

科学技術文献の「解析・可視化サービス」について

独立行政法人科学技術振興機構
文献情報事業本部本部長
門田 博文

PROFILE

情報提供部提供企画課長、経理部調査役、経理部
予算課長、総務部総務課長、研究支援部長、経理
部長、情報提供部長、平成19年より現職

✉ a-mg@jst.go.jp

☎ 03-5214-7983

1 はじめに

独立行政法人科学技術振興機構（以下「JST」）では、2008年2月から科学技術文献の解析・可視化サービス「AnVi seers (the Analysis and Visualization tool for seers : アンビシアーズ)」の提供を開始した。これは、特許文献において提供されている「解析・可視化サービス」を科学技術文献に適用したものである。

科学技術文献の解析・可視化については、特許化されていない情報をカバーする重要な情報入手方法ではあるが、一方で実現にあたっては特許文献の可視化には無い独自の課題を解決する必要があった。本稿では、解析・可視化の意義を述べたうえで、課題及び解決策を説明する。

2 科学技術文献の解析・可視化の意義、事例

2.1 解析・可視化の意義

特許文献が主に企業の活動に基づいているのに対して、科学技術文献は将来の特許の元となる基礎的な技術内容が含まれている。このため、科学技術文献を解析・可視化することにより、特許文献の形になる前の情報をいち早く入手することができる。

また、特許文献の場合は出願後公開されるまでに18ヶ月というタイムラグがあるが、科学技術文献の場合は、こうしたタイムラグがないことも、将来を予見するうえ

で重要なファクタといえる。

2.2 解析・可視化の実例（科学技術文献と特許文献との比較）

特許文献の解析・可視化サービスについては、以前より特許ベンダー各社より製品化されている。こうした特許文献の解析結果と科学技術文献の解析結果を比較することにより、新たな観点での分析が可能である。以下にその一例を記載する（引用した特許マップは富士通ATMS/Analyzerのデータ）。

2.2.1 連携先の違いやテーマの違いが分かる（図1.スケルトンマップ比較）

スケルトンマップは、主に機関同士のつながりを解析・可視化するツールである。

特許文献の解析結果を見ると企業と企業の連携が多く、大学と連携している例は少ないが、科学技術文献を見ると企業と大学との連携が多いことが分かり、基礎研究の段階での機関同士の連携が把握可能である。また企業が重視しているテーマは特許文献と科学技術文献では異なり、基礎研究の段階でどのようなテーマを重視しているかを見ることができる。これによりライバル機関の動向を確認したり、新規の連携先を探したりする場合に利用できる。

2.2.2 各機関の特許戦略の違いが分かる（図2.バブルグラフ比較）

バブルグラフは、各機関の研究動向を視覚的に時系列で確認するためのツールである。

科学技術文献の解析結果には特許文献には登場しない機関が数多く出てくるため、どの機関が基礎研究を重視

しているか把握可能である。また、特許文献では数多く出てくる機関が科学技術文献では出てこない場合があり、各機関の特許戦略の違いを分析することができ、自社の研究課題策定に活かすことができる。

3 科学技術文献の解析・可視化における工夫

特許文献と異なり、科学技術文献の世界では異表記が多く存在し、多くの科学技術文献データベースでは検索の際の大きな問題となっている。この問題は主に「著者所属機関」と「科学技術用語」の2点に大きく表れている。

ここでは、JST文献データベースにおいて、この問題を解決するためにどのように工夫したかを記述する。

3.1 著者所属機関における工夫

特許文献の場合は、出願人は登記簿の表記を使うことで機関名の異表記の問題は発生しないが、科学技術文献（JST文献データベース）では書誌情報、特に機関名については、記述言語の違いによるものや、入力工程を伴わないメタデータによる投入などの理由から、異表記が存在している。例えば、「東大 工学部」といった部門や学部を含んだものや「Univ. Tokyo JPN」のように国名などを含んだものなど多様なものがある。機関名の異表記が存在するまま分析を行った場合、同一の機関で分析されなければならない機関が別々で分析されることになる。このまま可視化処理へ回してしまうと、同一機関が分散することで誤った判断を与えかねない。

この機関名の異表記を解決する方法としてはあらゆる表記を集めて辞書を作成することが考えられる。JSTでは科学技術文献に出現した表記を中心にあらゆる異表記を考慮した機関名辞書を整備しており、この機関名辞書を活用し、名寄せを行うことで異表記による分析精度の低下を回避した。但し、異表記は膨大な数になり、名寄せの精度を上げるために全ての異表記を整備することは容易ではない。

そこでJSTでは、前方最長一致（キーワード前方より一番長く辞書とマッチングするものを検索する方法）をとることによって、一致率の一層の向上を可能とした。^[1]

3.2 科学技術用語における工夫

特許文献や科学技術文献においては、同一の概念に対して異なる用語を使用することがあり、機関名と同様の問題が生じている。JSTの文献データベースでは、個々の文献情報に対して、体系化された約4万語の概念代表語「JST科学技術用語シソーラス」を付与している。JSTではこれを用いて解析・可視化を行うことにより、科学技術用語の異表記による問題を解決している。

4 今後の課題

今後の課題として、特許文献の解析と科学技術文献の解析を統合的に行うことが考えられる。このためには、特許文献と科学技術文献で互いに表記が異なる機関名や科学技術用語間の関係づけを行う必要があり、現在、JSTでは機関名辞書や科学用語辞書の拡充を計画している。特に、科学技術用語辞書については、JST科学技術用語シソーラスを含む約45万語の大規模用語辞書に対して、今後数年間で特許用語約30万語を追加することを予定しており、将来的にはこうした辞書等を特許文献と科学技術文献の統合的な解析・可視化に役立てていく予定である。

以上

参考文献

- [1]小林義英,西田正 (科学技術振興機構), 渡部勇 (富士通研), 「文献情報を可視化するための機関名名寄せ手法」、情報メディア学会研究大会発表資料、Vol.6th、pp.51-54、2007

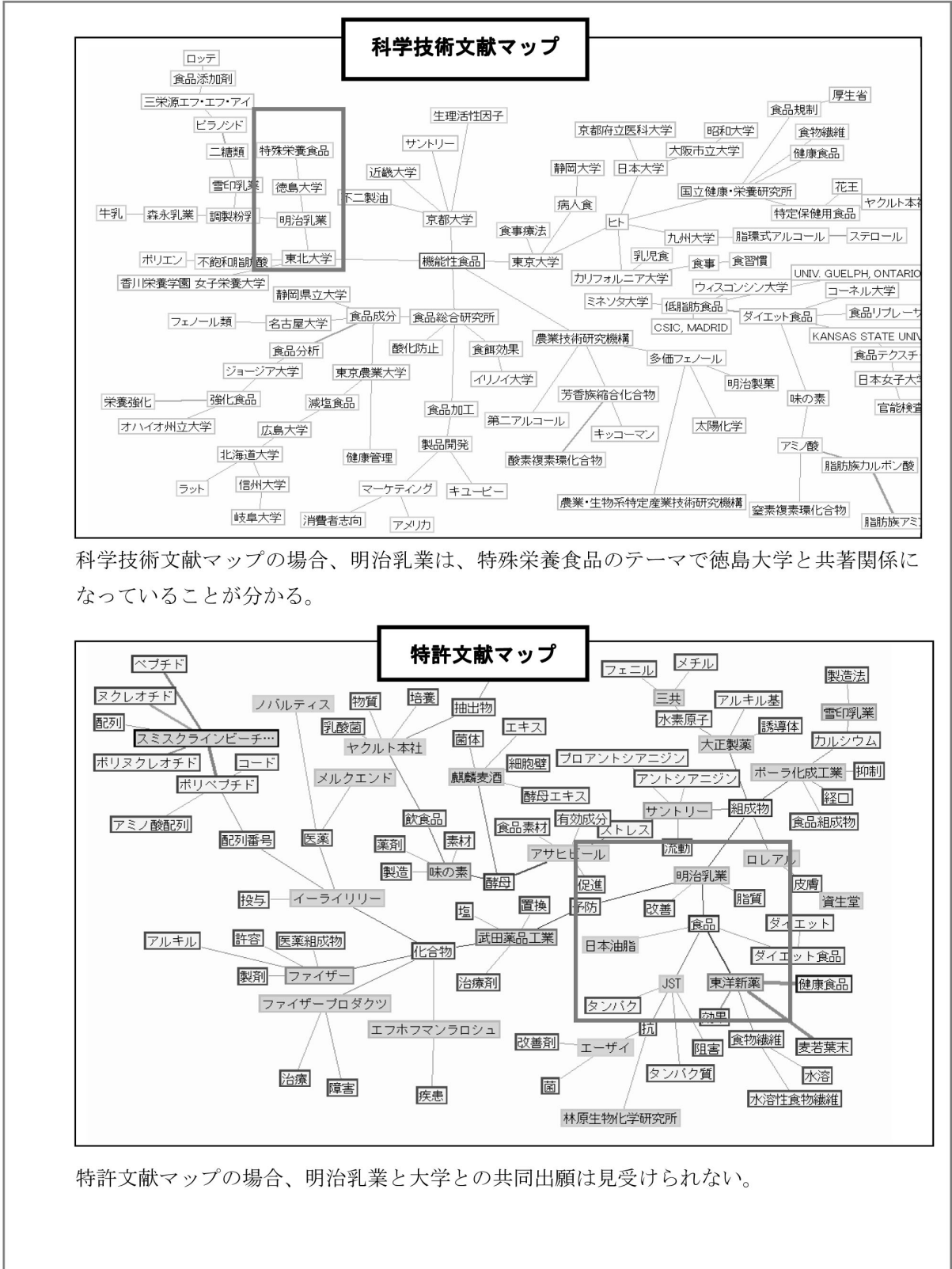
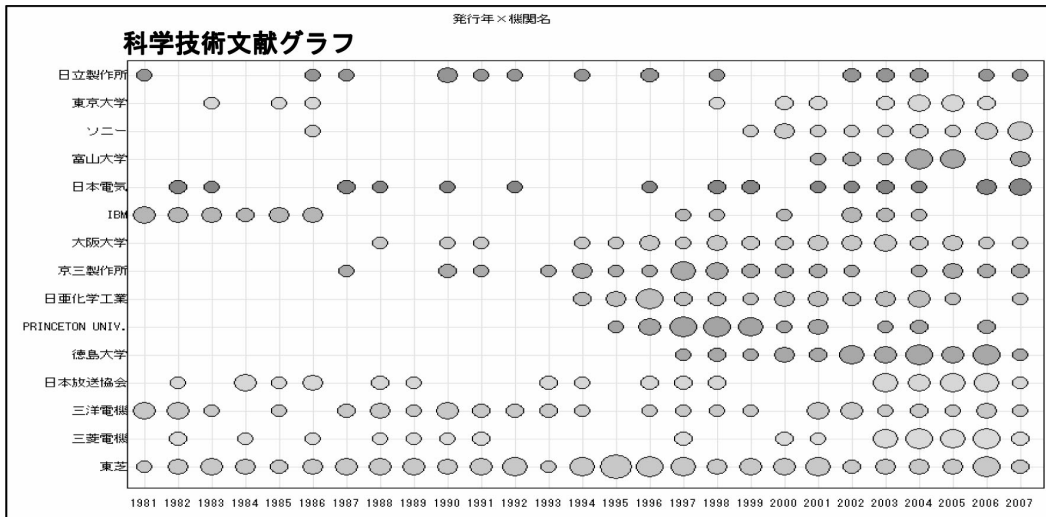


図1 スケルトンマップ比較 (テーマ「文献：機能性食品 特許：健康食品」)

科学技術文献 検索式：発光ダイオード/CT*表示装置/CT 2,916 件

特許文献 検索式：F ターム=5F041FF01 (発光ダイオード→用途→表示) 3,123



日立製作所、京三製作所で論文投稿がより盛んである。また、日本電気は 1990 年代後半より特許出願が減っているが、論文はコンスタントに書かれている。

豊田合成、日立電線、キヤノン、ローム、スタンレー電気、シャープにおいて、特許出願がより盛んである。

特許文献グラフ

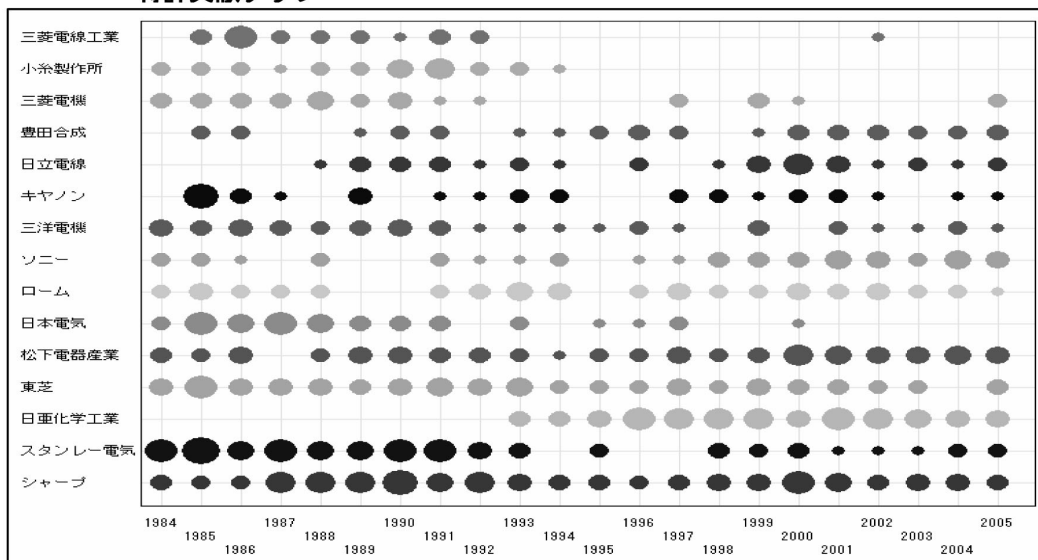


図2 発行年×機関名バブルグラフ比較 (テーマ「発光ダイオードの表示装置」)