

ビジュアルテキストマイニング技術 - 特許検索・分析支援への適用 -

Visual Text Mining Technology for Patent Mining

あらまし

近年、ビジネス資産としての特許の価値が重要視されるようになってきている。特許を企業経営の中で戦略的に活用し、収入の増加や競争力の強化へと結び付けていくためには、自社保有特許の位置付けや他社の技術戦略などを正確に把握することが不可欠であるが、そのための検索・分析作業に多大なコストと時間がかかるという問題があった。

そこで、富士通研究所は、このような問題を解決するため、特許検索・分析作業を効率化するためのナレッジマネジメントツールとして、ビジュアルテキストマイニング技術を適用した特許マイニングツールを開発し、社内利用を推進するとともに、製品化を進めている。

本稿では、特許マイニングツールの概要を説明するとともに、具体的な利用シーンに沿って機能を紹介する。

Abstract

In recent years, patents have gained recognition as an important business property. To fully utilize patents in business management and thereby increase income and strengthen competitiveness, companies must understand the position of their own patents and the technological strategies of their competitors. However, in order to do this, companies must spend a lot of time and money searching and analyzing patent information. To make patent mining more efficient, Fujitsu Laboratories has developed a knowledge management tool that mines patent information using visual text mining technologies. This tool is used within the Fujitsu Group and has also been released as a commercial product. In this paper, we give an overview of this tool and describe the practical use of its functions.



渡部 勇（わたなべ いさむ）
ITメディア研究所知能システム研究
部 所属
現在、テキストマイニング、ナレッ
ジマネジメント、特許検索・分析の
研究に従事。

ま え が き

テキストマイニングとは、文書情報から有益な知識を発見・抽出するための技術である。情報検索システムが、利用者の目的に合った文書を探し出すことを目的としているのに対し、テキストマイニングでは、文書を個別に調べても分からない、文書群全体に内在する知識（パターンやトレンド）を発見することを目的としている。また新しい研究領域ではあるが、この数年の間に実用化も急速に進み、大量のテキスト情報にアクセスするための新しい道具として、ビジネスの場面でも活用されるようになってきている^{(1),(2)}

富士通研究所では、可視化技術（分析結果を視覚化する技術）を用いることによりユーザの知識発見を支援するビジュアルテキストマイニング技術⁽³⁾⁻⁽⁵⁾を開発し、フリーアンサーのアンケート分析、コールセンタに寄せられる顧客からの問合せ情報の分析⁽⁶⁾などに適用する研究を行ってきた。近年では、新しい適用分野として、障害情報分析^{(7),(8)} 特許分析⁽⁹⁾⁻⁽¹³⁾への適用研究を進めている。

本稿では、ビジュアルテキストマイニング技術の特許検索・分析への適用について紹介する。まず特許検索・分析の問題について述べる。つぎに、富士通研究所で開発している特許マイニングツールの概要を説明する。続く二つの章では、具体的な利用シーンに沿って、特許マイニングツールの機能を紹介する。

特許検索・分析の問題

近年、ビジネス資産としての特許の価値が重要視されるようになってきている。例えば、IBMの場合、特許によるライセンス収入を、1990年度の3,500万ドルから2000年度の17億ドルへと、10年間で50倍に急増させている。また、米国では特許関連の侵害訴訟が日常化しており、コダックからポラロイドへの10億ドル（インスタント写真）、ミノルタからハネウェルへの1億3,000万ドル（自動焦点カメラ）、マイクロソフトからDECへの1億6,000万ドルというように、巨額な損害賠償が必要となるケースも多い⁽¹⁴⁾

企業の競争力を維持していくためには、「モノを売る」から「サービスを売る」へ、そして、その次

のステップとして「知識を売る」ことが求められる時代になってきている。このような流れは国内にも広がりつつあり、知的財産立国の実現を目指した知的財産戦略大綱が発表されるなど、特許を中心とする知的財産への関心が、非常に高まってきている。

特許情報をビジネスに活用し、特許によるライセンス収入の増加や企業の競争力の強化へと結び付けていくためには、自社保有特許の位置付けや他社の技術戦略などを正確に把握する必要がある。そのためには関連特許情報を検索し、収集された特許群の関連性を詳細に分析することが不可欠となる。しかし、このような分析には多大なコスト・時間と高度な専門知識が必要とされるため、十分な分析が行われずに、ビジネス的に価値のある特許が活用されずに休眠特許となっていたり、逆に不要な特許を維持するために無駄なコストをかけていたりするケースもしばしば起きていた。

富士通研究所は、このような問題を解決するため、特許に関する専門知識のない研究者・技術者でも簡単に特許検索・分析が行えるようにするための特許マイニングツールを開発し、社内利用を推進するとともに製品化を進めている。

特許マイニングツールの概要

本章では、まず特許マイニングツールの検索・分析処理で使用するインデックスDBの作成と利用について紹介し、つぎにインデックスDBをもとに特許マイニングツールが実現する機能を紹介する。

検索・分析処理の概要

特許マイニングツールの検索・分析処理の概要を図-1に示す。

特許マイニングツールで特許情報の検索・分析を行うためには、まず検索・分析処理で使用するインデックスDBを作成する必要がある。

インデックスDB作成の過程では、特許情報のテキスト部分（名称・要約・請求項・詳細な説明）に対して単語切出し、頻度集計、複合語構成・分割処理、係り受け解析（主語・述語、修飾語・被修飾語などの単語間の関係を抽出）を行い、キーワードを抽出する。抽出されたキーワードには、統計計算により重要度が付与される。特定の特許にしか出現しない特徴的なキーワードには大きな値が、どの特許にも出現するような一般的なキーワードには小さな

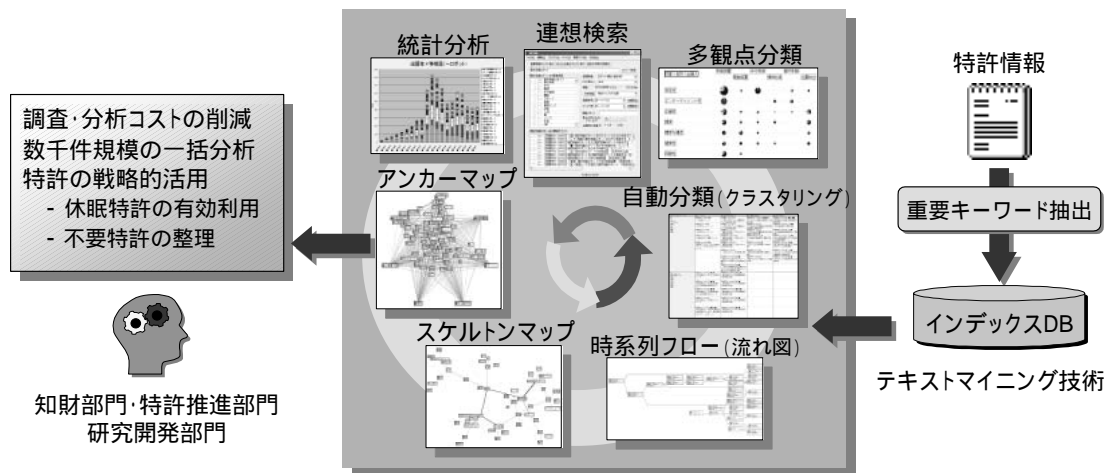


図-1 特許マイニングツールの検索・分析処理の概要
Fig.1-Searching and analyzing process of patent mining tool.

値が設定されることになる。上記の処理により、特許ごとに重要度付きのキーワード群が登録されたインデックスDBが作成される。

なお、出願人（特許を出願した組織名）・出願日・IPC（国際特許分類）・FI（ファイルインデックス）・Fターム（特許分類）などの書誌情報も種別ごとにインデックスDBに登録され、検索・分析に利用することが可能である。

検索・分析時には、インデックスDBを用いて、単語間・特許間の関連度（関連性の強さ）が計算される。単語間の関連度は、単語の共起度（二つの単語が互いに同一特許中に出現する度合い）を用いて計算され、同一特許の中で同時に現れる回数の多い単語ペアほど関連度の値が大きくなる。特許間の関連度は、単語の共有度（二つの特許が同一単語を共有する度合い）を用いて計算され、共通の単語を多く含む特許ペアほど関連度の値が大きくなる。

なお、関連度の値は前処理の段階であらかじめ決まっている固定的なものではなく、検索・分析実行時に計算される動的な値であり、分析対象となる特許群を絞り込むことによって変化していく。

検索・分析機能の概要

特許マイニングツールには、特許情報の検索・分析を支援する以下の機能が実装されている。

(1) 連想検索

通常のキーワード検索機能（キーワードを入力して特許をランキング検索）に加え、「関連単語検索機能」「類似特許検索機能」などがある。検索・分

析対象となる特許集合の絞込みに使用する。

(2) 統計分析

検索結果をリアルタイムで集計し、グラフ化する。書誌情報やキーワードの出現傾向の分析（IPCや出願人の経年変化、トレンドキーワードの分析など）に使用する。

(3) アンカーマップ

指定した単語を頂点に、その関連語を多角形の中に配置した概念マップの表示機能であり、単語の位置関係により、頂点に指定した単語間の特徴を表示する。比較分析（例えば出願人ごとの特徴比較など）に使用する。

(4) スケルトンマップ

骨格となる強い単語間関連情報だけを表示した概念マップの表示機能であり、単語間のつながりにより、主要な概念を表示する。特許集合全体の概要把握に使用する。

(5) 自動分類（クラスタリング）

特許を内容の類似性により自動分類（クラスタリング）し、出願人などの書誌情報によって表形式に整理する。特許集合全体の概要把握、特許集合の絞込みなどに使用する。

(6) 時系列フロー（流れ図）

内容の類似性や、引用・参照関係などを用いて、特許間の時間関係を可視化した流れ図を表示する。技術動向調査、基本特許の発見などに使用する。

(7) 多観点分類

係り受け解析と情報抽出の技術を用いて、特許の

目的や対象を抽出する。特許を目的別に分類したり、目的と対象の対応分析を行ったりする際に使用する。

(8) 引用分析

特許の明細書（書誌情報と本文）から、ほかの特許・論文への引用情報を抽出する。時系列フローの基礎情報として利用したり、被引用数（ほかの特許から何回引用されているか）を計算することにより有力特許発掘の基礎情報として利用したりする。

上記八つの機能群は相互に連携しており、ある機能の結果から別の機能を呼び出すことができるようになってい

特許検索の支援：公知例調査での利用

本章では、公知例調査において関連する単語や特許の効率的な検索を可能にする関連単語検索機能、類似特許検索機能を具体的に紹介する。

関連単語検索機能

公知例調査を行う際には、適切な検索式を組み立てて（あるいはキーワードを入力して）特許検索を行う必要がある。検索結果が粗すぎると内容チェックにコスト・時間がかかり、逆に絞り込みすぎると検索漏れが出てしまう可能性がある。特許検索のエキスパートは、同義語や特許分類（IPC・FI・Fタームなど）を活用することによって、検索効率を高めているが、一般の研究者・技術者にとっては効

率的な検索を行うことは容易ではない。ここで紹介する「関連単語検索機能」は、対象技術分野に関する同義語や特許分類の発見を支援し、効率的な検索を行う。

関連単語検索機能による検索例を図-2に示す。画面の最上段は検索キーワードを入力する領域であり、中段左側には入力単語に対する関連単語が、下段には入力単語を含む特許がランキング表示されている。関連単語の表示領域には、左側の図では「アーム」の関連単語が、中央の図では「アーム」の関連IPCが、右側の図ではIPC「H01L 21/68」の関連単語が、それぞれ表示されている。図-2の例のように、アームの「回転」の同義語・類義語として「回転」「旋回」といった単語を見つけたり（左側の図）、また、特定のキーワードに関連したIPCを探し（中央の図）、そのIPCの関連語を調べることによって（右側の図）、IPCの意味を推定したりすることも可能である。関連語としては、インデックスDBに入っているキーワード・書誌情報を種別ごとに表示することができ、また、特定の文字列パターンにより、表示単語の絞り込みを行うことも可能である。

以上のように、関連単語検索により、検索対象を絞り込んだり広げたりするための同義語や特許分類を見つけることができ、特許検索のエキスパートでなくても効率的な検索を実行することが可能となる。

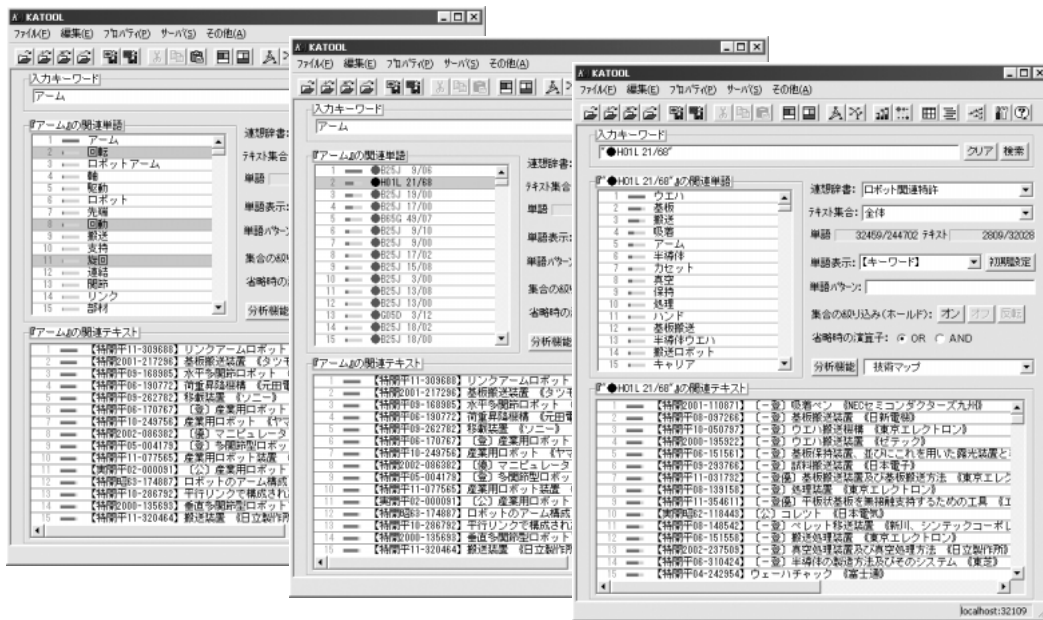


図-2 関連単語検索機能による検索例
Fig.2-Example of associative word search.

類似特許検索機能

入力キーワードを指定する代わりに、特定の特許を指定して、その特許に内容が類似する特許を検索するのが「類似特許検索機能」である。公知例調査の場合には、調査対象となる特許や、検索中に見つかった公知例などを入力特許として指定する。

図-3は、類似特許検索機能を使用し、ある特許を指定して、その類似特許の検索を行った検索例である。左側の図で番号指定された特許に対する類似特許が、中央の図の下段のリストにランキング表示されている。この類似特許から更に関連がありそうなものをピックアップして（グレーの網掛けで表示）、その特許群をキーに再度類似特許検索を行ったのが右側の図である。

以上のように、類似特許検索により、キーワードを一切指定せずに、ある特定の特許を出発点として、その類似特許を次々と見つけていくことが可能である。

そのほかの検索支援機能

公知例調査を行う場合には、「関連単語検索機能」「類似特許検索機能」を中心に使うことになるが、「自動分類（クラスタリング）」「時系列フロー（流れ図）」などの分析系の機能を、検索の補助に利用することもできる。「自動分類」では、内容の類似性による特許が自動分類されるので、調査対象が含

まれる分類を中心に調べていくことで、調査効率を上げることが可能である。また、時系列フローでは、時系列的な関係性が表示されるので、調査対象特許の上流に位置する特許（先願の類似特許・引用特許）を中心に調べていくことで、やはり調査効率を上げることが可能である。

特許分析の支援：動向調査での利用

本章では、特許分析に関して特許マイニングツールが持っている分析機能を具体的に紹介する。

統計分析

統計分析は、検索結果をリアルタイムで集計して、グラフ化する機能である。グラフの横軸・縦軸には、書誌情報とキーワードを自由に組み合わせて指定することが可能であり、出願年×出願人（出願人の経年変化）、出願年×キーワード（トレンドキーワード）、出願人×キーワード（出願人ごとの特徴キーワード）、出願人×出願人（共同出願人の分析）など、様々なグラフを作成することができる。グラフの縦軸の計算に、分布の偏りを表す統計量を利用することにより、変化がある部分を強調して表示する特徴量グラフを作成することも可能である。

図-4は、ロボット関連特許（約3万件の集合）に対して、横軸に出願年を、縦軸に「～ロボット」という文字列パターンのキーワード（「ロボット」で

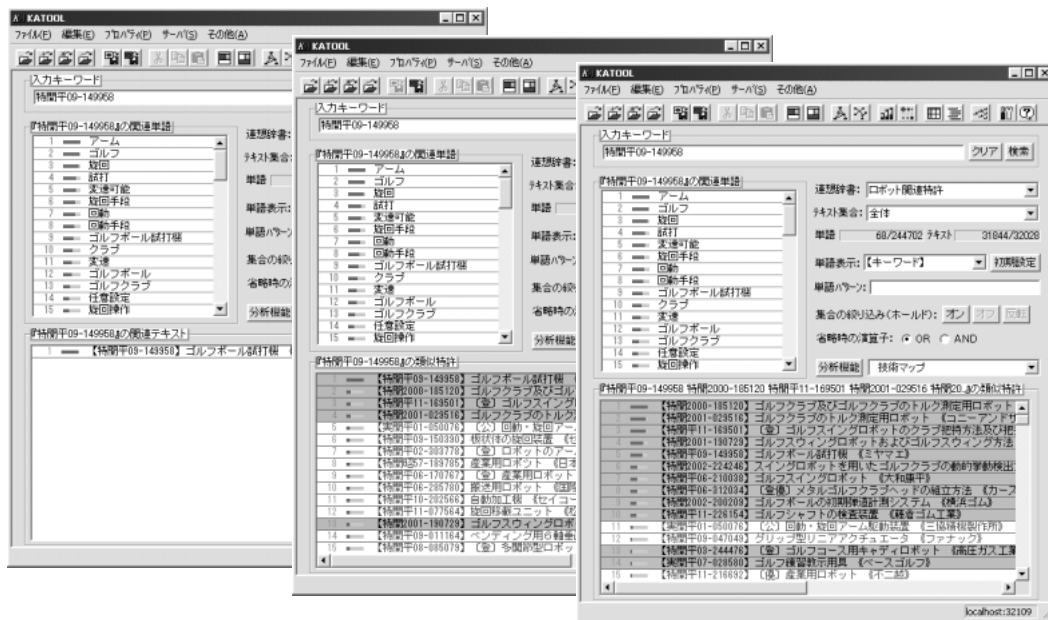


図-3 類似特許検索機能による検索例
Fig.3-Example of similar patent search.

終わるキーワード)を指定して作成したトレンドキーワードグラフである。グラフからは、ここ数年の傾向として「移動する手段を持ったロボット」の特許が増加傾向にあることを容易に読み取ることができる。

統計分析は特許分析の基本機能であり、まず全体としての特徴・傾向・変化などを概略としてとらえるために使用する。詳細な分析は、次節以降に説明する諸機能を用いて行う。

アンカーマップ

移動ロボット関連特許の集合に対して作成したアンカーマップを図-5に示す。図では、出願人の上位7社がアンカーとして7角形の頂点の位置に固定されており、7角形の内部には移動ロボット関連のキーワードが配置されている。7角形内のキーワードは、各頂点から単語間の関連度に応じた力で引っ張られており、その位置は各頂点からの引っ張り力のバランスによって決まる。したがって「センサ」や「移動」といった移動ロボットの共通キーワードは真中に、各出願人を特徴付けるキーワードは各頂点の近くに配置される。このようにアンカーマップでは単語の位置関係を見ることによって、出願人ごとの特徴比較を直感的に行うことが可能である。

スケルトンマップ

図-6は、移動ロボット関連特許のうちA社が出願人になっている特許だけで作成したスケルトンマップである。スケルトンマップでは、関連度が小さい関係を削除することにより、骨格となる構造(主要な関係)を表示する。中心的なテーマとなる重要な単語が、放射状の中心(ハブ)になる傾向があり、このハブを順に見ていくことで、全体の概観を把握

することができる。

多観点分類

図-7は、歩行ロボット関連特許の集合に対して作成した多観点分類のグラフである。横軸には出願年、縦軸には各特許から抽出された特許の目的・課題が表示されている。グリッド上に配置された円は、特許の出願件数を表しており、出願人によって色分けされている。

図からは、1992~1997年にかけては、B社が中心となり「安定性」「正確性」「簡単な構造」「精度」といった、歩行ロボットが有すべき基本的な性質に関する特許が多く出願されていたことが分かる。また1998~2001年にかけては、メインプレーヤがA社に代わり、「エンタテインメント性」「小型化」「安全性」といった、歩行ロボットが家庭に入ったときに求められる高度な性質に関する特許が多く出

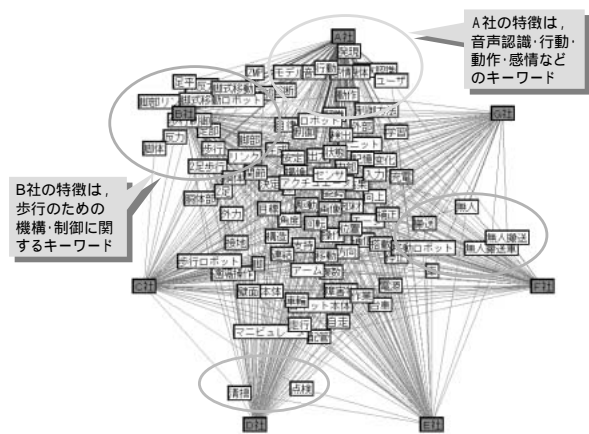


図-5 アンカーマップ
Fig.5-Anchored map.

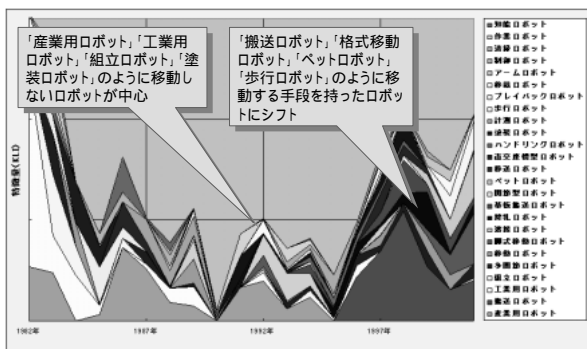


図-4 トрендキーワードグラフ
Fig.4-Trend keyword graph.

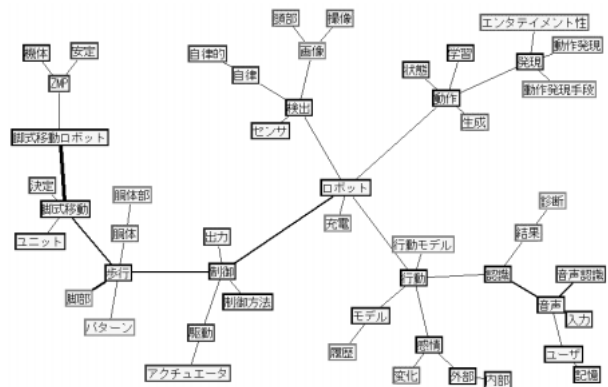


図-6 スケルトンマップ
Fig.6-Skeleton map.

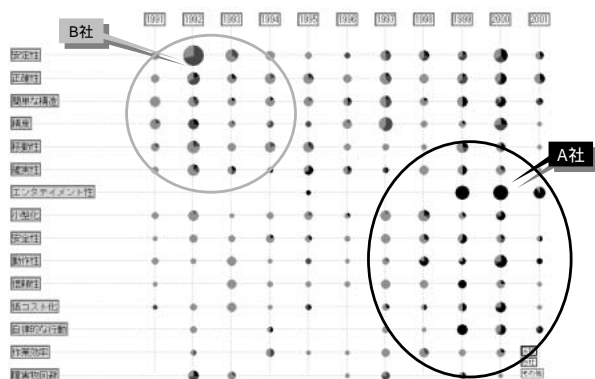


図-7 多観点分類
Fig.7-Multi-viewpoint clustering.

願されていることが分かる。

このように、多観点分類を用いることにより、特許分類やキーワードのグラフからは得られない詳細なトレンド・特徴をとらえることが可能であり、直感的に解釈しやすい結果を得ることができる。

そのほかの分析支援機能

動向調査を行う場合には、このほかに「自動分類（クラスタリング）」「時系列フロー（流れ図）」などの機能を使うことができる。「自動分類（クラスタリング）」は、特許群を、内容の類似性を用いて階層的に分類する機能であり、特許群の全体概要を俯瞰したり、人手で付与された特許分類（IPC、FI、Fタームなど）とは異なった観点で分析したりする際に有効である。「流れ図」は特許間の類似性・引用関係を時系列的に整理した図解であり、技術の流れを調べたり、基本特許・周辺特許の判断を行った際の基礎情報として利用できる。

む す び

本稿では、ビジュアルテキストマイニング技術を適用した特許マイニングツールについて紹介した。特許マイニングツールを用いることにより、大量特許情報の概要把握・比較分析を効率的に行うことが可能となり、調査・分析にかかるコスト・時間を大幅に削減・短縮することができる。また、手作業では困難だった数千件規模の一括分析が可能となるため、戦略的な知財活用を実現することができるようになる。

現在、このツールは富士通グループ内で実務に利用されており（2005年4月時点で、インストール台数：約1,200台、登録ユーザ数：860人）、一部の機

能が特許調査製品ATMS/IR.netに搭載されている。社内の知財部門で行った検証では、ツールの利用により、従来に比べて2～5倍の生産性向上が見られたとの結果が得られている。

今後は、引用情報を用いた特許評価支援機能を強化していく予定である。

参 考 文 献

- (1) 市村由美ほか：テキストマイニング - 事例紹介．人工知能学会誌，Vol.16，No.2，p.192-200（2001）．
- (2) 渡部勇：テキストマイニングの技術と応用．情報の科学と技術，Vo.53，No.1，p.28-33（2003）．
- (3) 渡部勇ほか：単語の連想関係によるテキストマイニング．情報処理学会 研究会報告 FI-55-8 DD-19-8，1999，p.57-64．
- (4) 三末和男ほか：テキストマイニングのための連想関係の可視化技術．情報処理学会 研究会報告 FI-55-8 DD-19-8，1999，p.65-72．
- (5) 渡部勇：ビジュアルテキストマイニング．人工知能学会誌，Vol.16，No.2，p.226-232（2001）．
- (6) 岡本青史ほか：カスタマーセンター支援システム．人工知能学会誌，Vo.15，No.6，p.1027-1034（2000）．
- (7) 齊藤孝広ほか：障害情報からのマイニング．情報処理学会 研究会報告 FI-61-20 NL-142-20，2001，p.145-152．
- (8) 齊藤孝広ほか：航空安全情報からのトラブル発生パターンの抽出について．言語処理学会 第11回年次大会，C1-1，2005，p.69-72．
- (9) 渡部勇：富士通研究所による特許検索・分析システム「ACCENT」．情報科学技術協会 INFOSTAシンポジウム2002予稿集，2002，p.7-12．
- (10) 小川 知也ほか：技術引用文献の可視化．情報処理学会 研究会報告 FI-70-10 DD38-10，2003，p.67-74．
- (11) 渡部勇：テキストマイニング技術による公知例調査の支援．ロボット学会誌，Vol.22，No.3，p.24-27（2004）．
- (12) 田中一成：特許文書の多観点分類について．情報処理学会 研究会報告 NL-161-10，2004，p.69-74．
- (13) 小川知也ほか：引用情報に基づく基本特許抽出．情報処理学会 研究会報告 FI-78-6 DD49-6，2005，p.41-48．
- (14) ケビン G．リベットほか：ビジネスモデル特許戦略．NTT出版，2000．