

# 自然言語処理技術を活用したパテントマップ自動生成システムの提案

Automatic Generation of Patentmap by using Natural Language Processing

IRD 国際特許事務所 / 株式会社アイ・アール・ディー **太田 貴久**

2006年豊橋技術科学大学大学院博士前期課程知識情報工学専攻修了。2006～2015年同大学研究員。2016年～IRD国際特許事務所。修士（工学）、言語処理学会会員、日本知財学会会員。2014年度から特許版産業日本語委員会委員。

✉ tota@ird-pat.com ☎ 06-6944-4530

IRD 国際特許事務所所長（弁理士） / 株式会社アイ・アール・ディー **谷川 英和**

1986年神戸大学工学部システム工学科卒業。同年、松下電器産業（株）に入社し、中央研究所等において、データベース管理システム等の研究開発に従事。1999年弁理士試験合格。2002年1月、IRD国際特許事務所を開設。所長、弁理士。2003～2007年3月京都大学COE研究員、2007年4月～京都大学非常勤講師、2009年4月～東京工業大学客員教授。博士（情報学）。弁理士会、日本知財学会、情報処理学会各会員。2007年度から特許版産業日本語委員会委員。

✉ htanigawa@ird-pat.com ☎ 06-6944-4530

## 1 はじめに

特許調査では、「自社・他社の強みや弱みの分析」や「特定の開発目的や特定の技術に関する出願動向の把握」といった調査が行われることがある。これらの調査では、各特許がどのような目的で開発され、どのような技術が用いられているかを抽出し、その分析を行う。この際、特許調査の現場では、パテントマップと呼ばれる図がしばしば用いられる。パテントマップとは、「特許情報を整理・分析・加工して図面、グラフ、表などで表したものの」<sup>[1]</sup>である。特に前記調査の場合では、特許に記述されている発明の課題（発明の効果／目的、「リサイクル性向上」など）と、構成（解決手段、発明が使用する特徴的な技術）の2つを軸として、軸上に配置された項目（以下、軸項目とよぶ）に対応する特許の件数をバブル等で表したパテントマップが役立つ。本稿では、このようなタイプのパテントマップを課題－構成マップと呼ぶ。課題－構成マップの例として、特許庁が作成したアルミリサイクルに関するパテントマップ<sup>[2]</sup>を図1に示す。

図1のように、課題－構成マップは、どのような技術

が活発に開発されているかが一目でわかり、さらに、バブルの内容を確認することで、どのような組織がその技術を保有しているのかが容易にわかる。

現在、このような課題－構成マップは、人手で作成されている。大量の特許から適切な課題や構成に関する表現を抽出することは非常に手間のかかる作業である。規模や技術分野にもよるが、1つのパテントマップを作成するために2～3ヶ月かかる場合もある。そこで、本稿では、自然言語処理を活用し、このような課題－構成マップを自動的に生成するシステムを提案する。

## 2 従来技術

従来、自然言語処理技術を利用したパテントマップ自動作成に関する研究<sup>[3]</sup>や実用的なシステムとしてNRI True Teller<sup>[4]</sup>等が提案されている。しかしながら、従来の研究やシステムは、特許全体をベクトル化し、そのベクトルに対して、多次元尺度構成法（MDS）や主成分分析等を施すことで次元数を縮減（多くは2次元）し可視化するものである。このようなタイプのマップは、軸の意味の解釈が自明でない等の問題がある。例えば、

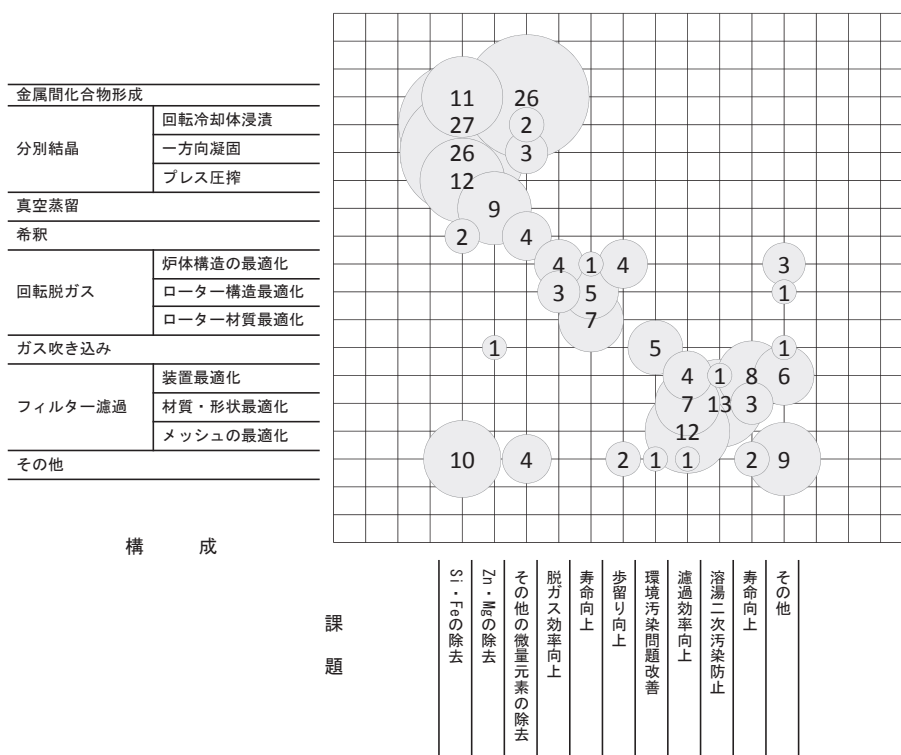


図1 課題-構成マップ

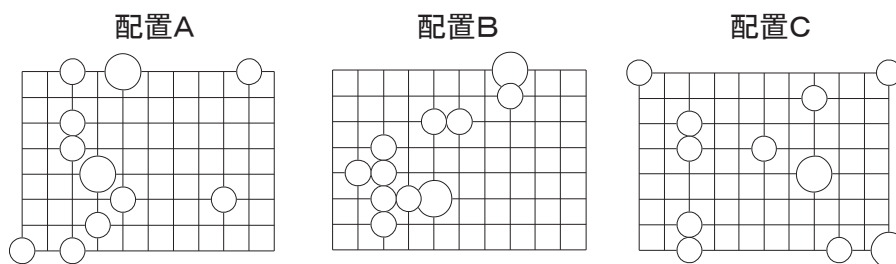


図2 バブルチャートの配置

このようなマップにおいて、特許が1つも配置されていない「空白地域」の意味を人間が理解することは非常に困難である。

これに対して、課題-構成マップは、課題と構成という観点で特許を分類しているため、各軸の解釈が明確であるという利点がある。

しかしながら、課題-構成マップにも問題はある。それは、軸項目の配置が「自由」な点である。出願年や公開年といった量的な軸項目の場合、大小関係が存在するため、軸項目の並びは一定（昇順もしくは降順）である。一方、課題-構成マップのような言語の軸項目の場合、そのままでは順序関係を決定できないため、様々な形のマップを作ることができてしまう。その結果、生成したマップが分析を妨げることや、恣意的な分析結果を導く可能性がある。ここで、同一のデータから生成した3つバブルチャートを図2に示す。

図2のように、同一のデータであっても軸項目の並びによってその印象は全く異なることが確認できる。

そこで、本稿で提案するシステムは、「出願人ごとに極力集めて配置」や「出現数に基づいて時系列順に配置」といったユーザが指定した条件に基づいて、その条件を最大限に満たすように軸項目を配置する機能を有する。

### 3 システム概要

本稿で提案するシステムは、ユーザから任意の特許群を受け付け、その特許群についての課題-構成マップを自動的に生成するシステムである（図3）。

そして、本システムが出力する課題-構成マップは、図4のような、ユーザから受け付けたデータをもとにブラウザ上で操作可能な課題-構成マップを生成する一種のWebアプリケーションである。図4において、ユー

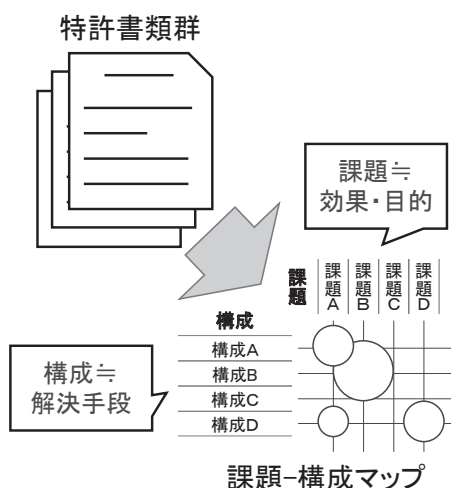


図3 システムの入出力

ザは巨大なマップでも自由に表示範囲を移動することができ、また、軸項目の種類や配置を自由に変更することができる。さらに、ユーザは入力した特許群の中から、その一部を自由に選択して、新たなマップを生成することもできる。

## 4 課題-構成マップ生成手法

本システムは、ユーザが入力した特許群の各特許に対して、様々な自然言語処理等を適用し、課題-構成マップを生成する。本システムのマップ生成処理の概要を図

5に示す。

マップ生成処理では、最初にユーザが入力した特許群の各特許から課題と構成に関する表現を抽出する（書式解析～課題／構成表現抽出）、次に、抽出した表現から軸項目を生成し（軸項目生成）、最後に、統合した表現をどのように各軸上に配置するかを決定する（軸項目配置）。また、本システムでは、課題-構成マップの生成と並行して、ユーザが入力した特許群に関する技術動向予測を行う（動向予測）。以下、各処理の概要について説明する。

### 4.1 書式解析

マップ生成処理では最初に、入力された特許群の書式を解析する。具体的には、日本の特許の場合、書類中に現れる隅付き括弧や、明細書中において、隅付き括弧の直後に現れる短い文字列（見出しと推定される文字列、例えば<実施例>等）を抽出し、その構造を解析する。本処理によって、入力されたデータを「書誌事項」、「特許請求の範囲」、「明細書」、「要約書」、「図面」等に分割し、さらに、特許請求の範囲の中の各【請求項】や、明細書の【発明を実施するための形態】といった見出しの構造を解析する。

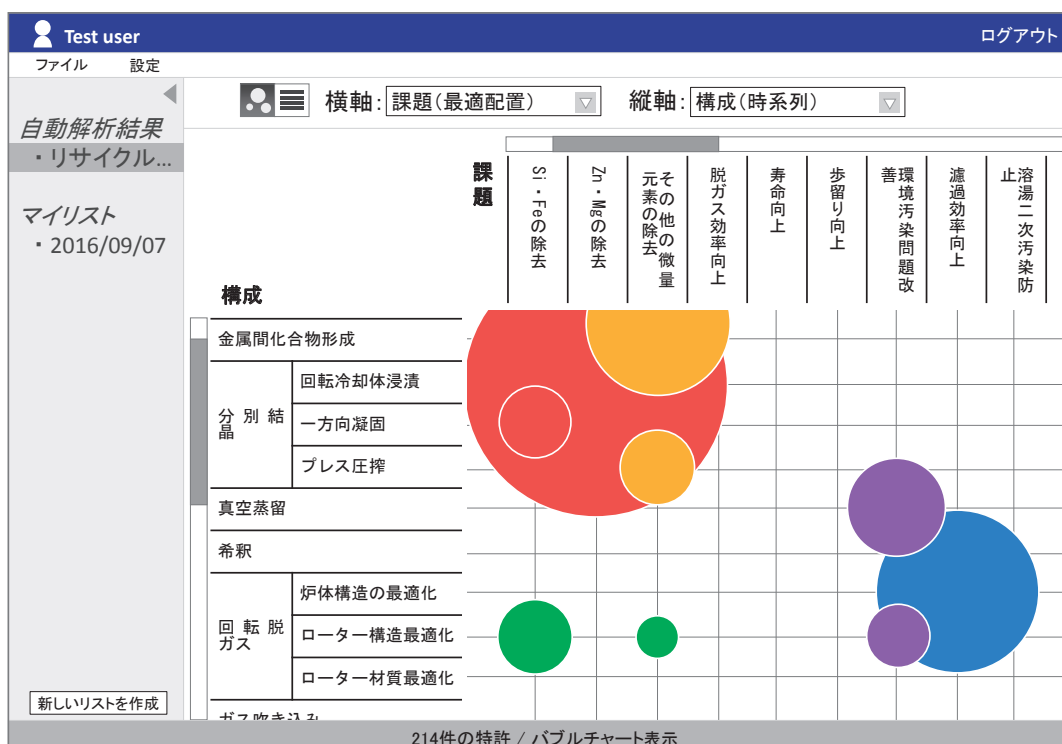


図4 システムのイメージ

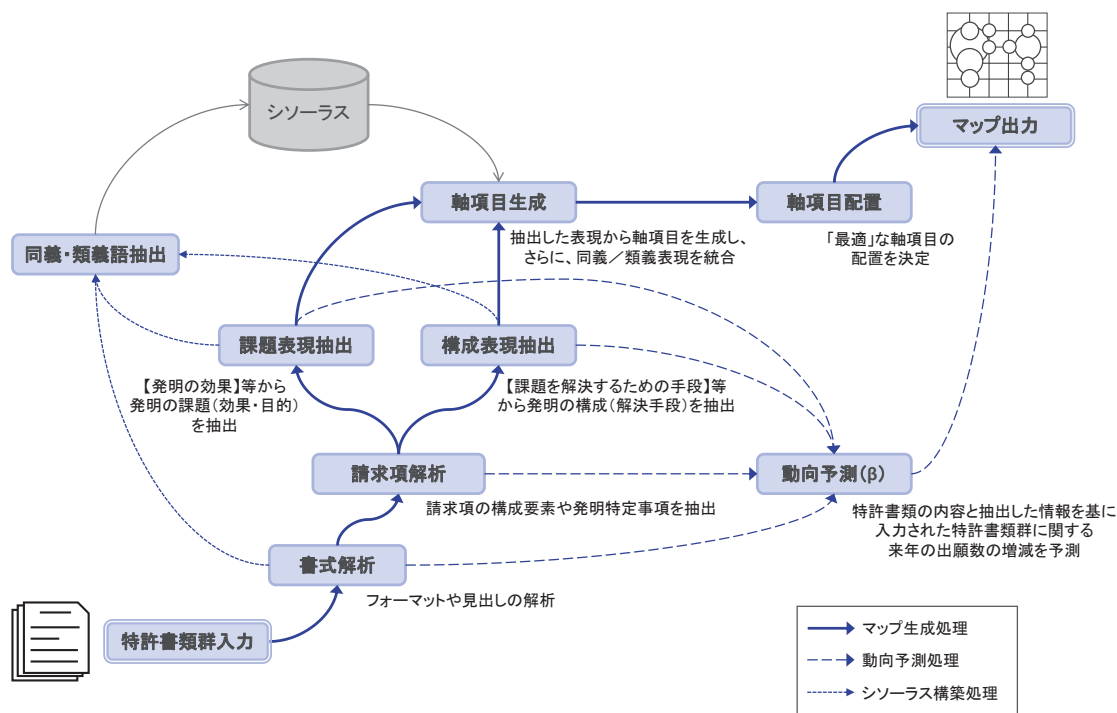


図5 処理の概要

## 4.2 請求項解析

次に、入力された各特許の請求項の構造を解析する。本システムにおける請求項の解析では、請求項の構成要素と、それを修飾する発明特定事項を抽出し、各構成要素がどの構成要素と関連するかを判別する。

## 4.3 課題表現抽出／構成表現抽出

請求項の解析が終わった後、課題－構成マップの軸項目の元となる課題に関する表現と構成に関する表現を各特許から抽出する。詳しい説明は省略するが、文献 [5,6] の手法や、先の請求項解析で得られた請求項の情報を用いた手法（例えば、請求項に現れる構成要素やその他の発明特定事項は課題（効果・目的）を表す表現ではない等）といった手法を併用して課題と構成に関する表現を抽出する。この際、書式解析によって得られた見出しに関する情報も活用する（例えば、課題（効果・目的）に関する表現は【発明の効果】に現れる等）。

## 4.4 軸項目生成

課題／構成表現抽出によって取得した各表現は、多くの場合、図上に表示するには長すぎるため、課題－構成マップの軸項目として適切ではない。そこで、文献 [7]

の手法等を用いて、表現を短くすることで、表現から軸項目を生成する。

さらに、同文献 [7] の手法を用いて、課題に関する軸項目を構成する形態素を素性としたクラスタリングによる、意味のまとめ上げを行う。これにより、表現が多少異なる軸項目を同じ軸項目として扱えるようになる。

一方、構成に関する表現は、特許が権利範囲を可能な限り広く取ろうとする関係上、課題に関する表現より抽象的で多様な表現が用いられる傾向がある。本システムでは、このような構成に関する表現をまとめ上げるために、文献 [8] の手法と同様の手法をシソーラスを用いて実行する。具体的には、シソーラスを利用し、各語の間の上位・下位関係を推定した後、異なる特許間で語を比較することで統合する。これは、同一特許内で、請求項に記載された語は、明細書に書かれた語の上位語であることが多いという特許文書の特徴を利用した手法である。

## 4.5 軸項目配置

これまでに説明した処理によって課題－構成マップをプロットすることは可能である。しかしながら、先にも述べたように、課題－構成マップの軸項目の配置は自由

であるため、恣意的な結果を導いてしまうマップを生成する恐れがある。そこで、本システムでは、各特許から抽出した軸項目を適切に並べる機能を提供する。

簡単にその手法を説明する。提案手法では、ユーザが指定した「2つのバブルA、Bが接したときの利得」に基づき、課題－構成マップ（バブルチャート）全体の総利得が可能な限り大きくなるように軸項目を配置する。ここで、バブルが接するとは、図6に示すいずれかの位置にバブルが存在することを意味する。これらの位置にバブルが存在するとき、2つのバブルのいずれかの軸項目が共通するため、2つのバブルが接する軸項目間には関係があると考えられる。

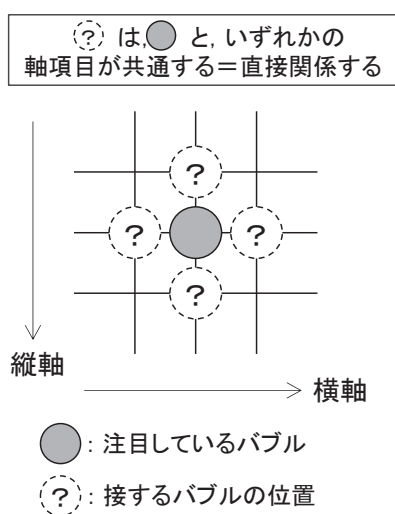


図6 接するバブルの位置

なお、最もシンプルで基本的な「2つのバブルA、Bが接したときの利得」は、「バブルAのサイズ×バブルBのサイズ」である。本システムでは、この利得を基本として、以下の式で定義される出願人を考慮したコスト等を利用することができる。

$$C(A, B) = \sum_i |a_i| |b_i|$$

ここで、 $i$ は出願人を表し、 $|a_i|$ および $|b_i|$ は、それぞれバブルAの出願人 $i$ の出願件数、バブルBの出願人 $i$ の出願件数を表す。

なお、詳しい説明は省略するが、総利得が最大となるような軸項目の配置を求める問題はNP困難に属するため、規模が大きな課題－構成マップの場合、最適解の導出が困難である。本システムでは、現実的な時間で計算可能な範囲で、総利得が可能な限り大きくなるように軸項目の配置を決定する。

さらに、本システムでは、上記の利得の最大化以外にも、「軸項目が最大出願数に達した年が早い順番」などの時系列に基づいた軸項目の配置も可能である。

## 4.6 動向予測 ( $\beta$ )

次に、本システムが備える簡単な技術動向予測機能について説明する。本システムは、ユーザから入力された特許群に関連する要素（出願人や特許分類コード（IPC等））について、来年度の出願数が「10%以上増加」「10%以上減少」「それ以外（現状維持）」を推定する機能を有する。

## 4.7 同義・類義語抽出

最後に本システムが出力する課題－構成マップとは直接関係ないが、本システムが備えるシソーラスの自動構築機能について説明する。先の軸項目生成では、シソーラスを用いて同義・類義語の判定を行った。本シソーラスは人手で用意することも可能だが、本システムは、入力された特許書類の情報と、抽出した課題／構成表現から自動的に生成する機能も有する。

# 5 おわりに

本稿では、自然言語処理技術を活用して課題－構成マップを自動的に生成するシステムを提案した。本システムにおいて、ユーザは、任意の特許群を入力するだけで、課題－構成マップを得ることができる。

## 参考文献

- [1] 特許庁, "特許マップとはー技術分野別特許マップについて (機械 6 焼却炉技術)", [http://www.jpo.go.jp/shiryous/s\\_sonota/map/kikai06/map/map.htm](http://www.jpo.go.jp/shiryous/s_sonota/map/kikai06/map/map.htm)
- [2] 工業所有権情報・研修館, "資源循環型社会の構築に向けて", <http://www.inpit.go.jp/blob/katsuyo/pdf/chart/fippan08.pdf>
- [3] Hideyuki Uchida, Atsushi Mano and Takashi Yukawa, "Patent Map Generation Using Concept-Based Vector Space Model", Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, 2004.
- [4] NRI True Teller, <http://www.trueteller.net/>
- [5] 酒井浩之, 野中尋史, 増山繁, "特許明細書からの技術課題情報の抽出", 人工知能学会論文誌, Vol.24, No.6, pp.531-540, 2009.
- [6] 坂地泰紀, 野中尋史, 酒井浩之, 増山繁, "Cross-Bootstrapping: 特許文書からの課題・効果表現対の自動抽出手法", 電子情報通信学会論文誌, D, Vol. J93-D, No.6, pp.742-755, 2010.
- [7] Hirofumi Nonaka, Akio Kobayashi, Hiroki Sakaji, Yusuke Suzuki, Hiroyuki Sakai and Shigeru Masuyama, "Extraction of the Effect and the Technology Terms from a Patent Document", IEEE The 40th International Conference on Computers & Industrial Engineering (CIE40), cie214jp-1, 2010.
- [8] Akio Kobayashi, Hirofumi Nonaka, Hiroyuki Sakai and Shigeru Masuyama, "An Automatic Thesaurus Construction Method for Technical Terms in Patent Maps", IEEE The 40th International Conference on Computers & Industrial Engineering (CIE40), cie248jp-1, 2010.