱2サイクルエンジン	株式会社いすゞセラムックス研究所
	15	1996-200073	直噴式ディーゼルエンジン	アイシン精機株式会社
	16	1994-042351	直噴式ディーゼルエンジン	日野自動車工業株式会社
	17	1995-071255	直噴噴射式ディーゼルエンジン	三菱自動車工業株式会社
	18	1993-296044	直噴式ディーゼルエンジン	河野通方
	19	1995-139352	滴流室構造	三菱自動車工業株式会社
	20	1993-156943	滴流室式ディーゼル機関の燃焼室	日産自動車株式会社
	21	1996-277715	自給二輪車用2サイクル水冷エンジンの燃焼室	ヤマハ発動機株式会社
	22	1993-296046	直噴式ディーゼルエンジン	河野通方
	23	1994-193449	直噴式ディーゼルエンジンの燃焼室	日野自動車工業株式会社
	24	1993-141243	内燃機関	日野自動車工業株式会社
	25	1994-257441	ディーゼルエンジンの燃焼室	株式会社いすゞセラムックス研究所
	26	1996-226330	副室式エンジン	いすゞ自動車株式会社
	27	1995-166808	滴流室式ディーゼルエンジンの燃焼室構造	株式会社豊田自動織機製作所
	28	1994-317154	滴流室付きエンジン	三菱自動車工業株式会社
	29	1994-212973	直噴式ディーゼルエンジン	日野自動車工業株式会社
	30	1993-106441	直噴噴射式ディーゼル機関の燃焼室	株式会社新燃焼システム研究所
	31	1994-221163	直噴式ディーゼルエンジンの燃焼室構造	マツダ株式会社
	32	1994-050151	ディーゼルエンジン	河野通方
	33	1994-241049	副室式エンジン	いすゞ自動車株式会社
*	34	1994-336932	直噴噴射式ディーゼル機関	株式会社新エイシー
	35	1993-296047	直噴式ディーゼルエンジン	河野通方
	36	1995-189701	内燃機関用ピストン	日野自動車工業株式会社
	37	1995-119466	副燃焼室部材	浜セツ株式会社
	38	1993-256136	直噴式エンジンの燃焼室	株式会社いすゞセラムックス研究所
	39	1996-218876	直噴噴射式ディーゼルエンジン	日野自動車工業株式会社
	40	1993-306619	直噴式ディーゼルエンジン	いすゞ自動車株式会社

※正解とは、検索した文献に対し、引用文献として使われた文献である

検索結果（クラスタリング後）

正解	Group ID	順位	公開番号	発明の名称	出願人
*	0	1	1994-081652	直噴式ディーゼルエンジンの燃焼室構造	トヨタ自動車株式会社
	0	22	1993-296046	直噴式ディーゼルエンジン	河野通方
	0	33	1994-241049	副室式エンジン	いすゞ自動車株式会社
	0	34	1994-336932	直噴噴射式ディーゼル機関	株式会社新エイシー
	0	35	1993-296047	直噴式ディーゼルエンジン	河野通方
	0	44	1995-269253	ディーゼルエンジンの燃焼室構造	株式会社いすゞセラムックス研究所
	0	53	1995-019052	副燃焼室を備えたピストン構造	株式会社いすゞセラムックス研究所
	0	109	1993-209524	副室式エンジン	いすゞ自動車株式会社
	0	126	1995-332091	ピストンに動重を持つ副室式エンジン	いすゞ自動車株式会社
	0	141	1996-135448	燃料噴流の衝突位置を制御したディーゼルエンジン	有限会社コイルポーター
	0	161	1994-185365	筒内直噴噴射式ディーゼル機関	トヨタ自動車株式会社
	0	222	1995-259567	直噴式圧縮着火内燃機関	株式会社日本クレンジエンジン研究所
	0	266	1994-307246	副室式エンジン	いすゞ自動車株式会社
	0	306	1995-158450	副室を備えた燃焼室	株式会社いすゞセラムックス研究所
	0	465	1996-144845	直噴式エンジンの構造	株式会社いすゞセラムックス研究所
	0	653	1993-010209	燃焼室を有するピストン	いすゞ自動車株式会社
	0	947	1995-119577	副室付2サイクルディーゼル機関の燃料噴射制御装置	トヨタ自動車株式会社
	1	2	1996-093480	ディーゼルエンジンの燃焼室	日野自動車工業株式会社
	2	3	1993-001543	直噴式ディーゼルエンジン	マツダ株式会社
	3	4	1996-074578	副燃焼室付きディーゼルエンジン	三菱自動車工業株式会社
	4	5	1995-042562	ディーゼルエンジンの副燃焼室	株式会社いすゞセラムックス研究所
	4	570	1995-189848	内燃機関の燃焼方法	学校法人五島育英会
	4	638	1995-054656	エンジンの燃焼室	株式会社いすゞセラムックス研究所
	4	645	1995-293344	直噴式エンジン	株式会社いすゞセラムックス研究所
	4	703	1994-146890	ガスエンジンの制御装置	株式会社いすゞセラムックス研究所
	4	727	1995-042563	ガスエンジンの燃焼室	株式会社いすゞセラムックス研究所
	4	728	1994-159058	副燃焼室を持つガスエンジン	株式会社いすゞセラムックス研究所
	4	777	1994-323211	副室を備えたアルコール改質エンジン	株式会社いすゞセラムックス研究所
	4	795	1995-127454	副室式ガスエンジン	いすゞ自動車株式会社
	4	818	1994-229318	高圧縮比副室式ガスエンジン	株式会社いすゞセラムックス研究所
	4	836	1995-054717	エンジンの燃焼室	株式会社いすゞセラムックス研究所
	4	912	1994-323139	燃焼方式の熱を利用するアルコール改質エンジン	株式会社いすゞセラムックス研究所
	4	949	1994-074047	副室式ガスエンジン	いすゞ自動車株式会社
	6	6	1996-105325	ディーゼル機関の燃焼室	日野自動車工業株式会社
	6	7	1993-098965	直噴式ディーゼルエンジン	日野自動車工業株式会社
	7	8	1995-317542	筒内直噴式エンジン	トヨタ自動車工業株式会社
	8	9	1994-264745	滴流室付きエンジン	三菱自動車工業株式会社
	9	10	1996-144766	直噴式ディーゼルエンジン	いすゞ自動車株式会社
	10	11	1993-296045	直噴式ディーゼルエンジン	河野通方
	11	11	1995-208176	直噴ディーゼルエンジンの燃焼室	トヨタ自動車株式会社

図4-2 概念検索結果に対するクラスタリング実施イメージ

検索結果の上位に集まっている。

概念検索にて単語の出現頻度の分布から計算される類似性に対し、人間の目で見えた内容の類似性（引用関係）を用いて補正することにより、内容が似ている文献が上位に集まってくる事が確認されている。これより、スクリーニング効率の向上が期待できると考える。

4-4. 評価ワークショップの紹介

以上で紹介したようなアプローチの有効性を評価するには、課題と正解からなる大量の評価用データが必要になる。本研究事例では、審査請求された特許とその引用文献を使ったため、容易に評価用データを収集することができたが、一般にこのようなケースは稀である。

そこで米国では、1992年にTRECと呼ばれる検索コンテンツ形式の評価ワークショップが始まった。TRECでは、参加者が同じ課題で検索精度を競い合う。これにより技術が加速度的に進歩し、10年間で検索精度が2倍にも向上した。その過程で大規模な評価用データも整備されてきた。

日本では、1997年より、国立情報学研究所がNTCIR（エンティサイル）という同様の評価ワークショップを開催している。NTCIRの特徴の一つとして、特許を検

索対象にしたタスクを早期から運営していることが挙げられる。このタスクでは、実務と乖離しないために日本知的財産協会等の知財専門家と協力してタスクを設計している。タスクの内容は、先行技術文献調査といった基本的なものに加え、新聞や論文を検索クエリとして特許文献を検索する横断検索、日中英の間の多言語検索、と多岐に渡る。また、検索だけでなく、特許自動分類や特許翻訳に関するタスクも行われている。NTCIRで作成した評価用データや参加システムの詳細は全て公開されている。

5. 概念検索精度に対する今後の課題と解決方針

4章では、特許を対象とした概念検索精度を向上させるための研究事例について紹介した。今後、概念検索精度をさらに向上させるために必要であると考えられる研究アプローチの一例について本章で述べる。

(1) 技術分野別の傾向を踏まえたアプローチの組合せの最適化

概念検索を特許検索に適用する場合は、技術分野によって有効性が異なっている可能性が高いと考えられる。ある技術分野に対しどのアプローチがどの程度有効であるかに応じて、その技術分野に最適なアプロー

チを組合せるという研究が必要である。これにより、検索精度をさらに向上させることが期待できる。

## (2) ユーザの視点や意図を反映させたインタラクティブな概念検索

これまで研究されてきた検索方式は、計算機が全自動で検索する方式が多い。しかし、実際には、ユーザとのインタラクションを通じて、ユーザの視点や意図が検索に反映されるような仕掛けを持つことが、精度向上につながると考える。例えば、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、今どこに着目しているか、検索対象となる分類や年範囲はどのあたりか等をユーザが指定し、検索結果を見ながら段階的に検索を繰り返すことができる検索の仕組みが必要であると考え。これにより、所望の文献をより早く見つけることが期待できる。

## 6. おわりに

本稿では、概念検索の仕組みと、特許検索への適用例、検索精度向上のための研究事例について述べた。概念検索を特許検索に適用するための研究は盛んであり、今後は特許検索の有効な手段の一つになると考えることができる。また、概念検索の結果得られる類似特許文献に付与されている分類を統計的に解析し、新規文献に対する分類付与を支援するなど、検索以外の場面でも利用できる可能性がある。概念検索技術は、研究の段階から実用化の段階に移行しつつあると考えられる。

本稿が概念検索の適用に対する、より活発な議論を喚起し、更に忌憚のないご意見を戴ければ幸いである。

### 【参考文献】

- [1] 岩山真、今一修「汎用連想計算エンジンGETAを用いた特許連想検索システム」Japio 2007 YEAR BOOK
- [2] 間瀬久雄「特許文章構成と分類情報を用いた類似特許検索方式の精度評価」Japio 2007 YEAR BOOK
- [3] N. Kando, Overview of the Sixth NTCIR Workshop, Proceedings of NTCIR-6 Workshop Meeting, <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/NTCIR/NTCIR6-PREFACE-NUM.pdf>
- [4] A. Fujii, M. Iwayama and N. Kando, Overview of the Patent Retrieval Task at the NTCIR-6 Workshop, <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/NTCIR/78.pdf>