

Title	エビデンス・ベースのJST事業成果可視化(2)
Author(s)	治部, 眞里; 橋本, 定幸; 山崎, 正和
Citation	年次学術大会講演要旨集, 25: 383-386
Issue Date	2010-10-09
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9319
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

エビデンス・ベースの JST 事業成果可視化(2)

○治部眞里・橋本定幸・山崎正和
(科学技術振興機構)

JST 事業成果可視化分析その 1 における背景の中、独立行政法人科学技術振興機構（以下、JST とする）の論文・特許統合検索システム（以下、PATLISYS-J とする）を使用した、政策のための科学としての「ポリシーマップ」を紹介する。

1. 分析方法

1.1 データベース

PATLISYS-J において、搭載している科学技術文献や特許文献は、以下のとおりである。

論文：Elsevier 社 Scopus Custom Data 1996 年から 2007 年までの約 15,000 万件

特許：EPO PATSTAT (Worldwide Patent Statistical Database) 800 万件

Elsevier 社の Scopus Custom Data は世界 5,000 以上の出版社の 18,000 誌以上のジャーナルを収録しており、3,800 万以上の書誌、抄録レコードが存在している。

1.2 人名名寄せ

PATLISYS-J の文献の人名を様々な分析に使用することができるようにするためには、人名を精度高く名寄せし、ユニークな ID で人物を特定する必要がある。

Scopus Custom Data 及び PATSTAT においても、さまざまな表記（ローマ字、ローマ字略記）があり、表記の違い同一人物や表記の同じ別人を区別することが非常に難しくなっている。また、研究者の異動についても、これを追うことは難しく、結果として精度が悪く、信頼性のないものになってしまっている。

このため、本調査では科学技術文献及び特許公報につ

いても名寄せが可能な自動名寄せシステムの開発を行った。JST が所有する科学技術文献データ JSTPlus、JMEDPlus、医学・薬学系データベース MEDLINE、特許公報データ（公開広報、公表広報、再公表候補）らの著者を名寄せしたデータをすべて使い、JST が運用する J-Global（科学技術総合リンクセンター）のために開発された名寄せ自動プログラムを Scopus Custom Data 及び PATSTAT 用にチューニングしたものである。

自動名寄せの結果を確かめるために、10 人の著者を用いて、再現率と適合率を調べた。Scopus Custom Data に関しては、再現率及び適合率はそれぞれ平均 94.4%、100%を達成している。PATSTAT に関しても、それぞれ平均 91.7%、100%を達成している。

1.3 特許の NPL と論文情報のマッチング

特許の引用情報にみられる被特許文献（NPL：Non Patent Literature）と論文情報のマッチングについては、Scopus Custom Data の論文情報について、それぞれのテキストに含まれている異表記の変換や大文字・小文字の変換等を行ったうえで、以下のような処理をしている。

- ① マッチングの前処理として、JST の運営する科学技術文献検索サービス「JDreamII」の略記辞書を使った変換処理、ストップワードの削除、雑誌名辞書の作成を行った。これらにより、マッチング処理の精度を高めるようにした。
- ② ページ情報を抽出した。情報が抽出できた場合は、論文情報と最初のページと雑誌名のイコールマッチングを行い、ページ情報が抽出できなかった場合

は最初のページ、巻、出版年、雑誌名、タイトルテキストについて全文検索マッチングを行った。

マッチング結果の精度を確かめるため、特許情報 50 件を無作為に抽出し、引用上の NPL とマッチングされた論文情報について目検によるチェックを行った結果、ほぼ 100% の確率でマッチングされていることが確かめられた。

1.4 共引用分析

本調査分析においては、第 1 期・第 2 期・第 3 期科学技術基本計画期間のポリシーマップを比較するため、それぞれ 1996 年から 2000 年、2001 年から 2005 年、2006 年以降までに発行された論文の中で、各年・各分野（生化学・遺伝学・分子生物学、化学工業、化学、地球科学、工学、環境科学、免疫学、物質科学、医学、学際分野、神経科学、薬理学・毒物学・薬剤学、物理学・天文学）の被引用数が上位 1% である論文を抽出し、集合 M を作成した。それらの論文を 2 つ以上抽出し、それらの論文の類似度を計算した。類似度の計算方法として、以下のようなコサイン関数（Salton's normalization of co-citation）を使用した。

$$K_{ij} = M_{ij} / \sqrt{M_i \cdot M_j} \quad (1.1)$$

式(1.1)の K_{ij} が 0.3¹ 以上の論文集合に対して、クラスター分析を行った。

1.5 可視化ツール

分析を可視化しマップ化するために、あらゆる複雑ネットワークを可視化、解析することが可能なプラットフォーム Cytoscape²を使用した。これはオープンソースプロジェクト（LGPL）で、各種ネットワークファイルの読み込み、プラグインによる自由な拡張が可能なことから、可視化のためのプラットフォームとした。

1.6 可視化方法

ポリシーマップとして、PATLISYS-J システムから以下のものが出力可能である。

- ① 特定の分野・特定の技術の発展の源泉となった業績の表示
- ② 共著者関係の表示
- ③ 被引用回数が高い論文の表示

2. 分析結果

2.1 特定の分野・特定の技術の発展の源泉となった業績を表示

図 1 は、キーワード“Stem Cells”で、データを絞り込み、Scopus Custom Data から抽出した論文群の引用・被引用関係をネットワーク図に示したものである。図中黄色 2006 は、2006 年 Cell 誌に掲載された京都大学の山中伸弥教授の「Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors」という論文を示している。この論文を引用している論文がいつ、誰によって引用されているかを可視化している。

ひし形のノード（◇）は、JST がファンディングした研究者を示し、JST がファンディングしていない研究者の論文は、四角のノード（□）となっている。特定のノードをクリックすると、下の Data Panel に論文タイトル、分野コード、雑誌名、著者キーワード、国コード、著者、機関、特許審査官に引用された論文か否かが表示される。ノードの大きさは被引用回数の大きさを示している。図 1 は論文を対象として分析したが、特許を対象として分析することも可能である。

¹ Thomson Reuters “Research Front Methodology”, <http://sciencewatch.com/about/met/rf-methodology/> (参照 20100518)

² Cytoscape は、<http://cytoscape.org/index.php> からダウンロード出来る。

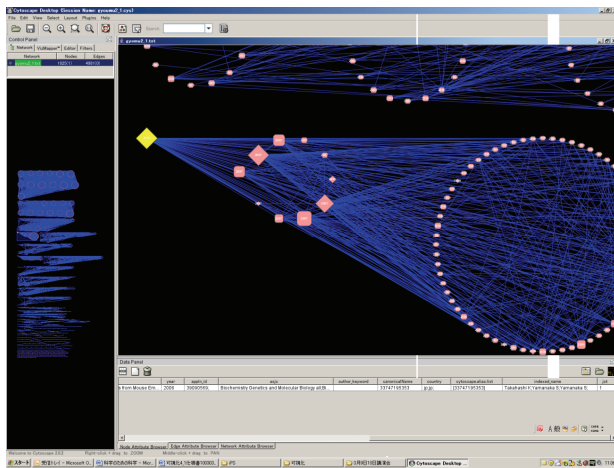


図1 Stem Cells 技術の発展

2.2 共著関係の表示

図2は、先ほど抽出した論文の共著関係をすべて図示したものである。黄色ひし形のノード（◇）が、京都大学の山中伸弥教授を示している。JSTがファンディングしたプロジェクトの代表研究者は◇で表示している。

左側の縦長のWindowに表示により、Stem Cellsのすべての研究者を一望することができる。右側で拡大された大きな塊は山中教授を中心において論文の共著関係示している。

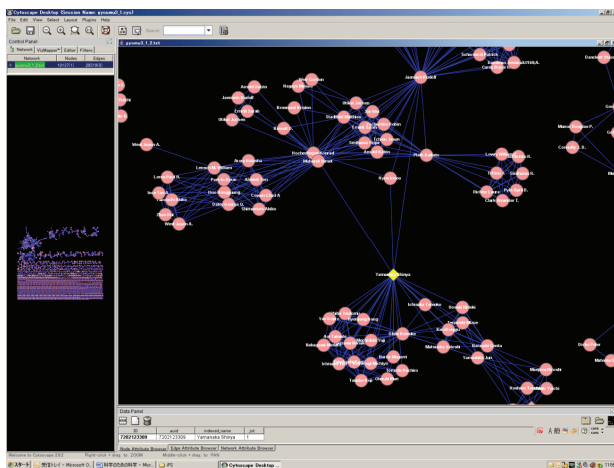


図2 共著関係の可視化

2.3 被引用回数が高い論文の表示

図3の左側のウィンドーが、第1期科学技術基本計画期間中の世界の科学技術の構造を示したものである。右

側は、左側のウィンドーの特定の部分を拡大した図である。四角（□）が一つのクラスターを示している。任意の四角（□）をクリックすると、右下のウィンドーにそのクラスターを形成する論文のタイトル、分野、著者キーワード、著者の国、著者名、クラスターを形成する論文数、それらの論文が特許の審査官に引用されているかどうかを確認することができる。

また、ひし形（◇）のノードは、JSTがファンディングした研究者の論文が含まれているノードである。左側の図で見ると赤のノードが、JSTがファンディングした研究者の成果である。構成される論文の被引用回数を足し合わせたものが、ノードの大きさとなっている。

図3、図4、図5の左側の図を比較すると、JSTがファンディングした研究者の論文を含む赤のノード数はそれぞれ全体の2.94%、3.86%、2.94%である。分析をした時点は、第3期科学技術基本計画中であるため、今後の伸びが機体される。分野別にみると、第1期期間中では化学工業分野、第2期期間中では免疫学の分野、第3期期間中は神経科学における赤のノード多くなっており、全体の中で占める論文の数が最も多いことが明らかになる。神経科学において心理学と関わる分野、免疫学と関わる分野、獣医学と関わる分野等、多岐にわたっていると特徴的である。図5に示すように融合していく分野を分析することも可能である。

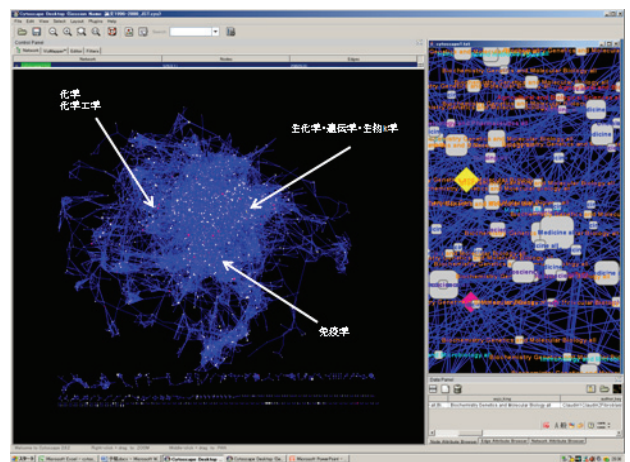


図3 第1期科学技術基本計画におけるポリシーマップ

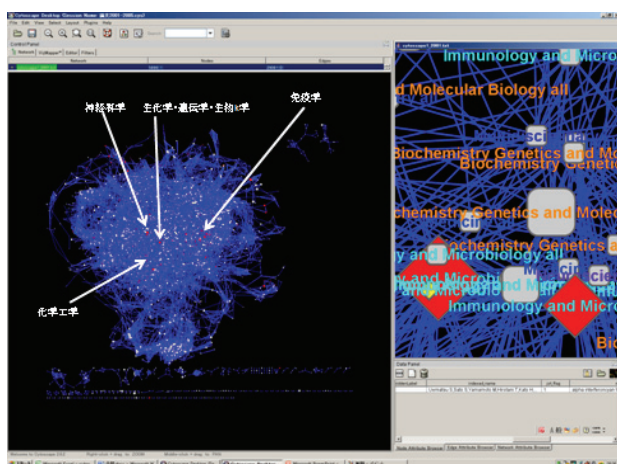


図4 第2期科学技術基本計画におけるポリシーマップ

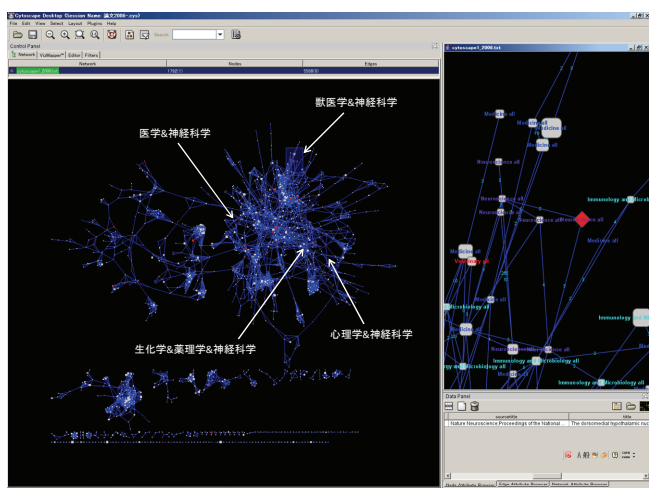


図5 第3期科学技術基本計画におけるポリシーマップ

3. まとめ

このようなポリシーマップシステムにより、世界の動向を俯瞰しながら、ファンディングしていく分野及び研究者を探索することができるようになると、科学的根拠（エビデンス）に基づく戦略的な政策立案の実現へと近づくと考えられる。また、このポリシーマップは、科学と技術の間の密接な関連の現れである特許の審査官に引用されている論文の俯瞰も可能なことから、今後技術に影響を与えられる分野の探索も可能である。

参考文献

(1) 落合 圭, 小林 義英, 橋本 定幸, 塩尻 栄美子, 山崎 雅和, 栗原 正昭, 浜中 寿, 坂内 悟, 國谷 実, 治部 眞里. “サイエンスリンケージによる JST 事業成

果分析（下） 可視化の具体的手法”. 情報管理. Vol. 52, No. 11, (2010), 651-659.

(2) 小林義英, 落合圭, 橋本定幸, 塩尻栄美子, 山崎雅和, 栗原正昭, 浜中寿, 坂内悟, 國谷実, 治部眞里.

“特許の引用情報にみられる論文情報の定量的分析のためのシステム開発”. デジタル図書館. No. 37, (2009), 78-84.

(3) 治部 眞里, 小林 義英, 落合 圭, 橋本 定幸, 塩尻 栄美子, 山崎 雅和, 栗原 正昭, 浜中 寿, 坂内 悟, 國谷 実. “サイエンスリンケージによる JST 事業成果分析（上） 国別・機関別の分析”. 情報管理. Vol. 52, No. 10, (2010), 601-609.

(4) 藤垣 裕子, 平川 秀幸, 富澤 宏之, 調 麻佐志, 林 隆之, 牧野 淳一郎 「研究評価・科学論のための科学計量学入門」丸善株式会社 平成 14 年 p70. (2004) .