

テキストマイニングを用いた効率的な特許調査方法

Effective patent research method using the Text mining

花王株式会社 知的財産センター **安藤 俊幸**

PROFILE 1985年現花王株式会社入社、研究開発に従事、1999年研究所の特許調査担当、2009年より現職。2012年よりアジア特許情報研究会にて活動。

✉ ando.t@kao.co.jp

1 はじめに

最近の新興国を含めた企業活動のグローバル化に伴い世界の特許出願も増加している。表1にファミリー中の最初の優先日が2012年の世界の特許文献を示す。発明をファミリー単位にまとめたデータベース(Questel社 Orbit.com)のデータを使用して最初の優先国と発行国の上位10ヶ国をマトリックス集計したものである。トータル1942693件のファミリーの内56%が中国(CN)を優先国としてCNで発行されている。以下同様に日本(JP)12%、米国(US)8%、韓国(KR)7%、ドイツ(DE)3%である。CNの件数が多すぎて小さな件数は0%に丸められている。

表1 世界の特許文献(実用新案含む)

		最初の優先国 →									
		CN	DE	EP	GB	JP	KR	RU	TW	US	WO
発行国 ↓	CA	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
	CN	56%	0%	0%	0%	2%	0%	0%	0%	2%	0%
	DE	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%
	EP	0%	1%	1%	0%	1%	0%	0%	0%	2%	0%
	JP	0%	0%	0%	0%	12%	0%	0%	0%	1%	0%
	KR	0%	0%	0%	0%	1%	7%	0%	0%	1%	0%
	RU	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%
	TW	0%	0%	0%	0%	1%	0%	0%	2%	0%	0%
	US	1%	0%	0%	0%	2%	1%	0%	0%	8%	0%
	WO	1%	1%	0%	0%	2%	0%	0%	0%	3%	1%

最初の優先日: 2012.01.01 ~ 2012.12.31 トータル数: 1942693件

このような状況下で日本特許庁は2015年1月に「中韓文献 翻訳・検索システム」^{1),2)}を、2015年8月には「外国特許情報サービス FOPISER (Foreign Patent Information Service)」³⁾を稼働させている。

2 特許調査への情報要求

R. S. Taylor の1968年の論文において、人間の情報要求 (information need) が下記4レベルに分類されている⁴⁾。

①直感的要求 (visceral need)

現状に満足していないことは認識しているが、それを具体的に言語化してうまく説明できない状態

②意識された要求 (conscious need)

頭の中では問題を整理できるが、あいまいな表現やまとまりのない表現でしか言語化できない状態

③形式化された要求 (formalized need)

問題を具体的な言語表現で言語化できる状態

④調整済みの要求 (compromised need)

問題を解決するために必要な情報の情報源が同定できずくらい問題が具体化された状態

前述の情報要求に習い特許調査への情報要求をなるべく明確になるように以下にまとめる。特許調査においては調査目的により程度の差はあれ再現率(網羅性)重視で検索集合を作ることを前提としている。ただし再現率重視だとノイズが増えて精度(適合率)が低下する。網羅性を重視する侵害防止調査では検索集合が数万件に達することもある。この数万件をたとえば出願番号順に初めから確認していくのではなく何らかの優先順位をつけて効率的に確認したい。また出願前の先行技術調査では出願する発明に最も近い先行技術を効率的に見つけた。このような目的には検索者のスキルが重要であるのはもちろんである。また特許データベースや市販のペテ

ントマップソフトの有用な機能は限界を知りつつ使えるものは使うことを前提とする。更に筆者は最近のIT技術を活用して個人レベルで効率的に特許調査支援を可能にする方法としてテキストマイニングに注目して検討を行ってきた。特に調査のスクリーニングプロセスをテキストマイニングで使われる手法で効率化したいと日々考えてきた。検討目的を端的に表すと人が読まずに、人が読むべき特許をなるべく効率的に優先順位付けしたいということになる。換言すると玉石混交の公報の中から調査目的に合った玉を選択することである。

特許調査においては引用/被引用情報の有効利用や調査結果の再利用、検索履歴の活用等も重要である。また最近の中国出願の急増にも対応するため中国語のテキストマイニングで上記目的に応えたい。

Taylor の情報要求は検索すべき問題の明確化に主眼がおかれている。特許調査における問題を明確にしてさらに例えば下記のように問題の所在を明らかにして問題性質に応じた解決を図る必要がある。

- ①データベースの検索テクニックで解決できる問題
- ②調査に使用している具体的なデータベースの機能では解決できない問題
- ③データをダウンロードして処理することでユーザー側の工夫で解決できる問題

①は検索スキルの向上に努める。②は他力本願で時間が掛かるがデータベースのベンダーに要望を明確に伝えることで解決できる場合がある。ベンダー側で問題であることを認識して確認/修正できるように明確にかつ再現できるように伝えることが重要である。問題ではなく、新機能を追加することでデータベースユーザーにメリットがある場合はそのことをアピールすることも同様に重要である。筆者もこの方法で過去にデータベースの問題を修正あるいは、バージョンアップ時に新機能を追加していただいたおかげで特許調査が楽になった経験は何度もある。

本稿では③の問題を主に扱う。より具体的にはテキストマイニングを応用して効率的な特許調査支援を目的に下記の検討を報告する。

- ・テキストマイニングを用いた精度重視の特許調査方法
- ・調査目的に合った検索結果のソートと分類

- ・特許マップへの応用を意図した公報の俯瞰可視化

3 テキストマイニング関連ツール

テキストマイニングには下記ツール類を使用した。

- ・PatAnalyzer 中国語/日本語解析ツール (自作)
- ・MeCab: 日本語形態素解析器⁵⁾
- ・IKAnalyzerNet: 中国語分詞ライブラリ⁶⁾
- ・SimCalc1 類似度計算プログラム (自作)
- ・R 言語: 統計解析⁷⁾
- ・Cytoscape: ネットワーク分析⁸⁾

テキストマイニング処理の概要を図1に示す。詳細は参考文献9~12が詳しい。

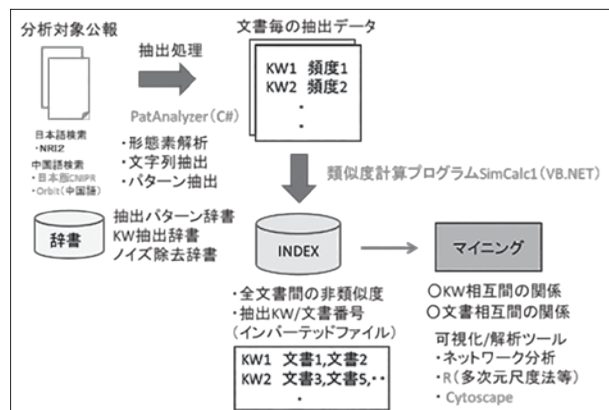


図1 テキストマイニング処理の概要

4 キーワードを用いたテキストマイニングによる特許調査支援

大量の公報をなるべく読まずにある程度の概要を把握する方法として図2に示すコンセプトマップがある。コンセプトマップは世界特許のデータベース Orbit.com に組み込まれた標準的な機能である。

図2は W02007069666 公報 (iPS 細胞) の頻度が高い英文キーワードを大きなフォントサイズで示している。表示される位置に意味はない。同様な図はタグクラウドとかワードクラウドとも呼ばれている。

日本語キーワードを同様に表示したのが図3および図4である。

図3は W02007069666 の請求項より

418 RESISTANT COLONY | ABNORMAL PHENOTYPE | ADULT MOUSE | ADULT MOUSE TAIL | ADULT SKIN | AF EMBRYONIC STEM | ANTIGEN | ARTIFICIAL CHROMOSOME | AUCTION | BCF | BETA CATENIN | BETA FUSION GENE | BETA GALACTOSIDASE GENE | BETA GALACTOSIDASE GENE FUSION GENE | BETA GLO KNOCKIN | BPGF | BPGF MANIFESTATION VECTOR | BSGO KNOCKIN FBK | BLASTOCYTES | BLASTODERMIC VESICLE GONAD | BLASTODERMIC VESICLE LUMP | BOOK DEPARTURE DISCERNMENT | C MYC | CANDIDATE GENE | CANDIDATE GENE KND | CANDIDATE REPROGRAMMING FACTOR | CARDIAC MUSCLE | CATENIN | SELECTION | CELL | CELL COLONY | CELL DIFFERENTIATION INDUCTION | CELL DIFFERENTIATION POTENCY | CELL ESTABLISHMENT | CELL GROWTH ABILITY | CELL MARKER GENE | CELL NUCLEUS | CELL PRODUCTION | CELL PROLIFERATION | CELL STRAIN | CELL STRAIN EXTRACT | CELL TRANSPLANTATION | CELL TRANSPLANTATION MEDICAL THERAPY | CELL TRANSPLANTATION MEDICAL THERAPY RESOURCE | CELL TRANSPLANTATION RESOURCE | CELL TRANSPLANTATION THERAPY | CELLULAR AUTO TRANSPLANTATION | CERTAINTY | CF EMBRYONIC STEM CELL | CHARACTER SEVEN | CHIMERA MOUSE BREED | CHIMERA MOUSE COMPILE | CHIMERIC MOUSE | CHIMERIC MOUSE MATING | CHROMOSOME TERMINAL | CLONE | CLONED EMBRYO | COLON BACILLUS CULTURE | COLONY | COLONY FORMATION | COLONY FORMATION OBSERVATION | COMPACT COLONY | COMPOUND TOXICITY | CULTURE | CYTOKINE | DAY NUTRIENT | DEFORMED SWELLING | DEPARTURE DISCERNMENT | DEPARTURE REALITY | DERMAL FIBROBLAST | DIABETES | DIFFERENTIATED CELL | DIFFERENTIATED CELL MORPHOLOGY | DIFFERENTIATED CELL NUCLEUS REPROGRAMMING | DIFFERENTIATED CELL REPROGRAMMING | DIFFERENTIATION | DIFFERENTIATION ABILITY | DIFFERENTIATION CELL INITIALIZATION |

図2 Orbit.com のコンセプトマップ (一部抜粋)



図3 形態素レベルのワードクラウド

RMeCab¹³⁾ + wordcloud¹⁴⁾ を用いて抽出した形態素の名詞とその頻度情報による形態素レベルのワードクラウドである。文字の向きに意味はない。

図4は形態素解析ツール MeCab の品詞と隣接頻度情報より求めた専門用語レベルのワードクラウドである。枠で囲んだ「誘導多能性幹細胞」、「Myc ファミリー遺伝子」、「核初期化因子」は調査において重要な特徴語である。

形態素と専門用語では形態素を用いると網羅性が向上し、専門用語を用いると調査の精度が高まる。

キーワード間の関係の解析には図5に示す語のネットワーク分析や係り受け解析がある。図5は WO2007069666 明細書より抽出した日本語の専門用語上位 100 語の Cytoscape によるネットワーク表示の一部である。楕円で囲んだキーワードはネットワーク的に他のキーワードとの関係上影響が大きいものである。

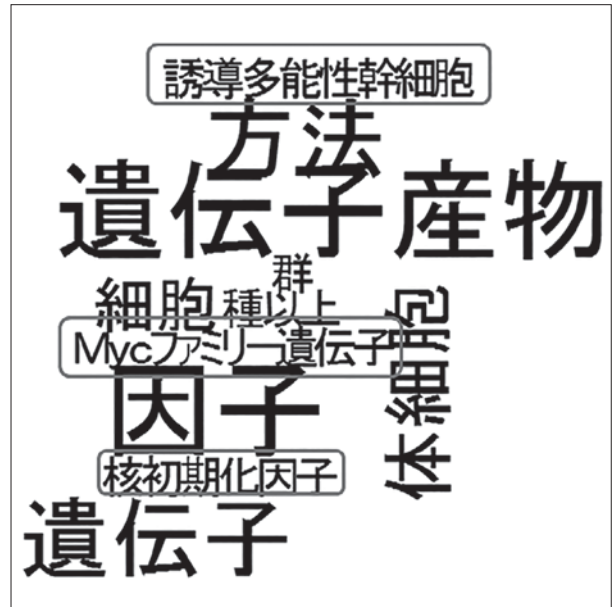


図4 専門用語レベルのワードクラウド

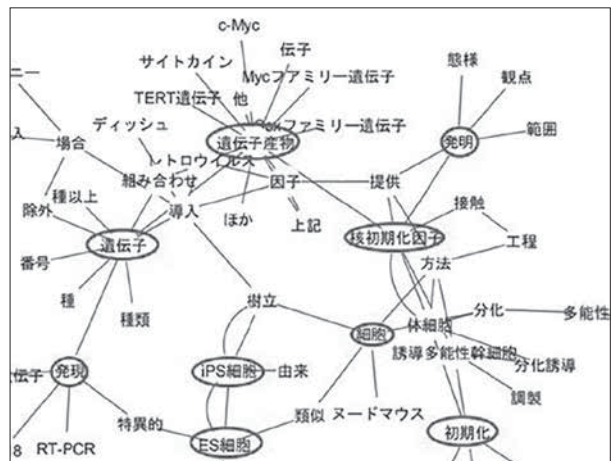


図5 Cytoscape による語のネットワーク

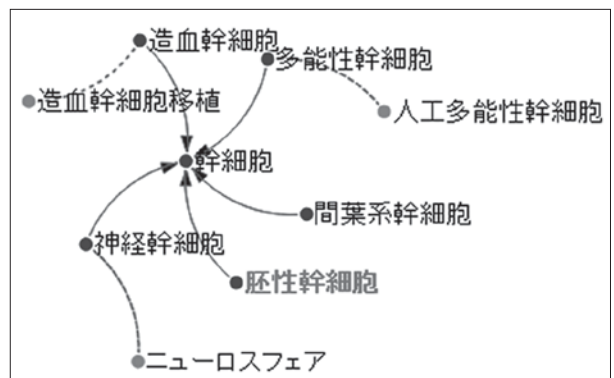


図6 Concept Map Search の例

語のネットワーク表示の応用例として図6に示す 2014年に公開された WIPO Pearl の Concept Map Search¹⁵⁾ の語のネットワーク表示がある。図6は ES 細胞の日本語表示である。英語や中国語での表示も可能

である。インタラクティブに表示されキーワードから公報にリンクされる等非常に興味深い。ただまだ収録語数が少なく iPS 細胞では 2 語しか表示されなかった。今後の収録語数アップが望まれる。

5 調査目的に合った検索結果のソートと分類

検索集合の公報間の関係に注目して公報のソートと抽出を検討した。公報を適切なソートキーでソートすることにより類似の公報同士を集めたり、いくつかのサブ集合に分類することができる。公報のソートについては並び替えに使用するソートキーで下記のような種類が考えられる。

検索集合のソートキー

- ①特許分類 (IPC, FI 等) → 大まかな技術分野
- ②出願人 (競合社) → 実務的に権利調査 (クリアランス調査) でよく使われる
- ③日付 (優先日、出願日、公開日、登録日等)
- ④リーガルステータス (侵害防止調査)
- ⑤類似率 (概念検索、類似検索) のスコア
- ⑥類似率 (キーワード検索のスコア)

例: Orbit.com の関連性 (Relevance score) 等

⑦独自定義の項目

例: 社内分類等

①~③は商用特許データベースや市販の Patent Map ソフトの機能として備わっていることが多く、ダウンロードデータを Excel でソートすることも容易である。

③日付は調査の種類により特定の日にちの前後で明確に意味が異なる場合がある。例えば先行技術調査において先願主義を採用している国では出願日を基準にして先行技術かどうか異なる。中国特許データベースの CNIPR では類似検索結果を新規性検索と侵害性検索結果に分けて表示される。

④リーガルステータスや生死情報は侵害防止調査時に使えたと調査効率上有用である。

⑤類似率 (概念検索、類似検索) は特定の公報を対象とした類似率で商用特許データベースの概念検索 (類似検索) のスコアをイメージすると理解しやすい。インパ

テック社の Patent Map EXZ の類似率ソート機能⁹⁾も同様である。Patent Map EXZ はキーワードを使用した類似率だけでなく IPC や独自に定義した項目 (例えば技術分野等) を使用して類似率を計算してソートすることも可能である。

⑥類似率 (キーワード検索のスコア) はキーワード検索の検索クエリに対する類似率である。例えば Orbit.com に 2015 年 7 月から導入された Quick Search のスコアである関連性 (Relevance score)。同じような機能としてアイ・ピー・ファイン社 THE 調査力の深度マイニングの専門用語によるソート機能 (日本語のみ)、同社 THE 調査力_クラウドによるワードソート機能を挙げることができる。

特に⑤⑥は内部的にテキストマイニングでよく使われる単語文書行列を用いて類似率を算出している。

⑦独自定義の項目は社内分類等独自に定義した項目である。

6 テキストマイニングを利用した特許調査

6.1 先行技術調査における類似率ソート

適合フィードバック (relevance feedback)¹⁰⁾ と呼ばれる仕組みを備えた検索システムでは検索回答の中から検索者の情報要求に合致 (適合) した回答を選択してシステムにフィードバックして再検索することで適合している回答を上位に持ってくるができる。

本稿執筆時 (2015.08.25) にベータ版が公開された Orbit.com 1.9.4 ベータのセマンテックサーチ (自然文検索) ではフリーテキストボックスに自然文を入力して NEXT ボタンを押すとシステムがテクニカルコンセプトを抽出して提示する。その中から目的に合ったコンセプトを選択して更に NEXT ボタンを押すと公報のタイトル、要約、出願人、Relevance score を score の降順に表示する。そこで適合している公報を選択して検索ボタンを押すと類似結果の検索結果が表示される。ここでさらに公報を選択して再検索することも可能である。このプロセスを繰り返すことで適合した公報を上位に集めることができる。複数公報を選択した場合はそれらの公報の合計あるいは平均に対して Relevance

score の降順に表示される。ブーリアン検索の検索結果と比べると score の大きい順に表示されるのが特長である。自然文の入力は英語がデフォルトであるが日本語や中国語の非英語の自然文を入力することもできる。この場合自動的に言語を検出して入力した言語のコンセプトとその-google 翻訳の英語を表示する。ここで必要なコンセプトを選択して検索すると日本語で入力した場合は日本文の公報と英語の公報が混ざった状態で score の降順で表示される。

特許データベースの概念検索（類似検索）は既に 10 年以上の実績がありその性質を理解して使用すると有用である。ただし次に述べるように限界もある。

6.2 引用 / 被引用と類似検索

いろいろな特許データベースで引用 / 被引用関係を分かり易い引用マップとして表示する機能は何年も前から実装されておりそれぞれ特徴がある。図 7 は NRI サイバーパテントデスク 2 の引用 / 被引用マップである。注目特許に対して引用 4 世代 / 被引用 5 世代の最大 9 世代まで表示させることができる。引用 / 被引用関係を示す公報間の線の色が下記の状態を表している。

- ・ 拒絶理由通知（拒絶理由の引用文献情報）
- ・ 拒絶査定（拒絶査定の引用文献情報）
- ・ 特許 / 登録査定（特許 / 登録査定の参考文献情報）
- ・ 先行技術調査（先行技術調査結果の参考文献情報）

注目特許を例えば特許 3872179（車両の衝突防止装置）とすると最大 9 世代で国内：178 件、海外：27 件である。

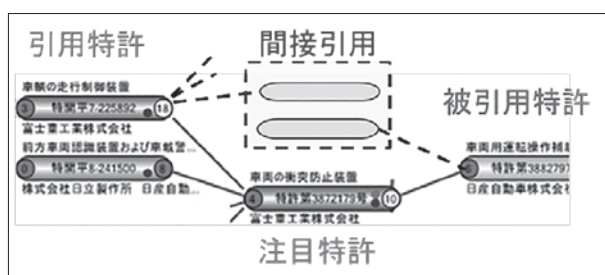


図 7 引用 / 被引用マップ（一部抜粋）

ここで間接引用は注目特許の引用特許から被引用方向に 1 世代、被引用特許からは引用方向に 1 世代を加えた合計である。図 7 の間接引用は説明のために加えたもので実際には表示されない。

国内特許に注目して間接引用と注目特許の類似検索を求めると下記の件数になる。

- | | |
|------------------|----------|
| ①注目特許の引用マップ | 国内 178 件 |
| ②間接引用 1 世代（①を含む） | 960 件 |
| ③注目特許の類似検索（請求項） | 上位 500 件 |

図示したのが図 8 である。注目特許を様々な分野から選択し直したり、検索条件を変えても類似検索結果に含まれる引用 / 被引用特許は予想外に少ない。図 8 の例では類似検索結果上位 500 件に含まれる引用 / 被引用 9 世代 178 件の内 8 件である。類似検索結果を分母にとると 1.6% である。間接引用 980 件まで広げても 6.4% である。逆に 1 件の審査で引用される引用 1 世代を考えると非常に少なくなる。この例では 4 件引用されていてそのうちの 1 件が類似検索の 235 位に現れる。パーセントを計算すると 0.2% である。ちなみにあとの 2 件は 1145 位と 7359 位に、最後の 1 件は類似検索上位 1 万件の圏外である。審査の過程で引用される特許件数は国や分野によっても傾向は異なる。また類似検索と比べてそれほど多い数ではないという性質を踏まえても一般論としてこれからの特許検索システムには内部の仕組みは問わずに類似検索（概念検索）結果の上位に引用特許が含まれることが特許調査の効率からも強く望まれる。

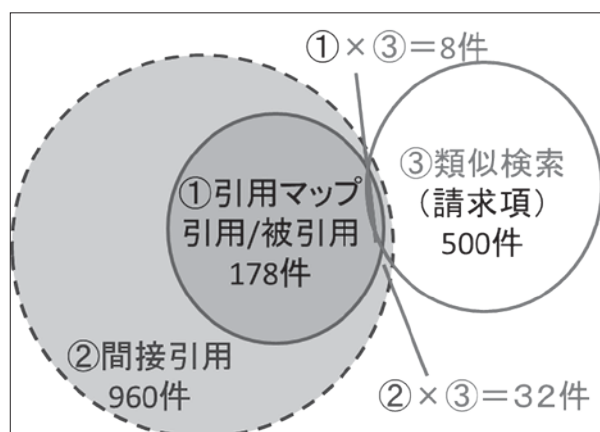


図 8 引用 / 被引用と類似検索の関係

6.3 類似率による公報関係の可視化と引用 / 被引用

図 9 の左側に NRI サイバーパテントデスク 2 の引用 / 被引用マップの全体図を示す。中心の注目特許の左側が引用特許で右側が被引用特許である。右側は国内の 178 件についてタイトル、要約、請求項までのテキスト

トデータより相互の類似度を計算して3次元表示したものである。類似度が大きい公報同士が近い距離にプロットされるように3次元空間上の位置を多次元尺度法で求めている。引用を赤ライン、被引用を青ラインにして、注目特許と直接引用 / 被引用関係にある第2世代を太いラインで示している。類似度はテキストマイニングや類似検索でよく使われているコサイン類似度を使用した。ブラウザで3次元コンピュータグラフィックスを表示させるための標準仕様である WebGL¹⁷⁾ で出力するとブラウザを用いて3次元表示マップを閲覧できる。3次元表示された引用 / 被引用マップをマウスで回転させて異なる視点から眺めることができる。

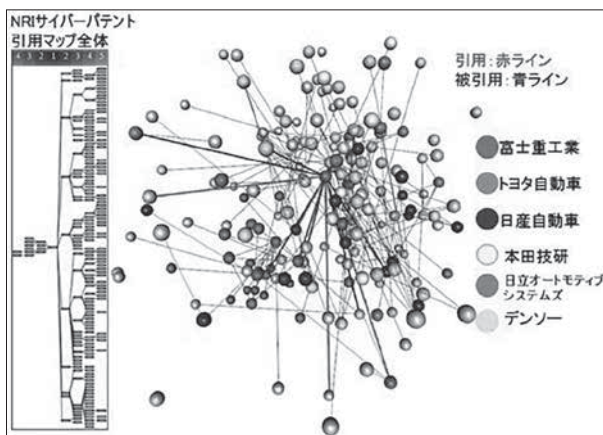


図9 類似率による公報関係と引用の可視化

左側の引用 / 被引用マップは審査の過程で発明のポイントが近いとして実際に使われた公報である。先行技術調査に類似率を使用して近い公報から確認する目的にはラインで結ばれた公報同士が近くなるように類似率の計算方法に更なる工夫が必要である。類似率計算は請求項を例にしてごく簡単に言うと請求項に含まれる単語の構成比率が近いと類似率は大きくなる。審査官引用は単に請求項に含まれる単語だけを見ているのではなく明細書のなかの一部分に違う表現で記載されていても発明のポイントが同じであれば引用される。

また引用情報（審査官引用）は審査が進む過程で発生するためかなりのタイムラグがある情報であり公報発行直後の段階ではあまり使えない。

6.4 引用 / 被引用情報のネットワーク解析

図10は図9と同じ集合の Cytoscape による引用 / 被引用のネットワーク解析である。左側が全体図で右側

は拡大図である。

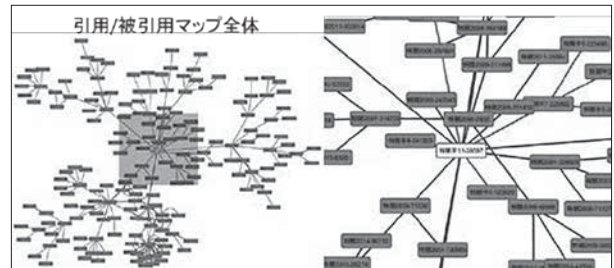


図10 Cytoscape による引用 / 被引用の解析

ネットワーク的な「重要度」を表す指標として次数中心性や媒介中心性等各種の中心性の指標がある。次数中心性 (Degree Centrality) は、ある点 (ノード) に接している辺の数である。媒介中心性 (Betweenness Centrality) は、ある点がその他の2点を結ぶ最短経路である度合であり、値が大きいほど中心性が高い指標である¹⁸⁾。中心性は重要語 (特徴語) を抽出する際の指標となる。Cytoscape ではこれらの中心性の指標に応じた可視化¹⁰⁾も可能である。

2014年の特許情報フェアに出展されたデータベースに引用 / 被引用関係の解析機能を備えたツールである AmberScope¹⁹⁾ は引用 / 被引用的に重要な特許を優先的にネットワーク表示する。このとき重要度の指標として「AmberScore 値」を独自に使用している。

6.5 侵害防止調査における検索結果のソート

権利調査の特殊性とその対応方法については文献20が詳しい。侵害防止調査における検索結果のソートあるいは分類はリスクの高いものの優先順位が高くなるように公報のリスクを見積もり、降順にソートするとリスクの高い順にスクリーニングできる。生死情報を使用して「生」と「死」に大別して生の優先順位を高くする。諸々の要因でデータベース上の生死情報が間違っている場合や一度死んだ特許が復活する可能性もある。復活の可能性等生死情報の信頼性を考慮して「死」だから即リスクゼロと判定できない場合もある。

また生死情報だけでなく登録、出願継続中、取消、失効・消滅、期限切れ等のリーガルステータスを使ってより細かく分類することもできる。整理標準化データの間コード、INPADOCの法的状況を示すPRSコード等を利用するのも有効である。特に中国特許の権利調査を



行う場合は生死情報の利用を最初に考慮するとよい。テキストマイニングを用いて技術内容の構成要件（エレメント）の有無を抽出する前に生死情報あるいはリーガルスタータスを使用できる場合は使った方がスクリーニング効率は良い。

予め定めたキーワードを使用して注目する構成要件の有る特許を積極的に抽出して優先順位を上げることは可能であるが、オールエレメントルールを念頭に構成要件が無いことを証明することは困難である。特定の用語あるいは用語の組み合わせがあるものをノイズとして除去することができる場合はスクリーニング効率を上げることが可能である。

特許のリスクだけでなく影響度合いを考慮することも重要である。競合社を正しく認識して優先順位に反映させることも大事なポイントである。

7 中国語のテキストマイニング

本稿の冒頭の表 1 に示したように最初の優先国が CN であり発行国も CN は 56% である。最初の優先国が CN であり発行国の言語が日本語あるいは英語の割合は非常に少ない。このことも中国語のテキストマイニングに取り組む一つの理由である。

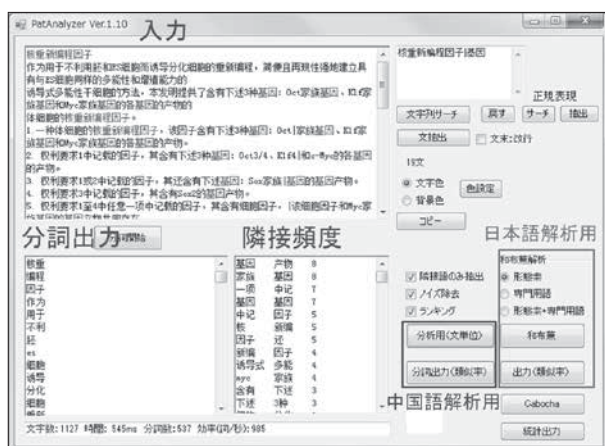


図 11 中国語 / 日本語キーワード解析ツール

図 11 に自作の中国語 / 日本語キーワード解析ツール PatAnalyzer の画面を示す。中国語キーワード抽出には IKAnalyzerNet を利用している。

図 12 は CN101356270 請求項 1 の形態素レベル

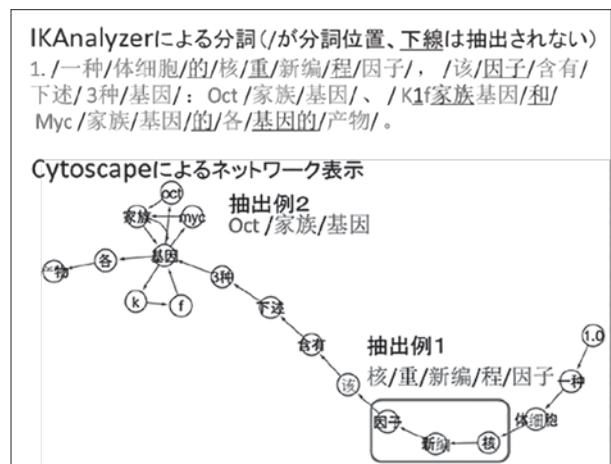


図 12 形態素レベルのネットワーク

のネットワークである。IKAnalyzerNet により分詞し各形態素の隣接頻度を求め Cytoscape でネットワーク表示した。中国語の専門用語抽出にはまだ課題がある。更に適切な抽出に向けて検討を進める予定である。

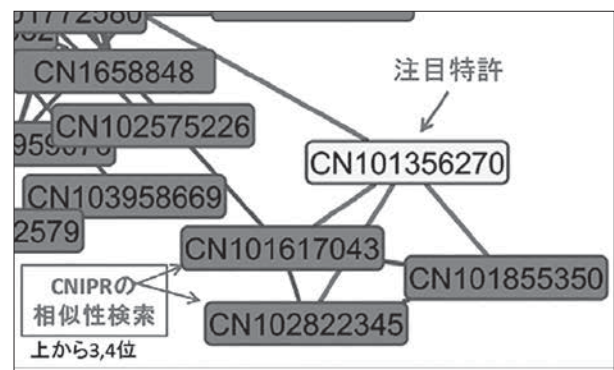


図 13 情報の類似度によるネットワーク表示

図 13 は情報の類似度(非距離)を使用してネットワーク表示したものの抜粋である。CNIPR の相似性検索(類似検索)の上位から 3,4 位の公報が近くにきている。注目公報の近くにある公報から確認することで精度重視の調査が可能になる。精度向上に向けて特徴語抽出や用語の重み付け、ネットワーク分析による公報の注目特許抽出等をさらに検討する。

8 おわりに

これまで特許調査の種類に応じたテキストマイニングによる調査の効率化の一端を述べてきた。

出願前の先行技術調査では概念検索(類似検索)を使

用できるが検索結果の上位に審査官引用されるような文献は意外に少ない。適合フィードバックのような手法で上位に持ってくることは可能ではあるがもう少し簡単に近い公報を抽出したいという情報要求がある。既存の検索システムではこのような要求に対してすぐには応えられないと思われる。このような時にテキストマイニングの手法を用いて類似率計算を下記のように工夫することも考えられる。

- ・ 発明のポイントに対する用語の重みを重くする
- ・ 発明のポイントを含む文あるいは段落抽出する

現在これらの点を考慮して審査官引用されそうな公報が上位にくるように類似率計算方法の検討を進めている。この場合の発明のポイントの抽出は当面人手による抽出を考えている。

将来的には技術動向調査時に発明のポイントの抽出を機械学習の手法等を用いて自動化して解析精度の向上を図りたいと計画している。

参考文献

- 1) 中韓文献 翻訳・検索システム
<http://www.ckgs.jpo.go.jp/>
- 2) 櫻井健太ら. 中国・韓国語の特許文献を日本語で検索可能なシステムのご紹介 (tokugikon 2015)
<http://www.tokugikon.jp/gikonshi/276/276kiko04.pdf>
- 3) 外国特許情報サービス FOPI SER (Foreign Patent Information Service)
<https://www.foreignsearch.jpo.go.jp/>
- 4) 徳永健伸. 情報検索と言語処理 (言語と計算 5). 東京大学出版会, 1999, p.3.
- 5) MeCab (和布蕪) 日本語形態素解析器
<http://taku910.github.io/mecab/>
- 6) IKAnalyzerNet. 中国語分詞ライブラリ
<http://www.piaoyi.org/c-sharp/IKAnalyzerNet.html>
- 7) R 言語
The R Project for Statistical Computing
<https://www.r-project.org/>
- 8) Cytoscape
<http://www.cytoscape.org/>
- 9) 安藤 俊幸ら. “中国語キーワードによる中国特許情報解析” 第 10 回情報プロフェッショナルシンポジウム
- 10) 安藤 俊幸ら. “中国特許解析・テキストマイニングによるKW分析” 第 11 回情報プロフェッショナルシンポジウム
https://www.jstage.jst.go.jp/article/infopro/2014/0/2014_31/_pdf
- 11) 安藤俊幸. テキストマイニングと統計解析言語 R による特許情報の可視化. 情報管理. Vol. 52, No. 1, (2009), 20-31.
https://www.jstage.jst.go.jp/article/johokanri/52/1/52_1_20/_article/references/-char/ja/
- 12) 今津均. Cytoscape による特許情報のネットワーク解析とビジュアル化. 情報管理. Vol. 54, No. 8, (2011), 463-475.
- 13) RMeCab
<http://rmecab.jp/wiki/index.php?RMeCab>
- 14) 末吉正成ら. “テキストマイニングを行う”. R ではじめるビジネス統計分析. 翔泳社, 2014. p. 304-334.
- 15) WIPO Pearl
<http://www.wipo.int/wipopearl/search/home.html>
- 16) Christopher D.Manning ら. “適合フィードバックとクエリー拡張”. 情報検索の基礎. 共立出版, 2012, p. 157-172.
- 17) 杉本雅広. 明解 WebGL. リックテレコム, 2015, 312p.
- 18) 金 明哲. “テキストにおけるネットワーク分析”. テキストデータの統計科学入門. 岩波書店, 2009, p. 63-72.
- 19) Ambercite 社 AmberScope
<http://www.ambercite.com/>
- 20) 静野健一: “特許調査, 特に権利調査における現状と課題”. 情報の科学と技術. 2015, Vol.65, no.7,p284-289.

上記 URL はいずれも 2015 年 8 月 25 日に確認したものである。