

# 特許情報検索の現状と今後の可能性

特許およびオープンな非特許情報を活用した特許分析手法、テキストマイニング技術の活用

The current status and future possibilities of intellectual property information search

一般社団法人日本知的財産協会 情報検索委員会委員長

戸田 敬一

オムロン株式会社入社後、開発部門を経て2002年より知財部門にて知財全般を担当。2015年から当委員会に所属。2017年より現職。

✉ Keiichi.toda@omron.com

## 1 はじめに

ここ数年で、知財を取り巻く環境が大きく変化しつつある。特に、知財情報の活用という観点では、IP ランドスケープというキーワードがきっかけとなり、企業の経営層は知財活用の有用性に対する期待を高めている。このような知財に対する期待に応えるには、知財の専門知識だけでは到底対応できず、多岐にわたる知識の引き出しをより多く持つことが求められる。しかし、一企業、一組織でこのような優秀な人材を確保することは困難であり、企業間や組織間における協業体制を整えることが必要不可欠となる。

日本知的財産協会（JIPA）情報検索委員会では、このような知財環境の変化や企業が求める課題に逸早く対応し、多くの企業から多様かつ優秀な人材が集まる利点を最大限に活用しつつ、個社では困難な課題の解決に向けた研究を行っている。具体的には、主に次の4つの観点からテーマの選定を行っている。①特許情報の活用、②新技術の活用、③経営・事業への貢献、④知財担当の必要スキル

本稿では、その中から以下2つの研究テーマについて、成果の一部を紹介する。

- ・特許およびオープンな非特許情報を活用した特許分析手法
- ・テキストマイニング技術の活用

## 2 特許およびオープンな非特許情報を活用した特許分析手法

IP ランドスケープでは、経営・事業への貢献が重要とされるが、非特許情報を組み込んだ提言であっても、提示内容が知財を主体にしたものや、知財専門用語を多用する内容では、受け手の理解を得ることは難しい。そこで、本稿ではこれまであまり特許分析に活用されていない、もしくは、誰もが身近に感じる非特許情報を活用した特許分析の手法と、これまでとは異なる、特許に馴染みのない部門への分析結果の提供について検討した。

### 2.1 プレスリリースを活用した事例

プレスリリースは、特許情報のように「情報の母集団」として分析するという視点ではあまり利用されていない。しかし、分析ツールの進化が目覚ましく、プレスリリースに「母集団」としての分析視点を入れることが現在では可能である。そこで、プレスリリースを利用して企業動向が窺えるかを考察した。本事例では、分析対象企業の注力領域に着目して探索を行った。また、特定期間における時系列変化をみることで、今後の注力領域がどこかについても検討した。

#### (1) 特定企業の注力領域探索事例

##### 1) 分析手順

今回は無料でプレスリリース情報の取得ができるポータルサイト JPubb（ジェイパブ）<sup>1)</sup> を使用した。JPubb では企業・官公庁別、業種別、本社所在地別、上場市場別で横断的にプレスリリースを検索可能であ



まうような出願も、このプレスリリースの分析から得られたキーワードを考慮した特許分析を行うことで、異なる視点からの分析が可能と考えられる。実際、図3に現れた「飼料・工場」のプレスリリース内容を確認すると、飼料原料製造工場の衛生管理の強化という内容であった。こういった体制変化も企業の意志が反映されるので、少ない特許情報と組み合わせることで、企業動向が窺える可能性がある。また、通常であればプレスリリース公開より早い段階で関係する特許を出願するが、出願公開までのタイムラグを考慮すると、タイミングによっては、プレスリリースで事前に予兆を拾うことも可能であろう。

特徴語の変化を時系列で見る事で、単一企業の母集団であればその企業の変化をキャッチでき、競合企業や業界全体の母集団であれば、ある特定領域に対する企業毎の取り組み時期の違いや、注力領域の違い、業界全体のトレンド等を確認するなどの応用も可能である。

### (3) 分析結果からの考察

目的に応じて、どのようなプレスリリースを母集団とするかを選択することが重要である。また、同じテキストマイニングを使った分析でも、プレスリリースのどの部分を分析対象とするかにより分析結果が左右される。今回の2つの事例は、プレスリリースの「見出し」がそのリリースで最も言いたいことを表していると考え、見出しの部分を分析対象とした。

このように、プレスリリース情報を纏まった集合として分析することで、特許情報とは違った気づきや予兆を発見できる場合がある。その観点から、競合企業や注目分野の技術の深堀に繋げることや、プレスリリースと特許情報を絡めた分析により、研究開発部門以外へも新たな提案が見込まれるのではないだろうか。特に、特許に馴染みのないマーケティング部門や営業部門など技術系以外の部門に対しては、身近な情報であるプレスリリースによる分析を提供することで円滑なコミュニケーションが期待されるであろう。

## 2.2 求人情報を活用した事例

転職情報サイトに掲載されている求人情報は、主に即戦力となりうる人材確保を目的としたものである。従って、転職情報はその企業の人的リソースへの投資意欲と

捉えることができ、直近で注力しようとしている未来予測的な事業動向が得られる可能性が高い。また、転職情報を掲載している企業は、求人掲載のために相応の費用を費やしているため、投資情報としての確度もある程度担保されている。

そこで、本事例では、「ブロックチェーン分野における最新事業動向の取得」を題材として分析を試みた。

### (1) 非特許情報による分析手順

まず、転職情報サイトDおよびRで「ブロックチェーン」をキーワード項目で検索し、ブロックチェーン分野の求人情報を収集した。転職情報サイトDでキーワード項目に「ブロックチェーン」を含む求人を検索した結果、該当求人数は114件となった(2018年11月16日検索実行)。また、転職情報サイトRで、キーワード項目に「ブロックチェーン」を含む求人の検索では、該当求人数は37件となった(2018年11月16日検索実行)。ヒットした求人において事業分野が明記されているものについて確認すると、表2のように転職情報サイトDではフィンテック仮想通貨に次いで、電力需給管理およびドローンに関する求人情報が多く掲載されていた。転職情報サイトRでも表3のようにフィンテック仮想通貨に次いで電力需給管理に関する求人情報が多く掲載されていた。ブロックチェーン分野においてフィンテックや仮想通貨に関連する事業は活用拡大が既に加速している分野であるが、電力需給管理およびドローンについては今後、事業拡大が進む可能性がある分野と推察した。

表1 転職情報サイトDにおける「ブロックチェーン」関連求人の業種ランキング

事業分野	業種	件数
明記有	フィンテック仮想通貨	32
	電力需給管理	7
	ドローン	6
	医療	1
不明確	Web/スマホアプリ開発	12
	システムインテグレーター	33
	コンサル	18
	その他	5
	合計	114

表2 転職情報サイトRにおける「ブロックチェーン」関連求人業種ランキング

事業分野	業種	件数
明記有	フィンテック仮想通貨	3
	電力需給管理	15
不明確	Web/スマホアプリ開発	5
	システムインテグレーター	12
	コンサル	1
	その他	1
合計		37

(2) 特許情報による分析手順

ここまでの結果を踏まえて、電力需給管理については「電力需給管理におけるブロックチェーン技術」を、ドローンについては「ドローン制御におけるブロックチェーン技術」として特許分析を行った。

本稿では PatentSQUARE (パナソニック) を使用して、「電力需給管理におけるブロックチェーン技術」については電力需給管理の観点(特許分類およびキーワード)とブロックチェーンの観点(特許分類およびキーワード)を掛け合わせた検索式でヒットした集合(日本: 183 件(図4および表3参照)、世界: 1,497 件(図5および表4参照))を母集団とした。世界についてはPCT 特許、欧州特許、米国特許、中国特許を対象に検索した。

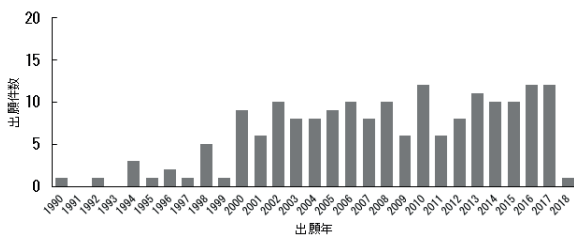


図4 電力需給管理におけるブロックチェーン技術 (日本) 出願件数推移

表3 電力需給管理におけるブロックチェーン技術 (日本) 出願人ランキング (上位9位まで)

出願人名 (名寄せ)	件数
国内電機メーカー T社グループ	22
国内電機メーカー S社グループ	18
国内電機メーカー P社グループ	10
国内電機メーカー F社グループ	8
国内電機メーカー H社グループ	6
国内電機メーカー N社グループ	5
国内電機メーカー M社	4
国内自動車メーカー H社	4
海外再生エネルギーベンチャー T社	4

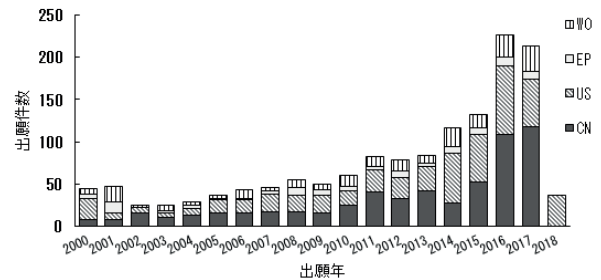


図5 電力需給管理におけるブロックチェーン技術 (世界) 出願件数推移

表4 電力需給管理におけるブロックチェーン技術 (世界) 出願人ランキング (上位11位まで)

出願人名 (名寄せ)	件数
ORACLE グループ	46
INTEL CORP.	39
国内電機メーカー S社グループ	38
国内電機メーカー P社グループ	37
QUALCOMM INC	22
GOOGLE INC.	22
IBM	20
SIEMENS グループ	19
FIRESTAR SOFTWARE, INC.	16
国内電機メーカー H社グループ	15

「ドローン制御におけるブロックチェーン技術」についてはドローン制御の観点(特許分類およびキーワード)とブロックチェーンの観点(特許分類およびキーワード)を掛け合わせた検索式でヒットした集合(日本: 39 件(図6および表5参照)、世界: 942 件(図7および表6参照))を母集団とした。世界についてはPCT 特許、欧州特許、米国特許、中国特許を対象に検索した。各母集団を作成してマクロ分析を行ったところ、両技術共に2015年頃から特に米国および中国の出願件数が急増しており(図5および図7参照)、ここ数年が技術の立ち上がりの時期であることが確認できた。

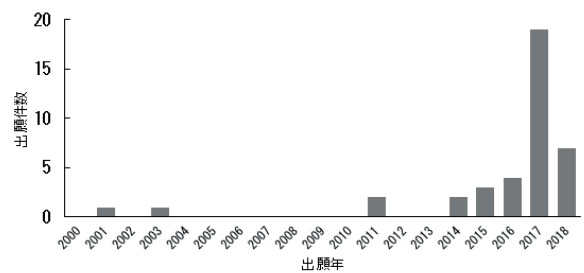


図6 ドローン制御におけるブロックチェーン技術 (日本) 出願件数推移

表5 ドローン制御におけるブロックチェーン技術（日本）出願人ランキング（上位4位まで）

出願人名（名寄せ）	件数
中国民生用ドローンメーカー S社	12
米国航空宇宙関連企業 B社	5
国内光学機器メーカー R社	2
国立研究機関，国内ドローンベンチャー P社	2

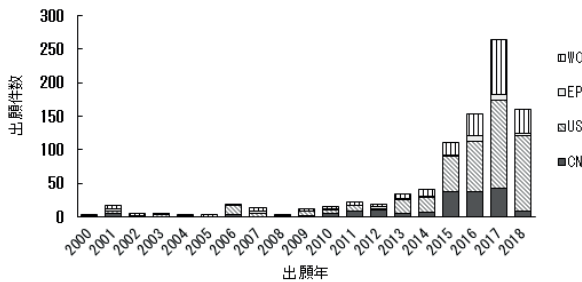


図7 ドローン制御におけるブロックチェーン技術（世界）出願件数推移

表6 ドローン制御におけるブロックチェーン技術（世界）出願人ランキング（上位10位まで）

出願人名（名寄せ）	件数
WALMART グループ	60
STRONGFORCE IOT PORTFOLIO 2016, LLC	42
INTEL グループ	29
IBM	22
ALTR SOLUTIONS, INC.	18
QUALCOMM INCORPORATED	16
CHRONICLED, INC.	13

出願人ランキングを確認すると、電力やドローンそのものに関連する企業は少なく、IT や自動車に関連する企業が上位に入っており、更にウォルマート、IBM、アマゾンといった米国巨大企業も注目していることが窺える（表6参照）。

### (3) 分析結果と考察

「電力需給管理におけるブロックチェーン」および「ドローン制御におけるブロックチェーン」の実用化状況を確認するために、Web 検索で情報を収集した。「電力需給管理におけるブロックチェーン」については環境省モデル事業に関する情報ははじめ、実証実験に関する情報が得られた。「ドローン制御におけるブロックチェーン」についてはウォルマートおよび IBM の特許出願、ドローンとブロックチェーンを融合した管制システム構築を目指す国際非営利組織および実証実験に関する情報が得ら

れた。いずれも実用化を目指す段階の動きであり、求人情報を利用した本事例において、実用化の期待が高い事業動向をいち早く確認することができた。このように求人情報を利用することで事業動向の「兆し」を抽出することが可能になるとと思われる。最新事業動向の「兆し」は知的財産部門だけでなく事業部門等でも広く活用できる情報である。また、求人情報と発明者情報を組み合わせることで、他社の人的リソース配分を予測することも可能になると考えられ、このような情報は人事部門にも提供可能であると思われる。

## 2.3 特許およびオープンな非特許情報を活用した特許分析手法の可能性

以上、特許およびオープンな非特許情報を活用した特許分析手法の一部を紹介したが、JIPA 情報検索委員会では、この他にも種々の事例について分析を行っている。その研究成果の詳細は、JIPA から発刊されている「知財管理」<sup>2)</sup> を参照されたい。

## 3 テキストマイニング技術の活用

知財情報の活用において、非特許情報が重要であることはおわかりいただけたと思うが、やはり基本となる特許情報の分析は疎かにすることはできない。本研究テーマでは、特許情報のマクロ分析を行う上で必要不可欠とされるテキストマイニング技術を取り上げ、テキストマイニングによるマクロ分析の結果を、迅速かつ質の高い特許情報分析に繋げる上での課題を検証し、テキストマイニングの質を高めうるテキストマイニングツールの活用方法を検討した。

### 3.1 課題の検証

特許庁が公開している特許出願技術動向調査等報告書（技術動向調査報告）に示された概要を、テキストマイニングツールを用いて分析し、同報告書に示された俯瞰図等との比較を行った。

### 3.2 課題の検証方法

検証対象とする技術動向調査報告のテーマとしては、平成 25 年度に報告された「ロボット」を選定した。

母集団は、かかる技術動向調査報告に記載された検索

式に基づいて抽出した日本出願 5,847 件とした。この母集団について、テキストマイニングを実施し、キーワードマップを作成し、技術動向調査報告の概要中に示された俯瞰図と比較し、テキストマイニングによるマクロ分析の結果について、迅速かつ質の高い知財情報分析に繋げる上での課題を検証した。

検証に用いたテキストマイニングツールは、Derwent Innovation (ThemeScape)、Biz Cruncher、KH Coder、Orbit Intelligence、TechRader、Text Mining Studio、CyberPatent Desk テキストマイニング (旧「TRUE TELLER パテントポートフォリオ」(以下、「CPD」)) の7つである。本稿では、このうち2つのツールを用いた事例を紹介する。

### 3.3 課題の検証結果

参照した技術動向調査報告の概要<sup>3)</sup>には、技術俯瞰図が示されており、観点毎にキーワードがまとめられている。

すなわち、まず技術全体を応用技術と要素技術に分け、応用技術を、「産業用ロボット分野」と「サービスロボット分野」に分け、さらに「サービスロボット分野」を「特殊環境用ロボット」と「サービスロボット」に分けた上で、それぞれの具体的な用途が列挙されている。

また、要素技術は、「全体構造技術」、「安全技術」、「制御技術」、「知能化技術」、「認識・コミュニケーション技術」の5つに分けた上で、それぞれの具体的な技術が列挙されている。

一方、テキストマイニングツールを用いて作成したキーワードマップについて、説明する。Biz Cruncher で描いた図8では、左側にアーム関連、右下に歩行関連、右上に制御関連のキーワードが配置されているが、それらのキーワードの観点は統一されていない。

すなわち、技術動向調査報告に示された俯瞰図が、用途、技術などの観点毎にキーワードを整理しているのに対して、キーワードマップは様々な観点のキーワードが混在している。

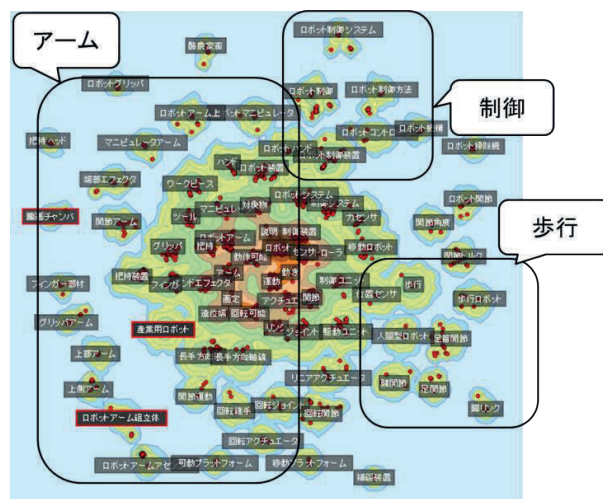


図8 Biz Cruncherによるキーワードマップの例

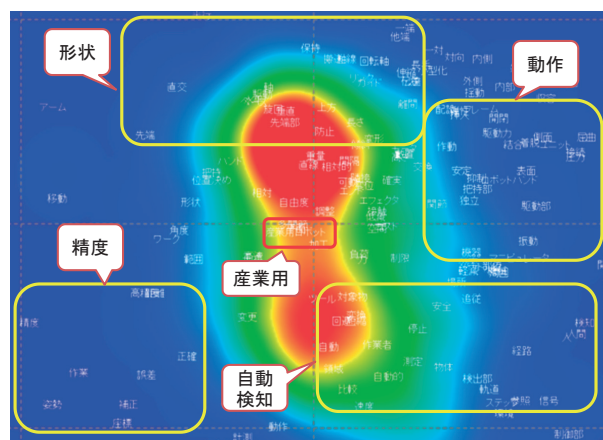


図9 CPDによるキーワードマップの例

我々は、技術動向調査報告のように、観点毎にキーワードを整理して俯瞰することが、我々の目的において求められると仮定し、テキストマイニングツールを用いて作成したキーワードマップ中のキーワードを観点毎に整理しようと考えた。しかしながら、様々な観点のキーワードが描画されたキーワードマップでは、特定の観点における特徴的なキーワードが、描画されていない場合があることに気づいた。

例えば、CPDで描いたキーワードマップを示す図9においては、用途における特徴的なキーワードとしては、「産業用」が描画されているのみである。結果として、かかるキーワードマップから、ロボット分野における用途を俯瞰することは困難である。

なお本稿では、「用途における特徴的なキーワード」とは、「ロボット自体の用途に関するキーワード(例えば「産業用」)を意味し、ロボットを用途とする材料や部材等に関するキーワード(例えば「ロボットアーム」)は、含まないものとする。

### 3.4 設定した課題

以上のことから本稿では、特定の観点における特徴的なキーワードがキーワードマップ上に描画されないという点を課題として設定した。

次に、設定した課題の原因を検証した。図9に示したキーワードマップの元になる、テキストマイニングで抽出されたキーワードリストの上位300件を確認したところ、表7に示すように、150位以降にキーワードマップ中に描画されなかった用途における特徴的なキーワードが複数見つかった。

表7 キーワードリスト中の用途における特徴的なキーワード

順位	単語
31	産業用
173	搬送ロボット
196	歩行補助ロボット
255	移動ロボット
262	医療用

このことから、用途における特徴的なキーワードの多くがキーワードマップ中に描画されない理由は、キーワードとして抽出されていないわけではなく、様々な観点のキーワードが混在して抽出されるため、全体としては使用頻度が低いキーワードと判断され、描画できるキーワードの数（または現実的に確認可能なキーワードの数）の上限によって描画対象から外れているためと考えた。

### 3.5 検討結果

そこで、本稿では、テキストマイニングによるマクロ分析の質を向上させるべく、様々な観点のキーワードの混在が少なく、注目する観点における特徴的なキーワードが描画されたキーワードマップを得る方法を検討した。

なお、母集団としては、上記した課題の検証に用いたロボットに関する日本出願5,847件を用いた。

#### (1) 明細書の分析事例1

前記母集団について、KH Coderを用いて明細書を分析した事例を示す。

明細書における「要約」あるいは「技術分野」をテキストマイニングによる分析対象とし、テキストデータは

インターネット特許検索サービスであるCKSWebを用いて準備した。キーワードマップは、関連性の高いキーワード同士を距離と実線のネットワークとして描画できる共起ネットワーク図を選択した。

#### 1) 分析箇所検討事例

まず、「要約」を分析対象にして、分析、描画を実行した例を示す(図10)。

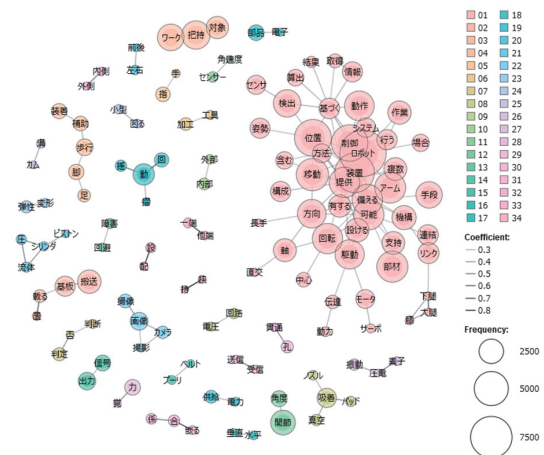


図10 「要約」の共起ネットワーク図

当該図10の右側には比較的広大なネットワークを確認できるが、具体的な課題までは捉えられず、「要約」中に同じく含まれる解決手段と紐付いた状態として認識することもできない。

これは、「要約」には課題と解決手段という複数の観点からの文章が含まれているにもかかわらず、テキストマイニングツールではそれらを区別したり紐付けたりする機能を持たないことが原因であると推測される。

また、技術的な内容を表すものではない一般的な動詞も多く見られ、それにより用途における特徴的なキーワードの一部が埋没して、描画されていない可能性が考えられる。

次に、「技術分野」を分析対象にして、分析、描画を実行した例を示す(図11)。

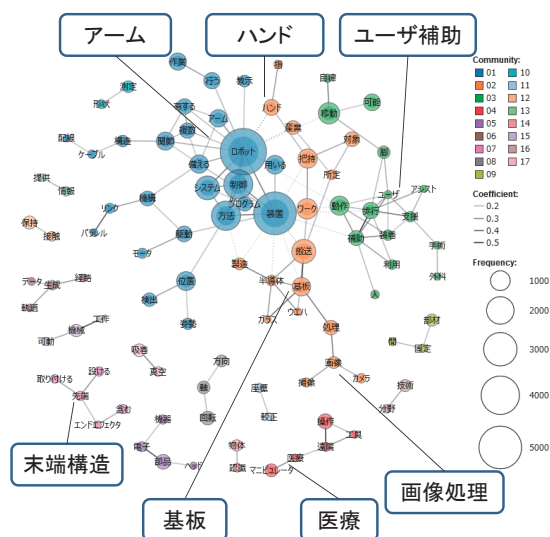


図 11 「技術分野」の共起ネットワーク図

この例では、特定の用途及び構造を示すと考えられるネットワークを確認することができる。

これらは、「要約」を分析対象とした場合には確認されなかったもので、「技術分野」を分析対象として選択したことにより、単一の観点から分析が行われたためであると考えられる。

このように、当事例においては、集合に含まれる主な技術内容を俯瞰するような大まかな分析の場合には、「技術分野」が分析対象として適していることが示唆される結果となった。

なお、「要約」及び「技術分野」の共起ネットワーク図の作成に際して、描画結果の見やすさを最大化した上で両者のネットワーク分布の粗密を均一にすべく、設定に差異を設けた。具体的には、「要約」に対しては、ストップワードとして、「図」、「少なくとも」、「それぞれ」、「前記」、及び「発明」を設定し、かつ最低出現数が 170 以上のキーワードを使用した共起のうち、上位 120 を表示する設定とした。「技術分野」に対しては、ストップワードとして、「特許」、「出願」、「米国」、「本明」、「細書」、「参照」、「開示」、「実施」、「形態」、「発明」、「特に」、及び「係る」を設定し、かつ最低出現数が 100 以上のキーワードを使用した共起のうち、上位 130 を表示する設定とした。最低出現数は、任意回数以上出現したキーワードのみを描画させるための足切り設定である。

## 2) ストップワード設定方法と事例

「技術分野」を分析対象として、用途における特徴的なキーワードをより積極的に描画するため、ストップ

ワード自動判定を検討した。

ストップワード自動判定には、各キーワードに以下のように定義するスコアを用いた。

DF (「技術分野」) / (max {DF (「請求項」), DF (「課題」), DF (「効果」)} + 1)

上記定義においては、DF (X) は、対象とするキーワードが、記載箇所に記載された特許文献の数を表す。

本定義では、キーワードのスコアが 1.0 以上であるとき、そのキーワードが他の記載箇所（請求項、課題、効果）と比べて技術分野により多く出現することを意味する。一方スコアが小さいほど、技術分野以外の記載箇所により多く出現することを意味する。

ストップワードを設定しなかった場合 (図 12)、スコアが 0.4 以下のキーワードをストップワードに設定した場合 (図 13)、及びスコアが 1.0 以下のキーワードをストップワードに設定した場合 (図 14) の共起ネットワーク図を示す。

それぞれの共起ネットワーク図は、「技術分野」における出現数が多い順に上位 200 語のキーワードを描画した。

共通設定として、特許明細書の定型句である「出願」、「米国」、「本明」、「細書」、「参照」、「開示」、「実施形態」、「発明」、「特に」、「係る」を予めストップワードに設定した。

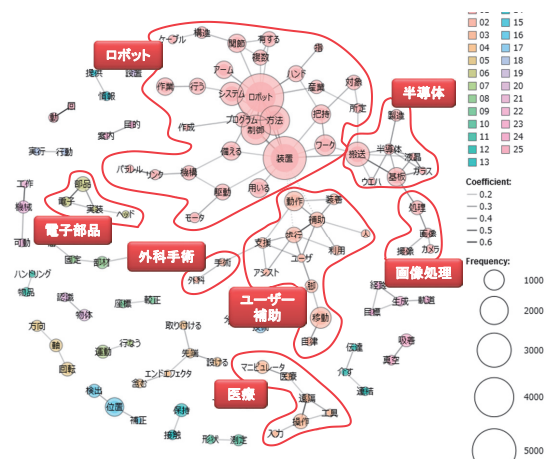


図 12 ストップワードを設定していない「技術分野」共起ネットワーク図



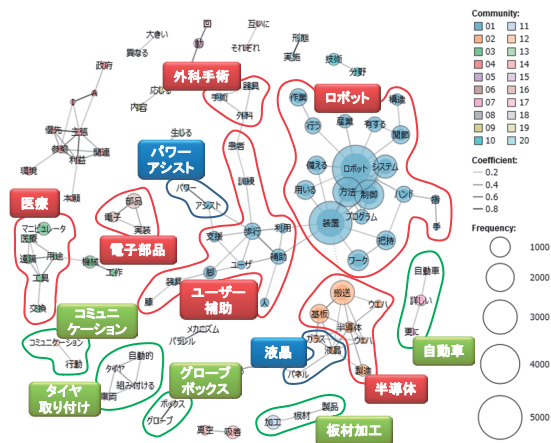


図 13 スコア 0.4 以下のキーワードをストップワードとした「技術分野」共起ネットワーク図

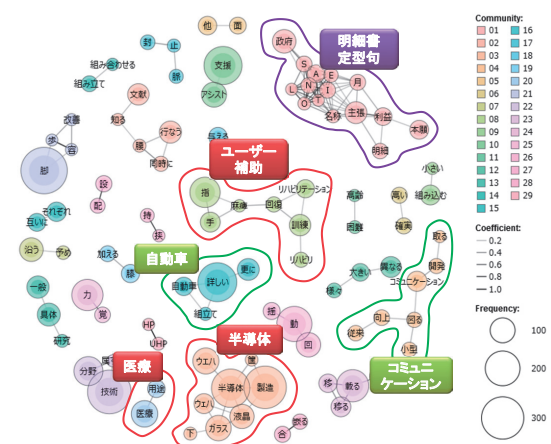


図 14 スコア 1.0 以下のキーワードをストップワードとした「技術分野」共起ネットワーク図

スコアが 0.4 以下のキーワードをストップワードとした場合 (図 13) では、「自動車」や「コミュニケーション」など用途における特徴的なキーワードのネットワークがストップワードを設定しなかった場合 (図 12) と比較して新たに出現した。

一方、スコアが 1.0 以下のキーワードをストップワードとした場合 (図 14) では、「外科手術」など用途における特徴的なキーワードの一部がストップワードに含まれて消失し、特許明細書の定型句である「本願」や「明細」などのノイズワードが新たに出現した。

このように、自動判定によってストップワードを設定して、描画することで、図 13 のように用途を俯瞰しやすいキーワードマップを作成できることを確認した。

一方で、ストップワードに設定するスコアのしきい値の設定には注意が必要である。参考までに、スコアのしきい値とキーワードの出現状況との関連を表 8 に示す。表中、それぞれのスコアのしきい値において出現した

キーワードについて「○」で示す。

表 8 ストップワードに設定したスコアと、キーワードの出現

ストップワードに設定したキーワードのスコア	画像処理	ロボット全般	半導体	ユーザー補助	外科手術	液晶	パワーアシスト	電子部品	自動車	コミュニケーション	板材加工	タイヤ取り付け	グローバルボックス
標準条件 (図 12)	○	○	○	○	○				○	○			
スコア 0.3 以下		○	○	○	○	○	○	○	○	○	○	○	○
スコア 0.4 以下 (図 13)		○	○	○	○	○	○	○	○	○	○	○	○
スコア 0.5 以下		○	○	○	○	○	○	○	○	○	○	○	○
スコア 0.6 以下		○	○	○	○	○	○	○	○	○	○	○	○
スコア 0.75 以下		○	○	○	○	○	○	○	○	○	○	○	○
スコア 0.1 以下 (図 14)		○	○						○	○	○		

## (2) 明細書の分析事例 2

次に、Biz Cruncher を用いて分析した事例を示す。

このテキストマイニングツールでは、各キーワードの希少性と多様性という 2 つの尺度に基づいて非公開の方法で算出される、重要度という独自のスコアを用いて、その公報内での特徴を示すキーワードが抽出される。

希少性とは、そのキーワードが対象の公報群内にどれだけ出現するかの尺度であり、出現数が少なければ希少性が高くなる。一方、多様性とは、公報内におけるそのキーワードの使われ方のバリエーションの尺度であり、バリエーションが多ければ多様性が高くなる。

### 1) 記載箇所による観点の絞り込み

前述の重要度は、明細書内の記載箇所毎に算出することもできるので、記載箇所毎に特徴的なキーワードを抽出できると考えられる。

例えば、「請求項」であれば構成、「技術分野」であれば用途や技術における特徴的なキーワードを抽出できると予測した (表 9)。

表 9 記載箇所毎に含まれうると予測した観点の一覧

記載箇所	含まれうる観点
発明の名称	構成 / 用途
要約	構成 / 課題
請求項	構成
技術分野	用途 / 技術
背景技術	技術
課題	課題 / 用途 / 技術
発明の効果	効果 / 用途
利用可能性	用途

この予測に基づいて分析対象とする記載箇所を選択を行い、複数の記載箇所を組み合わせる方法を検討した。

この方法によって、複数の記載箇所に記載されている注目する観点における特徴的なキーワードの重要度が上がり、それ以外のキーワードの重要度が相対的に下がることで、結果的にキーワードマップ上に、注目する観点における特徴的なキーワードが多く描画されると考えた。

ここでは、用途における特徴的なキーワードが含まれそうな記載箇所として、「技術分野」、「課題」、「発明の効果」、「利用可能性」を分析対象とした。

Biz Cruncherにおいて初期設定されている「発明の名称」、「要約」、「請求項」を分析対象として、重要度上位 80 語を描画してキーワードマップ化すると、用途における特徴的なキーワードが 13 語描画された（表 10 及び図 15）のに対し、分析対象を変更し、「技術分野」、「課題」、「発明の効果」、「利用可能性」とした結果では、「歩行補助装置」や「人型ロボット」等、新たなキーワードが抽出され、用途における特徴的なキーワードが 26 語に増加した（表 11 及び図 16）。

表 10 初期設定で分析対象となる記載箇所と観点

記載箇所	含まれる観点
発明の名称	構成／用途
要約	構成／課題
請求項	構成



図 15 初期設定での分析

表 11 変更後の分析対象（記載箇所）と観点

記載箇所	含まれる観点
技術分野	用途／技術
課題	課題／用途／技術
発明の効果	効果／用途
利用可能性	用途

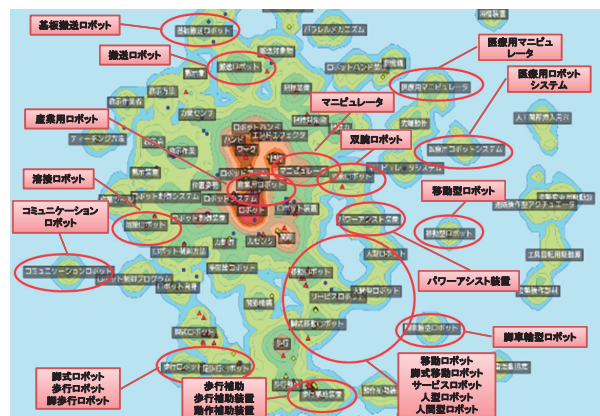


図 16 分析対象を変更

しかし依然として、用途以外の観点のキーワードも多い上、単一の記載箇所にのみ記載されているキーワードの中には、用途における特徴的なキーワードでありながら、記載箇所を組み合わせることで逆に埋没したのもあった。

## 2) 複数記載箇所の減算組み合わせ

次に、複数の記載箇所を引き算の形で組み合わせる方法を検討した事例を報告する。

「技術分野」のみを分析対象として、同様に重要度上位 80 語を描画してキーワードマップ化すると、用途における特徴的なキーワードが 23 語描画されたものの、逆に用途との関連性が低い「ロボットアーム」や「マニピュレータ」といった、技術に関するキーワードも多く含まれていた（表 12 及び図 17）。

表 12 技術分野に含まれると考えられる観点

記載箇所	含まれる観点
技術分野	用途／技術

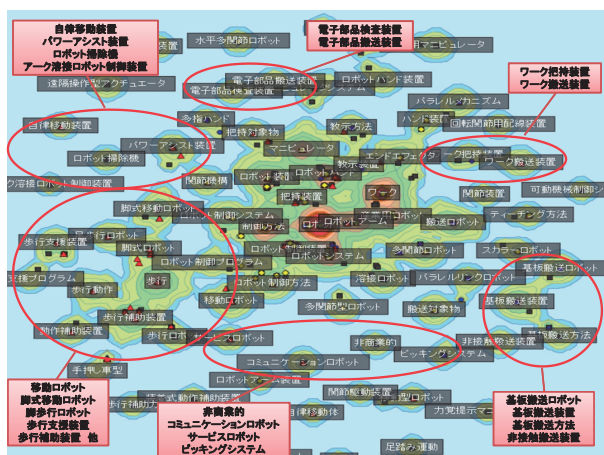


図 17 技術分野のみの分析

そこで、技術に関するキーワードがより多く含まれそうな「背景技術」を分析対象として算出した重要度を、「技術分野」の分析で算出された各キーワードの重要度から減算し、重要度の再集計を行った（図 18）。

抽出キーワード	技術分野重要度	背景技術重要度	算出後スコア
ワーク	3	2.7	0.3
基板搬送	2.5	キーワード抽出なし	2.5
制御	2	2	0

各記載箇所ごとの重要度を減算する

図 18 重要度の再集計に関する概念図

つまり、「技術分野」で抽出されたキーワードのうち、「背景技術」でも抽出されたキーワードは、再集計された重要度が下がることになる。

この処理の後に作成したキーワードマップ（図 19）では、「荷搬送ロボット」などの埋没していたキーワードが描画され、重要度上位 80 語中 49 語が用途における特徴的なキーワードとなった。これらのキーワードには、「ワーク搬送装置」、「ワーク搬送方法」、「ワーク把持方法」等、より細かい用途における特徴的なキーワードも含まれていた。

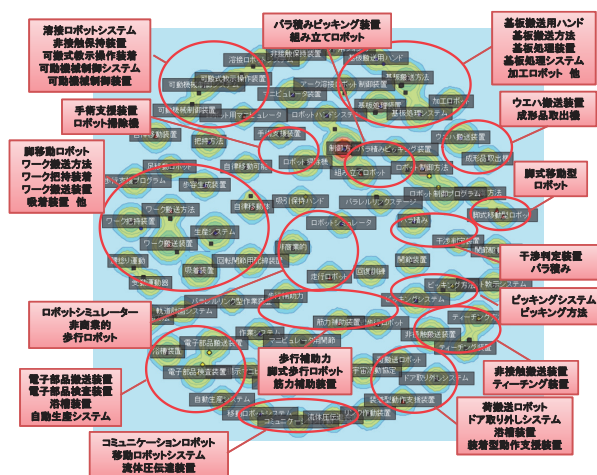


図 19 技術分野から背景技術を減じた分析

これらの方法は、重要度という Biz Cruncher 独自のスコアを用いて算出しているが、例えば記載箇所間のキーワードランキングの違いを用いるなど、同様の考え方を使得、他のテキストマイニングツールにも応用できるのではないかと考えている。

### 3.6 テキストマイニング技術活用の可能性

以上、特定の観点における特徴的なキーワードがキーワードマップ上に配置されていないという課題に対して、注目する観点の特徴的なキーワードを優先的に抽出する必要があると考え、第 1 に分析対象を選択すること、第 2 に選択した分析対象と他の箇所とのキーワードの使用頻度差に基づく特徴語の抽出やストップワードの設定が有効であることを確認した。

さらに、2 つの分析ツールを用いて上記を検証し、汎用性の高い方法であることを示すことができた。

JIPA 情報検索委員会では、上記 2 つのツールの他にも DWPI 抄録を分析した事例などを紹介している。その研究成果の詳細は、JIPA から発刊されている「知財管理」<sup>4)</sup> を参照されたい。

## 4 おわりに

JIPA 情報検索委員会の研究活動の中から、2つのテーマを抜粋し、成果の一部を紹介した。この他にも、第四次産業革命関連技術の特許分類に関する研究、オンライン調査による模倣品発見手法の検討、IP ランドスケープに関する実態調査と考察といった、情報検索・分析に関連するテーマについて研究を進めている。その成果については、知財管理誌等で公開していくので、期待されたい。

## 注記（引用文献、参考文献）

- 1) レッドクルーズ株式会社、プレスリリースポータルサイト JPubb（ジェイパブ）  
<http://www.jpubb.com/>（参照日：2019.3.21）
- 2) 情報検索委員会第4小委員会、特許およびオープンな非特許情報を活用した特許分析手法の検討、知財管理、Vol.70、No. 1（2020）
- 3) 平成25年度特許出願技術調査報告書概要（ロボット）  
[https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/document/index/25\\_robot.pdf](https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/document/index/25_robot.pdf)  
（参照日 2019年4月15日）
- 4) 情報検索委員会第3小委員会、テキストマイニング技術の活用に関する研究、知財管理、Vol.69、No. 10（2019）