

テキストマイニングを使用した 特許マップ作成手法の開発

会員 川上 成年

要 約

特許マップは、特許情報分析の結果を図表により「見える化」したものである。特許マップの内、課題・解決手段マップは、人間が明細書を読み、明細書から「課題用語」と「解決手段用語」を抽出する作業を行い作成される。しかし、この抽出作業に時間がかかるという課題がある。本稿は、テキストマイニングの技術を使用することにより、「課題用語」と「解決手段用語」の抽出作業を効率化することを目的とする。

目次

1. はじめに
 - (1) 特許マップについて
 - (2) 課題・解決手段マップについて
 - (3) 課題・解決手段マップの作成法について
 - (4) 本稿の目的
2. テキストマイニングについて
 - (1) テキストマイニングとは
 - (2) 使用ソフトについて
3. テキストマイニングを使用した課題・解決手段マップの作成手順
 - (1) 抽出語による処理
 - (2) コーディングによる処理
4. 本手法のマクロ分析への拡張について
 - (1) マップの試作
 - (2) 本手法のマクロ分析への拡張
5. まとめ

マップやランキングマップ等が用いられる。セミマクロ分析には、課題-解決手段マップや、課題-用途マップ等が用いられる。ミクロ分析には、クレームマップや構成要件対比表等が用いられる。

分析	目的	分析対象	分析件数	マップ
マクロ分析	自社他社 動向分析	書誌的事項	1000件程度	-時系列マップ
				-ポートフォリオマップ
				-ランキングマップ
セミマクロ分析	自社他社 技術分析	明細書から抽出された用語	100件程度	-課題・解決手段マップ
				-課題・出願人マップ
				-課題・用途マップ
ミクロ分析	権利関係分析	請求項の文言	10件以下	-構成要件対比表

(図1 特許情報分析に使用されるマップ)

1. はじめに

(1) 特許マップについて

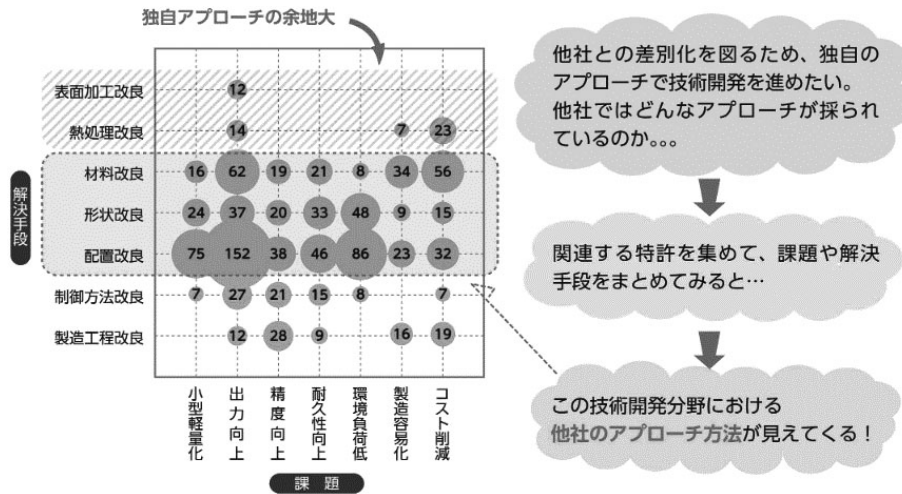
図1は、特許情報分析に使用されるマップを示す。特許マップとは、特許情報分析の結果を図表により「見える化」したものである。図1に示すように、特許情報分析には、書誌的事項（特許分類、出願番号、公開番号、特許番号、出願日、出願人名等）に基づき、1000件程度の出願を対象とした分析（マクロ分析）と、課題、解決手段等に関するデータ・コーディングを行い、100件程度の出願を対象とした分析（セミマクロ分析）と、権利判断に関する分析（ミクロ分析）とがある。

マクロ分析には、時系列マップやポートフォリオ

(2) 課題・解決手段マップについて

図2は、課題・解決手段マップの例を示す。課題・解決手段マップは、セミマクロ分析の結果を「見える化」したものである。課題・解決手段マップは、「課題用語」と「解決手段用語」を縦軸、横軸とした二元表である。

課題・解決手段マップによれば、二元表中の特許出願件数を比較することにより、課題と解決手段の関連性の強弱を判断できる。さらに、二元表中の課題と解決手段に対応する特許出願を分析することにより、自社技術の特許性や抵触性の判断や、自社他社の出願状況を把握できる。それらに基づいて、自社の技術開発計画や特許出願計画を策定できる。



(出典：平成 30 年度 (独) 工業所有権情報・研修館委託事業 はじめての特許情報分析活用に向けて パンフレット 「特許マップってなんだろう？」 事務局 一般社団法人発明推進協会 調査管理チーム)

(図 2 課題-解決手段マップの例)

(3) 課題・解決手段マップの作成法について

図 3 は、課題-解決手段マップの作成手順を示す。課題・解決手段マップは、特許明細書から抽出した課題や解決手段の用語に関するコーディングを行い二元表の各軸とする。コーディングの手法は以下の通りである。

まず、特定商品について特許情報調査を行う (STEP1)。ヒットした出願の明細書等から解決手段用語を抽出して KJ 法を使用してまとめる (STEP2)。ヒットした出願の明細書等から課題用語を抽出して KJ 法を使用してまとめる (STEP3)。特許マップソフトを使用して「課題用語」及び「解決手段用語」に対応する特許出願の件数を二元表にまとめる (STEP4)。これにより課題-解決手段マップは完成する⁽³⁾。

KJ 法は、1960 年代に考案された手法である。KJ 法では、用語をカードに記述し、それらカードをグループごとにまとめて複数のグループを作り、各グループの関連性に応じて系統図等を作成する作業を行うもの

である⁽⁴⁾。

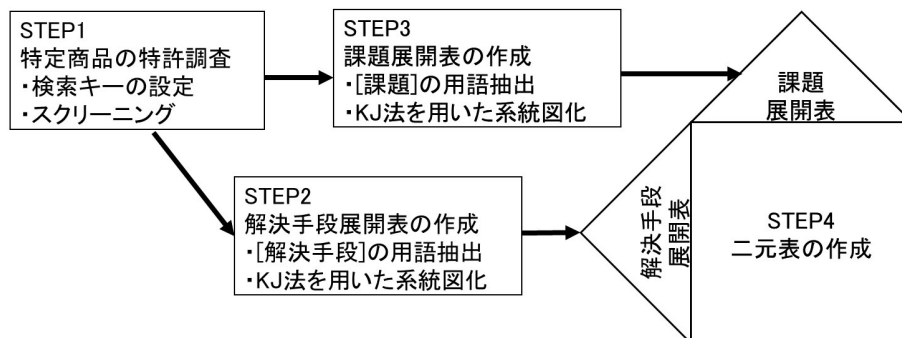
なお、社内に整理した製品の課題や解決手段の分類 (社内分類) がある場合には、KJ 法によらず、それらをそのまま各軸に使用してもよい。

(4) 本稿の目的

課題・解決手段マップを作成するためには、人間が 1 件 1 件明細書を読み込んで、その意味内容を理解したのち、「課題用語」と「解決手段用語」を明細書から抽出する作業を行うことになる。この人手による抽出作業には、以下の 2 点の問題がある。

①作業時間の問題

人間が明細書を読み込んで、「課題用語」と「解決手段用語」を抽出する場合には、熟練者でも明細書 1 件当たり、5~10 分程度の時間が必要である。そうすると明細書 100 件処理するには 9~18 時間程度の時間が必要となる。これでは、多忙な知財担当者等が、業務



(図 3 課題-解決手段マップの作成手順)

として行うのは難しい。作業時間に制限がある場合には、分析できる明細書の件数も少なくせざるを得ない。

②複数の作業者によるばらつきの問題

作業時間を短縮するために、複数人で作業を分担することが考えられる。しかし、各人の用語抽出処理の得手不得手により、抽出された用語に、ばらつきが生じやすい。したがって、単純に各人の作業結果を組み合わせることはできない。このことから、各人の作業結果をすり合わせして統合する作業がさらに必要となる。

このような課題を解決するために、本稿は、「課題用語」と「解決手段用語」の抽出作業を、テキストマイニングの技術を利用して、ある程度自動化することにより作業を効率化することを目的とする。

2. テキストマイニングについて

(1) テキストマイニングとは

テキストマイニングとは、文書形式のデータを定量的な方法で分析することをいう。テキストマイニングでは、コンピュータの処理によってデータの中から自動的に言葉を取り出し、さまざまな統計手法を用いた探索的な分析を行う。これにより、パターンやルール、ひいては知識の発見を目指すものである。

(2) 使用ソフトについて

本稿では、KH Coder を使用してテキストマイニングを行う。KH Coder は、樋口耕一氏により作成されたテキストマイニング用のフリーソフトウェアである⁽¹⁾。フリーソフトであることから、無料で、学術的な目的であれ商業的な目的であれ、自由に使用できる。

KH Coder は、<http://kncoder.net> からダウンロードできる。KH Coder については、KH Coder に添付されるマニュアルや、さらに詳しい解説書⁽¹⁾や、やさしく書かれた解説書⁽²⁾が存在する。これらを参照することにより、容易に操作を理解できる。

なお、本稿では、KH Coder を使用することを前提に記載されているため、操作について不明な点があれば、これら解説書を適宜参照いただきたい。

3. テキストマイニングを使用した課題・解決手段マップの作成手順

ここでは、分析対象を「おむつ」とした場合の課題・解決手段マップの作成手順について説明する。分析対象を「おむつ」としたのは、前稿⁽⁵⁾でのコーディングを利用できるためである。

(1) 抽出語による処理

①特許調査の実施

特許情報は、J-PlatPat や各種有料特許 DB を利用して収集できる。本稿では、検索対象を、出願人：A 社、対象製品：おむつとして検索した。検索によりヒットした特許出願のうち、40 件の特許出願を分析の対象とした。

本稿では特許情報として、「要約書」の情報を収集した。要約書は、「課題」と「解決手段」が項目立てて記載されているため、課題と解決手段の関係を把握するために都合が良い。

②データの加工

図 4 は、特許調査で抽出された 40 件の要約書の内の、任意の 3 件の要約書のデータを示す。これらデータは KH Coder 用に、下記の加工がおこなわれている。

図 4 に示すように、要約書から、不要な語句である、項目名【課題】、【解決手段】及び【選択図】を削除する。さらに、【課題】と【解決手段】の間の改行を削除し、要約書全体として一段落化する。なお、これらの加工は、表計算ソフトの機能を利用すれば容易にできる。このような加工を行った 40 件の要約書データを csv 形式で保存する。

使い捨てのパンツ型着用物品における前胴周り域と後胴周り域との接合域による肌への刺激を防止する。使い捨てのパンツ型着用物品 1 の前胴周り域 6 と後胴周り域 7 とが、互いの側縁部 8、9 で重なり合い、溶着部 13 において一体化して胴周り側部域 10 を形成する。胴周り側部域 10 における着用者の肌に対する当接面には、前後胴周り域 6、7 の間をまたいで前後胴周り域 6、7 に接合しており、溶着部 13 の肌への接触を防ぐことが可能な隠蔽用シート片 31 が設けられる。

尿取りパッドとして使用するのに適したものであり、少量の尿が排泄されたとき、および比較的覆う尿が排泄されたときの双方において、尿を確実に捕捉しやすい吸収性物品を提供する。バルブと高吸水性樹脂とで形成された吸収体 13 の表面に液透過性シート 20 が設けられ、この液透過性シート 20 により、弾性部材 25 を有する肌当て部 21 と、側方固定部 22 と、立ち上げ部 23 が形成されている。この立ち上げ部 23 には撓み余裕部 23a が形成されている。少量の尿が与えられたときには、肌当て部 21 が肌に密着することで吸収でき、比較的多く与えられた尿は空洞部 26 で貯められた後に吸収体 13

で吸収される。
前後パネル部を切り離した後にそれらを再び連結する締結手段を、使用時まで使用可能、かつ安全な状態に保つことができるパンツ型の使い捨て着用物品の提供。使い捨ておむつ10が、連結された前パネル部16および後パネル部17と、それら前後パネル部16、17を切り離すことが可能であって、胴周両側部18、19に設けられた切断部22と、切り離された前後パネル部16、17を再び連結させることが可能であって、切断部22に近接するように後パネル部17の内面14の側に設けられた締結部材23とを有し、切断部22と締結部材23とを覆う被覆部材24が、後パネル部17の内面14の側に設けられている。

(図4 テキストマイニングに使用する要約書データ)

③前処理の実行

加工した要約書データを KH Coder へ読み込み、前処理を実行する。前処理を実行することにより、要約書データについて形態素解析が行われる。そして、テキスト全体から分析対象となる品詞別の抽出語が得られる。

図5は、前処理により得られた抽出語を示す。図に示すように品詞ごとに抽出語が得られる。各抽出語の右にある数字は、出現回数である。なお、後の説明のため「吸収」の部分に灰色で塗りつぶした。

④複合語の処理

形態要素分析によれば、複合語も複数の用語に分解される。しかし、複合語として特有の意味を持つ用語もある。KH Coder には複合語を抽出できる機能があるため、これを利用して複合語を抽出する。

名詞		サ変名詞		形容動詞		副詞可能		未知語		タグ		動詞	
方向	107	吸収	75	可能	21	前後	20	ファスニングテープ	25			延びる	28
領域	56	形成	39	主	5	それぞれ	6	バリアシート	15			設ける	27
胴回り	53	前記	28	切	5	場合	6	ファスニング	12			有す	23
使い捨て	48	接合	27	确实	4	少量	2	吸	10			有する	18
ウエスト	43	提供	27	安らか	3	全体	2	クロッチ	9			備える	15
シート	38	位置	25	簡易	3	当たり	2	カフ	7			織る	10
部材	31	着用	25	容易	3	その後	1	ファスナ	7			覆う	10
弾性	29	伸縮	14	自由	2	一部	1	ファスニングシステム	7			含む	9
部分	22	伸長	12	適切	2	一方	1	ウエブ	6			沿う	8
物品	22	展開	12	安全	1			凹	5			及ぶ	7
両側	18	配置	11	安定	1			設	4			取り付ける	6
長手	17	圧縮	10	好適	1			シャーシ	3			曲げる	5
周り	16	使用	10	自在	1			テープファスナ	3			欠く	5
一対	15	切り取り	10	自然	1			該	3			出る	5
状態	15	連結	10	十分	1			MD	2			並ぶ	5
要素	14	構成	9	縦長	1			インジケーター	2			与える	5
コア	13	排泄	9	不快	1			レッグ	2			重ねる	4
パネル	13	固定	8	不揃い	1			圧	2			係る	3
パンツ	13	フィット	7	密	1			11.755gf	1			受ける	3

(図5 品詞別の抽出語)

複合語	スコア
吸収性物品	659.6804811
ウエスト域	539.8857054
着用者	402.6039928
胴回り部	320.5345554
吸収体	262.1964789
幅方向	253.4272268
長手方向	230.388388
吸収性コア	209.5700255
パネル部	119.170253
胴周り域	119.0037093
縦方向	118.4803773
胴回り域	117.934022
横方向	115.6486178
横方向X	108.9569452
パンツ型着用物品	96.31447099

(図6 抽出された複合語の図)

図6からわかるように、「吸収性物品」という用語の出現スコアが大きい。要約書には、請求項1をそのままコピー & ペーストすることが一般的である。請求項の末尾は「発明の名称」であることが多い。したがって、「吸収性物品」は「発明の名称」と推測される。これにより、「吸収性物品」が多く出現していると思われる。このままでは、「吸収」という用語の出現数が過大となるおそれがある。

これを防ぐためには、「吸収」と「吸収性物品」とを、別の用語として抽出する必要がある。これには、前処理において、「吸収性物品」を強制抽出する指定を行い、再度、前処理を実行すればよい。

体は文脈から切り離されているため、共起ネットワーク図が示す意味を理解することは難しいと思われる。

(2) コーディングによる処理

① コーディングルールの作成

コーディングとは、仮説、テーマに基づいて、抽出語を組み合わせることをいう。コーディング作業を行う前には、いくつかのカテゴリー（以下、コードという）を作成し、「特定の記述がデータ中にあるならばそのデータを特定のコードに分類すること」といった基準を作成する。この基準のことをコーディングルールという。

図9は、コーディングに使用したコードを示す。コードは、テキストマイニングにより抽出した用語をKJ法を使用してまとめてもよい。テキストマイニングソフトのクラスター分析を用いて特徴語を抽出し、コードとしてもよい。社内に自社技術分類がある場合には、それをコードに使用しても良い。

本稿では、前稿⁽⁵⁾で抽出された、図9に示す課題と解決手段をコードに使用する。なお、本稿では「解決手段」を「構成要件」としているが、分析の目的に応じて、「解決手段」を「機能」としてもよい。

課題	解決手段
・柔軟性	・脚回部
・フィット性	・吸収体部
・通気性	・シート部
・取扱性	・股下部
・吸収性	・デザイン部
・感触性	

(図9 コーディングに使用したコード)

図10は、コーディングルールを示す。コーディングルールの作成は、図7に示す形態素解析により抽出された多数の語を、各コードに当てはめてゆけばよい。コーディングルールは、csv形式で保存する。

図10に示すコーディングルールは、単純に抽出語の「or」を取っている。コーディングルールには、他の算術演算子や論理演算子も使用できるので、より複雑なコーディングルールを作成することも可能である。

抽出語を各コードに当てはめるには、個々の抽出語の意味を把握する必要がある。意味が明確な場合には、そのまま当てはめを行えばよい。この場合、作業者が明細書を読み込んで意味を確認する作業を省略できる。意味が不明な場合には、明細書の記載に戻って

確認したり、技術文献やインターネット検索により確認すればよい。

図10のコーディングルールは、あくまでも一例である。複数の関係者により、使用するコードや、コードと抽出された用語の対応関係の検討を行って、コーディングルールの精度を高くすることが好ましい。

*柔軟性	弾性 or 変形
*フィット性	伸縮 or 伸長 or フィット or 密着 or 屈曲
*通気性	透過 or かぶれる
*取扱性	使い捨て or 取り外し or 視覚 or 装着 or 視認 or 接着 or 廃棄 or におい
*吸収性	吸水 or 吸収 or 漏れ
*感触性	感触 or 痛み or 接触 or 刺激 or 圧迫 or 不快
*脚回部	胴回り or ウエスト or ギャザー or 大腿 or サイドフラップ
*吸収体部	コア or パッド
*胴回部	腹部 or テープ or ファスニングテープ or ファスニング or ファスナ or ファスニングシステム or テープファスナ or テープファスナタブ
*シート部	シート or バリアシート
*股下部	股下 or 股間 or クロッチ
*デザイン部	外観 or PANTONE or インジケーター

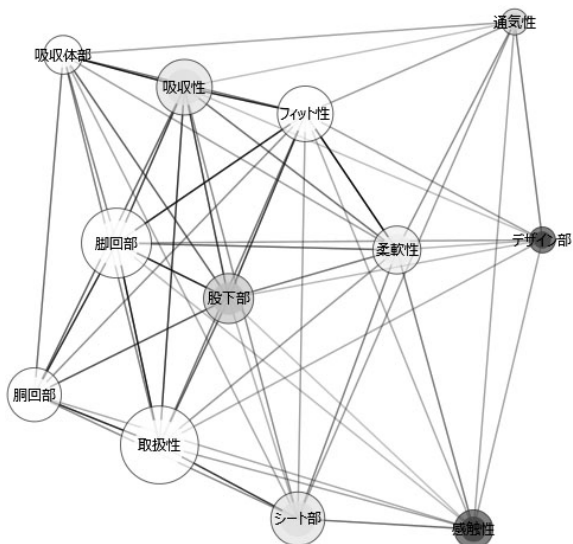
(図10 コーディングルール)

② 共起ネットワーク図の出力

図11は、課題コードと解決手段コードの共起ネットワーク図である。共起ネットワークは、図10のコーディングルールに基づいて、課題コードと解決手段コードのJaccard係数を計算したものである。共起ネットワークは、課題コードと解決手段コードの類似度行列を視覚化したものである。

図8の抽出語に基づく共起ネットワーク図と比較した場合には、各コード自体の意味が明らかであり、用語が整理されていることから、比較的理解しやすい図となっている。

共起ネットワーク図は、従来からの特許情報分析を行っている者には、なじみが薄く、従来の二元表とした方が分析に使いやすい場合もあると思われる。さらに、QFD等の他の分析手段と組み合わせた分析^{(3),(5)}を行う場合にも、二元表とすることが好ましい。



(図 11 コードに基づく共起ネットワーク図)

③コードに基づく類似度行列の出力

図 12 は、課題コードと解決手段コードの類似度行列を示す。類似度行列は、図 10 のコーディングルールに基づいて、課題コードと解決手段コードの Jaccard 係数を計算して生成されたものである。類似度が高いほど、高い共起関係にあるといえる。

類似度行列は、課題コードと解決手段コードとの間のみならず、各課題コード間、及び、各解決手段コード間においても計算される。二元表化にあたっては、各課題コード間、及び、各解決手段コード間の類似度は不要であるため、削除する。これにより、課題コードと解決手段コードとの間のみの類似度を示す行列となる。

	柔軟性	フィット性	通気性	取扱性	吸収性	感触性
脚回部	0.25	0.444	0.038	0.459	0.345	0.069
吸収体部	0.125	0.222	0.111	0.194	0.467	0
胴回部	0.042	0.16	0	0.419	0.261	0.105
シート部	0.19	0.115	0.133	0.375	0.16	0.235
股下部	0.211	0.35	0	0.273	0.35	0.056
デザイン部	0.077	0.125	0.2	0.1	0.059	0.111

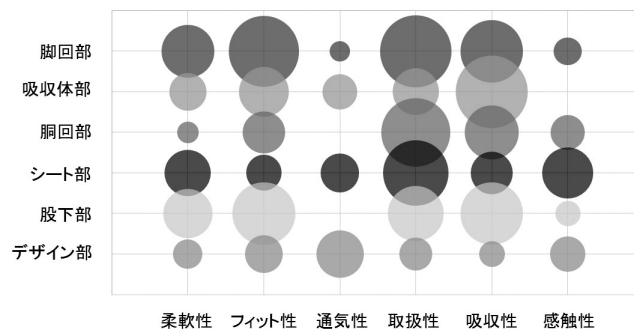
(図 12 課題コードと解決手段コードの類似度行列)

④二元表の作成

図 13 は、課題コードと解決手段コードの類似度行列から泡グラフ（バブルチャート）を作成したものである。本マップは、所定の作成法⁶⁾により、市販の表計算ソフトで作成した。このようにして、課題・解決手段マップを作成できる。

本手法によれば、人間による作業はコーディングの

部分だけであり、他の処理はコンピュータが実行するため、課題・解決手段マップを効率よく作成できる。



(図 13 課題・解決手段マップ)

4. 本手法のマクロ分析への拡張について

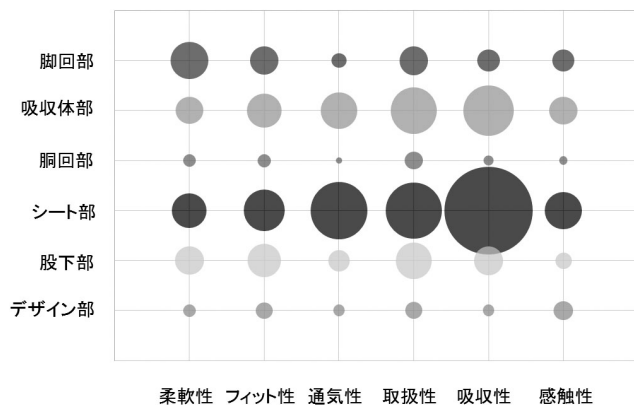
(1) マップの試作

KH Coder は、上限で文書数 10,000・抽出語数 100 といった程度の処理も可能とされる。ソフトによる処理時間を確認するために、40 件より要約書の件数を多くして課題・解決手段マップを試作した。

図 14 は、FI: A61F13/511.100 (おむつ) にてヒットした特許公開公報 1287 件の要約書について、上記手法にて作成した課題・解決マップを示す。同様の製品について分析を行うことから、コーディングルールは図 10 で示したものをそのまま使用した。

1287 件の要約書の前処理や類似度行列生成等のコンピュータによる処理に要した時間はほぼ 2 分であった。このように、件数を 1287 件とマクロ分析領域まで多くしても無理なくマップを作成できる。

従来のセミマクロ分析は、100 件程度の明細書を対象とすることが通常であった。これは、人手による抽出作業に時間を要し、件数を増やすことができないからである。本手法によれば、セミマクロ分析でも、分析件数をマクロ分析の領域まで拡張できる。



(図 14 課題・解決マップ)

(2) 本手法のマクロ分析への拡張

マクロ分析では、技術内容を分析するために、IPC、FI、Fターム等の特許分類が使用される。これは、特許分類は書誌的事項として明細書から容易に抽出できるからである。

しかし、特許分類は、先行技術調査のために作成されたものであり、技術内容の理解に必ずしも適したものではない。さらに、特許分類は、記号で記載されるから、一目見て技術内容を理解できるものではない。特に、経営層や投資家等の特許を専門としない人間に説明する資料に、特許分類は使用しにくい。

本手法を用いることにより、特許マップに使用する用語は、課題用語や解決手段用語等の意味を把握できる用語となるから、マクロ分析の内容も、より理解しやすくなる。特に、特許を専門としない人間に説明する資料に使いやすい。

今後は、テキストマイニング技術及びコーディング技術の発展に伴い、特許分析に特許分類を使用することが少なくなってゆくことが予想される。

5. まとめ

(1) 特許マップの作成にテキストマイニングを活用することにより、人間の作業はコーディングルールの作成のみとなり、他の処理はソフトウェアが実行することとなる。これにより、特許マップ作成の工数を削減することが可能となる。

(2) 作成したコーディングルールは、次回の特許マップ作成にも使用できるため、回次の作業からは、コーディングルールの作成は不要となり、ソフトウェアによる処理のみで、特許マップを作成できる。

(3) 人手で課題・解決手段マップを作成する場合には、明細書 100 件を処理するのがせいぜいであったが、テキストマイニングの技術を活用することにより、より大きな件数を処理できる。

(4) コーディングルールの作成は、「コード」と「抽出された用語」の関連のみ把握できれば良いため、従来のコーディングのように、作業者が個々の明細書を読み込んで課題や解決手段を見出す作業を減らすことができる。

(5) 本稿における処理は、KH Coder というフリーソフトと市販の表計算ソフトのみで実行できるため、誰でも低コストで容易にテキストマイニングを活用した特許マップを作成できる。

(参考文献)

- (1) 樋口耕一, 「社会調査のための計量テキスト分析 内容分析の継承と発展を目指して」(2014), 株式会社ナカニシヤ出版
- (2) 牛澤賢二, 「やってみようテキストマイニング 自由回答アンケートの分析に挑戦!」(2018), 朝倉書店
- (3) 鶴見隆, 川上成年, 谷津維則, 「戦略的商品手法の開発 QFD と特許情報の融合」, 『知的財産イノベーション研究の諸相』, pp.76-112 (2014), 日本知財学会知財学ゼミナール編集委員会編, コンテンツ・シティ出版
- (4) 川喜田二郎, 「発想法 - 創造性開発のために」(1967), 中公新書
- (5) 川上成年, 谷津維則, 「ブランド QFD を活用した調査手法の開発」, パテント, Vol.70, 2017.3, pp.82-89 (2017)
- (6) 野崎篤志, 「経営戦略の三位一体を実現するための特許情報分析とパテントマップ作成入門」, pp.185-193 (2016), 一般社団法人発明推進協会

(原稿受領 2018. 12. 10)