

## 特許拒絶理由通知書を利用した複合語類似の検証の一考察

柳堀恭子<sup>†1</sup> 津田和彦<sup>†1</sup>

本論では、特許出願に対して特許法 29 条 1 項の新規性を拒絶理由とした特許審査官が発行する拒絶理由通知書を対象とし、本願と引用文献の間で対比が明示されている名詞および複合名詞を抽出する。その語の類似が 1)文字列照合、2)同義語・類義語、3)シソーラス（上位・下位概念、全体・部分）のいずれでも抽出できない語に対して構文解析の手法による類似性の向上を検証した。さらに、この結果を特許文における類似語として意味を拡張していくことにより特許用知識辞書を構築することが可能となり、出願における先行調査や技術調査に役立つことを示した。

### A Study of verification of similar compound noun using the patent notification of reasons for refusal

KYOKO YANAGIHORI<sup>†1</sup> KAZUHIKO TSUDA<sup>†1</sup>

In this study, we extracted clearly oppositional nouns and compound nouns from the application itself and the citations, issued by patent examiners as a reason for refusing patent applications based on the new rule of patent act 29, article 1, applicable to patent application issues. We examined to what extent they are similar by using the parsing method because we were unable to extract neither 1) word string matches, nor 2) synonyms, nor 3) thesaurus (high or low order concept, whole or part). In addition, we were able to construct a patents specialized dictionary, and were able to show that it will be useful for prior and technical researches before submitting a patent application by expanding these results with the meanings of synonyms in the patent texts.

#### 1. はじめに

特許出願者は、出願の前に既に類似特許が存在していないかどうかを調査するために多くの時間と資金をかける。しかし、膨大に存在するすべての特許を漏れなく調査することは困難である。出願前に完全に調査したつもりが、最終的に存在していた類似特許を見逃したことで拒絶査定を受け、特許を取得できないこともある。そうなれば、出願者にとっての損失は非常に大きいものとなる。

このような先行する特許の見逃してしまう要因に、特許文中に多々出現する複合語がある。特許は、新規性を問われるため世に存在しない物について記述することが多い。そのため発明物を示す固有名詞が存在しないため、その機能や性能を示す単語を複雑に組み合わせられて作られている場合が多い。特許文中の複合語には、このような特徴があるため、検索キーワードの設定が困難であるという特徴がある。

本研究では、先行研究調査の一助となる材料を提供するために、出願に対する拒絶理由通知書を利用した複合語の

類似辞書を作成することを最終目的として取り組んでいる。先行研究を調査する際、検索の範囲を拡張するためには、検索の際のキーワードが多いほど、再現率の高い検索が可能となる。けれども、単純に検索キーワードを増やしても検索効率の向上は僅かである。

特許として成立するか否かは、拒絶査定により決定する。本研究は、特許成立を拒む拒絶理由通知書からダイレクトに検索キーワードとなる複合語を抽出するものであり、有効な複合語を抽出できる可能性がある。さらには拒絶査定において、拒絶される審査の過程を解き明かすための情報の提供を示唆できれば、より有効となるであろう。

#### 2. 特許審査のフロー

特許審査の流れを図 1 に示す。

出願者が特許出願してからのち、1 年 6 ヶ月後に、その出願内容が公開特許公報という形で公開される。ここまでの審査では、出願内容の審査は行われず、形式的、手続き的な要件を満たしているかどうかの方式的な審査のみが実施される。

特許審査は自動的には行われず、出願者は、出願後 3 年

<sup>†1</sup> 筑波大学大学院ビジネス科学研究科  
University of Tsukuba Graduate School of Business Science

以内に、当該出願を審査してもらうための審査請求を行わなければならない。この間に、出願者は先行技術調査を進め、実際に審査請求するか否かを再判断することもできる。あらためて審査請求が行われた出願は、実体審査の段階へと進んでいく。ここで審査された出願内容が、特許審査官により審査され、特許査定、あるいは拒絶査定との判断がなされる。

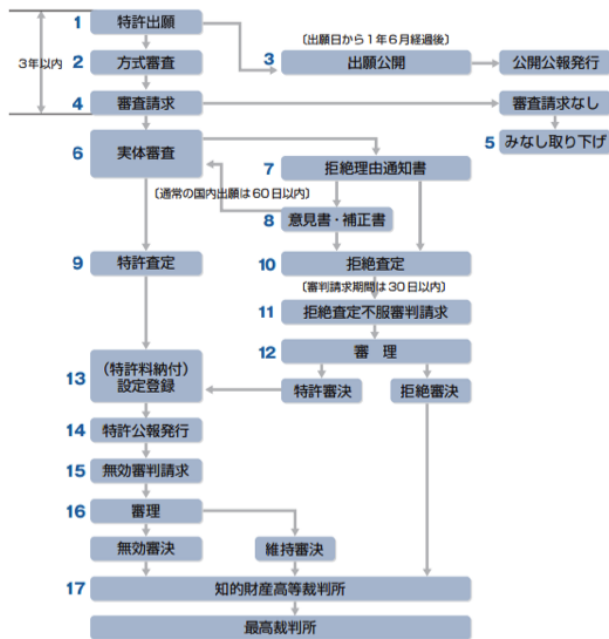


図1 特許審査の流れ (特許庁 HP より)  
 Figure1 Flow of patent examination

近年、特許検索の精度は向上してきているが、図2に示す通り、特許出願における拒絶査定割合は依然として少なくない。

実績	2006年	2007年	2008年	2009年	2010年	前年比
特許査定件数	129,071	146,383	159,961	178,227	205,652	115%
拒絶査定件数	129,400	147,678	154,163	171,396	164,639	96%
(うち応答なし拒絶査定件数)	68,879	78,246	85,443	105,004	100,951	96%
FA後取下げ・放棄件数	7,915	5,567	4,779	5,169	4,600	89%
特許査定率	48.5%	48.9%	50.2%	50.2%	54.9%	-
拒絶査定率	51.5%	51.1%	49.8%	49.8%	45.1%	-

図2 最終処分実績の推移  
 (特許行政年次報告書 2011年版 特許庁)  
 Figure 2 Changes in final disposal performance

### 3. 複合語の利用

拒絶理由通知書より抽出された複合語（本研究では複合名詞をメインに抽出した）は、表層上は異なった表現で書かれているが、意味が同じというような単語をどのようにして類似と判定するか、その手法について検証を行った。

特許出願の文書中、特に発明の肝となる請求項の部分に

は、多くの複合語が散見する。これは、特許審査では発明の新規性と進歩性に焦点が当てられるために、それまで存在しないような新しい造語の形で新規性と進歩性を表現する複合語を使用することが多いということが理由の1つと考えられる。

本研究で、この点に着目し、複合語の利用可能性について考察した。

最終的には複合語を検索語とする類似特許検索のシステムへの応用を目的に、まず、複合語で検索する際の問題の有無と、複合語利用可能性に関して検討した。

#### 3.1 複合語利用の課題

ある文書と他の文書を類似と判定する方法として、文書の中に出現する単語の出現頻度や文書中の出現単語の重要度などに依るものが一般的に知られている。例えば、一文の中で“情報表示手段”のような複合語が頻出する割合は、“情報”、“表示”、“手段”といった複合語を構成する名詞単語が頻出する割合に比べて少ない。このような解析には形態素解析[a]が使われる。形態素解析した結果、“手段”、“装置”、“特徴”など特許文書中に特に頻出するような名詞が多く抽出される。それらが頻出したからといって比較する文書同士が類似とはならないが、機械的に抽出された形態素[b]をもとに類似度が計算されると、比較する文書同士の多くが類似と判断される場合が多い。

一方、文書の内容を的確に現す複合語を使って類似文書を探すとすると、複合語自体の出現率が少ないため、類似度は低い値となる。その結果、比較する文書同士が類似しないという結果になり得るという課題がある。

また、上記の事例である“情報表示手段”と“データ提示手段”のように同義であるが、複合語を構成する単語が変われば検出漏れするという課題もある。特許においては、複合語を構成する単語数が多いため、構成する単語の順序が異なるような複合語も見受けられる。

#### 3.2 複合語利用の可能性

複合語利用については課題もあるが、一方の形態素解析を用いた類似文書検索についても課題がある。

3.1 で述べたとおり、形態素解析では、特許文書のような特徴的な名詞が多く出現するような文書群においては、ほとんどの文書で類似と判定されてしまう。つまり、本当に類似である文書以外の多くも類似とされてしまうことから、類似文書検索をする場合、不必要な文書についても目を通さねばならず、通常の類似文書検索で使われるような形態素解析の技術は、特許文書検索のような精密な検索には適切ではないといえる。

a) 文法の知識（文法のルール集まり）や辞書（品詞等の情報付きの単語リスト）を情報源として用い、自然言語で書かれた文を形態素に分割し、それぞれの品詞に区切ること  
 b) 言語学用語で意味を持つ最小の単位のこと

複合語を用いた類似文書検索では、文書と他の文書群との類似度をみたときに、再現率は低い、適合率は形態素解析での類似文書検索より高いという特徴がある。さらに形態素解析の類似文書検索で問題となる誤って類似と判定される文書の多さ、つまりノイズは低い。

しかし、漏れのない検索が重要となる特許において、先行技術調査では再現率の高さが重要となる。そのため、再現率を高めるための工夫が必要である。

出現する複合語と同義あるいは類似と定義される複合語を見つけることができれば、再現率を高めることができる可能性がある。

## 4. 拒絶理由通知書からの辞書作成手段

### 4.1 拒絶理由通知書から複合名詞の抽出

特許電子図書館(IPDL)[c]より、特許出願状況の審査書類情報照会を行い、公開特許公報の公開番号 2003-000001 から 2003-007500 の 7,500 件の審査状況を概観した。審査書類情報照会は、特許審査の状況を示すものである。そのうち、既に拒絶査定を受けた出願の拒絶理由通知書のみを抽出した。

その拒絶理由通知書の中で、特許審査官が拒絶理由として、特許法第 29 条第 1 項[d]の新規性を有しない発明に対して、その要件を満たさないことを理由として拒絶を受けたもののみを複合語抽出対象の文書とした。

もちろん拒絶理由には、他にも特許法第 36 条の発明の明確性違反や第 29 条第 2 項の発明の進歩性の要件を満たしていないなどがあるが、本研究では、拒絶されるべき対比部分が明確に示されていることから、第 29 条 1 項での拒絶理由となっているもののみを抽出文書対象とした。審査官によっては、明確に対比部分を示していない場合もあるので、その場合は抽出文書対象外とした。

### 4.2 複合語の比較方法

4.1 で抽出した拒絶理由通知書のうち、さらに審査官が下記の図 3 の例で示すような書き方をしている部分に注目した。ここで抽出された複合語の対比を行っていく。

引用文献1の高分子ゲル、光検出部及びデータ処理部は、本願発明のプラスチック基板、損傷検出手段及び検出結果解析手段にそれぞれ相当する。そして、引用文献1のDNA鎖切断現象の検知部は、本願発明のプロープに相当する。

c) <http://www.ipdl.inpit.go.jp/homepg.ipdl>

d) 第 29 条 産業上利用することができる発明をした者は、次に掲げる発明を除き、その発明について特許を受けることができる。

1. 特許出願前に日本国内又は外国において公然知られた発明
2. 特許出願前に日本国内又は外国において公然実施された発明
3. 特許出願前に日本国内又は外国において、頒布された刊行物に記載された発明又は電気通信回線を通じて公衆に利用可能となった発明

図 3 審査官が示す対応例

Figure 3 Corresponding cases the examiner indicated

図 3 では、出願された本願発明において、引用文献に書かれた“高分子ゲル”と本願発明で書かれた“プラスチック基板”が同等ということが明記されている。IPDL を利用し、検索条件を 2003 年 1 月 1 日から 2003 年 12 月 31 日に公報が刊行された出願のうち、請求の範囲に“プラスチック基板”のみのキーワードをもつ特許公開公報を検索すると、155 件がヒットする。一方、“高分子ゲル”のみのキーワードだと 47 件がヒットする。“高分子ゲル”∩“プラスチック基板”は 0 件なので、本願で書かれた“高分子ゲル”のみのキーワードで検索した結果 155 件よりも 2 つのキーワードで検索した結果 202 件へ検索範囲が広がる。ただし、審査官が同等とするとした複合語でも、その文書の中においてのみ同義であると読み取れるような場合がある。例えば、“第二部材”と“取付部材”が抽出されたとして、それぞれを形態素解析した結果、“第二”と“取付”を比較するが、これらは当該抽出元文書の中では同義になるかもしれないが、一般的に同義であるとは考えられない。このように、すべての抽出複合語が類似と見なされるわけではない。そのため、それぞれの複合語を比較していく方法を検討した。

#### (1) 文字列照合による判定

ここでは、複合語を形態素解析し（この場合は複合名詞から、名詞に分割）、同一の形態素を持つかどうかで類似を判定する。

例えば、“処理割当装置”と“処理管理装置”の 2 つが抽出された場合、“処理割当装置” = “処理” + “割当” + “装置”、“処理管理装置” = “処理” + “管理” + “装置”と、それぞれ 3 つの形態素に分けることができる。この例では、3 つの形態素のうち“処理”と“装置”の 2 つの形態素で一致しているので文字列照合の結果では一致とする。ただし、同一の形態素が存在したからといって、複合語すべての文字列が一致しているとは判定できない。○○装置と××装置のような“装置”という特許文書の中では極めてありふれている語を利用した語において、○○と装置、あるいは××と装置で 2 つの形態素から成る語でそのうちの 1 つの形態素である“装置”がそれぞれの複合語内で一致したからといって、○○装置と××装置で複合語の文字列照合の結果を一致とはしない。これをルール化したものが表 1 である。

以下の表 1 に複合語を形態素解析したときの形態素数による文字列照合の一致・不一致のルールを定めた。本願と引用文献双方から抽出した複合語を形態素解析した形態素数を  $n$  および  $m$  とした。

表1 形態素数による一致判定ルール

Table 1 Rules of the match determination by the number of morphemes

n=mの場合
n=m=1 n個の一致で類似
n=m=2 n-1個の一致で類似(但し、内容も検討)
n=m≥3 n-1個の一致で類似(但し、内容も検討)
n<mの場合
m-n=1 n=2のときはn個の一致で類似 n>2のときはn-1個の一致で類似(但し、内容も検討)
m-n≥2 n=2のときはn個の一致で類似 n>2のときはn-1個の一致で類似(但し、内容も検討)

(2) 同義語・類義語による判定

(1)で類似判定できなかった複合語群に対して複合語の一致判定を行う。

ここでの同義語・類義語は、全体-部分の関係あるいは上位概念-下位概念のものとして定義する。

(1)のルールを用いて、形態素を、「類語大辞典」[e]、「類語新辞典」[f]、「日本国語大辞典」[g]などで類義語と定義されているものをここでの同義語・類義語として各形態素の判定を行った。

(3) シソーラスによる判定

ここでのシソーラス[h]とは、(2)に近い物であるが、より抽象的な表現での概念を含むものとした。

(2)の同義語・類義語判定との違いとして、言葉的分類と意味的分類に分けた。

表2では、その考え方の例をあげておく。

(2)の「車」と「牽引車」は、上位概念-下位概念の関係となる。「公開制度」と「情報公開制度」は、全体-部分の関係となる。それぞれの中で「車」および「公開/制度」と同じ文字列(形態素)をもっている。(2)の場合は、形態素の文字列での同義・類義で複合語全体としての類義の判定をするので、言葉的分類では同義・類義と判定する。

一方、(3)で定義するシソーラスは、文字列の一致ではなく、形態素の意味を考える。「柑橘/類」の文字列と「ミカン」の文字列は一致しないけれど、意味的分類により類似とする。このように、「データ/獲得/手段」と「情報/入手/方法」のように形態素解析をした結果、同じ形態素文字列は含まないが、類似と判断できるものをシソーラスとした。

表2 本論での同義語・類義語とシソーラスの考え方

Table 2 Concept of synonyms and thesaurus in this study

e) 「類語大辞典」(講談社 2002)  
f) 「類語新辞典」(三省堂 2005)  
g) 「日本国語大辞典」(小学館 2001)  
h) 通常シソーラスとは、単語の上位 / 下位関係、部分 / 全体関係、同義関係、類義関係などによって単語を分類し、体系づけた辞書を指す

例:		
	(2)同義語・類義語	(3)シソーラス
	<言葉的分類>	<意味的分類>
上位概念	車	果物、柑橘類
下位概念	牽引車	バナナ、みかん
全体	公開制度	
部分	情報公開制度	

(4) 構文係り受け解析を用いた判定方法

上記(1)から(3)の複合語類似の判定方法を用いても、類似と判定できない場合の方法を提案する。

本研究では、複合語を含む文全体(例えば請求項1の全文)を構文解析[i]し、複合語の前後に出現する他の品詞(複合名詞、名詞、動詞など)との関わりから、比較する複合語の類似を判定する。

複合語に直接係る形容詞が同じ(類似)である場合、複合語が最終的に係る動詞が同じ(類似)である場合を精査して、最終的に目的の複合語が類似であるかどうかを判断する手法である。

図4に判定方法の流れを示す。

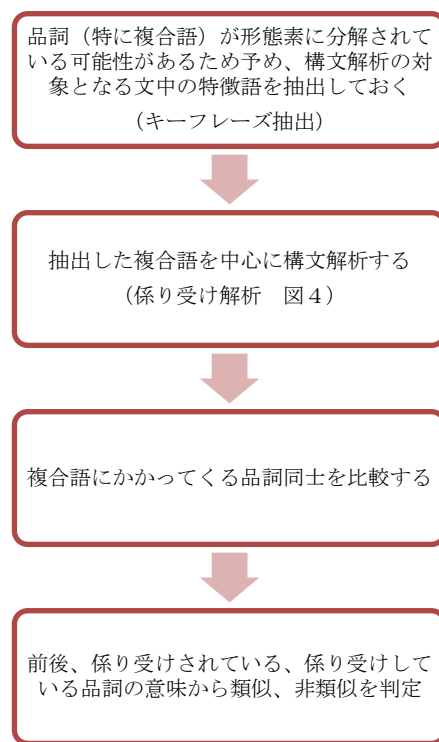


図4 判定方法の流れ

Figure 4 Flow diagram of a judgment method

構文解析には、Yahoo! JAPAN が提供している WEB API のうち、テキスト解析 API を利用した[j]。今回は、この中

i) 単語や字句で構成される文を、定義された文法に従って解釈し、文の構造を明確にすること  
j) <http://developer.yahoo.co.jp/webapi/jlp/> (2013年10月22日参照)

で以下の API を使い、係り受け解析を行った。解析例を図 5 に示す。

- 形態素解析
- 係り受け解析
- キーフレーズ（特徴語）抽出

	chunk.id	dependency	surface	reading	baseform	pos
38	12	14	識別	しきべつ	識別	名詞
39	12	14	情報	じょうほう	情報	名詞
40	12	14	と	と	と	助詞
41	13	14	前記	ぜんき	前記	名詞
42	13	14	記録	きろく	記録	名詞
43	13	14	媒体	ばいたい	媒体	名詞
44	13	14	に	に	に	助詞
45	14	15	対応	たいおう	対応	名詞
46	14	15	する	す	す	助動詞
47	15	16	タイトル	たいとる	タイトル	名詞
48	15	16	と	と	と	助詞
49	15	16	を	を	を	助詞
50	16	25	対応	たいおう	対応	名詞
51	16	25	付け	づけ	付け	接尾辞
52	16	25	て	て	て	助動詞
53	16	25	、	、	、	特殊
54	17	18	特定	とくてい	特定	名詞
55	17	18	の	の	の	助詞
56	18	19	記録	きろく	記録	名詞
57	18	19	手段	しゅだん	手段	接尾辞

元の文書

「記録媒体に記録された画像データを管理する管理プログラムであって、コンピュータを、前記記録媒体に記録されている画像データと前記記録媒体の識別情報と前記記録媒体に対応するタイトルとを対応付けて、特定の記録手段に記録させる手段、前記特定の記録手段に記録されている前記画像データに基づく画像を表示手段に表示させる際に、前記画像データに対応付けられた前記識別情報と前記タイトルとを前記表示手段に表示させる手段、として機能させることを特徴とする管理プログラム。」

図 5 文の係り受け解析  
 Figure 5 Dependency parsing

まず、抽出した複合語の chunk.id から dependency となる chunk.id を追っていき、シンプルな一文を作る。

抽出した複合語は元の文書にある下線の「識別情報」である。この「識別情報」の chunk.id にかかっている dependency から係り受けをたどっていくと、以下のように、多少文章に違和感があるものの、対象となる複合語の前後の係り受けが明確になり、比較が容易な文が得られることがわかる。

前記記録媒体の  
識別情報と  
 対応する

タイトルとを  
 対応付けて、  
 基づく  
 画像を  
 表示させる  
 際に、  
 管理プログラム。

同様に引用文献から抽出した複合語に対しても同じ作業を行い、比較したい複合語にかかってくる、複合名詞、名詞、動詞と比較することで、複合語の類似判定を行った。

## 5. 評価結果

まず、検証対象の 7,500 本の公開特許公報の審査情報照会のうち、特許法第 29 条 1 項が拒絶理由の拒絶理由通知書が出ているもので、特許審査官が複合語の対比を明確にしている公報のみを抽出した。

その中で対比されている複合語を抽出し、複合語の対比表を作ったところ、453 ペアの複合語対比表が得られた。

この複合語対比に対して、4.2 節で示した複合語の比較方法の手順(1)文字列照合による判定、(2)同義語・類義語による判定、(3)シソーラスによる判定、(4)構文係り受け解析による判定にて複合語同士のペアが類似するか否かを検証した。

表 3 対比複合語の類似判定

Table 3 Similarity determination of comparison compound

	一致数	一致率 (%)
(1)文字列照合	179	39.51
(2)同義語・類義語	221	48.79
(3)シソーラス	311	68.65
(4)構文解析	366	80.80

※一致数、一致率とも累積数

その結果を表 3 に示す。(1)の文字列照合では、約 40% のヒット率となる。文字列照合とは、現在利用可能なキーワード検索システムで採用されているアルゴリズムである。即ち、通常の検索システムでは、40% の検出率となる。

これに(2)同義語・類義語や(3)シソーラスを組み合わせ、単語の意味的な拡張を行うと、70% 近い精度にまで向上してきた。さらには(4)構文解析を用いて対象複合語の前後の単語まで考慮すると、約 80% の割合で対比される複合語の類似が証明できた。

したがって、このような複合語の対応表から意味的類似の側面をもつ類義語拡張辞書を構築できる可能性が示された。さらに、そのことにより特許出願前の先行技術調査や

技術動向調査などにおいて、検索をする場合に1つのキーワードを用いるよりも、より意味の近い(類似の)複数の複合語をキーワードとした検索に利用することも考えられる。

## 6. おわりに

5.の結果を受け、拒絶理由通知書を利用した複合語抽出に一定の結果を示すことができたと考え、同時に考慮すべき点もいくつか発見した。これらへの対応は今後の課題である。

まず、今回調査した複合語の中には、“第二部材”、“第一金属層”など、当該の文献内のみ固有である複合語があった。今回は、上記のように、“第一”、“第二”など、その文書中のみ示されるような序数詞も含めて検討した。例えば、“筒状プラグ”と対応する“第二継手部材”は(1)の文字列照合判と(2)の同義語・類義語判定では類似と判定しなかったが、“プラグ”が“継手部材”の一種であるため、(3)シソーラス判定(表2で示したような意味的分類)にあたるとして、類似と判定した。しかし、この例では、“プラグ”、“継手部材”に関連性があるために類似判定できたが、シソーラス判定でも類似と判定できないような名詞であった場合は、(4)の構文係り受け解析の判定を試すこととなる。その場合、係り受けの前後判断から、最終的に複合語が類似となる可能性が高い。

しかし、類似と判断できたとしても、この複合語が出現した分野すべてにまたがる検索に使えて、有効な複合語であるとは言い切れない。そのため、序数詞を含む複合語については、(1)~(3)の判断ができない場合は、対象から外すことにするか、あるいは、あらかじめ類似判断から外しておくなどの措置を講ずる必要がある。

次に、出願された本願からの複合語抽出は請求項からのものであるが、引用文献からの複合語抽出については、対象複合語が書かれている箇所が指定されている場合がほとんどであるが、請求項に限らず、実施例やその他の箇所に書かれている場合がある。

引用文献の中には、“支持金具 30”、“ヒータ水槽 130”など、複合語の最後に数字が書かれているものがあり、これらは、特許文書中の請求項以外の実施例などに記載される場合の図の番号となるため、これら数字ははじめから除外して考える必要があるなど、複合語抽出のルールも細かく定めなければならない。

さらに、今回の研究では分野を特定せずに特許公開公報の番号順に公報を抽出していった。そのため、発明対象分野がIPC分類でみても多岐にわたっている。

類義語拡張辞書を作っていくにあたって、あらかじめIPC分類などによる分野別に作成することにより精密な辞書を作ることが可能である。特定する分野によっては拒絶

率が高い分野と低い分野などがあるため、まず分野を特定して調査すべきである。

最後に、本手法を適用した実際の検索精度の向上について実験による評価を行う予定である。

## 参考文献

- 1) Uchiyama, K. Aihara, S. Ishizaki, S.: Identifying semantic relations in Japanese compound nouns for patent documents analysis, in: Proceedings of the 3rd international conference on Large-scale knowledge resources: construction and application, Springer-Verlag, Tokyo, Japan, pp. 75-81(2008).
- 2) 寺田昭, 吉田稔, 中川裕志: 同義語辞書作成支援システム, 自然言語処理, vol.15, No.2, pp39 - 58(2008).
- 3) Larkey, L.S.: A patent search and classification system, ACM DL, pp. 179-187(1999).
- 4) Tseng, Y.-H. Lin, C.-J. Lin, Y.-I.: Text mining techniques for patent analysis, Information Processing & Management, vol.43, pp1216-1247, (2007).
- 5) 佐藤一誠, 中川裕志: 係り受け関係を考慮したテキストマイニングのための半構造マイニング手法の提案, DBSJ Letters, vol.5, No.2, (2006).
- 6) 橋本泰一, 藤井敦: 特許文書のための形態素解析辞書の構築, 言語処理学会第18回年次大会発表論文集, pp789-792(2012).
- 7) 滝川諒, 後藤智範: 特許抄録に出現する多字種複合語に対する字種に基づく解析 part. 1-多字種複合語の抽出と構成字種の解析, 情報処理学会研究報告. 自然言語処理研究会報告 2011, No.2, pp1-15(2011).
- 8) 藤井敦: 特許情報を用いた辞典検索システム, 情報処理学会研究報告, pp9-15(2008).
- 9) 特許庁, <http://www.jpo.go.jp/indexj.htm>