

Title	特許情報を利用した技術動向分析技術に関する調査研究 [課題研究報告書]
Author(s)	竹森, 久美子
Citation	
Issue Date	2010-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8952
Rights	
Description	Supervisor: 島津 明 教授, 情報科学研究科, 修士

課題研究報告書

特許情報を利用した技術動向分析技術 に関する調査研究

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

竹森 久美子

2010 年 3 月

課題研究報告書

特許情報を利用した技術動向分析技術 に関する調査研究

指導教員 島津 明 教授

審査委員主査 島津 明 教授

審査委員 飯田 弘之 教授

審査委員 白井 清昭 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0810704 竹森 久美子

提出年月：2010 年 2 月

概 要

企業・研究機関にとり、研究開発する技術分野における、これまでの流れを知ることとは、今後の研究開発の方向性を予測し、テーマを決める上で重要である。特許公開公報は、新規な技術を公表するものであるため、論文と同様、研究開発動向を知るための資料となりうると考えられる。今回の課題研究では、特許情報を元に、研究開発テーマを決めるために参考となる、技術動向を示す情報を検出し可視化するシステムについて、（１）専門用語の検出、（２）特許用語の簡易な用語への置き換え、（３）検出した専門用語を使用しての技術動向の可視化の３工程に分けて調査研究を行うこととする。調査研究を行った技術動向を可視化する技術に関し、有効な点、改良すべき点、技術を組み合わせることによる効果等について考察し、研究結果をまとめることとする。

目 次

第 1 章	序論	6
1.1	背景	6
1.2	本研究の目的	6
第 2 章	技術動向を分析する技術に関する調査研究	8
2.1	専門用語を検出する技術に関する従来研究	12
2.2	簡易な用語への置き換え技術に関する従来研究	17
2.3	技術動向を可視化する技術に関する従来研究	19
第 3 章	ツールの検証	27
3.1	ネットワーク機能ツールの検証	27
3.2	データマイニング・ツールの検証	30
3.2.1	特許情報の収集	30
3.2.2	形態素解析	33
3.2.3	サーモグラフの作成	34
3.2.4	単語群と出願年の相関マップの作成	41
3.2.5	データマイニング・ツールの評価	43
第 4 章	技術動向を分析する技術に関する考察	49
4.1	従来技術に関する考察	49
4.1.1	専門用語を検出する技術に関する考察	49
4.1.2	簡易な用語への置き換え技術に関する考察	50
4.1.3	技術動向を可視化する技術に関する考察	50
4.2	ツールに関する考察	51
4.3	まとめ	51
第 5 章	今後の課題	52
第 6 章	結論	53
	参考文献	54

図 目 次

2.1 引用関係を利用したネットワーク図	19
2.2 共通キーワードを上位ノードとした Semantic network 図	20
2.3 主成分分析を用いたパテントマップ	21
2.4 Semantic Web によるマップ	22
2.5 図面によるパテントマップ	23
2.6 新規に付与された USC による技術動向マップ	24
2.7 単語群と出願年との相関マップ	25
2.8 特許の件数分布を等高線で表したマップ	26
3.1 ネットワーク機能項目選択のインターフェイス	28
3.2 ネットワーク機能図（出力結果）	29
3.3 単語マップ	35
3.4 サーモグラフ（単語「ロボット」読み込み時）	35
3.5 サーモグラフ（2000 年，ロボット）	36
3.6 サーモグラフ（2001 年，ロボット）	37
3.7 サーモグラフ（2002 年，ロボット）	37
3.8 サーモグラフ（2003 年，ロボット）	38
3.9 サーモグラフ（2004 年，ロボット）	38
3.10 サーモグラフ（2005 年，ロボット）	39
3.11 サーモグラフ（2006 年，ロボット）	39
3.12 サーモグラフ（2007 年，ロボット）	40
3.13 サーモグラフ（2008 年，ロボット）	40
3.14 単語群と出願年の相関マップ	41
3.15 ロボット分野における分類別論文発表件数	46

表 目 次

2.1 技術動向分析技術に関する従来研究	8
3.1 検証ツールの名称と内容	27
3.2 FI 分類ランキング	32
3.3 F ターム分類ランキング	32
3.4 単語頻度ランキング	33
3.5 検証用単語選択リスト	34
3.6 サーモグラフと相関マップにおける特徴語	42
3.7 特許出願件数の伸び率、及び累積出願件数	44
3.8 ロボット分野の技術分類表	45
3.9 サーモグラフと相関マップにおける特許・論文該当語	48

第1章 序論

1.1 背景

企業・研究機関は、新規な技術の開発を行うと、学会にて論文発表を行うと同様に、特許権を取得すると独占排他的実施の権利を有することができるため、特に企業では、特許出願を行うことが多い。特許法では、特許出願人に対して、独占排他的権利を認めることの代償として、その新規な技術を公開することを要求している。この特許出願を行った技術を公開する制度として、特許庁は、通常、特許出願から1年半後に、特許公開公報を発行している。

企業・研究機関にとり、研究開発する技術分野における、これまでの流れを知ることは、今後の研究開発の方向性を予測し、テーマを決める上で重要である。特許公開公報は、新規な技術を公表するものであるため、論文と同様、研究開発動向を知るための資料となりうると考えられる。なお、新規な技術情報を得るソースとしては、特許公開公報の他に、特許の登録後に権利取得したことを公表するため発行される特許公報もある。(以下、「特許公開公報」と「特許公報」から得られる情報のことを「特許情報」という。)

1.2 本研究の目的

今回の課題研究では、特許情報を元に、研究開発テーマを決めるために参考となる、技術動向を示す情報を抽出し可視化する技術について、調査研究を行う。

特許情報を元に技術動向を可視化するため、3つの工程に分けて、それぞれについて調査研究を行うこととする。

- (1) 特許情報からの専門用語の検出
- (2) 特許用語の簡易な用語への置き換え
- (3) 検出した専門用語を使った技術動向の可視化

(1) については、特許情報から技術動向を表す専門用語を検出する技術、更に検出した専門用語の分類技術等について調査研究する。

(2) については、特許用語は難解な場合が多いため、特許用語をわかりやすい用語へ置き換えることができないか調査研究する。

また、(3) については、特許情報から検出した専門用語（又は、置き換えた用語）を利用して、技術動向を可視化する技術について調査研究する。

次に、文献で紹介された使用可能なツールを用いて検証を行う。最後に、調査を行った技術について有効な点、改良すべき点、技術を組み合わせることによる効果等について考察し、3つの工程を合わせたどのようなシステムが有効であるか検討して研究結果をまとめることとする。

第2章

技術動向を分析する技術に関する調査研究

本章では、第1章で述べた技術動向を分析する技術について、特許情報から、技術動向を表している技術用語を検出するための専門用語の検出技術、特許用語は分かりにくい用語であるため、検出した特許用語を簡易な用語へ置き換えるための技術、検出した専門用語（又は、置き換えた用語）を使った技術動向を俯瞰することのできる技術動向の可視化技術について、3つの工程に分けて調査を行った。以下、述べていく。

	分類	技術名	開発者	年	出典
1	専門用語の検出技術	論文の表題解析 (コード割当)	今井 俊	1999	“表題解析による科学技術論文の自動分類” 北陸先端科学技術大学院大学修士論文
2		論文の表題解析 (要素技術の一覧表示)	難波英嗣, 谷口裕子	2006	“学術論文データベースからの研究動向情報の抽出と可視化” 言語処理学会 第12回年次大会
3		論文の表題解析 (要素技術のグループ分け)	近藤友樹, 難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山	2007	“論文データベースからの研究動向情報の抽出” 言語処理学会 第13回年次大会
4		特許公報の自動分類	安藤 俊幸	2007	“特許情報の分析・評価支援:「termmi」と統計解析言語 R による特許情報の可視化” 第4回情報プロフェッショナルシンポジウム
5		単語間の階層関係の判定	大石康智, 伊藤克亘, 武田一	2006	“単語の共起関係と構文情報を利用した単語階層関係

			哉，藤井敦.		の統計的自動識別” 情報処理学会研究報告
6		特許データベースからのシソーラス自動構築	難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山.	2007	“特許データベースからのシソーラスの自動構築” 言語処理学会 第 13 回年次大会
7		特許分類と、特許検索式の作成	有賀康裕, 内藤和幸.	2005	“パテントマップソフトの特許検索への活用 特許調査に不慣れな技術者または初心者でも精度の良い調査結果を得る手法” 第 2 回情報プロフェッショナルシンポジウム
8		Semantic Web による特許情報の収集と分析	Mark Giereth, Thomas Ertl.	2008	“Visualization Enhanced Semantic Wikis for Patent Information” 12 th International Conference Information Visualization
9		特許の引用関係を利用した技術動向分析	Zhiqiang Liu, Donghua Zhu.	2009	“Web Mining based Patent Analysis and Citation Visualization” School of Management and Economics, Beijing Institute of technology
10		手掛かり語からの特許課題検出	渡辺勇, 小川知也, 田中一成.	2005	“特許情報の分析・評価支援－多観点分類と引用分析－” 第 2 回情報プロフェッショナルシンポジウム予稿集
11	簡易な用語への置き換え技術	論文と特許のデータベースを統合した検索環境の構築（引用関係の利用）	安善奈津美, 難波英嗣, 相沢輝昭, 奥村学.	2005	“特許、論文データベースを統合した検索環境の構築” 情報処理学会 研究報告
12		論文と特許のデータベースを統合した検索環境の構築（論文 ID と特許分類コードの対応付け）	難波英嗣, 藤井敦, 岩山真, 橋本泰一.	2009	“学術論文の国際特許分類への自動分類：第 7 回 NTCIR ワークショップ特許マイニングタスク成果報告”
13		論文用語と特許	釜屋英昭, 難波	2008	“論文用語の特許用語への

		用語の自動変換	英嗣, 竹澤寿幸, 奥村学.		自動変換” 言語処理学会 第14 回年次大会
14		特許用語と、対応する簡易な用語の検出	Mase H, Matsubayashi T, Ogawa Y, Yayoi T, Sato Y. and Iwayama M.	2005	“NTCIR-5 Patent Retrieval Experiments at Hitachi, “Proc. of NTCIR-5 Workshop Meeting
15		書誌情報を基にした特許出願推移のグラフ化	Chen-Yuan Liu, Shenq-Yih Luo.	2007	“Analysis of developing a specific technological field using the theme code of Japanese patent information” Scientometrics
16	技術動向の可視化技術	特許の引用関係を利用したネットワーク図	Edward Badger	2008	“Techniques for analyzing literature search results ”World Patent Information 30
			Shann-Bin Chang, Shu-Min Chang, Wei-Yuan Guh.	2007	“Exploring the Technology Diffusion Trajectories and Groups of Basic Patents of Business Methods: Using the Patent Citation Network” PICMET 2007 Proceeding
17		共通キーワードを上位ノードとした Semantic network 図	Jong Hwan Suh, Sang Chan Park.	2006	“A New Visualization Method for Patent Map: Application to Ubiquitous Computing Technology”, Advanced Data Mining and Applications, Springer Berlin / Heidelberg
18		キーワードの出現頻度を基にした可視化	宮田道生, 福川忠昭, 高橋正子.	2006	“特許戦略のための特許情報の計量化ーテキストマイニングを活用した企業の注力技術比較の試みー” 日本オペレーションリサーチ学会, 2006 年秋季研究発表会
19		特許の引用関係を利用した重力モデル	伊神正貫.	2007	“ナノテクノロジー関連特許出願のマッピング、引用関係を用いた技術群の同定と

					知識の流れの計測” The Japan Society for Science Policy and Research Management
20		Semantic Web によるマップ生成	Mark Giereth, Thomas Ertl.	2008	“Visualization Enhanced Semantic Wikis for Patent Information” 12 th International Conference Information Visualization
21		図面による技術動向の可視化	野崎篤志.	2008	“図面情報を用いたパテントポートフォリオ分析” 知財学会・第6回年次学術研究発表会
22		新規に付与された特許分類を利用した技術動向分析	有賀康裕, 新井大樹.	2001	“US クラス 705 に関するパテントマップ解析” 情報管理 Vol. 43
23		検索条件に該当する特許件数を可視化したネットワーク図	桐山勉, 田中宣郎, 川島順, 都築泉, 大山勝弘, 長谷川正好, 玉置研一.	2005	“分析 Tool を利用した Rapid Technology Intelligence Process 研究 (1): プロフェッショナル・ディベロップメントとテック・マイニング” 第2回情報プロフェッショナルシンポジウム
24		データマイニングを利用したサーモグラフ図	中居 隆.	2008	“テキストマイニングによる知財ポートフォリオ分析” 情報管理. Vol. 51
25		データマイニングによる単語群と出願年との相関マップ	山本外茂男.	2009	“産学連携のマッチング性分析におけるテキストマイニングの有効性” 情報の科学と技術 59 巻 6 号
26		単語間の関係に Naïve Bays モデルを利用した地図状マップ	Edward Badger	2008	“Techniques for analyzing literature search results” World Patent Information 30

表 2.1: 技術動向分析技術に関する従来研究

2.1 専門用語を検出する技術に関する従来研究

技術動向を分析するためには、まず、キーとなる専門用語を抽出することが必要である。そこで、特許情報から専門用語を切り出し、更に専門用語間の階層を判定したり、専門用語を分類する技術について調査を行い、考察を行った。

(1) 論文の表題解析 (コード割当)

論文の表題解析により、論文の技術分類が行われている[1]。論文の表題を機能語、動詞に形態素解析して論文表題の構造を把握し、目標に対する役割であるのか、動作の対象範囲であるのか、実現のための道具であるのか、専門用語集（岩波情報科学辞典、用語の木）を参照して名詞間の意味的關係を推定し、コード割当を行う研究がなされている。

(2) 論文の表題解析 (要素技術の一覧表示)

キーワード検索で、特定分野の論文を収集し、論文の表題から要素技術に関する情報を収集し、要素技術を使う研究分野を一覧表示する研究がなされている[2]。要素技術を抽出する手法は、表題の記載が「Aに基づいた」、「Bを用いた」である場合、AやBに要素技術を示す用語が一般的に含まれていることに着目したものである。「～に基づいた」、「～を用いた」等を手がかり語として、表題に、HEAD（主題、一番末尾の名詞句）、METHOD（手がかり語・・・「に基づいた」、「を用いた」等）、GOAL（手がかり語・・・「のための」等）の構造タグを付し、手がかり語と構造タグの対応リストを作成し、要素技術を抽出した研究である。

(3) 論文の表題解析 (要素技術のグループ分け)

論文データベースから動向情報を自動的に抽出し、可視化する研究がなされている[3]。(1)と異なる点は、表題解析に機械学習を取り入れた点と、主題だけでなく要素技術にも着目した点である。また、(2)と異なる点は、要素技術が属する技術分野について、類似した分野を1つのグループにまとめた点である。この類似した分野のグループ化とは、①ヲ格、ニ格の入出力が似た用語をグルーピング、②上位・下位関係のある用語をグルーピングすることにより行った。また、主題には形式的な主題があるとして、真の主題に置き換え、形式的な主題を不要語リストとして機械学習を行った。

以上(1)～(3)の研究は、論文の表題を解析することにより、キーとなる用語

の検出を試みたものである。論文の場合は、表題から何についての論文なのか推定でき、適切な分類細目を決定できることが多い。しかし、特許情報の場合は、権利と関わることから、論文の表題に対応する「発明の名称」について、権利範囲を限定しないように、広い意味に取れる一般的な名称とすることが多く、発明の名称から分類細目を決定することはできない。そこで、特許情報を対象とする場合は、権利範囲を示し、権利書として取り扱われる部分である「特許請求の範囲（以下、「請求項」という）」、「要約」、全文を解析することにより、キーとなる専門用語の検出を行うことが考えられる。特許情報については、例えば、以下のようにして専門用語を検出する試みがなされている。

（４）特許情報の自動分類

テキストマイニングツールにより専門用語を抽出し、自動分類を行う研究がなされている [4]。和布蕪にて特許公開公報の全文を形態素解析し、テキストマイニングツール（「termmi」）で専門用語を抽出し、ベクトル空間法で類似度を計算、特徴キーワードをクラスター分析し、更に、キーワードの語彙頻度と文書頻度をかけた重み付けを行うためのパラメータを計算して専門用語抽出リストを生成し、当該専門用語抽出リストを参照して人手で主観自動分類用 L U T（Look Up Table）を編集し、自動分類を行ったものである。

上記（４）の技術は、専門用語の抽出対象を特許公開公報の全文としている。自動分類用 L U T のある技術分野であれば、専門用語を抽出し、自動分類を行うことができるが、どの分野の特許についても専門用語の抽出と分類を行うことができるというわけではない。自動分類用 L U T を作成するために行う、テキストマイニングツールを使用しての専門用語の抽出は、操作者の勘に頼ることになり、対象とする技術分野に対して精通していないと難しい。また、自動分類用 L U T も人手で編集しており、自動分類用 L U T は、対象分野に詳しい者でないと作成は困難である。

更に特許情報の場合も、論文に関する試みと同様に、手がかり語、上位・下位関係を利用した構造分析が可能であると考えられる。構造分析が可能かに主眼を置き、調査を行ったので、以下に述べて検討する。

（５）単語間の階層関係の判定

Web ページについて、2 つの単語間に階層関係があるか否か出力する自動識別器に関する研究がなされている [5]。見出し語とその説明文に出現する単語との共起関係

を考慮した出現頻度に基づくモデルと、単語に隣接する形態素に着目した構文情報の統計量に基づくモデルから上位語、下位語を判定する。例えば、説明文には上位語を含んでいるが、下位語が含まれているとは限らないため、まず階層関係の判定を行い、更に局所的な構文情報、例えば、上位語であれば「科」、下位語であれば「など」等、単語の周辺に頻出する構文情報を利用して、階層関係を判定するものである。

上記（５）の技術は Web ページについて研究されたものであり、特許に関するものではないが、特許情報についても同様のアプローチが可能であると考えられる。例えば、見出し語に対応するものは「発明の名称」とし、権利範囲を記載した「請求項」、特許について説明した明細書部分から検出した専門用語の上位語を「発明の名称」とすることが考えられる。また、「AなどのB」とした表現は、Bを上位語、Aを下位語と捉えることができると考えられる。この上位・下位語の関係について特許情報を対象として、以下のような研究がなされている。

（６）特許データベースからのシソーラス自動構築

特許データベースから、シソーラスの自動構築を行う研究がなされている [6]。特許明細書から同義語対を検出する場合、同義語対は、同一文書内に共出現することが殆どなく、共起を使った手法での抽出は困難である。そこで、数多くの上位語、下位語を持っている用語を同義語対であるとする引用分析手法を用いて、同義語対の抽出を行った。また、「AなどのB」との記載である場合、Bを上位語、Aを下位語として検出するが、違うものも検出することがわかった。この場合に、「などの」を「の」で置き換え、自然であれば上位、下位関係にないことがわかった。

「AなどのB」という表現から上位・下位関係を検出するにあたり、（６）の技術を取り入れることで、精度を増すことができると考えられる。また、特許情報の書誌情報には、分類項目がある。この分類項目を利用して、専門用語を検出するための特許公開公報又は特許公報の絞り込みや分類が可能であると考えられる。

また、共通した上位語、下位語を持っている用語を同義語対とすることで、有効に分類が可能であると考えられる。

（７）特許分類と、特許検索式の作成

特許情報には、IPC（国際特許分類）又は F ターム、FI（日本の特許庁が独自に付与）、US クラス（US Patent & Trademark Office (USPTO) が独自に付与）等が記載され、既に技術分類がなされている。そのため、これら特許分類により、ある特定の技術領

域を絞ることができ、当該分野における技術動向を分析することが可能になる。この特許分類を検索式に取り込んで、目的とする特許公開公報又は特許公報を特許データベースから検索する方法について研究がなされている [7]。まず、キーワードにて検索し、検出した特許公開公報又は特許公報から FI の情報を取得し、FI ランキングを生成する。この FI ランキングの上位 FI に存在する特許公開公報又は特許公報の内容を手で確認し、高い確率で該当特許が存在する場合には、検索式として、キーワードに FI を OR で追加する。ノイズを多く含む場合は、上位 FI に存在する特許公開公報又は特許公報から別の特許分類である F タームのランキングを生成し、上位 F タームに存在する特許公開公報又は特許公報の内容を手で確認し、高い確率で該当特許が存在すれば、属する FI に当該 F タームを AND でかけ、全体の検索式には OR で追加するといった方法で検索式を作成していく。この方法で作成した検索式により、高い精度で目的とする特許公開公報又は特許公報を検出することができるというものである。

特許公開 (特許) 公報に付されている特許分類は、一定の基準で付与されておらず、ばらつきがある。特許分類によって自動的に分類することはできないが、(7) の技術により母集団に対して、ランキングの上位にある分類の内容を確認しつつ、特許公開 (特許) 公報の検索の精度を高めることができる。ただし、当該方法では、内容を確認しつつ検索式を作成していくため、操作者は専門知識が必要であり、時間もかかってしまう。

ヨーロッパ特許庁 (The European Patent Office (EPC))、USPTO、日本特許庁は、特許公開 (特許) 公報を XML で公開していることを利用して、Resource Description Framework (RDF) により検索する研究が、ヨーロッパを中心に行われている。

(8) Semantic Web による特許情報の収集と分析

EC プロジェクトとして、特許関連コンソーシアム PATExpert により、特許をナビゲートし、可視化するインターフェース PATViz の研究がなされている [8]。XML によって記述した文書に、World Wide Web Consortium (W3C) により規格化された RDF を用いてメタデータのタグを付け加え、自動的な情報の収集や分析のアプローチを可能とする Semantic Web について研究されている。文書の変換、要約化、更には、形式意味表現によるマッピングを行っている。RDF でタグを付けることにより、テキストデータ、図面、IPC 言語の相違を統合して検索することが可能となる。タグとして、ユー

ザにより意味注釈（semantic annotation）を付けて、データベースに反映し、自動検索を可能とする。

上記（８）の技術は、テキストデータのみならず、図面や分類を合わせ、総合的に類似度を判定するのに有効と思われる。

（９）特許の引用関係を利用した技術動向分析

特許の引用関係を用いて、技術動向を把握する研究がなされている [9]。特許出願時の明細書には、特許を中心に先行技術文献が記載されている。また、審査時に審査官によって、先行文献として特許が示されることが多い。この引用、被引用の関係をj用いて、特許出願の変遷をたどることができる。

殆どの特許に引用、被引用関係が見られるため、上記方法は、特許に表れている技術の変遷を辿るのに有効であると思われる。該当する特許公開（特許）公報から、キーとなる用語をうまく検出することができれば、引用・被引用を利用して技術動向を把握することが可能となる。

（１０）手掛かり語からの特許課題検出

表現パターンから、特許の目的、課題の特徴を検出する研究がなされている [10]。「～を向上する」という表現パターンには、目的、課題について記載されていることが多いことから、当該表現パターンを手掛かり語として、特許公開公報の「要約」、「発明が解決しようとする課題」という課題について記載されている項目から特徴情報を検出するものである。

上記（１０）の技術は、特許公開公報における課題記載部分に注目し、キーワードを検出するものである。課題について動向を把握するには有効であると思われるが、その課題を解決するための技術を検出するものではないため、技術動向を把握する目的には不十分である。また、「～を向上する」という表現を用いていない場合は、検出できないことになる。

2.2 簡易な用語への置き換え技術に関する従来研究

特許情報から切り出した特許用語は、論文で用いられている用語と異なることが多く、難解である。そのため、研究者が技術動向を把握しやすいよう用語を置き換えることができないか調査し、考察を行った。

(1 1) 論文と特許のデータベースを統合した検索環境の構築（引用関係の利用）

論文と特許のデータベースを統合した検索環境構築の研究がなされている [11]。まず、①論文中の特許への引用箇所の抽出、②特許間の引用関係の抽出を行い、更に、③特許中の論文への引用箇所の抽出を行い、④同じ内容の特許と論文を対応づけるものである。特に、②と③を中心に研究が行われた。②に関しては、番号の表記方法がほぼ一定であるため、抽出は容易である。③については、特許データベースから手がかり語により抽出した論文の記述がある文に、人手でタグを付したコーパスを訓練データとし、更に手がかり語を収集し、コーパスに再度機械学習を行い、抽出するものである。手がかり語としては、ポジティブな手がかり語、「例えば」、「記載」、「参照」、ネガティブな手がかり語「新聞」等を使用している。ネガティブな手がかり語とは、その用語があると、探している文献とは違うと判断して弾く手がかり語のことである。

上記（1 1）の技術では、論文と特許公開（特許）公報、特許公開（特許）公報どうしを対応づけることが可能であるが、用語どうしの対応づけまでは示唆していない。

(1 2) 論文と特許のデータベースを統合した検索環境の構築（論文 ID と特許分類コードの対応付け）

論文の表題、特許の請求項を解析して統合データベースを構築する研究もなされている [6]。また、K-Nearest Neighbor (K-NN)法を適用し、論文抄録に特許分類 IPC を付与、論文 ID を入力すると、IPC コードを出力する検索環境構築の研究もなされている [12]。

上記（1 2）の技術においても、論文の表題と特許の請求項又は IPC は対応しているが、用語の対応ではないので、置き換えにそのまま適用することはできない。

(1 3) 論文用語と特許用語の自動変換

論文と特許の引用関係は、数パーセントであり、引用関係をたどるには限界がある。

そこで、引用手法とシソーラス手法を合成した方法について研究がなされている[13]。あるテーマに関する特許と論文を網羅的に収集することを目的に、論文用語と特許用語を自動変換するものである。まず、引用手法を取り入れ、①論文用語を入力、②用語を表題に含む論文を検索し、③論文と引用関係にある特許を収集し、④③の特許の請求項から用語を抽出、頻度順に並べ出力する。④については、第1請求項とその従属項として主要箇所を特定し、予め用意している不要語句リストと照合、削除して、残った用語を頻度順で出力する。更に、シソーラス手法を取り入れ、特許中の「AやBなど（等）のC」という構造に着目し、この構造となっている文から上位語・下位語のセットを抽出し、シソーラスを構築、⑤専門用語を入力、⑥入力した専門用語を下位語としている特許用語を収集、⑦用語セットを頻度順に並べ出力することを可能とする。また、特許は請求範囲をなるべく広く確保するため、一般性の高い用語を用いて記述されていることから、論文用語に対応する特許用語を論文用語の上位語と捉えている。

上記（13）の技術は、論文用語から特許用語への変換についての研究であるが、逆に特許用語から論文用語への変換も可能であると思われるため、本調査研究の置き換え技術と目的が近い。論文用語は技術者に理解しやすいため、簡易な用語として論文用語を採用し、特許用語を論文用語に置き換えることは有効であると考え。また、論文・特許間の引用関係が少ないことに対して、シソーラスを構築して補強し精度を高めている点においても本調査研究に有益な研究であると考え。

（14）特許用語と、対応する簡易な用語の検出

特許公開（特許）公報の明細書中には、図面の「符号の説明」という項目がある。「符号の説明」には特許用語と簡易な用語が併記されているものがあるため、これらに関連用語対として抽出する研究がなされている[14]。

上記（14）のように「符号の説明」の項目に特許用語とカッコ書きで簡易な用語が併記されているものがあり、簡易な用語への置き換えに有効であるが、このように併記される場合は多くなく、一部の用語の置き換えができるにすぎない。

2.3 技術動向の可視化技術に関する従来研究

2.1にて抽出したキーとなる専門用語、又は2.2にて置き換えた用語を利用して、技術動向を把握することができないか、特許情報を使った技術動向を可視化する技術について調査を行い、考察を行った。

(15) 書誌情報を基にした特許出願推移のグラフ化

特定技術領域について年度別に出願件数推移を示すマップについて研究がなされている[15]。キーワードによりキーワードが使用されている該当特許公報を選択し、当該公報のFIを確認し、特定FIについての検索を行った。出願人情報から企業別に、企業を列、出願年を行とした出願件数表、X軸を出願年、Y軸を累積出願件数とした出願推移グラフ、X軸を出願年、Y軸を発明者数とした企業別の従事する研究者数の推移を可視化したグラフ、Fタームには、手段、目的を定義づけたタームがあることから、母集団の特許公開（特許）公報より手段、目的を表すFタームを抽出し、手段のFタームを列、目的のFタームを行とした特許の件数表等について記載されている。

上記(15)の技術は、グラフで、どのような技術の特許出願が多いかを可視化して技術のトレンドを把握することができる。しかし、新たな技術が生まれていく過程を可視化することができず、技術動向の可視化には不十分であると思われる。

(16) 特許の引用関係を利用したネットワーク図

2-1(9)で述べたような特許の引用関係を利用して、数世代にわたってネットワーク図にしたマップ[16]、ネットワーク図を時系列にマッピングしたもの[17]、更に、被引用回数が多い特許群で作成した引用フロー図[10]について研究されている。

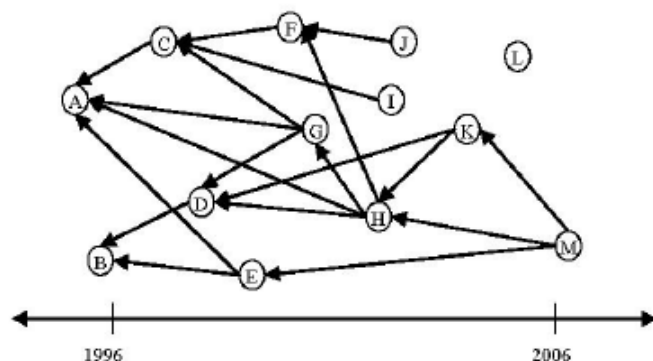


図 2.1: 引用関係を利用したネットワーク図[17]

上記(16)のように引用関係を可視化したマップは、引用関係にある特許公開(特許)公報からキーとなる専門用語を検出して当該専門用語をマップのノードとして配置することができれば、技術動向を把握するのに有効であると考えられる。

(17) 共通キーワードを上位ノードとした Semantic network 図

統一された書誌情報と統一されていないテキスト情報の両方を利用して、特許情報へのバランスの取れたアプローチを行い、可視化する研究がなされている[18]。①技術分野を絞り、キーワードを選択、②検索された特許公開(特許)公報から、キーワード(テキスト情報)を検出してキーワードランキングを出力し、キーワードリストを再定義する。③更に、キーワードを行、特許公開(特許)公報を列としてマトリクスを作成し、④K-means 法でクラスタリングを行い、特許公開(特許)公報をグループ分けする。そして、⑤各グループに含まれているキーワードを特定し、⑥共通キーワードを上位ノードとする Semantic network を作成する。⑦ ④～⑥を繰り返して、⑧キーワードの関係を最もよく表している Semantic network を選択し、⑨キーワードが含まれている最先の出願日(書誌情報)を抽出、⑩キーワードと出願日が記載されているノードを年順に配置して、技術動向を可視化したものである。ユビキタスについて実験が行われたが、当該マップにより技術の分岐の変遷を見ることができた。

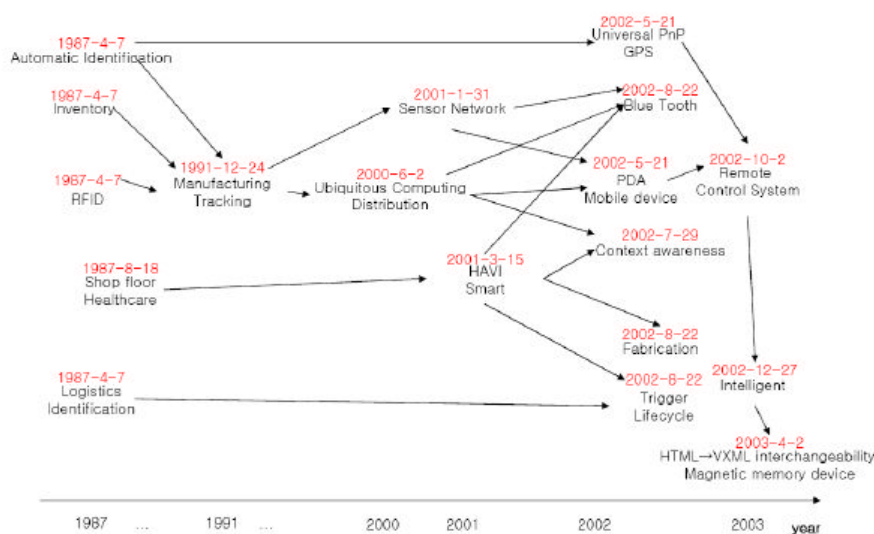


図 2.2: 共通キーワードを上位ノードとした Semantic network 図[18]

上記（１７）の技術は、グループ間で共通するキーワードを上位ノード（概念）として、時系列で前の技術であるとしている。確かに、複数の文献で取り上げられている用語は、研究者の間で既知になっており使われている可能性が高いので、機械的にマップ化するための着眼点として、有効なものであると考える。なおかつ、当該用語が最初に使われている特許公開（特許）公報の出願日情報を取得して、年度別に時系列で表すことにより、技術動向を俯瞰することができる。

（１８）キーワードの出現頻度を基にした可視化

基本キーワードと共起頻度が高いキーワードの出現頻度からマップ化する研究がなされている [19]。基本キーワードを決め、同一文章中に、当該キーワードと同時出現頻度の高い名詞（キーワード）を特定、当該キーワードの出現数をカウントし、基本キーワード×明細書数によるキーワード出現数行列を生成し、基本キーワード相関行列の主成分分析を行い、パテントマップとするものである。

例えば、第１主成分の電話内部の技術と、第２主成分の電話外部の技術を軸とし、各部品の名詞が使われる回数が多いことによる重要度について可視化されている [19]。電話の内部では「回路要素部」、電話の外部では「筐体部」のキーワードの頻度が高く、また、企業別にも表示されており、富士通は電話の内部、日本電気では電話の外部が重視されており、松下電器（パナソニック）は、両方バランスを取った状態であることを見てとることができる。

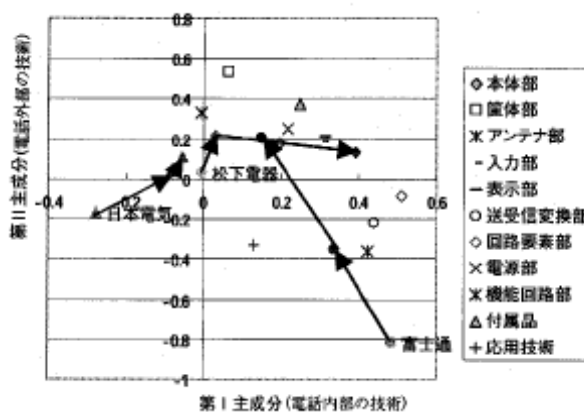


図 2.3: 主成分分析を用いたパテントマップ[19]

(19) 特許の引用関係を利用した重力モデル

引用関係を利用した重力モデルによるマッピングの研究がなされている [20]。重力モデルでは、マップの1つの点1つの出願を表し、引用によって結びつけられている出願間にもみ、2点の距離 r に比例する引力が働く。また、出願の点を中心として、半径 r_c 内に存在する全ての出願から $0.1/r^2$ の斥力が働くとしてマッピングを行ったものである。

上記(19)の技術では、引用関係にある出願どうしが近い位置にマッピングされるので、見た目にわかりやすい。しかし、引用関係にない特許間でも、同じ技術分野に属する場合、属さない場合がある。引用がないと一様に斥力が働くというのは、実際の特許間の関係を反映しているとはいえない。

(20) Semantic Web によるマップ生成

PATExpert による 2.1(8) で述べた PatViz では、濃さで特許件数を表現した IPC 分類マップ、出願国マップ、出現数の多い特徴語ほど大きく表示する特徴語のマップ等の生成を研究している [8]。

上記(20)の技術のマップは、どの技術分野(要素)が注目されているのか、また、どの国が特許に注力しているのか、俯瞰することができ、有効であると考えられる。ただし、時系列での技術動向を見ることはできない。

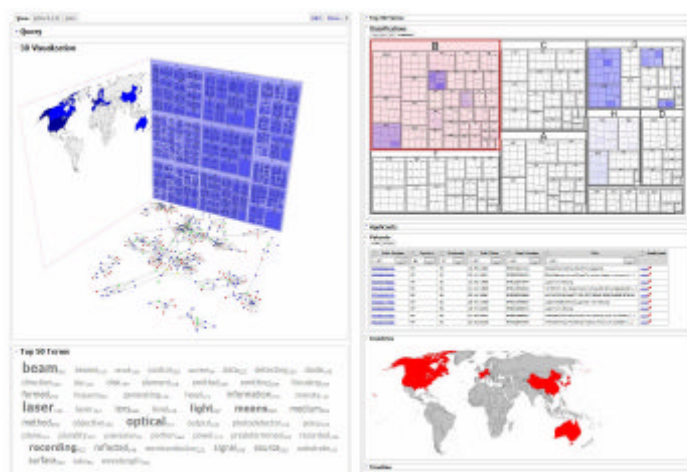


図 2.4: Semantic Web によるマップ[8]

（２１）図面による技術動向の可視化

特許の図面をグルーピングし、可視化する研究がなされている [21]。目視によるグルーピングであるが、機械系分野の特許には図面が付いていることが多く、全体を表したシステム構成図が、選択図（代表図）とされている場合は、全体の構成そのものが発明の特徴であると判定し、システム構成図が図面にあるが、選択図となっていない場合は、部品や制御回路に特徴があると判定し、ある程度まとまった件数のあるグループの図面をマップに配置し、可視化したものである。

上記（２１）の技術は、図面に着目したことがユニークである。また、要約に選択された代表図面が、システム図であるのか否かで、全体の構成についての特許か、部品の特許か判定して分類する技術もユニークである。図面も表示するのでわかりやすいマップとなる。

（２２）新規に付与された特許分類を利用した技術動向分析

USPTO が独自に付した分類 USC を利用して技術動向を可視化する研究がなされている [22]。Y 軸を新規な技術分野として USPTO により新たに付された USC、X 軸を USC が付された年として最新の技術動向を表したグラフである。時系列で新たに分類が付される様子がわかり、これによっても技術動向を可視化することができる。

選択図種別	グループ名	件数	選択図種別	グループ名	件数
システム構成図		85	要素・部品	T-A-ELE-A	1
回路図	T-A-CIR-A	1		T-A-ELE-B	3
	T-A-CIR-B	13		T-A-ELE-C	1
	T-A-CIR-C	5		T-A-ELE-D	7
	T-A-CIR-D	1		T-A-ELE-E	1
	T-A-CIR-E	1		T-A-ELE-F	1
フローチャート	T-A-FLOW-A	1		T-A-ELE-G	1
	T-A-FLOW-B	1		T-A-ELE-H	1
	T-A-FLOW-C	2		T-A-ELE-I	9
	T-A-FLOW-D	1		T-A-ELE-J	1
				T-A-ELE-K	1
				T-A-ELE-L	1
				T-A-ELE-M	1
				T-A-ELE-O	1

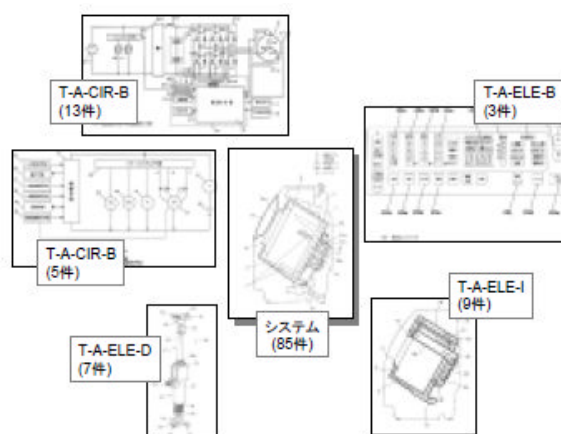


図 2.5: 図面によるパテントマップ[21]

(22)の技術は、技術動向を把握する上で有効であると考え。また、簡易な方法で工数が少ないため、時間がない場合で、特定分野の技術動向を把握する必要がある時に、よい方法であると考え。

(23) 検索条件に該当する特許件数を可視化したネットワーク図

Matheo Patent 社が開発した分析ツールが紹介されている[23]。年代別、企業別、IPC 別等にネットワーク機能を利用し、頻度をネットワーク図で表現したものである。MathoPatent は、デモ版を試すことができるため、次章で検証する。

(24) データマイニングを利用したサーモグラフ図

データマイニングによる可視化の研究がなされている[24]。特許公開(特許)公報にはいくつかの項目がある。例えば、要約では、簡潔に記載されているものの、書き方が厳密にルール化されていないため、バラツキがある。請求項では、権利範囲が明確に記述されているが、請求項により発明の本質との関係性、重要性が異なる。そこで、当該研究では、テキストマイニングの対象範囲を第1請求項としている。手順としては、IPC、キーワードによる検索式で、一次母集団データを生成する。この一次母集団について、テキストマイニング・ツールにより、請求項の文章を形態素解析し、請求項に使用されている単語とその頻度分布を出す。ユーザは、単語と頻度分布等を参照しながら、想定外のノイズとなっている単語を見つけ除外したり、気付いた関連用語を検索条件にフィードバックしながら、検索式を修正する。修正検索式により二次母集団を生成し、二次母集団から、対象となっている請求項の文章に対し、形態素解析を行い、検出された単語群よりユーザが単語を選択し、単語マップを自動生成する。

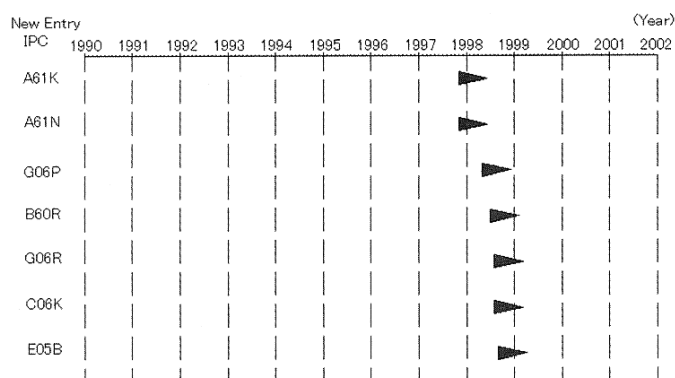


図 2.6: 新規に付与された USC による技術動向マップ[22]

単語マップでは、2つの単語どうしが同時に出現（共起）する文献数が多いほど単語間の距離を近くして自動配置を行う。当該マップの縦軸、横軸に意味はない。この単語マップを元に、更に、単語の用いられている文献数の多さを文献の分布密度として温度表示してサーモグラフを生成し、可視化するものである。このテキストマイニングにより、従来と異なる分野への適用可能性、新技術への取り組みの萌芽を把握することが可能となる。当該ツールは、野村総合研究所（NRI）の TrueTeller というデータマイニングソフトであり、次章で検証を行う。

（25）データマイニングによる単語群と出願年との相関マップ

上記 TrueTeller を利用し、単語群と出願年との相関関係を可視化した相関マップの研究がなされている [25]。重要技術用語の経年変化を分析することができ、実験では、右上から中央下、左上へとU字型となった年度ごとの単語との相関関係を表すマップが生成された。U字型の内側にある単語は、どの年代でも使用される技術用語であり、外側にある単語は、比較的にその年代に特化して使用されている技術用語であることがわかった。

上記（24）の技術は、単語の近さ（共起性）と、頻度を可視化することができたが、（25）の技術では、単語マップに時系列要素を加えることができ、技術動向を可視化するためには有効であると考ええる。当該技術も次章で検証することとする。

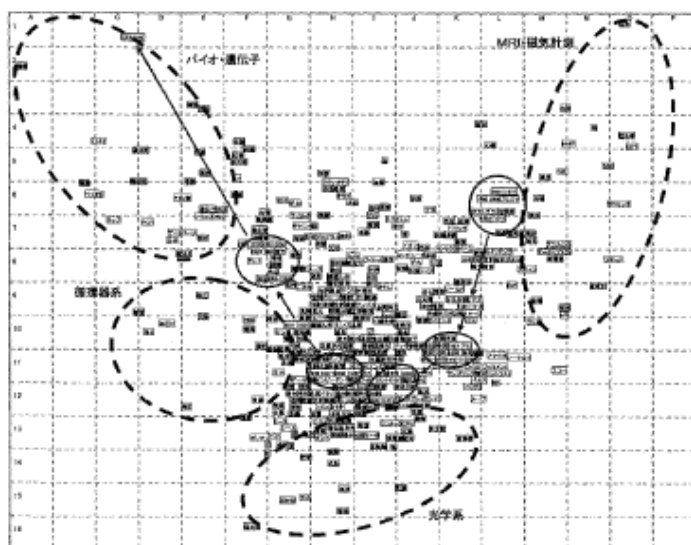


図 2.7: 単語群と出願年との相関マップ [25]

(2 6) 単語間の関係に Naive Bays モデルを利用した地図状マップ

地図状に可視化する技術としては、言葉、句の関係を Naive Bays モデルにより単語マップ状に配置し、件数が多いと等高線で山となるような可視化技術がある[16]。上記(2 6)の技術は、(2 4)、(2 5)と同様、単語どうしの近さ、頻度を地図で表現し、可視化できるものである。

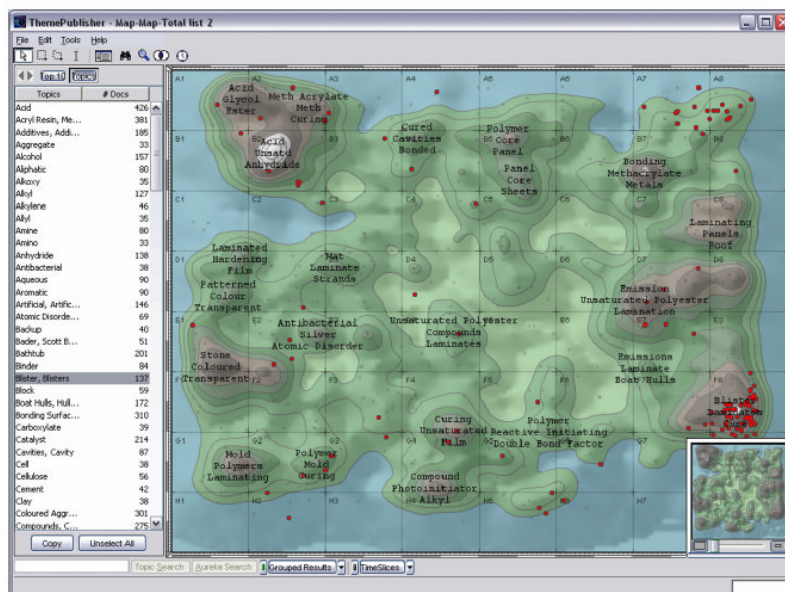


図 2.8: 特許の件数分布を等高線で表したマップ [16]

第3章 ツールの検証

本章では、第2章で紹介されている使用可能な可視化ツールの検証を行うこととする。

第2章 2.3 (23) で紹介されている検索条件に該当した項目をネットワーク機能を利用して紐状につなげ可視化する”MatheoPatent”、同じく第2章 2.3 (24)、(25) で紹介されているデータマイニングを利用した TrueTeller を使用し、検証を行う。

	システム名	企業名	内容
1	MatheoPatent (デモ版)	MatheoPatent	技術動向の可視化
2	TrueTeller	野村総合研究所 (NRI)	形態素解析
3	同上 (サーモグラフ)	同上	技術動向の可視化
4	同上 (単語群と出願年との相関マップ)	同上	技術動向の可視化

表 3.1: 検証ツールの名称と内容

3.1 ネットワーク機能ツールの検証

第2章 2.3 (23) の MatheoPatent は、年代別、企業別、IPC 別等にネットワーク機能を利用し、関連のある情報どうしを結びつけ、頻度とともにネットワーク図で表現したものである。

デモ版を使い、技術動向を把握するための可視化ツールとして、可視化の観点における有効性を検証することにした。デモ版では既に入っているデータ (スノーボード分野) でしか試すことができない。データは、ヨーロッパ特許庁 (EPC) より取得した特許公開公報の情報から、出願年、出願人名、IPC 等の書誌事項、発明の名称及び要約から取得したキーワード等がそれぞれ関連づけられて格納されている。

これら項目の中から、図のように、2つの項目を選択して、ネットワーク図を表示することができる。

本調査研究では、技術動向の把握について主眼を置いているので、時系列情報と技術を表していると思われるキーワードから、どのように技術動向を把握できるかを見ることとし、PR.(Year) (優先年) と Abstract Keywords (要約のキーワード) を選択した。

PR.(Year) (優先年) と Abstract Keywords (要約のキーワード) を選んだ場合、該当する優先年とキーワードが、該当する出願件数を表す数値とともにタグ状に紐でつながれている状態で表示された。

タグをカーソルでドラッグすると、動かすことができるが、重なって表示されるため、見にくい。頻度の高い単語は色が付されている。例えば、頻度の高い単語”snowboard”は、引用されている件数が 48 件であることを示す”48”の数値とともに、頻度が高いことを示す緑色で表示されている。”snowboard”はタグ状になっており、紐状のもので、複数の優先年を表すタグとつながっている。その中で、件数の多い優先年”2006 年”は、引用されている特許の件数”43”の数値とともに件数が多いことを示す赤色で表示されている。

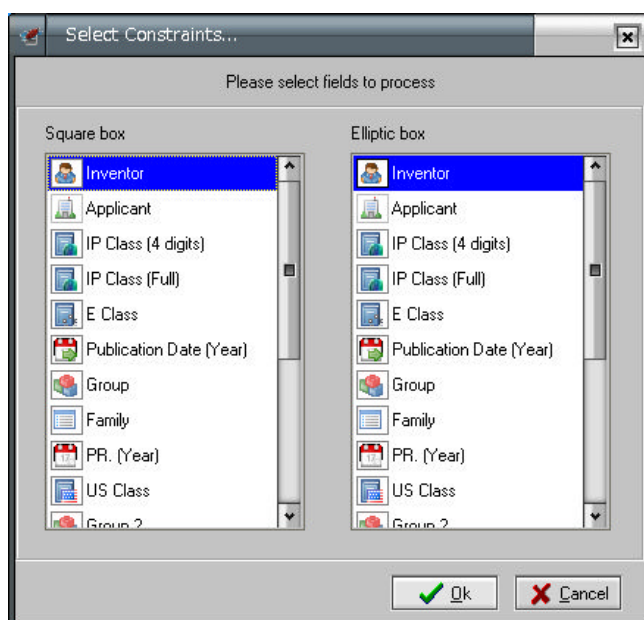


図 3.1: ネットワーク機能項目選択のインターフェイス

3.2 データマイニング・ツールの検証

第2章 2.3 (24)、(25)で紹介されている野村総合研究所(NRI)のテキストマイニングツール、TrueTellerを使用して、検証を行った。特許データベースから取得した特許情報のデータより、指定した所定の項目について形態素解析し、その単語の中から、ユーザが必要なキーワードを選択する。TrueTellerは、選択したキーワードを自動配置して単語マップを生成する。単語マップは、2つの単語どうしが共起する文献数が多いほど単語間の距離を近くして自動配置する。当該マップの縦軸、横軸に意味はない。そして、この単語マップを下敷きとして、キーワードが使われた文献の分布密度を温度表示するサーモグラフを生成して技術動向を可視化するものである。

3.2.1 特許情報の収集

今回の検証のテーマを「ロボット」とした。

最初に、特許データベースより、ロボットの技術分野に該当する特許公開公報を収集する。データベースから特許公開公報を収集するための検索式の作成を行った。検索式を作成するにあたって、最初に、検索式に使うキーワードとして、どのような言葉が適切であるか、検討をした。

労働安全衛生法に基づく産業用ロボットの使用等の安全基準に関する技術上の指針によると、「ロボット」の定義は、以下の通りとなっている。

(1) 産業用ロボット

(2) マニプレータ・・・人間の上肢に類似した機能を有し、次のいずれかの作業を行うことができるものをいう

イ) その先端部に当たるメカニカルハンド(人間の手に相当する部分)、吸着器等により物体を把持し、空間的に移動させる作業

ロ) その先端部に取り付けられた塗装用スプレーガン、溶接用トーチ等の工具による塗装、溶接等の作業

以上の定義も考慮し、検索式に使うキーワードは、「ロボット+マニプレータ+マニピュレータ」とした。

次に、検索式の精度を向上させるために、キーワードに特許分類を追加することを

検討した。IPC におけるロボットの分類について検討した結果、B25J が、マニプレータの分類で、該当することがわかった。当該クラスの注釈¹としては、以下のよう
に記載されている。

「このサブクラスにおいては、下記の用語は以下に示す意味で用いる： — “マニプレータ” は、空間内において有形の運動および方向の変更ができ、そのような運動および変更は、ヘッドから離れた手段によって随意に制御されるような把持部、またはワークヘッドを有するハンドリング器具、装置または機械を含む、例. プログラム制御の産業用ロボット」

先に検討したキーワードにロボットを示す、IPC 分類、「B25J」をかけ、検索式を「(ロボット+マニプレータ+マニピュレータ) * B25J」とした。
なお、キーワードによる検索対象となるテキスト情報の範囲としては、特許の権利範囲を記述する項目である「請求項」とした。

上記検索式にて、1995 年以降に特許出願され、公開された特許公開公報を特許データベースにて検索すると、16,685 件ヒットした。形態素解析するには、当該件数は多すぎるため、対象範囲を絞り込むため、特許出願日を 2000 年以降として再検索した。特許出願日が 2000 年以降は、5,734 件であったため、当該範囲を分析対象範囲とし、最終的に検索式を以下の通りとした。

[キーワード (請求項) : (ロボット+マニプレータ+マニピュレータ)] * [IPC : B25J] * [出願日 : 2000 年 1 月 1 日～]

特許データベースから検索式にて取り出した母集団がロボットを表すものとして適切であるのか検証するため、IPC 以外の他の特許分類にて、当該母集団における分類のランキングを出した。

当該母集団において、他の特許分類 FI では、付与されている分類は、以下の通りで、ロボットを表しており、適切な母集団であることを確認した。

同様に、F ターム分類についても、ランキングを出した。結果は、以下の通りで、ロボットを表しており、適切な母集団であることを確認した。

¹特許庁の電子図書館 IPDL より

	FI 分類	分類定義	件数
1	B25J 13/00 Z	その他のもの	19
2	B25J 5/00 C	歩行装置をもつもの	18
3	B25J 9/10 A	軌跡制御・補間・位置補正〔速度の制御，原点復帰を含む〕	14
4	A63H 11/00 Z	その他のもの	12
5	H01L 21/68 A	移送〔ウェハ移送具、アーム、マニピュレータ、ハンド、ロボット、台車、コンベヤレイアウト設計〕	9
6	B25J 13/08 A	視覚〔光学〕による制御	8
7	B25J 9/22 Z	その他のもの	6
8	B25J 13/08 Z	その他のもの〔視覚手段以外の手段による位置・重さの検出〕	5
8	B25J 5/00 F	二足歩行ロボット	5
8	B65G 49/07 C	移動台、又はアームによるもの	5

表 3.2: FI 分類ランキング

	F ターム分類	分類定義	件数
1	3C007CS08	無軌道で移動するもの。車輪、クローラ、歩行装置などを有し、無軌道で移動するもの。	35
2	3C007HS27	電動機。	25
3	3C007AS36	エンタテインメント用。遊技、娯楽、愛玩に供するもの。	24
4	3C007WA04	犬、猫型。四足歩行する動物（トラ、キリン等も含む）	24
5	3F060AA00	マニプレータの用途で A A 0 1 ～ A A 1 0 に該当しないもの。 （例）遊戯用、商業サービス用、教育用、似顔絵描き用。	23
6	3C007WA14	四足	23
7	3F060CA14	歩行装置を有するもの	22
8	3F059AA00	マニプレータの用途で A A 0 1 ～ A A 2 0 に該当しないもの。 （例）遊戯用、商業サービス用、教育用、似顔絵描き用。	19

8	3C007KT01	一つのカメラをもつもの	19
8	3F059BB06	歩行装置をもつもの。壁面を昇るための吸着歩行脚なども含める。	18

表 3.3: F ターム分類ランキング

以上のように、抽出した母集団について、他の分類、FI、F タームの上位 10 位までのランキングを出して検証した結果、FI、F ターム共にロボットに関する分類であったことから、当該母集団は、ロボットに関する特許であることが、確認された。

3.2.2 形態素解析

検出した特許公開公報の母集団に対し、書誌項目、請求項及び要約のテキスト文書といった特許情報を取り出した。その後、TrueTeller の機能にて、請求項及び要約のテキスト文書に対し、テキストマイニングを行うため、形態素解析を行った。形態素解析を行った結果、取り出された単語、上位 10,000 語について、頻度、割合、件数から以下のように、ランキングを出した。

	単語	品詞	頻度	割合(%)	件数
1	ロボット	名詞	36887	48.43	13885
2	制御	名詞	14448	22.54	6461
3	備える	動詞	7784	19.78	5670
4	位置	名詞	15057	18.94	5430
5	特徴とする	名詞	4885	17	4874
6	移動	名詞	11087	16.48	4724
7	動作	名詞	10505	15.81	4533
8	株式会社	名詞	4650	15.59	4468
9	ロボット	名詞	4446	14.21	4073
10	有する	動詞	5283	12.9	3698
11	駆動	名詞	6801	11.62	3331
12	システム	名詞	4660	11.44	3279
13	方向	名詞	6238	10.67	3059
14	基づく	動詞	4426	9.92	2843
15	機構	名詞	5155	9.59	2748
16	設ける	動詞	4243	9.36	2684
17	複数	名詞	4712	9.23	2645
18	情報	名詞	7425	9.13	2618

19	回転	名詞	6004	8.96	2570
20	検出	名詞	6174	8.79	2519

表 3.4: 単語頻度ランキング

移動	名詞
ロボット	名詞
システム	名詞
方向	名詞
機構	名詞
回転	名詞
検出	名詞
センサ	名詞
アーム	名詞
作業	名詞
搬送	名詞
姿勢	名詞
ハンド	名詞
産業	名詞
画像	名詞
角度	名詞
モータ	名詞
認識	名詞
マニピュレータ	名詞
歩行	名詞

表 3.5: 検証用単語選択リスト

リストを見ると、例えば、ロボットに関する検索を行ったので、上位に「ロボット」、「ロボット」が入っている。また、技術分野に関係なく、特許によく用いられる表現である、「備える」、「特徴とする」、「有する」、「基づく」、「設ける」等も上位に入っていることがわかる。

TrueTeller の単語マップを作成するために単語の選択が必要である。単語マップを作成するため、技術分野に関係なく使用される特許用語等を除き、10,000 語のリストより、ロボット技術の特徴を示すと思われる単語を目視で 124 語選択した。

3.2.3 サーモグラフの作成

形態素解析した単語から、3.2.2 で選択して作成した単語リストを TrueTeller に読み込ませ、単語マップの作成を行った。

このマップを下敷きとして、当該単語の使用されている特許公開公報の件数が多いほど、温度が高く表示されるサーモグラフの作成を行った。

なお、第1回目の作成では、次のようなサーモグラフとなった。サーモグラフの温度は、単語の相対件数に基づき、濃淡が出る。単語「ロボット」の件数が多いため、「ロボット」のある箇所の温度が高くなってしまい、「ロボット」に対して件数の少ない単語は全て温度が低く、薄く表示され、ロボット以外の特徴語のわかりにくいことが判明した。そのため、単語マップから「ロボット」、「ロボット」、「マニピュレータ」の単語を外して、再度、作成することとした。

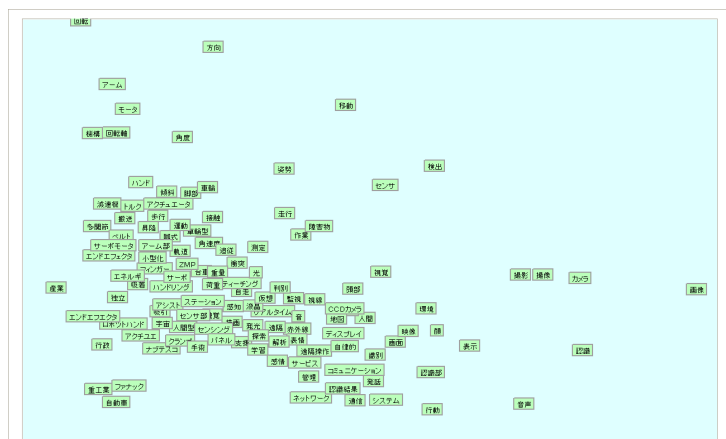


図 3.3: 単語マップ。

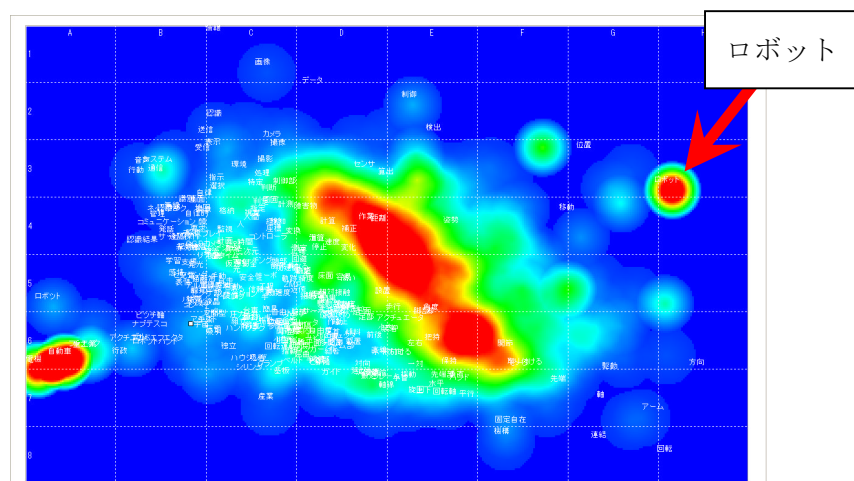


図 3.4: サーモグラフ (単語「ロボット」読み込み時)

時系列で技術動向を分析するため、サーモグラフは、特許の出願日を基準として、2000年より1年分ごとに作成した。分析方法としては、件数の多い技術（赤又は黄色で表示）、前年に比べ、件数の増えている技術を、その年の特徴ある技術とすることにより行った。

2000年のサーモグラフによると、「重工業」、「自動車」、「産業」といった単語に代表される産業用ロボットの特許出願と、「通信」、「システム」に代表される通信機能を有したロボットの特許出願の多いことが特徴である。「発話」、「認識」、「移動」、「歩行」、「多関節」、「ハンド」、「脚部」、「搬送」、「車輪」、「移動」といった単語の頻度も高い。

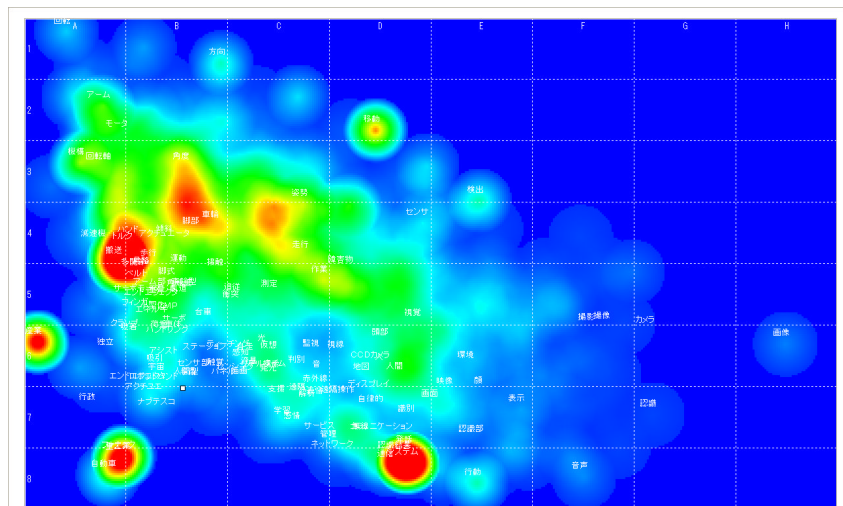


図 3.5: サーマグラフ (2000 年, ロボット)

2001 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2000 年に比較し、「アクチュエータ」、「アーム部」、「作業」、「回転軸」、「音声」といった単語の頻度が増えている。

2002 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2001 年に比較し、「接触」、「姿勢」、「画像」といった単語の頻度が増えている。

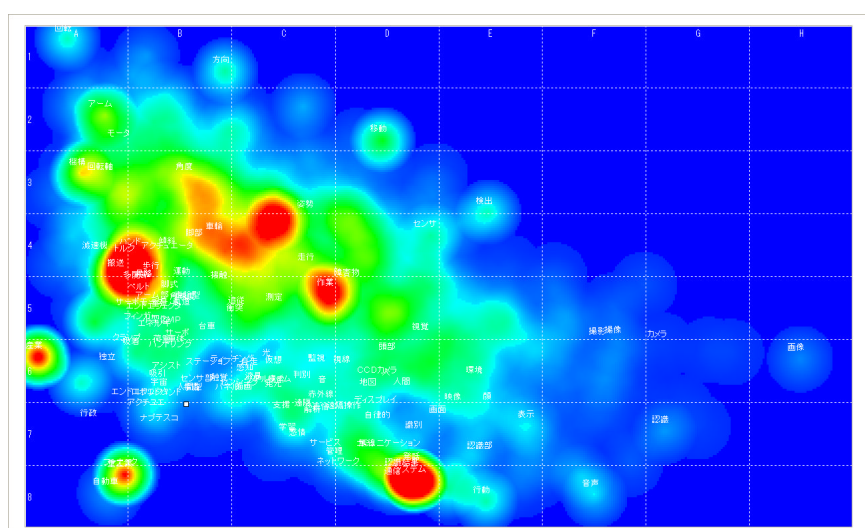


図 3.6: サーマグラフ (2001 年, ロボット)

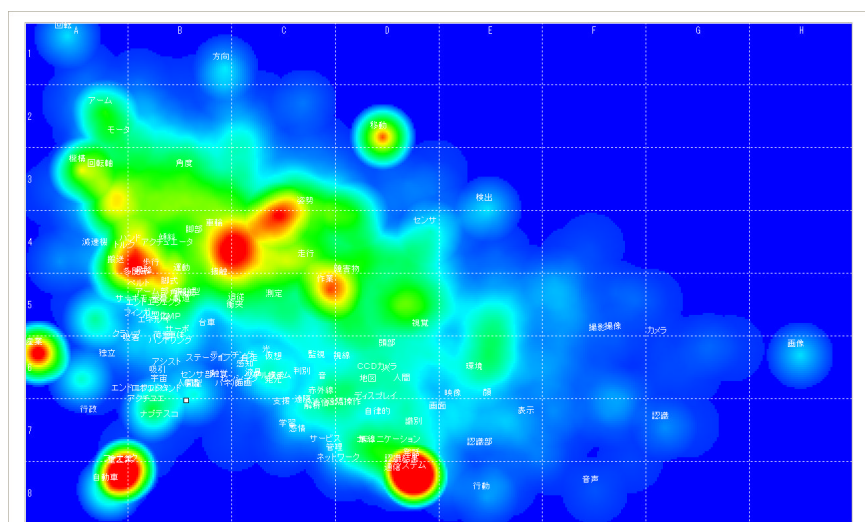


図 3.7: サーマグラフ (2002 年, ロボット)

2003 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2002 年に比較し、「行動」、「認識部」、「学習」、「感情」といった単語の頻度が増えている。

2004 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2003 年に比較し、「サービス」、「管理」、「ネットワーク」、「視覚」、「アーム」といった単語の頻度が増えている。

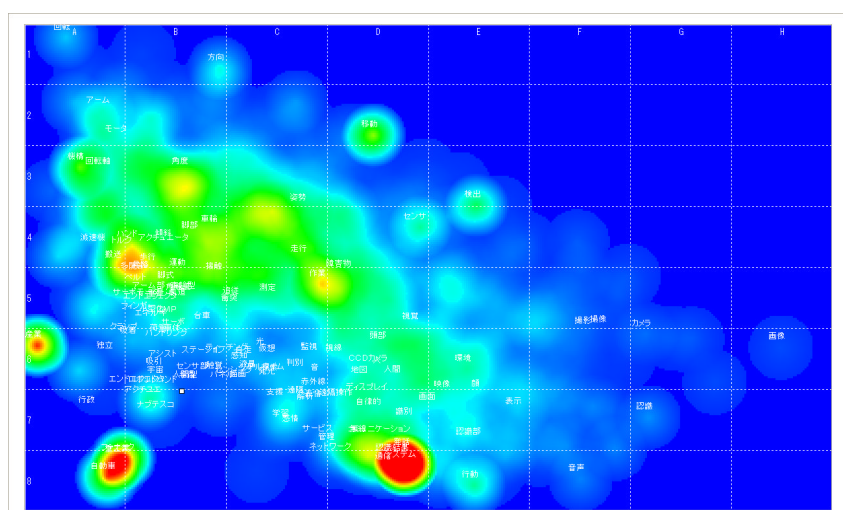


図 3.8: サーマグラフ (2003 年, ロボット)

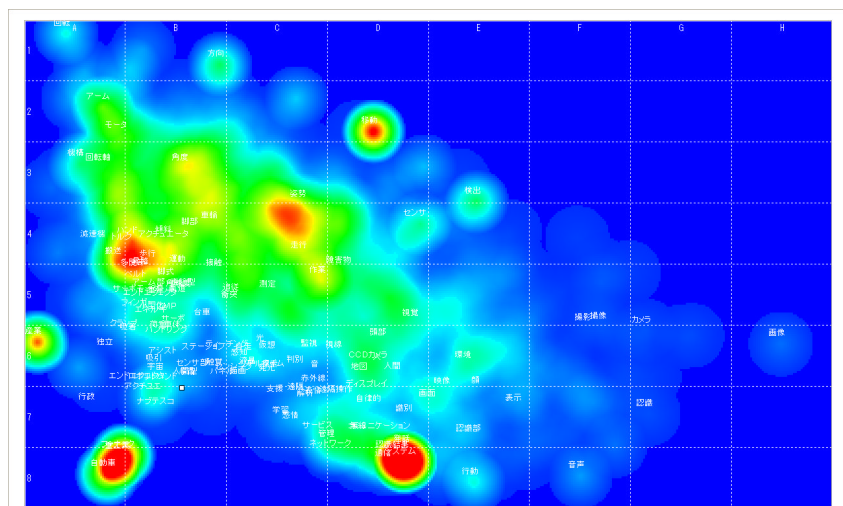


図 3.9: サーマグラフ (2004 年, ロボット)

2005 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2004 年に比較し、「車輪」、「ハンド」、「歩行」、「脚式」といった単語の頻度が増えている。

2006 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2005 年に比較し、「作業」の頻度が増え、「移動」が減少している。

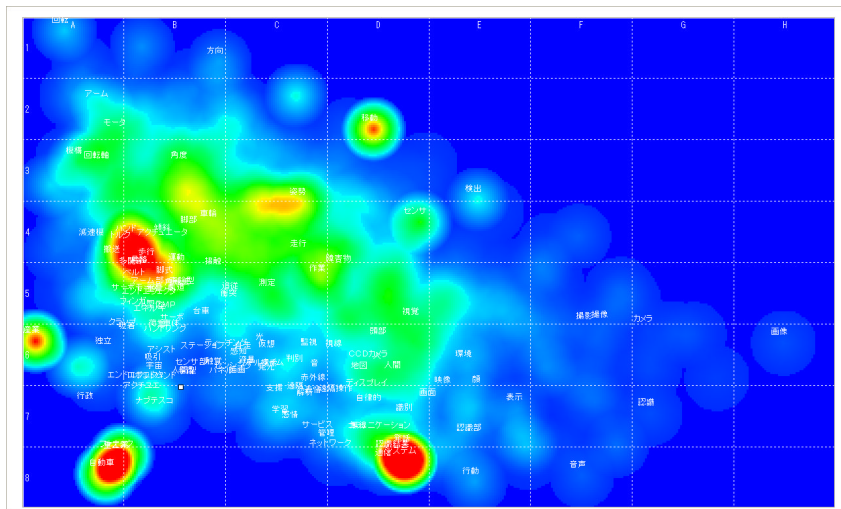


図 3.10: サーマグラフ (2005 年, ロボット)

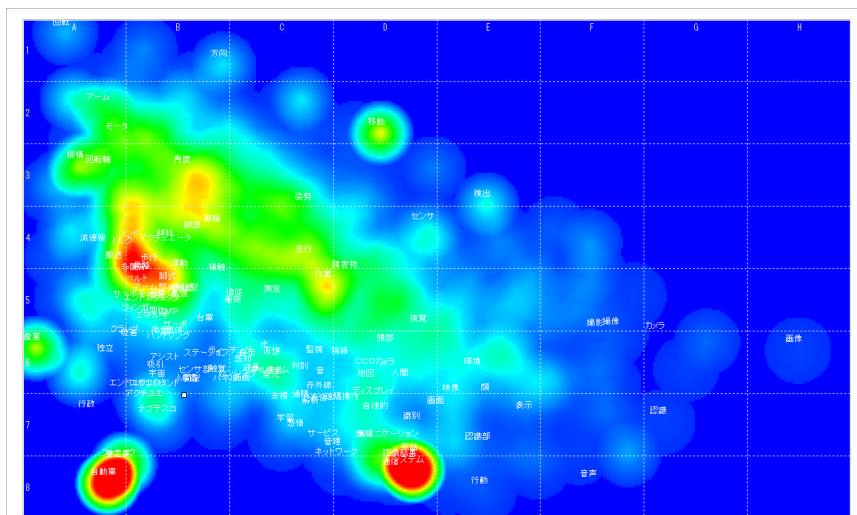
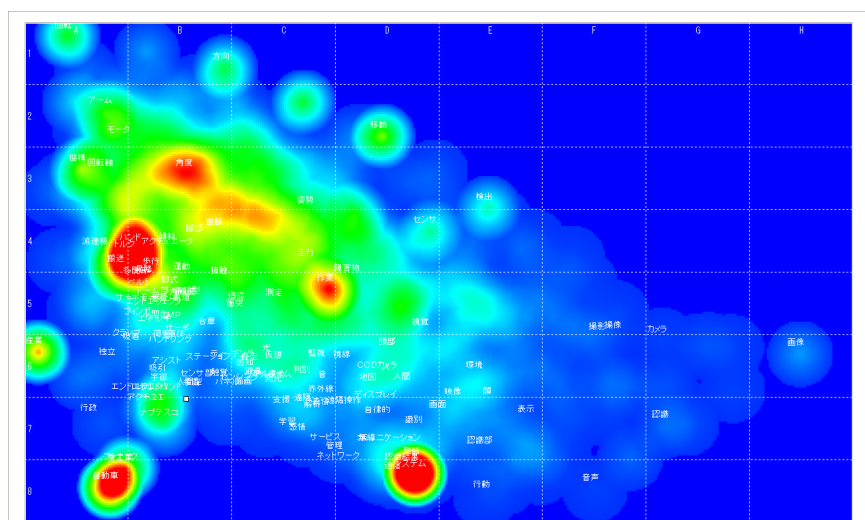


図 3.11: サーマグラフ (2006 年, ロボット)

2008 年のサーモグラフによると、産業用ロボット、ロボットの通信機能に関する特許出願が変わらず多いが、2007 年に比較し、「表示」、「画像」といった単語の頻度が増えている。



40

3.2.4 単語群と出願年の相関マップの作成

形態素解析した単語から、3.2.2 で選択して作成したリストを TrueTeller に読み込ませ、該当する単語と単語が使用された特許の出願年についての相関マップを作成した。

2. 1. 3 (25) の文献によると、単語群と出願年の相関マップの場合、出願年を結んだ線の外側にその年の特徴が現れるとのことである。

当該相関マップの出願年を結んだ線の外側に表示されている当該年の特徴語を検出し、3. 2. 3 のサーモグラフの分析結果と、当該相関マップの分析結果とを表に纏めた。時系列で、サーモグラフ、相関マップ双方で共通する言葉は「青」で表示した。結果を見ると、双方のマップで共通する単語の少ないことがわかった。

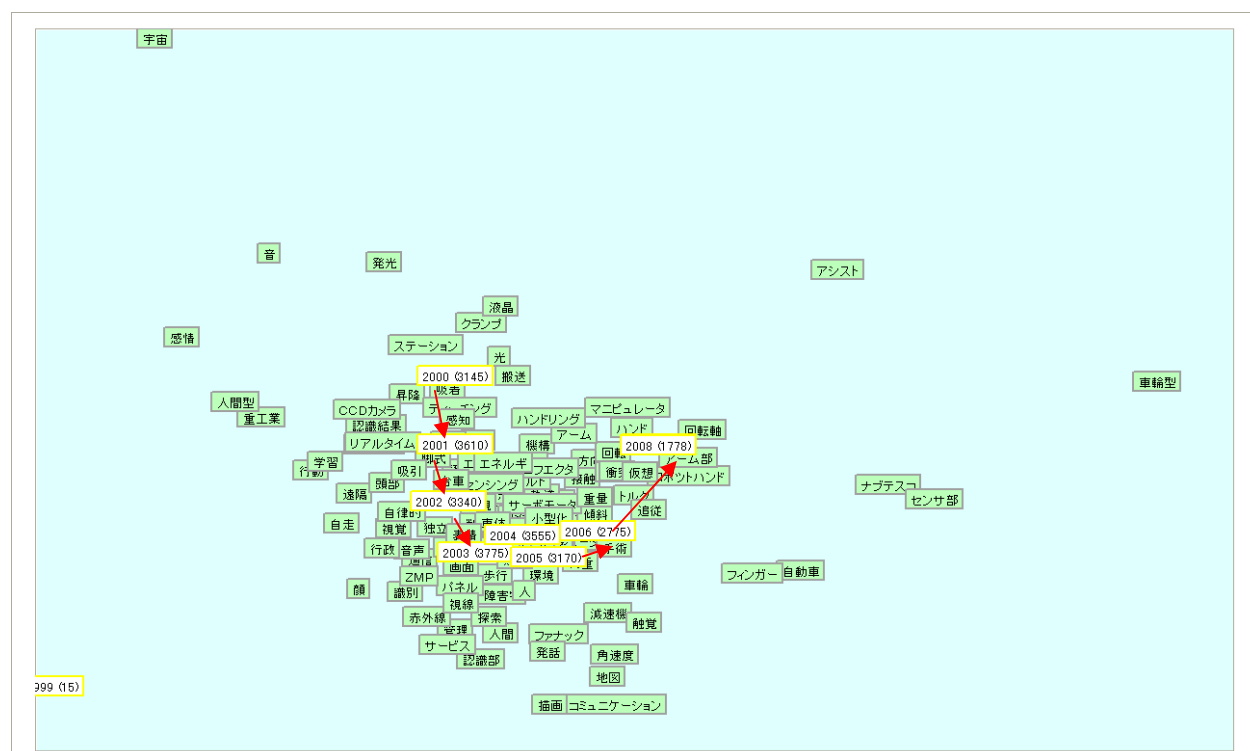


図 3.14: 単語群と出願年の相関マップ

	サーモグラフ	相関マップ
2000 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「ハンド」、「脚部」、「搬送」、「車輪」	「CCDカメラ」、「認識結果」、「リアルタイム」、「重工業」、「人間型」
2001 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「歩行」、「多関節」、「ハンド」、「アクチュエータ」、「アーム部」、「脚部」、「搬送」、「車輪」、「作業」、「回転軸」、「音声」	「学習」、「頭部」、「吸引」、「遠隔」、「作業」
2002 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「車輪」、「接触」、「姿勢」、「作業」、「画像」	「自律的」、「自走」、「視覚」、「音声」
2003 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「多関節」、「行動」、「認識部」、「学習」、「感情」	「画面」、「パネル」、「視線」、「赤外線」、「認識部」、「管理」
2004 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「多関節」、「歩行」、「移動」、「サービス」、「管理」、「ネットワーク」、「視覚」、「アーム」	「歩行」、「障害物」、「探索」、「人間」
2005 年	「重工業」、「自動車」、「通信」、「システム」、「発話」、「認識」、「移動」、「車輪」、「多関節」、「ハンド」、「歩行」、「脚式」	「環境」、「発話」、「描画」
2006 年	「重工業」、「自動車」、「通信」、「システム」、「発話」、「認識」、「歩行」、「多関節」、「	「手術」、「車輪」、「触覚」、「角速度」、「地図」、「コミュニケーション」、「自動車」

	「ハンド」、「脚式」、「車輪」、「作業」	
2007 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「多関節」、「ハンド」、「回転」、「作業」、「方向」「角度」	「追従」、「ロボットハンド」、「センサ部」
2008 年	「重工業」、「自動車」、「通信」、「システム」、「発話」、「認識」、「歩行」、「多関節」、「ハンド」、「脚式」、「表示」、「画像」	

表 3.6: サーマグラフと相関マップにおける特徴語

3.2.5 データマイニング・ツールの評価

TrueTeller の評価を行った。評価の方法としては、特許庁よりロボットの技術動向調査資料²が発行されているため、これらと照合することにより行った。

(1) 特許動向調査資料による評価

特許庁の資料に、1999 年～2004 年の間に出願された特許について、出願件数の伸び率、及び累積出願件数について報告された資料がある。

出願件数の伸び率算出の式は、以下の通りのものである。

$$(\text{2002} \sim \text{2004 年の出願件数}) / (\text{1999} \sim \text{2001 年の出願件数})$$

表の統計は、国別に出されたものであるが、今回の検証は、日本の特許を対象としているので、日本の部分のみ参照した。日本で最も伸びている分野は、紫（1.5 ポイント以上）の「移動型ロボット」、「指」、「ロボット協調・群ロボット」、次に伸びている分野は、薄紫（1.5 ～ 1 ポイント）の「マイクロロボット」、「安全装置」、「アクチュエータ」、「インタフェース」、「遠隔制御」、「パワーアシスト・ウェアラブル」、「自律移動制御」、「画像認識技術」、「音声認識技術」、「音声合成・対話技術」、「音声・画像に拠らないコミュニケーション」で、減少している分野は、青（1 ポイント未満）の「マスタースレーブ型ロボット」、「プログラム制御型ロボット」、「マニピレータ装置付小室」、「把持部」、「腕」、「センサ」、「位置決め制御」、「ティーチングシ

² 平成 18 年度 特許出願技術動向調査報告書「ロボット」平成 19 年 4 月特許庁

中区分	伸び率				累積件数（1999～2004）			
	日本 国籍	米国籍	欧州 国籍	全ての国 籍	日本 国籍	米国籍	欧州 国籍	全ての国 籍
マスタースレーブ型ロボット	0.86	1.08	0.60	0.86	95	25	16	138
移動型ロボット	2.10	1.63	0.94	1.85	867	242	136	1293
マイクロロボット	1.07	1.44	0.44	1.13	116	22	13	162
プログラム制御型ロボット	0.89	0.73	0.54	0.78	619	175	212	1038
マニプレータ装置付き小室	0.86	0.00	1.00	1.00	13	1	8	24
把持部	0.52	0.95	0.84	0.60	1399	260	234	1921
関節・手首	1.00	0.79	0.80	0.95	551	84	106	755
腕	0.61	0.99	0.69	0.70	427	147	140	731
安全装置	1.26	0.88	1.64	1.26	355	32	29	418
指	3.38	0.61	0.43	2.06	434	87	43	575
センサ（ロボット用途）	0.99	1.44	0.71	1.05	691	139	84	946
アクチュエータ	1.14	1.00	2.00	1.14	346	80	39	474
位置決め制御	0.94	1.90	1.35	1.12	670	168	94	959
その他プログラム制御	0.83	2.05	2.19	1.06	942	186	137	1291
ティーチングシステム	0.96	1.44	3.75	1.06	296	22	19	340
インタフェース	1.03	1.55	0.68	1.05	207	56	37	306
遠隔制御	1.10	2.55	1.86	1.70	164	181	60	426
パワーアシスト・ウェアラブル	1.50	2.75	1.00	1.91	25	30	8	64
人工知能技術	0.84	1.39	1.00	0.95	367	105	34	520
自律移動制御	1.27	1.75	1.58	1.45	605	184	98	926
ロボット協調・群ロボット	1.89	2.23	1.14	1.84	104	129	60	298
画像認識技術	1.06	1.98	1.00	1.23	536	152	54	761
音声認識技術	1.04	1.08	0.97	1.07	1004	152	71	1250
音声合成・対話技術	1.06	0.61	0.87	1.01	611	58	28	705
音声・画像に依らないコミュニケーション	1.25	1.06	2.33	1.29	115	33	10	160

表 3.7: 特許出願件数の伸び率、及び累積出願件数（特許庁資料）

ム」、「人工知能技術」であった。

2002～2004 年に、サーモグラフで増えた単語「重工業」、「自動車」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「アクチュエータ」、「アーム部」のうち、「移動」、「歩行」は、特許庁資料の「移動型ロボット」に当てはまり、2.1 ポイント増加、「アクチュエータ」は、そのままの表現で1.14 ポイント増加、「発話」は、特許庁資料の「音声合成・対話技術」に当てはまり、1.06 ポイントの増加となっている。しかし、「アーム部」は、「腕」で0.61 ポイントと減少となっている。「重工業」、「自動車」、「通信」、「システム」は該当がなかった。

2002～2004 年に、相関図で特徴語として検出された単語、「自律的」、「自走」、「視覚」、「音声」、「画面」、「パネル」、「視線」、「赤外線」、「認識部」、「管理」、「歩行」、「障害物」、「探索」、「人間」については、同様に検討した結果、「自律的」、「自走」、「音声」、「画面」、「パネル」、「認識部」、「歩行」については増加分野に該当し、「視覚」、「視線」、「赤外線」、「管理」、「障害物」、「探索」、「人間」については、該当がなかった。

大区分	中区分	小区分
全体構造技術	マスタースレーブ型ロボット	—
	移動型ロボット	二足歩行、多脚、車輪、クローラ、素状体、飛行、その他移動手段
	マイクロボット	—
	プログラム制御型ロボット	直交座標型、円筒・極座標型、多関節型、パラレル型、その他プログラム制御型ロボット
	マニピレータ装置付き小車	—
部分構造技術	駆動部	—
	関節・手首	—
	腕	—
	安全装置	—
	指	—
制御技術	センサ（ロボット用途）	視覚、聴覚、触覚、その他センサ
	アクチュエータ	—
	位置決め制御	—
	その他プログラム制御	—
	フィードバックシステム	—
知能化技術	インタフェース	制御装置、制御台、その他人による制御手段
	遠隔制御	—
	パワーアシスト・ウェアラブル	—
	人工知能技術	ニューラルネットワーク、確率モデル、知識ベース制御、行動ベース制御、模動・感情モデル、適化・適応、その他の数学的モデル、その他の学習・推論
	自律移動制御	自己位置同定、マップ生成、S L A M、経路計画、障害物回避、ナビゲーション
コミュニケーション技術	ロボット協調・群ロボット	分散機能制御、群ロボット、その他ロボット協調
	画像認識技術	物体認識、人物認識、動画認識、その他画像認識技術
	音声認識技術	音声認識技術、話者・音源の同定
	音声合成・対話技術	音声合成、対話
	音声・画像に依らないコミュニケーション技術	—
応用技術	サービスロボット	農林業、畜産、運輸・倉庫、土木・建築、医療、オフィスワーク（受付・案内・搬送等）、警備、介護・福祉、ゴミ処理・清掃、エンターテインメント、ホーム（掃除）、ホーム（留守番）、ホーム（子守）、ホーム（その他）
	特殊環境用ロボット	探査、水中・海洋、宇宙、原子力、メンテナンス、軍事、廃棄物処理、災害対応
	産業用ロボット	製造業（溶接、塗布・塗装、研磨・バリ取り、入出荷、組み立て、樹脂成形、金属加工、電子部品実装（半導体等）、計測・分析、搬送）、バイオ産業

表 3.8: ロボット分野の技術分類表（特許庁資料）

（２）論文動向調査資料による評価

以下は、同じく特許庁資料の中の、1999～2005 年の分類別論文発表件数である。

サーモグラフの 2000～2005 年の特徴語、「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「作業」、「アクチュエータ」、「アーム部」、「ハンド」、「脚部」、「搬送」、「回転軸」、「接触」、「姿勢」、「画像」、「行動」、「認識部」、「学習」、「感情」について、論文分類別件数にあったのは、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「アクチュエータ」、「アーム部」、「車輪」、「音声」、「画像」、「行動」、「認識部」、「学習」、「感情」であり、「二足歩行」4 位、「その他移動手段」が 6 位であった。

「重工業」、「自動車」、「産業」、「通信」、「作業」、「ハンド」、「脚部」、「搬送」、「回転

軸」、「接触」、「姿勢」は、該当がなかった。

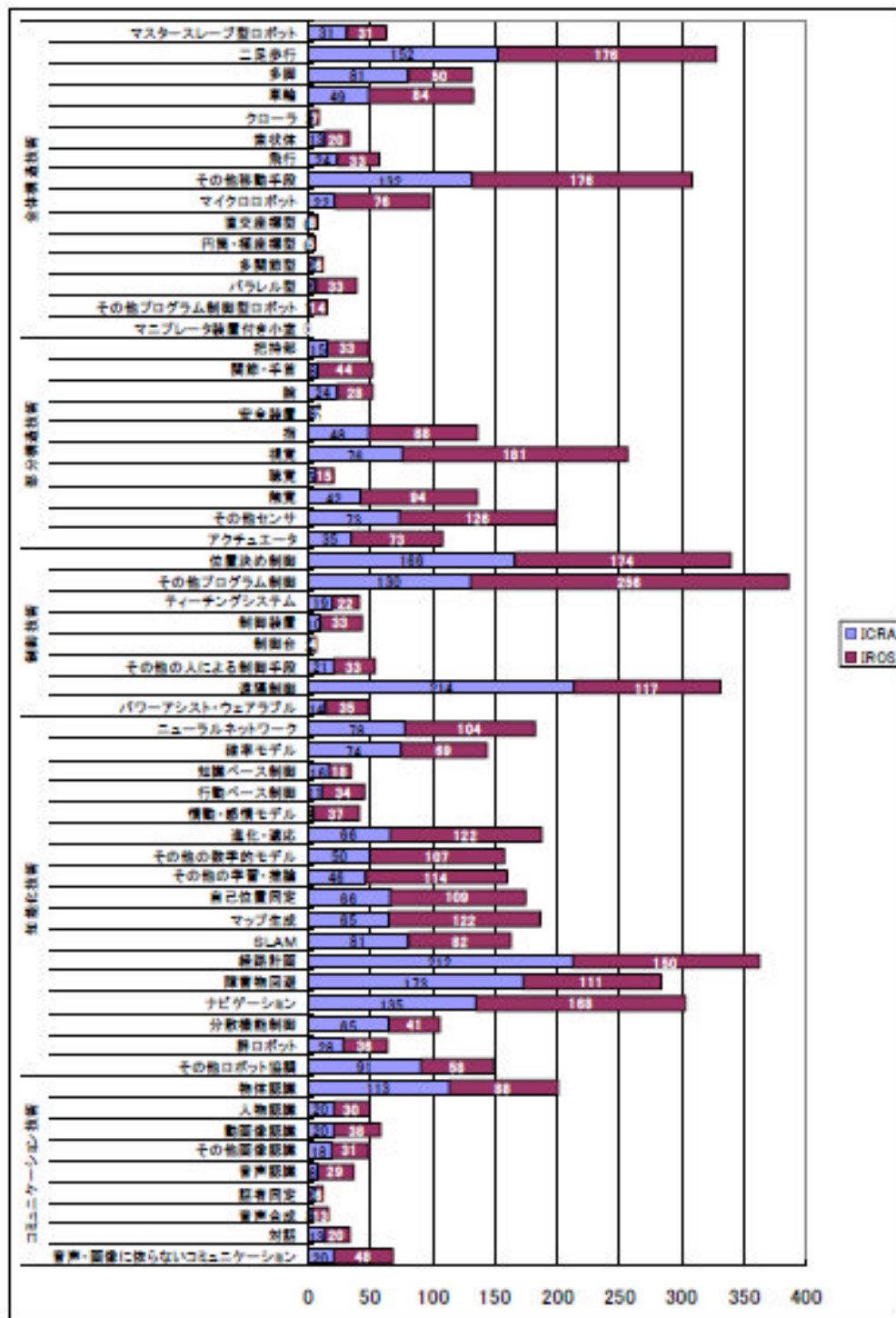


図 3.15: ロボット分野における分類別論文発表件数（特許庁資料，2006 年）

相関マップの 2000～2005 年の特徴語、「C C Dカメラ」、「認識結果」、「リアルタイム」、「重工業」、「人間型」、「学習」、「頭部」、「吸引」、「遠隔」、「作業」、「自律的」、「自走」、「視覚」、「音声」、「画面」、「パネル」、「視線」、「赤外線」、「認識部」、「管理」、「歩行」、「障害物」、「探索」、「人間」、「環境」、「発話」、「描画」、「サービス」、「管理」、「ネットワーク」について、論文分類別件数にあったのが、「学習」、「遠隔」、「視覚」、「音声」、「認識部」、「歩行」、「障害物」、「発話」、であり、特に、「二足歩行」4位、「遠隔制御」5位、「障害物回避」8位、「視覚」9位であった。

「C C Dカメラ」、「認識結果」、「リアルタイム」、「重工業」、「人間型」、「頭部」、「吸引」、「作業」、「赤外線」、「管理」、「環境」、「描画」、「サービス」、「管理」、「ネットワーク」について該当がなかった。

以上の評価の結果を表に纏めた。資料の関係により部分的な評価となったが、特許資料と合致する言葉は「緑」(2002～2004 年の評価)、論文資料と合致する言葉は「赤」(2000～2005 年の評価)で表した。特許、論文で重複するものについては、赤字に緑のマーカーで表現している。

サーモグラフ、相関マップで検出した単語に対し、他の資料でも引用されていた単語の割合を算出した。

サーモグラフで、 $13/42=0.31$ [2002～2004 年 (特許)]、 $47/86=0.55$ [2000～2005 年 (論文)] であった。

相関マップで、 $7/14=0.5$ [2002～2004 年 (特許)]、 $7/27=0.26$ [2000～2005 年 (論文)] であった。

この結果より、特許動向資料については、単語群と出願年との相関マップが、論文動向資料については、サーモグラフが、合致する単語数の多いことが判明した。なお、合致する割合は高くとも 5 割程度であった。

	サーモグラフ	相関マップ
2000 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「ハンド」、「脚部」、「搬送」、「車輪」	「CCDカメラ」、「認識結果」、「リアルタイム」、「重工業」、「人間型」
2001 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「歩行」、「多関節」、「ハンド」、「アクチュエータ」、「アーム部」、「脚部」、「搬送」、「車輪」、「作業」、「回転軸」、「音声」	「学習」、「頭部」、「吸引」、「遠隔」、「作業」
2002 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「移動」、「歩行」、「多関節」、「車輪」、「接触」、「姿勢」、「作業」、「画像」	「自律的」、「自走」、「視覚」、「音声」
2003 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「多関節」、「行動」、「認識部」、「学習」、「感情」	「画面」、「パネル」、「視線」、「赤外線」、「認識部」、「管理」
2004 年	「重工業」、「自動車」、「産業」、「通信」、「システム」、「発話」、「認識」、「多関節」、「歩行」、「移動」、「サービス」、「管理」、「ネットワーク」、「視覚」、「アーム」	「歩行」、「障害物」、「探索」、「人間」
2005 年	「重工業」、「自動車」、「通信」、「システム」、「発話」、「認識」、「移動」、「車輪」、「多関節」、「ハンド」、「歩行」、「脚式」	「環境」、「発話」、「描画」

表 3.9: サーモグラフと相関マップにおける特許・論文該当語

第4章 技術動向を分析する技術に関する考察

第2章で調査した技術動向を分析する技術に関する文献、第3章で行った使用可能な可視化ツールの検証結果を元に、以下、特許情報を利用して技術動向を分析、可視化するためには、どのような技術が良いのか考察を行った。

4.1 従来技術に関する考察

今回、1. 専門用語を検出する技術、2. 簡易な用語への置き換え技術、3. 技術動向を可視化する技術の3点において調査研究を行ったが、以下、各技術についての考察を行うこととする。

4.1.1 専門用語を検出する技術に関する考察

第1に、専門用語を検出する技術について、論文において研究されているように、特許情報からも、手がかり語、上位・下位関係を利用しての構造分析により検出が可能であると考ええる。特に、「請求項」の項目では、書き方のスタイル（文章構造）に一定の傾向があり、例えば、「Aを備え、Bを備え、Cを備えたことを特徴とするD」、「Eを有し、Fを有し、Gを有することを特徴とするH」といったスタイルが多い。この場合、A、B、Cは、Dの、E、F、Gは、Hの要素技術となるので、「備え」や「有し」を手がかり語として、手がかり語の直前の語を専門用語として抽出することが可能となる。

また、「請求項」の前半、「～において」、「～であって」等の手がかり語の前の「～」の部分（preamble、プレアンブル）は、発明の属する技術を説明する場合が多い。「～において」、「～であって」を手がかり語として、手がかり語の前に使用された単語群の中の共通語により特許を分類し、共通語が使用されているIPC等の特許分類をコードとして割当て、当該特許の属する分野を特定することが可能であると考ええる。

上記手法で分類した、分類を同じくする専門用語について、当該専門用語が最初に使用された特許公開（特許）公報の年月日の情報を取得することにより、時系列で特定分野における技術動向を把握することが可能であると考ええる。

4.1.2 簡易な用語への置き換え技術に関する考察

第2に、簡易な用語への置き換え技術について、特許・論文間の引用関係を利用し、特許が引用している論文の用語に置き換えることも従来研究より考えられるが、論文を引用している特許は数パーセントと言われており、非常に少ない点、特許で引用している論文を特定したとしても、論文の用語の中から特許用語に対応する用語を決定するには当該技術の専門知識が必要となる点より、自動置き換えは難しいと考える。また、同じ技術に対して、特許では明細書の書き手により表現が異なるため、辞書を作成することも難しい。狭い分野で、人手により種々ある表現を纏めて辞書としていくことは可能であると思うが、機械的に広い分野を網羅することは困難である。結論としては、特許用語の簡易な用語への置き換えは難しいと考える。

4.1.3 技術動向を可視化する技術に関する考察

第3に、技術動向を可視化する技術に関しては、例えば、引用関係を利用し、引用関係にある特許公開（特許）公報からキーとなる専門用語を検出して当該専門用語をマップのノードとして、引用関係により配置することができれば、技術動向を把握するのに有効であると考えられる。

また、共通キーワードを上位ノードとした **Semantic network** 図の研究がなされている。単語の頻度ランキングを出し、キーワードを行、キーワードが使用された特許公開（特許）公報を列としてマトリクスを作成し、**K-means** 法によるクラスタリングにて特許（公開）公報をグループ分けし、各グループに含まれているキーワードを特定し、共通キーワードを上位ノードとする **Semantic network** を作成、更に、キーワードが含まれている最先の出願日（書誌情報）を抽出して、キーワードと出願日が記載されているノードを年順に配置して、技術動向を可視化したものである。グループ間で共通するキーワードを上位ノード（概念）として、時系列で前の技術であるとしている。

確かに、複数の文献で取り上げられている用語は、研究者の間で既知になっており使われている可能性が高いので、機械的にマップ化するための着眼点として、有効なものであると考える。なおかつ、当該用語が最初に使われている特許公開公報の出願日情報を取得して、年度別に時系列で表しているため、技術動向について、視覚的にわかりやすい。ただし、単語の頻度ランキングは自動で出すものの、マップに使用する単語は、ユーザが選択するものであることと、ネットワーク図が多数生成された場

合は、その中から技術動向を表しているネットワーク図をユーザが選択しなければならないことから、当該技術による可視化マップの生成は、調査する分野の専門知識がないと難しい点において限界があると言える。例えば、未参入の分野で事業を行うことを考えている企業が、当該技術を使用して技術動向を把握しようとしても、当該分野の専門家がいない段階では、当該技術を使つての動向把握は難しいことになる。

4.2 ツールに関する考察

前章において、可視化ツールを使用し、技術動向を俯瞰できるかについて評価を行った。MathoPatent は、例えば要約に使用されている全てのキーワードが、各キーワードの使用されている特許公開（特許）公報の書誌情報と紐で結びつけられている。紐の数かなり多く、一見してわかりにくく、カーソルで紐を解くように動かさないと、関係を認識することができない。キーワードが使用された全ての年がわかる点では有効であるが、技術動向を俯瞰することは難しい。

TrueTeller は、まず、形態素解析を行った後の単語の選択に、調査する当該分野の専門知識が必要となる。また、サーモグラフは、単語の使用件数が多いかを見ることができ、年度ごとに1枚ずつ作成しなければならない、時系列で一度に見ることに適していない。単語群と出願年の相関マップは、時系列に表現されるので見た目に分かりやすいが、その年の特徴語を検出しているものの、単語が最初に使われた年を示すものでないため、技術がどのように開発されていったかを把握するには、必ずしも適切であるとは言えない。

また、精度について検証を行ったが、発表資料と照合した結果、検出した特徴語と発表資料で報告されている技術用語とが合致する割合は、5割程度であった。

4.3 まとめ

調査研究した3つの工程を含む一連のシステムとして、技術動向分析をするためにどのようなものが有効であるのか検討した結果を述べる。

まず、特許情報より技術動向を表すための専門用語を検出する対象として、特許データベースより必要な情報を切り出す目的で、キーワード及び特許分類により検索式を作成し、検索式にて検索することにより特許データベースより特定分野の特許公開（特許）公報の特許情報を取得する。

次に、得た特許情報より、例えば、請求項といった項目を特定し、項目の情報から手がかり語「備え」や「有し」などを使い、専門用語を検出し、専門用語リストを作成する。また、検出した専門用語の分類のため、「請求項」の前半、「～において」、「～であって」等の手がかり語の前の部分は、発明の属する技術の説明部分（**preamble**）であるため、当該部分より単語を検出し、検出した単語から他の特許公開（特許）公報の **preamble** と共通する語があるか否かで分類を行う。

そして、共通語を請求項の後半部分で検出できる特許公開（特許）公報を探し、見つければ、共通語を専門用語の親ノードとし、専門用語、共通語それぞれ請求項の後半部分にて検出した特許の出願日情報を取得し、時系列で配置を行い、同様のことを繰り返し、用語をつなげていく。

このようなシステムとすると、当該分野の専門知識を必要とせず、技術動向を可視化することのできる技術動向マップを作成できるのではないかと考える。

第5章 今後の課題

技術動向の可視化技術の課題として、共通語を親ノードとするネットワーク図を作成した場合で、多くのネットワーク図が生成されたとき、どのネットワーク図が技術動向を表しているのか選択する必要性が生じ、この場合には専門知識が必要となる。

また、技術動向マップにより、これまでの技術動向を俯瞰することはできるものの、今後の技術動向の予測まではできないため、技術動向を予測することのできるマップについて検討すべきであると考ええる。

第6章 結論

技術動向分析技術として、1．専門用語の検出技術、2．簡易な用語への置き換え技術、3．技術動向の可視化技術の3点に分けて、調査研究を行った。1．専門用語の検出技術に関しては、特許情報においても手がかり語を使って構造解析を行え、また、専門用語の分類も可能であると考ええる。2．簡易な用語への置き換え技術については、論文用語から特許用語への変換、特許情報の中からの簡易な用語の検出等考察したが、特許明細書の書き手により表現が異なり、1つの用語について種々表現があることから難しいとの結論に至った。3．技術動向の可視化技術については、特許間の共通語を親ノードとして、当該共通語を最初に用いた特許の出願日を時系列で配置するネットワーク図が有効であると考ええるが、当該技術分野の専門知識が必要であり、課題が残る。

参考文献

- [1] 今井 俊, “表題解析による科学技術論文の自動分類”, 北陸先端科学技術大学院大学修士論文, 1999.
- [2] 難波英嗣, 谷口裕子, “学術論文データベースからの研究動向情報の抽出と可視化”, 言語処理学会 第 12 回年次大会, p35-38, 2006.
- [3] 近藤友樹, 難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山, “論文データベースからの研究動向情報の抽出”, 言語処理学会 第 13 回年次大会, p470-473, 2007.
- [4] 安藤 俊幸, “特許情報の分析・評価支援: 「termmi」 と統計解析言語 R による特許情報の可視化”, 第 4 回情報プロフェッショナルシンポジウム, p13-17, 2007.
- [5] 大石康智, 伊藤克亘, 武田一哉, 藤井敦, “単語の共起関係と構文情報を利用した単語階層関係の統計的自動識別”情報処理学会研究報告, SLP-61, p25-30, 2006.
- [6] 難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山, “特許データベースからのシソーラスの自動構築”, 言語処理学会 第 13 回年次大会, p1113-1116, 2007.
- [7] 有賀康裕, 内藤和幸, “パテントマップソフトの特許検索への活用 特許調査に不慣れな技術者または初心者でも精度の良い調査結果を得る手法”, 第 2 回情報プロフェッショナルシンポジウム, p1-4, 2005.
- [8] Mark Giereth, Thomas Ertl, “Visualization Enhanced Semantic Wikis for Patent Information”, 12th International Conference Information Visualization , p185-190, 2008.
- [9] Zhiqiang Liu, Donghua Zhu, “Web Mining based Patent Analysis and Citation Visualization”, School of Management and Economics, Beijing Institute of technology, 2009.
- [10] 渡辺勇, 小川知也, 田中一成, “特許情報の分析・評価支援—多観点分類と引用分析—”, 第 2 回情報プロフェッショナルシンポジウム予稿集, p19-23, 2005.
- [11] 安善奈津美, 難波英嗣, 相沢輝昭, 奥村学. “特許、論文データベースを統

合した検索環境の構築” 情報処理学会 研究報告, NL-168, p21-26, 2005.

[12] 難波英嗣, 藤井敦, 岩山真, 橋本泰一, “学術論文の国際特許分類への自動分類: 第7回 NTCIR ワークショップ特許マイニングタスク成果報告”, 2009.

[13] 釜屋英昭, 難波英嗣, 竹澤寿幸, 奥村学, “論文用語の特許用語への自動変換” 言語処理学会 第14回年次大会, p801-804, 2008.

[14] Mase H, Matsubayashi T, Ogawa Y, Yayoi T, Sato Y, Iwayama M, “NTCIR-5 Patent Retrieval Experiments at Hitachi, “, Proc. Of NTCIR-5 Workshop Meeting, PP.318-323, 2005.

[15] Chen-Yuan Liu, Shenq-Yih Luo, “Analysis of developing a specific technological field using the theme code of Japanese patent information”, Scientometrics, Vol.75, No.1, p 51-65, 2008.

[16] Edward Badger, “Techniques for analyzing literature search results”, World Patent Information 30, p326-334, 2008.

[17] Shann-Bin Chang, Shu-Min Chang, Wei-Yuan Guh, “Exploring the Technology Diffusion Trajectories and Groups of Basic Patents of Business Methods: Using the Patent Citation Network”, PICMET 2007 Proceeding, p1784-1789, 2007.

[18] Jong Hwan Suh, Sang Chan Park, 2006, “A New Visualization Method for Patent Map: Application to Ubiquitous Computing Technology”, Advanced Data Mining and Applications, Springer Berlin / Heidelberg, Volume 4093, 2006.

[19] 宮田道生, 福川忠昭, 高橋正子, “特許戦略のための特許情報の計量化ーテキストマイニングを活用した企業の注力技術比較の試みー”, 日本オペレーションリサーチ学会, 2006年秋季研究発表会, p78-79, 2006.

[20] 伊神正貫, “ナノテクノロジー関連特許出願のマッピング、引用関係を用いた技術群の同定と知識の流れの計測”, The Japan Society for Science Policy and Research Management, p688-691.

[21] 野崎篤志, “図面情報を用いたパテントポートフォリオ分析”, 知財学会・第6回年次学術研究発表会, 2008.

[22] 有賀康裕, 新井大樹, “US クラス 705 に関するパテントマップ解析”, 情報管理 Vol. 43, No.11, p965-975, 2001.

[23] 桐山勉, 田中宣郎, 川島順, 都築泉, 大山勝弘, 長谷川正好, 玉置研一, “分析 Tool を利用した Rapid Technology Intelligence Process 研究 (1): プロフェッショナル・ディベロップメントとテック・マイニング”, 第 2 回情報プロフェッショナルシンポジウム, p133-137, 2005.

[24] 中居 隆, “テキストマイニングによる知財ポートフォリオ分析”, 情報管理. Vol. 51, No. 3, p194-206, 2008.

[25] 山本外茂男, “産学連携のマッチング性分析におけるテキストマイニングの有効性”, 情報の科学と技術 59 巻 6 号, p291-297, 2009.

以上