テキストマイニングによる知財ポートフォリオ分析

Patent portfolio analysis by text mining

中居 隆¹

NAKAI Takashi¹

.....
1 NRIサイバーパテント株式会社 企画営業部 (〒100-0005 東京都千代田区丸の内1-6-5) Tel : 03-5208-1570 Fax : 03-5208-5550
E-mail : t-nakai@mri.co.jp

1 NRI Cyber Patent, Ltd. (1-6-5 Marunouchi Chiyoda-ku, Tokyo 100-0005)

原稿受理 (2008-04-01)

(情報管理 51(3):194-206)

1. はじめに—知財ポートフォリオの的確な把握が、知財経営の第一歩

これまで日本では、研究開発の成果である特許について、出願・保有の「数」が少なからず重視されてきた。国内企業どうしでは、数に基づくクロスライセンスが頻繁に行われ、企業は、ライバル企業の権利行使を阻止するための布石として、一定の「数」を押さえる必要があった。例えば、他社が強力な基本技術を発明しても、自社が周辺の製造技術の特許として多く押さえれば、少なくとも相手企業は、事業を展開する上で、十分に独占的な権利主張はできなくなる。こうした背景から、企業による出願は増え続け、自社が保有する特許の状況すら的確に把握・管理することが難しい状況にある。

さらに、多くの企業において、経営層、研究開発の現場、知財部門の間で、自社の知的財産の状況認識にずれがあることが指摘されている。研究開発の現場は、経営層が技術を理解していないことに不満を抱き、経営者は、研究開発の現場は事

業化に向けた意識が足りないと嘆く。自社の知的財産を取り巻く環境、すなわち知財ポートフォリオを的確に把握し、社内の共通認識を醸成することは、知財経営の実現に向けた第一歩と言える。

近年、企業活動において無形資産の重要性が高まっている中で、研究開発投資の効率性、収益性が問われている。やみくもに出願件数を追求するのではなく、自社のコアコンピタンスに対しては、積極的に権利化を進め、それ以外の領域については、外部の資源も活用するといった、より柔軟で高度な判断が求められている。自社の事業戦略、研究開発戦略との関係において、いかに、知財ポートフォリオの最適化を図っていくかという視点が求められる。

現在、自社・同業他社の知財ポートフォリオを的確に把握するための手段として、テキストマイニングの技術・ソフトウェアを使って、大量の特許群の公報データを解析し、可視化する手法が注目されている。本稿では、その背景と経緯、具体例に基づく分析の流れ、特長、今後の課題、展望について記す。

図1 検索系と分析系のテキストマイニング

図2 テキストマイニング～可視化の技術構成

図3 文章の構造（構文）解析

2. テキストマイニングによる知財ポートフォリオ分析

2.1 分析系のテキストマイニングは、マーケティング分野から発展

日本において、テキストマイニング技術は、検

索系と分析系の2系統で発展してきた。検索系は、WWWの検索サービスなど、インターネットの広がりとともに、企業だけでなく、生活者にとっても不可欠の道具として定着している。また、特許や文献の検索サービスでは、1,000万件以上の公報の全文を対象に、指定された条件に基づき、極めて高速に、漏れのない抽出処理を行えることが求められる。

一方、分析系のテキストマイニングは、主にマーケティング分野を中心に、活用が広がってきた。顧客からの苦情や問い合わせを受けるコールセンターに寄せられる声や、アンケート調査のフリー回答を解析することで、顧客の声を直接、製品・サービスの開発・改善につなげるものである。

製造業においても、品質リスクの監視体制として、顧客の声の異変をいち早く察知し、迅速に対応するためにテキストマイニングを活用しているケースがある。こうした分析系のテキストマイニングでは、分析者の問題意識、仮説等に応じて、より深く文章の意味を解析するといった、検索系とは異なる性能・機能・操作性が求められる（図1）。

2.2 テキストマイニングによるポートフォリオ分析の技術構成

テキストマイニングによるポートフォリオ分析の技術は、大きく、テキストマイニング技術と可視化技術から構成され、さらにテキストマイニング技術は、自然言語処理のプロセスとデータマイニングのプロセスに分かれる（図2）。

2.2.1 テキストマイニング技術—自然言語処理

大量のテキスト情報について、そのままではコンピュータ処理しづらい自然文（文章）をその構成要素や構造、意味を抽出し、解析するものであ

図4 テキストマイニングによる特許ポートフォリオ分析

る。文章から単語を切り出し、さらに単語間のつながりの関係（構文）を解析することで、単語の品詞、単語と単語の関係（「係り受け」）をとらえ、詳細な分析を可能とする（図3）。

2.2.2 テキストマイニング技術—データマイニング

統計処理や人工知能等の手法を用いて、大量のデータから、一定の相関関係、パターンを見つけ出す分析技術である。有名な事例は、ウォルマートがPOSデータをマイニングしたところ、金曜日の夕方は、ビールとオムツを一緒に購入する男性客が多いことを発見したというもので、一見、無関係そうで、人の経験や勘では気づかない発見をもたらすものである。

自然言語処理とデータマイニングが組み合わさったテキストマイニングは、自然文をコンピュータ処理しやすいようにデータ化した上で、分析する技術と言える。特許分析の場合には、①出願人や出願年、発明者などの書誌的事項、法的ステータスを表す審査経過情報と、②特許公報の文章から切り出した単語と係り受けの相関を、データマイニングにより見いだすものである（図4）。

図5 知財ポートフォリオ分析の流れ

2.2.3 可視化技術

知財ポートフォリオに対する、経営層、研究開発の現場、知財部門の共通理解を支援するためには、テキストマイニングで分析した結果を、いかに分かりやすく、効果的に見せられるかが重要となる。単に、クロス集計の結果をグラフや表形式で示すだけでなく、ネットワーク図として表現したり、言葉をベースとした、テキストマイニングならではのマッピング手法も広がりを見せている。

3. 知財ポートフォリオ分析の実際

ここでは、注目が集まる有機ELに関する日本特許を対象に、株式会社野村総合研究所（NRI）グルー

ブが提供する知財ポートフォリオ分析専用ソフトウェア「TRUE TELLERパテントポートフォリオ」で実際に分析を行った内容をもとに、テキストマイニングによる分析のプロセスと特長を説明する(図5)。

3.1 分析対象文献の決定

テキストマイニングでは、分析の目的に応じて、特許や実用新案の公開データはもちろん、自社内で管理している研究開発の起案情報や、出願後の公開前情報、論文などの文献情報、WWW上に流布する他社の技術情報など、さまざまなデータソースを使い分け、統合して解析することができる。

今回は、日本における、当該分野に関する技術的な取り組みの状況と経緯を特許データから探るために、日本における特許出願情報として、公開公報(公表、再公表を含む)を分析対象とした。

3.2 解析対象テキストの決定

解析対象とするテキストについては、発明の内容・特徴が簡潔に記述され、関連する重要な言葉が網羅的に含まれつつ、一方で、発明の本質とは関係ない言葉は極力含まれないものが望ましい。例えば、公開公報の「要約」は発明の特徴、対象とする課題、解決手段が、簡潔に記載されているケースが多い一方、書き方が厳密にルール化されていないため、出願人・代理人によって、情報の内容にバラツキがある。「特許請求の範囲(請求項)」は、権利範囲が明確に記述されている一方、請求項によって、発明の本質との関係性、重要性が異なる。そのため、第一請求項、あるいは、独立請求項のみを解析対象とすることも考えられる。そのほか、発明の用途を分析したい場合には、「実施例」や「技術分野」も有効である。

特許庁が提供する公報データのXML(Extensible Markup Language)化により、明細書の中から、必要な項目のテキストデータを切り出すことが容易になり、分析の目的に応じたデータの使い分けが可能な環境が整いつつある。

今回の分析では、「発明の名称」+「要約」+「特許請求の範囲(第1項)」を連結したテキストを解析対象とした。

3.3 対象文献群(一次母集団)の抽出

分析の対象となる文献群のデータを、外部の検索システム・サービス、あるいは、企業内の知財管理のデータベースから、分析の目的に合わせた条件で検索・抽出する。

これまで、例えば、企業にとって新規分野への参入可能性を調査する場合など、当該技術に対する予備知識が十分ではない中で、手探りで検索式を立てており、極めて属人的な経験・ノウハウ頼みの部分があった。また、文献群を抽出できたとしても、人が目を通せる件数には限界があるため、「1,000件以内に絞り込むために、検索条件をどう設定するか」という本末転倒とも言える思考プロセスを取らざるを得なかった。

テキストマイニングによる分析では、大量の文献群に対して、ソフトウェアが母集団の文中に出てきた言葉の頻度分布を自動抽出し、この情報をもとに分析を進める。出現単語から、想定外のノイズを見つけ除外したり、逆に、発見し、気付いた関連用語を検索条件にフィードバックするといったプロセスを通じて、抽出条件を精査していくことができる。そのため、この段階では、従来の感覚からすると少し広めに、ある程度のノイズを許容するつもりで、一次母集団を抽出することが有効である。

今回は、国際特許分類やキーワードによる検索

A社(化学系)

B社(電機)

C社(精密機器)

図6 マッピング分析 「サーモグラフ」

式を設定し、民間の特許データベースサービスにて1万件弱の一次母集団データを抽出した。

3.4 分析ソフトウェアへのデータ取り込み—データ処理

テキストマイニングのソフトウェアは、母集団

データを取り込む際に、各文献のテキストに対して自然言語処理を行い、単語や係り受け（主語－述語、修飾語－被修飾語の関係）の抽出を行う。これを、文献データの属性情報（特許の場合、出願番号、出願年、出願人、発明者などの書誌的事項、法的ステータスを表す審査経過情報など）とひも付け、データマイニングが可能なデータベースと

図7 マッピング分析 コレスポネンスマップ

して整理・格納する。

3.5 マッピング等による概要分析

まずは、一次母集団に対して、出現単語・係り受けの頻度分布、言葉に基づくマッピング分析により、母集団全体の技術体系・領域と、出願人、出願年別の出願状況と推移を大枠で把握する。

図6のサーモグラフでは、初めに、母集団の文中に頻出する技術用語どうしが同時に出現する公報の件数が多いほど、マップ上で単語間の距離が近くなるように単語を自動配置し、分析のベースとなる「単語マップ」を作成する。こうして単語の分布によってできた技術的な領域分析を、共通の「下敷き」として、そこに発明者別の文献の分布密度を温度表示のように表すのが、「サーモグラフ」である。A社は電極、輸送層・注入層、発光層、B社は電気系統、C社は高分子の塗布、インクジェット関連にまとまった出願があることが、一目で確認できる。

図7のコレスポネンスマップでは、出願年と単語群の相関を、1枚のマップに表現することで、各年代における特徴的な取り組みを浮き彫りにしている。2002年前後は成膜、塗布技術、2004年以降は「フレキシブル (=柔軟性。電子ペーパー)」「照

明」「封止膜」「印刷法」などの新しい用途や、製造技術の出現が可視化されている。

3.6 対象母集団の見直し—二次母集団の抽出

3.5において、明らかに分析目的と無関係な技術領域や頻出単語が見つかった場合には、テキストマイニングのソフト上で、簡単に母集団から除外（母集団の絞り込み）することができる。一方、重要性が高く、本来、母集団抽出の検索式に加えておくべき単語を発見した場合には、検索サービス・システムにフィードバックし、二次母集団として抽出し、母集団データの入れ替え、あるいは、追加取り込みを行うことになる。

3.7 グループング等による詳細分析

文中に出現した言葉の頻度順リストから、言葉を仕分けすることで、文献群を仕分けし、出願傾向と動向を分析する。「画面が大きい」というように、「画面」「大きい」の1単語では意味が特定されない場合には、「画面—大きい」という「係り受け」を仕分けの条件として指定することで、より精緻な分類が可能となる。

今回、有機ELの「製造工程」と「課題」の二つの観点から仕分けを行い、出願人、出願年との相関を分析した。A社は、各工程、課題に幅広く出願している。一方、B社は「成膜」「封止」の製造工程、課題としては「耐熱性」、C社は、「フレキシブル性」への取り組みが特徴的である（図8）。

このように、分析者の問題意識・仮説に基づき、さまざまな観点（課題・要素技術・機能・用途・製造方法等）から文献群を仕分けすることで、意思決定に資する詳細なポートフォリオ分析が可能となる。これまで、主に発明審査のために付与されたFI (File Index) やIPC (International Patent

観点1 「製造工程」	単語・係り受け(O-O)	観点2 「課題」	単語・係り受け(O-O)
ITO陽極	ITO、陽極、透明電極・・／ (透明)-(電極)・・係り受け	消費電力	消費電力、省電力、低電力・・／ (電力)-(消費)・・係り受け
蒸着成膜	真空蒸着、加熱蒸着、ビーム蒸着、 シャドーマスク	大型化	大型化、大画面、大面積・・／ (画面)-(大きい)、(画面)-(広い)・・係り受け
封止	封止、乾燥剤、接着剤、 キャップ・・	フレキシブル性	フレキシブル、折り曲げ、柔軟性・・

図8 分類条件と分析結果

Classification) 等の特許分類による分析が、企業の意思決定の材料としてなじみにくかったのは、分析者の抱える問題意識・仮説を、分類の視点・切り口として反映しにくいことが一つの理由と考えられる。

3.7.1 テキストマイニングによる分類手法の特長

文献群の記載内容に基づき、仕分けして分析する手法は、企業によっては、これまでも「めくり」という言葉に表されるように、多くの人手と時間をかけて行われてきた。しかしながら、文献数に比例して、時間と労力がかかるため、現実的には、対象件数に限界があった。

もう一つの問題は、新規領域の調査など、分析担当者が当該技術に対してまだ十分な知識がない中でも、あらかじめ、仕分けのルール＝「箱」を決めなければならない点にある。また、いったん「箱」を決めても、途中まで読み進めたところで、新たな「箱」の必要性、あるいは、何度も出てきた言葉の重要性に気づいた場合には、再度、「箱」

の体系を見直し、1件目から改めて目を通していくといったこともしばしばであった。テキストマイニングによる仕分けの手法は、出現単語を頻度順に見ながら、仕分けの「箱」を作り、同時に仕分けを行っていくことに特徴がある。また、一つの出願が、複数の機能、課題、技術要素から構成されている場合もあるが、重複を厭わずに仕分けを行っていくことができる。さらに、どこの「箱」にも入らなかった文献群の頻出単語を確認しながら、新たな「箱」を作るといったことが簡単にできるのも、専用ソフトウェアならではの手法と言える。

さらに、一度作成した仕分けのルールは、企業にとって貴重な資産となる。例えば、研究開発を次のステップに進めるか否かの意思決定を行う際には、スタート時に行った分析に、その後の新着文献を追加データとして取り込むことで、自動的に仕分けを行い、差分のみを確認できる。新着文献に初めて出現した単語を自動抽出し、仕分けの条件を更新していけば、有形のノウハウ、企業にとっての蓄積として育てていくことができる。

利用	単語	頻度
☐	liquid crystal display device	10
☐	organic EL display device	6
☐	input data calculation result	6
☐	portal image display information	5
☐	liquid crystal display panel	4
☐	Al alloy electrode layer	4
☐	parallel light flux group	4
☐	multilayer film optical filter	3
☐	active matrix display device	3
☐	light shielding sealing agent	3
☐	stereoscopic image display screen	3
☐	internal total reflection condition	3
☐	optical waveguide base material	3
☐	pixel display light Ld	3
☐	real time electronic display	2

図9 英語の連語抽出

3.7.2 自動分類, 半自動分類

テキストマイニング技術には、単語の出現傾向の類似性から、文献群を完全自動で仕分けする「クラスタリング」と呼ばれる手法がある。極めて短時間で手軽に仕分けができる反面、ロジック・プロセスがブラックボックス化し、明解な説明・検証が難しくなるため、企業における意思決定の材料としては活用場面が限られる。

また、例えば、自社の特許群に対して、あらかじめ付与された仕分けコードに基づき、他社の特許群を仕分けしていく、いわば「半自動分類」の手法がある。これは、各分類別の自社特許群の文献中に特徴的に出現する特徴語と、特徴の程度の評点をソフトウェアが自動算出し、これをもとに、他社の特許群を、内容的に近い分類に仕分けするものである。ソフトウェアが自動抽出した特徴語に対して、分析者が一般的と思われる言葉を除外するなど、手を加えると、当該母集団において仕分けの精度はさらに向上する。

テキストマイニングによる仕分けの手法では、分析に掛ける時間・労力と、結果の精度はトレードオフの関係にある。分析の目的、求められる精

度、かけられる時間等に応じて、手法の使い分けが求められる。

3.8 ユーザ辞書のメンテナンス・分析精度の向上

テキストマイニングによる分析では、ソフトウェアが保有する言葉のデータ（辞書）が分析のベースとなる。しかしながら、極めて専門性の高いテクニカルタームや、新たに出現する技術用語をカバーすることには限界があり、これを補う形で、ユーザ固有の辞書を簡単に追加・メンテナンスできることが求められる。辞書の追加・メンテナンスを支援する機能として、「母集団の文中に何回も出てきた、ソフトウェアが保有する辞書にない文字列」を言葉の候補として自動抽出し、これを分析者の判断で登録することにより、ユーザ独自の辞書を育てていくことができる。これによって、より精度が高く、企業の意思決定に資する分析が可能となる。

なお、言語の違いによる精度向上のポイントの一つとして、日本語の場合には、文章をどこで区切って単語を抽出するかが問われるが、英語の場合には、複数の単語をどこまでつなげて連語として扱うかが問われる（図9）。

3.9 分析結果から元データを確認—マクロ⇄ミクロの分析

知財ポートフォリオの分析では、ソフトウェアが作成したマクロな分析結果に、即座に納得して分析完了ということではなく、逆に、ミクロに分析結果を確認・検証しながら、分析を深掘りしていくプロセスが不可欠である。分析ソフトウェアには、マクロ⇄ミクロな分析・確認を円滑に行うことができる機能性・操作性が求められる。特許検

索のサービス・システムや社内データベースとリンクすることで、最終的には、特許公報などの元データまで戻って、関連図面等も参照できるとな
お望ましい。

3.10 アウトプットの作成（絞り込み）

知財ポートフォリオの分析では、母集団全体でのクロス集計（出願人×出願年，出願人×課題）だけでなく，例えば，出願人を特定した上で，課題×出願年，課題×要素技術といったクロス集計を行っていけば，考えられる分析の組み合わせは大幅に増える。ソフトウェアを利用すれば，こうした大量のアウトプットを簡単に作成することができる。

しかしながら，分析の目的，問題意識や仮説に照らした時に，本当に必要なアウトプットは限られることが多い。経営の意思決定に資する分析結果を発信する場合には，本当に必要十分なアウトプットに絞り込まなければ，大量の情報にあふれている経営者の琴線に触れることはないだろう。

4. テキストマイニングによるポートフォリオ分析の課題と展望

4.1 現場と知財担当者のコワークがより重要に

誤解を恐れずに言えば，テキストマイニングによる知財ポートフォリオ分析は，一つボタンを押せば，すぐに有用な分析結果と示唆が得られるといったものではない。言葉をもとに分析を行うことから，当然ながら，分析担当者には「技術がわかる『目』」が求められる。3章で説明した各プロセスにおいて，分析担当者の知識・知見が反映されることで，初めて，真に役立つ分析が可能となる。

図10 発明者のネットワーク分析

分析結果を実際に活用し，自社の事業戦略，研究開発戦略の判断を行うのは企業の経営層や，研究開発の現場である。一方，多くの企業においては，こうしたソフトウェアを，知財業務の主管部署であり，知財データの扱いに慣れている知財部門，あるいは，研究開発の企画部門・知財担当で導入・利用している。しかしながら，技術が複雑化・多様化・短サイクル化する中で，知財担当者が最先端の技術を広く理解し続けることには限界があり，現場の担当者を巻き込んで分析を進めることが求められる。知財スタッフには，これまで以上に，アウトプットの活用者の問題意識や仮説を理解・整理し，最適な分析方法を提案し，技術者の知識・知見を引き出しながら，分析を進めるといった，コミュニケーション力やコンサルティング力が求められるようになる。知財担当者，調査担当者の，新たな腕の見せ所と言えよう。

4.2 テキストマイニング技術—今後の課題と発展の方向性

テキストマイニングの活用は，コンピュータの情報処理性能の向上とともに，企業のさまざまな業務の中に広がってきた。分析の根幹は，自然言

語処理の性能にあり、今後の課題・発展の方向性として、以下の点が挙げられる。

- ・大量データ分析のための情報処理速度のさらなる向上
- ・長文分析のための「要約」機能（重要文、重要語の抽出）
- ・文章の乱れ、企業固有の言い回し、単語の省略形などへの対応（ノイズ除去、照応解析）
- ・仮説に基づく半自動分類、分類支援のさらなる向上
- ・化学式、化合物名などの抽出精度の向上

4.3 可視化技術の進化—テキストマイニング×ネットワーク分析

図10は、一企業の出願データから、共同発明のつながりをネットワーク図として可視化したもので、テキストマイニング技術と、ネットワーク分析の技術を融合したものである。共同発明を特徴付けるキーワードが接続線の上に示されており、時間の経過とともに、どのように人がつながり、技術開発への取り組み、融合が進んだかを可視化している。

昨今、製品が複雑化し、多様な技術の融合が進む中で、自社の人材・技術をどう活かして、次のイノベーションを生み出すか、あるいは、同業他社の取り組み、目指す方向を把握する上で有効な手法と言える。

5. 知財業務を支えるインフラへの昇華

テキストマイニングによるポートフォリオ分析は、さまざまな目的、場面に広がりを見せている。以下にその一例を挙げる。

5.1 M&Aの戦略策定・デューディリジェンス

製造業界におけるM&Aの意思決定において、技術的な融合・シナジーは、極めて重要な判断要素である。しかし、対象企業の財務状況、有形資産については検討の初期段階から、詳細に調査するのに対して、技術については、客観的な評価が難しく、経営者を含めた関係者の共通認識を形成することは難しい。当然、ある程度、検討が進んだ段階では、詳細な評価（デューディリジェンス、Due Diligence）を行うことになるが、検討の初期段階では、担当者の主観的な評価に委ねられることが少なくない。

自社と対象企業の知財ポートフォリオを比較・分析することで、M&Aによる技術的な補完・囲い込みの可能性を評価、シミュレーションすることで、一つの有効な判断材料とすることができる。

5.2 自社技術の新たな用途、提供・提携先の発掘

自社が保有する要素技術や製造方法、化合物などのシーズに対して、特許情報をヒントに、新たな用途や、提携・提供先を発掘することができる。自社が保有する要素技術に関連する特許を、特許分類等で絞り込まずに幅広く抽出し、「実施例」「技術分野」等のテキストを解析することで、従来とは異なる分野への適用可能性を探ったり、異業種・業界の出願人による関連出願を発掘し、新たな提携・提供先のヒントを得ることができる。

5.3 同業他社の新技術への取り組みの徹底把握

同業他社の出願に出現する単語を、徹底的に抽出・確認することで、新技術への取り組みの萌芽をいち早く把握することができる。

その会社にとって、従来の事業領域とは、まっ

特許出願ポートフォリオと関連市場規模のギャップ分析

※1 公報の文中に含まれる単語・係り受けから、特許をフィルムの用途に分類。用途別の市場規模データ(※2)とマッチングした。
※2 「2006液晶関連市場の現状と将来展望」(株式会社富士キメラ総研)

図11 知財ポートフォリオと市場規模の統合分析

たく異なる分野への出願であれば、特許分類コードから当たりをつけて、発掘することができる。一方、既存技術の応用や、他分野への適用の場合には、既存技術に関する特許コードに埋もれてしまうため、テキストマイニングによる言葉の掘り起こしが有効となる。

5.4 自社の事業・製品と知的財産のひも付け—統合分析

企業が事業に関する意思決定をする際には、市場規模と将来予測、現在のシェアと競合状況、社内リソース（人材、設備投資）の投入状況といった情報をもとに総合的な判断が行われ、知財ポートフォリオの分析結果は、判断材料の一つに過ぎない。

本来、知的財産は、事業や製品・サービスと結びついて初めて価値を生み出すものである。例えば、自社・他社の特許群を、自社の事業・製品とひも付けることができれば、各事業の市場規模と予測、自社の売上・利益目標と重ね合わせた分析が可能となり、事業戦略、研究開発戦略の有効な判断材料となる。また、人材配置・設備投資といったリソースの投入状況と重ね合わせれば、自社の事業戦略・研究開発戦略と、知財戦略との相関・ギャップを評価することができる。通常、自社の

日本・海外、社内・社外の情報を、担当事業・技術、役職に応じたメニュー・レイアウト・参照権限に基づき、社内配信

図12 「TRUE TELLER知財ポータル」 画面イメージと概念図

出願・特許については、研究開発の初期段階から、事業・部署と明確にひも付けた管理が行われるが、他社の出願・特許を、自社の事業・製品と結びつけて分析することは容易ではない。テキストマイニングにより、言葉をもとに、他社の特許群と自社の事業・製品をひも付けることで、自社の事業・製品の相対的な強み、弱みを技術の側面から評価することが可能となる。

図11では、ある企業のフラットパネルディスプレイ向けフィルム関連の特許群について、テキストマイニングの分類手法により、関連する部材(用途)とひも付け、さらに部材別の市場規模データとマッチングすることで、各特許を関連部材の市場規模データと関連付けた。これを、同企業の出願領域をサーモグラフで表示した上に、市場規模によって色分けしたプロットとして重ね合わせることで、自社の知財ポートフォリオと、関連市場のポテンシャルの整合性、ギャップを見ている。

6. 分析ツールから業務インフラへ

企業の知財担当者や、研究開発の企画担当者は、それぞれの問題意識、仮説に基づき、自ら必要なポートフォリオ分析を行う立場にあるが、多くの技術者・研究者は、まずは、担当する事業、業務、

役職に応じて、必要な情報を、整理された形で分かりやすく提供してほしいというニーズを持つ。例えば、毎週発行される国内・海外の公報に関するSDI (Selected Dissemination Information) 検索の結果を確認することは、必要な業務ではあるものの、未加工の情報を、未整理の状態で読み込んでいくことは、現場の担当者にとって負担が小さくない。各社員が、担当する事業、業務、役職に応じてカスタマイズされた分析結果画面を、社内のイントラネットを通じて自動配信することで、担当者は、自分の関心の高い切り口の情報から確認していくことが可能となる(図12)。

7. おわりに

今後、企業は、情報武装を怠れば、自社の知財ポートフォリオを、同業他社のほうがよく把握しているといった事態を招くことになるだろう。かつて、インターネットによる特許検索のサービスは、10年ほどの間に、企業にとって当たり前の道具として急速に定着していった。テキストマイニングによる知財ポートフォリオの分析についても、近い将来、企業にとって当たり前の道具になっていくものと確信している。

参考文献

- 1) 三室克哉, 鈴村賢治, 神田晴彦. 顧客の声マネジメント: テキストマイニングで本音を「見る」. オーム社, 2007, 232p.
- 2) 三宅将之, 宗裕二. 知財ポートフォリオ経営: 研究開発戦略の評価と知財M & Aの考え方. 東洋経済新報社, 2005, 231p.

著者抄録

知的財産に関する自社・他社の状況を的確に把握することは、知財経営に向けた第一歩であり、知財ポー

トフォリオ分析が注目されている。テキストマイニング技術の活用により、従来、多くの労力と時間を要した、大量の文献群の内容まで切り込んだ分析が可能になるとともに、言葉をベースにしたマッピングやネットワーク分析、文中に出現する単語・係り受けに基づくグルーピングなど、分析・可視化手法も高度化している。今後は、単なるツールにとどまらず、企業・研究機関の知財業務を支えるインフラへの昇華が期待される。そして同時に、これを活用する、調査・企画部門や知財部門のスタッフには、新たなスキルが求められるようになる。

キーワード

テキストマイニング, 知財ポートフォリオ, 可視化手法, マッピング分析, 知的財産, 特許, ネットワーク分析

Author Abstract

To grasp the condition of intellectual property of this firm and competition companies precisely is a first step for IP management. And now, patent portfolio analysis by text mining is paid attention. By text mining software, it becomes possible to analyze contents of a massive patent specification documents in detail. Analyzing and visualizing methods are sophisticating, for example, mapping based on words, network analysis, grouping documents based on words in documents, and so on. From now on, text mining doesn't stay in a mere tool, it will be an infrastructure that supports the IP strategy of companies and research institutes. And staffs of a planning section and an intellectual property section will be asked for new skills.

Key words

text mining, patent portfolio, visualizing method, mapping analysis, intellectual property, patent, network analysis