

## 特許情報における発明者群の解析 2G3-4

### Analysis of Research and Development Structure of Inventors based on Patent Documents

飯野由里江<sup>\*1</sup>  
Yurie Iino

山田泰寛<sup>\*2</sup>  
Yasuhiro Yamada

廣川佐千男<sup>\*2</sup>  
Sachio Hirokawa

<sup>\*1</sup> 九州大学大学院システム情報科学府 <sup>\*2</sup> 九州大学情報基盤研究開発センター  
Graduate School of Information Science and Electrical Engineering, Kyushu University  
Research Institute for Information Technology, Kyushu University

This paper describes a method to analyze the structure of research and development division of companies. It is based on the co-occurrence analysis of inventors of a company described in their patent documents. Empirical evaluation is reported using the system of "concept graph" and "matrix graph", which the authors developed for general purpose text mining.

## 1. はじめに

特許情報は、権利情報としての先願特許調査の他に、市場(技術)調査、企業動向調査にも利用される。

特許は、事業発展のための技術開発の結果として出願され公開される。従って、特許情報を特許出願人(企業)の研究開発投資の結果として捕らえることができ、その明確な指標となるのが特許出願件数と発明者数である。一方、発明者情報は発明を行った研究技術開発者の情報であり、その分析により発明における研究技術開発者のつながりや変遷を知ることができる。

発明者情報を解析する意義には以下のものがあると考えられる。

- (1) 発明者数はその研究技術開発にかかった人員数と推測できる、
- (2) 発明者の研究技術分野別の分布は、ある企業がどの分野の研究技術開発に注力しているかを知る手がかりとなる、
- (3) 発明者間の関係は、その技術開発体制を知る手がかりとなる、
- (4) 発明者の変遷は、技術の流れを知る手がかりとなる、
- (5) 筆頭発明者としての出願が多い発明者は、その技術開発テーマにおけるリーダー的存在である。従って、発明者情報は、出願人(企業)の研究技術開発活動を知る上で手がかりとなる。

市川ら[市川 01]によれば、企業業績(一人当たりの売上高)と知的成果物(発明者一人当たりの特許出願件数)との間に高い相関関係がある。また、松尾ら[松尾 05]は、特定技術分野における研究活動を分析する上で、人間関係ネットワークの抽出が重要な基礎データになることを示唆している。このように、特許情報における発明者を群として捕らえることは、企業における研究技術開発活動を知る上で重要な指標となると考えられる。

本稿では、概念グラフ[廣川 05]及びマトリックス検索システム[関 07]を用いて、特許情報における発明者群を分析する手法を提案する。

## 2. 解析手法

### 2.1 概念グラフ

廣川ら[廣川 05]は、文書頻度を用いることにより、文書集合に現れる単語の上位下位関係の定式化する概念グラフを提案している。一つのクエリに関連する検索結果の文書集合について関連する単語群の階層構造を動的に構成する。概念グラフ

ではまず、検索対象全体の文書集合  $U$  を固定する。 $U$  の部分集合  $X$  と単語  $u, v$  について、 $|X|$  は  $X$  に含まれる文書の個数、 $df(u, X)$  は  $u$  を含む  $X$  中の文書集合の個数、 $df(u^*v, X)$  は  $u$  と  $v$  の両方を含む  $X$  中の文書集合の個数、 $r(v, u) = df(u^*v, X) / df(v, X)$  とする。単語  $w$  と文書集合  $X$  について、 $r(v, u) > 0.5$  となるとき  $w$  は  $X$  の特徴語であるという。二つの単語  $u, v$  と単語集合  $X$  について、 $df(u, X) > df(v, X)$  かつ  $r(v, u) > 0.5$  であるとき、 $X$  に関して  $u$  は  $v$  の上位であるという。概念グラフは、クエリ  $q$  に対する文書集合  $D(q)$  の特徴語を節、 $D(q)$  における上位関係を枝とする有向グラフである。具体的な可視化においては、単語  $v$  についてすべての上位語に枝を描くのではなく、その隣接上位だけとすることで枝の数は抑えられる。

### 2.2 マトリックス検索システム

マトリックス検索システム[関 07]は、文書集合に対して検索を行い、その検索結果をユーザが選択した二つの観点に基づいて分類し、2次元の表形式で表示する。類似した内容の文書群を一つのセルにまとめることができる。また、様々な観点を組み合わせることで分析を行い、異なる観点の対応を知ることができる。検索領域やマトリックス表示の縦軸横軸をユーザが自由に選択することにより、多面的な検索を実現している。それぞれのセルに該当する文書数が表示されるため、どのセルに多くの文書があるか知ることができる。従来のパテントマップ[新井 03]のマトリックスマップとの違いは、機能に関しては、検索結果から特定のセルを選択することで、そのセルに該当する文書集合のみをさらに分類することができる。これにより、一度得られた検索結果をより詳細に分析することができる。データに関しては、従来のパテントマップにおいて各セルに対応する基本単位は特許明細書であるが、本稿で実装したマトリックス検索システムでは発明者となっている。

## 3. 実験

本稿で使用するデータは、2000年から2005年に出願された化粧品分野の特許明細書として、IPC と FI のいずれかが A61K7 の特許明細書 16,375 件である。特許明細書は、出願日、出願人、発明者、発明の名称、要約、特許請求の範囲などの情報を一定の様式で記載されており、IPC や FI の分類コードが割り当てられている。

### 3.1 概念グラフによる解析

特許明細書群についての概念グラフを構築すれば、検索結果として得られる特許明細書群に現れる膨大な量の用語の関連解析が動的に可視化される。また、特許明細書の特定の部分、例えば、課題や解決手段等の項目だけに着目すれば、項目間の違い、項目間の関連を分析することができる。

この特性を利用して、企業別に概念グラフを作成したところ、発明者を群として捕らえることが容易であり、また、発明者群と特定の技術用語を同時に表示させることにより、特定技術分野における発明者人数の解析が可能であることがわかった。

大手化粧品企業 A 社を対象に、発明者名及び技術用語を特徴語として概念グラフを作成したところ、図 1 のような結果が得られた。この概念グラフでは、出現頻度の高い発明者名及び技術用語を抽出し、更に特徴語間の共起頻度の上位・下位解析を行うことにより、出願件数の多い発明者及び技術用語とその関係者(共同発明者)及び技術用語を効果的に可視化できることがわかる。図 1 の結果より A 社において複数の開発グループが存在すること、またその開発グループが出願した特許明細書に記載された技術用語が同時に表示されることから、各グループの開発技術を容易に把握することができる。なおこの解析においては、右側に示された発明者の出願のうちの過半数が左側に示された発明者との共同、すなわち関連度、 $r(v, u) > 0.5$  としている。

### 3.2 マトリックス検索システムによる解析

マトリックス検索システムにおいて、各特許情報を、明細書中に出現する発明者ごとの情報に変換して、マトリクスマップを作成した。各発明者の情報は、発明者が出願した特許群に関する出願年、筆頭発明者、要約、IPC、FI 等の情報である。

企業別に出願年別、特許技術分類別グラフを作成することにより、各企業における各技術分類に関する発明者数の変遷を容易に知ることが可能であることがわかった。すなわち、各企業の技術開発の変遷を知ることができる。また、筆頭発明者などの発明者とのマトリクスマップを作成することにより、発明者をグループ化して把握することが可能であることがわかった。このことにより、発明者群を詳細に知る上での大きな知見を得ることができ



図 1：発明者と技術用語の概念グラフ (A 社)

印の箇所には、発明者名の中に「染色」「色調」といった用語が表示され「染毛剤」の開発者と予測可能

## 4. むすび

### 4.1 本検討における効果

今回の検討により、概念グラフが、発明者間のかかわりを一目で表すことができ、また、企業のタイプによる発明者群の違いを明確にすることができ、研究開発体制についても示唆され、企業評価に重要な有用なツールとなりうることを示した。更に、技術用語を同時に表示させることによって、発明者群の開発技術を容易に類推することができ、企業における研究技術開発にかかった人員数(投資)を技術ごとに推測することが可能と考えられる。

一方、マトリックス検索システムによれば、発明者群の定量的な分析が可能であることを明示した。また、出願年と技術分類等をもちいてマトリクスマップを作成することにより、出願年による主要技術の違いを容易に類推することができ、企業における研究技術開発の注力度の変遷を推測することが可能と考えられる。

### 4.2 今後の課題

本検討により、特許情報より得られる発明者を群として捕らえ、企業の研究開発体制及び研究開発投資の変遷について知見が得られる可能性があることを示したが、いずれも定性的評価であり、定量的考察が課題となっている。

### 参考文献

- [新井 03] 新井喜美雄 :特許情報分析とパテントマップ, 情報の科学と技術, (社)情報科学技術協会, 2003 .
- [市川 01] 市川照久, 辻 秀一, 笠原久美雄, 片木孝至 :企業内研究所の知的生産性評価法に関する提案と実証, 経営情報学会誌, 2005 .
- [関 07] T. Seki, T. Wada, Y. Yamada, N. Ytow, and S. Hirokawa :Multiple Viewed Search Engine for e-Journal — a Case Study on Zoological Science. In Proceedings of the 12th International Conference on Human-Computer Interaction, 2007 .
- [廣川 05] 廣川佐千男, 下司義寛, 和多太樹 :文書群からの概念グラフの構成, 情報処理学会研究報告, 情報学基礎研究会報告, 2005 .
- [松尾 05] 松尾 豊, 友部博教, 橋田浩一, 中島秀之, 石塚満 :Web 上の情報から人間関係ネットワークの抽出, 人工知能学会論文誌, 2005 .

A61K	540	41	60	39	65	125	210
A61Q	250	0	0	0	0	55	195
A61P	235	17	18	14	25	56	105
C11D	80	3	9	5	9	18	36
G01N	57	1	4	0	6	12	34
C09K	56	1	2	0	5	16	32
A45D	48	4	11	3	5	14	11
C07D	47	3	4	0	5	13	22
C12N	44	0	1	0	5	10	28
C08L	41	2	6	2	1	8	22
縦の総数	1398	72	115	63	126	327	695
横の総数		2000	2001	2002	2003	2004	2005
		出願年					

図 2：出願年別技術分類のマトリクスグラフ (A 社)