

知財活用の実際 論文と特許データベースを統合した ジャンル横断検索および技術動向分析

広島市立大学 情報科学研究科 知能工学専攻
難波 英嗣
e-mail : nanba@hiroshima-cu.ac.jp

抄録

近年、特許出願が重要な研究活動のひとつとして考えられるようになってきた今日、大学研究者自身が関連論文とあわせて関連特許について情報を検索したり、特許を出願したりするという機会が増えている。2007年5月に政府の知的財産戦略本部が発表した「知的財産権推進計画 2007」においても、大学研究における特許情報の重要性が謳われており、大学研究者の利用を想定した特許・論文情報統合検索システムの整備がこの計画のひとつに挙げられていることから、この傾向は今後さらに強まっていくと思われる。そこで本稿では、特許と論文を対象にした検索や技術動向分析に関する諸研究を概観し、関連システムやサービスを紹介する。

キーワード

特許, 論文, 技術動向分析, ジャンル横断検索, 引用関係

1. はじめに

特許出願が重要な研究活動のひとつとして考えられるようになってきた今日、大学研究者自身が関連論文だけでなく関連特許について情報を検索したり、特許を出願したりする機会が増えている。2007年5月に政府の知的財産戦略本部が発表した「知的財産権推進計画 2007」¹⁾においても、大学研究における特許情報の重要性が謳われている。この計画で、大学研究者の利用を想定した特許・論文情報統合検索システムの整備が含まれていることから、このような傾向が今後さらに強まっていくことがうかがわれる。

また、特許と論文を検索するのは、大学研究者に限った話ではない。例えば、特許庁の審査官は、出願された技術が特許権の取得に該当するかどうか判断するために、過去に同様の特許が出願されたり論文が発表されたりしていないか調査する。これは、一般に無効資料調査と呼ばれているものである。同様の調査は、民間企業ではサーチャーと呼ばれる専門の担当者が審査官による審査を経た出願技術を再調査し、競合する他者の権利を無効化するために社内で行われることもある。そこで、本稿では、特許と論文を対象にした近年の情報検索や技術動向分析について紹介する。

近年、特許と論文を対象にした検索システムや動向情報分析ツールの開発や整備が急速に進んでいる。例えば、科学技術振興機構(JST)では、特許、論文検索システム JSTPatM を開発し、2007年3月から大学向けに試験的公開を行っている²⁾。

また、単に関連特許と論文を対象に検索を行うだけでなく、それらの技術動向分析に関するツールの整備も進みつつある。このような技術動向分析は、大学、企業、研究所等で、研究や開発の方向性を決めるために、戦略調査目的で行うだけでなく、投資や研究資金の分配といった政策決定の際の根拠としても利用することができることから、この分野への関心も高まっている。実際に科学研究活動について分析・評価を行う科学計量学(サイエントメトリクス)の分野では、特許と論文を対象にした様々な分析結果が報告されている³⁾。それによれば、一般に、基礎研究と産業の結びつきが強い分野では、特許が数多くの論文を引用していると考えられており、そこで特許中の引用論文数を数えることで、分野ごとの基礎研究と産業の関連度を定量的に測ることが可能になる。

本稿では、特許と論文を対象にした情報検索や技術動向分析に関する諸研究を概観し、著者らが開発中のシステムを含む関連システムやサービスを紹介する。本稿の構成は以下のとおりである。次節では、特許と論文を対象にした検索について、3節では、技術動向分析について、それぞれ述べ、4節で本稿をまとめる。

2. 特許と論文を対象にした検索

2.1 特許と論文を対象にした検索の課題

特許にあまり馴染みのない研究者が特許検索を行うのは容易なことではない。第一の理由として、特許では請求範囲をなるべく広く確保するため、一般性の高い特許用語を用いて記述する。第二に特許用語は、一般に学術用語よりも多様な表現が用いられる。例えば、「機械翻訳」という学術用語に対する特許用語は「機械翻訳」の他にも「自動翻訳」「言語変換」などがある。このため、単純に表層的な単語の一致度を用いる従来の検索モデルでは、同じキーワードで特許データベースと論文データベースを検索しても、用語の使われ方の違いから、そのキーワードに関する論文や特許を十分に収集できるとは限らない。また、特許を効率的に検索するには、IPCコード、FIコード、Fタームなどの特許固有の分類記号を使いこなす必要があるが、それには専門的な技術と経験が必要となる。それゆえに、特許に関する知識を持っていない人にも、より容易に特許を検索したり、特許と論文を対象にした技術動向分析を可能にしたりする仕組みが必要とされている。

2.2 利用可能な論文、特許検索システム

1節で述べたJSTのJSTPatM以外にも、論文・特許検索システムやサービスはいくつか存在する。その代表的なものは、トムソンサイエンティフィックが提供する一連のサービスで、その中に含まれるDerwent Innovations Indexでは、世界40特許発行機関が発行する2200万件以上の特許を検索することができる。さらに引用関係を辿って、特許中で引用されている文献を検索することも可能である。⁴⁾

また富士ゼロックスは、日米特許約1,800万件とJSTの科学技術文献データ2,000万件を検索できるサービスDocuPatを、2005年9月から始めている⁵⁾。DocuPatでは、特許検索で広く用いられる論理演算検索機能を提供しており、特許検索用の検索式でそのまま論文を検索することができる。

Googleは、Web上に存在する論文データを収集して構築した論文データベースGoogle Scholarを公開している⁶⁾。さらに、2006年12月から米国特許検索サービスGoogle Patent Searchも開始している。このサービスでは、1790年代からの米国特許約700万件を検索することができる⁷⁾。

これらのサービスは、大学研究者だけでなく知財の専門家にとっても有益であるが、現状では、2.1節で述べた問題を完全に解決するまでには至っていない。そこで、次節では、2.1節の問題点に関する著者らの取り組みについて述べる。

2.3 特許，論文データを統合した検索

筆者らは、Web上のPostscriptとPDF形式の日英論文約78,000件を収集し、引用論文データベースPRESRIを構築している⁸⁾⁹⁾。この論文データベースと特許データベースを、(1)特許，論文間の引用関係の解析と(2)論文用語の特許用語への自動変換(例えば、論文用語「DRAM」を特許用語「半導体記憶記録装置」に変換する)という2種類の方法で統合する技術を開発している。さらに、特許検索に不慣れなユーザ向けに、自動的に構築した特許用語シソーラスを用いて検索支援を行う機能も提供しており、これらが2.1節で述べた課題の解決方法になっている。以下では、システムの動作例および仕組みについて、簡単に説明する。

検索システムは、ユーザに2種類の検索機能を提供する。ひとつは、キーワード検索であり、もうひとつは特許，論文間の引用関係を用いた検索である。ユーザは、まずキーワード検索を行う。

図1は、キーワード検索の画面である。図のフォームに、論文の場合はタイトル中の語、著者名等を入力する。特許の場合、発明の名称に含まれる語、発明者、出願人等を入力することができる。さらに、論文検索フォーム中で「タイトル」フォームにキーワードを入力した状態で「関連特許用語」というボタンを押すと、関連する特許用語のリストがポップアップウィンドウ中に表示される。

図1では、「フロッピーディスク」に関連する特許用語として「記憶媒体」「記録媒体」「ディスク状記録媒体」といった用語が表示されており、ユーザがこれらの用語を選択すると、その用語が特許検索フォームに追加される仕組みになっている。

なお、この関連特許用語のリストは自然言語処理技術を用いて自動的に生成されたものであるため、特許データさえ追加すれば、人手で特許，論文用語対応辞書のメンテナンスをする必要がない、という特徴がある¹⁰⁾。特許検索フォームに追加された用語は、このフォームの下にある「関連特許用語」というボタンを押すと、さらに特許用語シソーラスを用いて検索キーワード拡張を行うことができる。このシソーラスも自然言語処理技術を用いて自動的に生成されたものであり、現在約200万語から構成されている¹¹⁾。

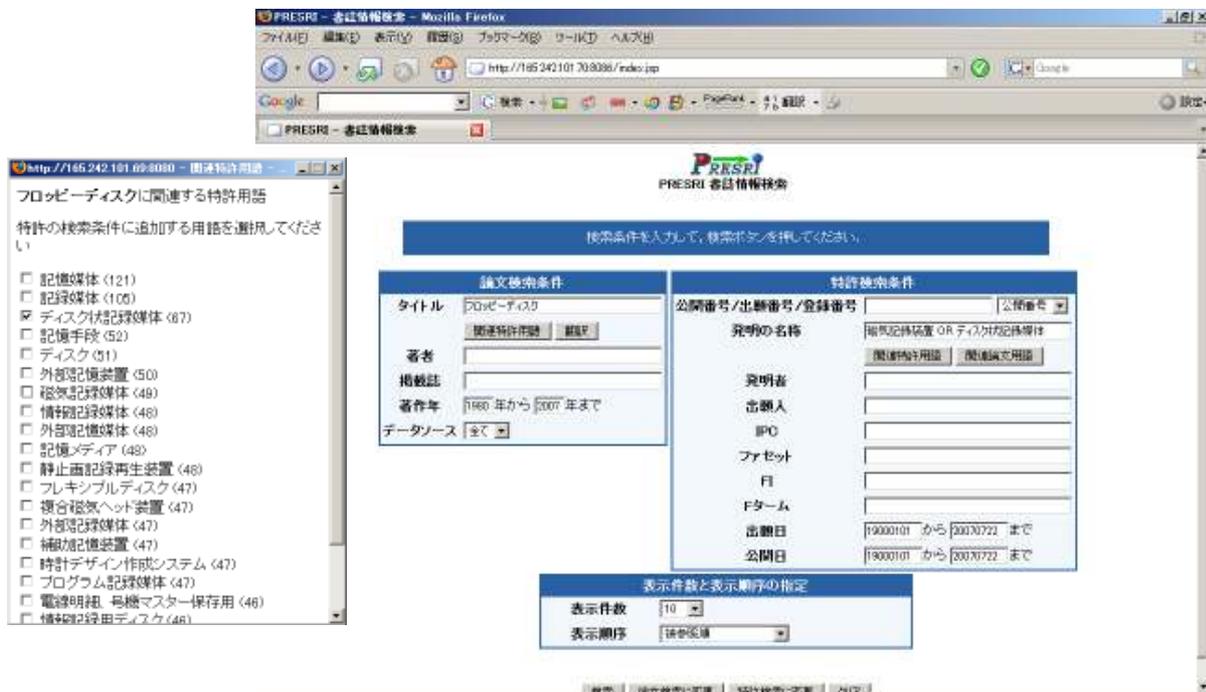


図 1 キーワード検索画面



図 2 キーワード検索の結果表示

次に、キーワード検索の結果の表示例を図 2 に示す。図 2 は、「フロッピーディスク」というキーワードで論文検索を行った場合の例である。検索結果の一覧表示画面では、文献ごとに書誌情報と共にチェックボックスが表示される。ユーザーが興味のある複数の文献にチェックし、「チェックした文献をグラフ表示」というボタンを押すと、チェックした文献およびこれらの文献と引用関係にある特許や論文が、図 3 に示すようなグラフとして表示される。

図 3 において、■は特許を、●は論文を示している。また引用関係が矢印で表示される。図 3 中で、■や●をクリックすると、特許や論文の詳細情報が表示される。このように、あるトピックに関連する複数の特許や論文をグラフとして提示することで、そのトピックに関する技術動向の流れを視覚的に捉えることができる。なお、特許、論文間の引用関係の解析も自動的にを行っている¹²⁾¹³⁾。

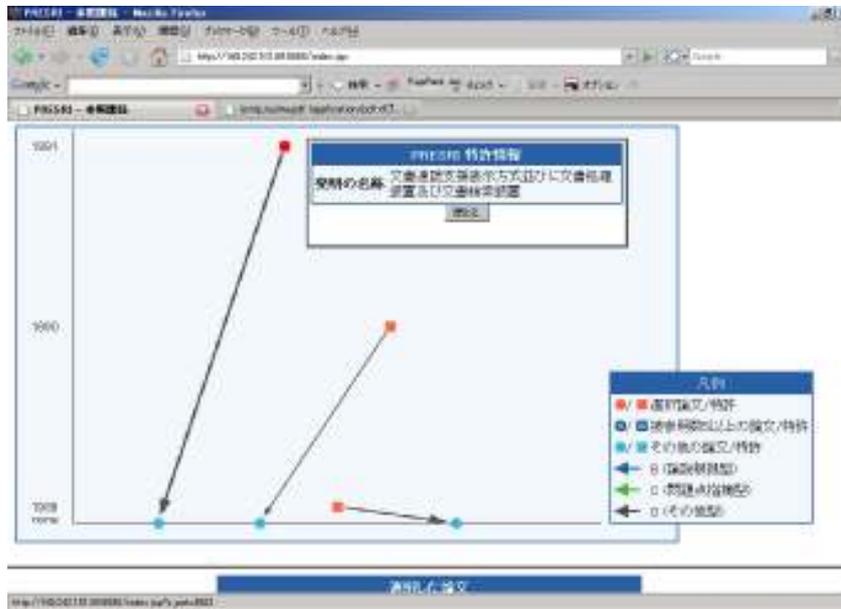


図3 特許，論文間の引用関係のグラフ表示

3. 特許と論文を対象にした技術動向分析

本節では、まず特許と論文を対象にした技術動向分析システムを紹介し、次に実際に技術動向分析を行った例として、科学計量学分野の研究を紹介する。

3.1 技術動向分析システム

1節でも述べたとおり、この分野の関心は非常に高いが、現在、実際に利用可能なシステムはまだ数少ない。そのひとつとして、トムソンサイエンティフィックのAurekaを紹介する¹⁴⁾。Aurekaは、基本的には特許の分析システムであり、与えられた特許集合に対し、引用関係をつリーで表示したり、特許をクラスタリングしてThemeScapeマップという俯瞰図で表示したりする機能がある。このシステムは、pdfやMS-Wordといった様々なフォーマットの論文データのインポートが可能であるため、論文を特許と共にThemeScapeマップ上にマッピングして分析することができる。

ある分野において、「どのような要素技術がいつ頃から使われているのか」という情報を網羅的に収集し整理することは、その分野の技術動向を概観するのに必要不可欠であるが、その作業には多くの時間と労力を要する。そこで、筆者らはこのような情報を自動的に抽出し、可視化するシステムを構築している¹⁵⁾。

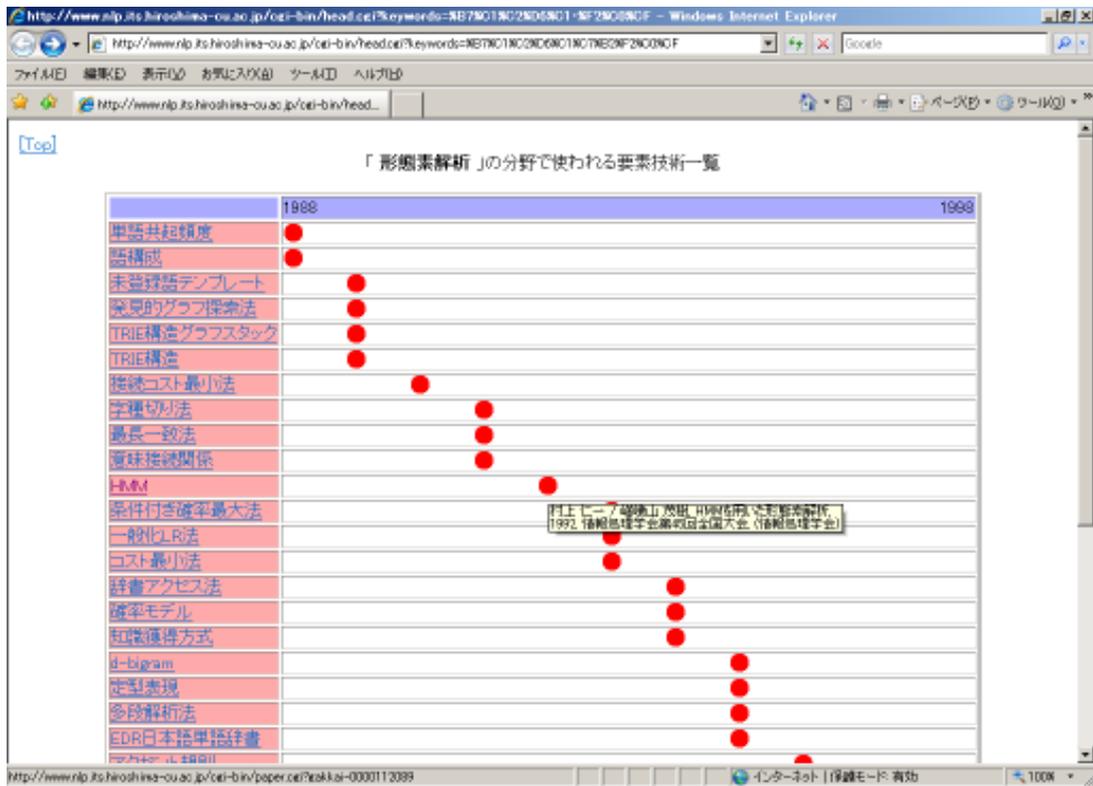


図4 「形態素解析」で使われる要素技術の一覧表示



図5 「HMM」(隠れマルコフモデル)を要素技術として用いている分野の一覧表示

以下では、システムの動作例および仕組みについて、簡単に説明する。図4は、「形態素解析」という用語をシステムに入力した時の解析結果を示している。図において、左端に「形態素解析」の要素技術名が列挙してあり、その用語が論文表題中で使われた年が、各技術の右側に示してある。例えば図4の「HMM」(隠れマルコフモデル)の場合、この用語を論文表題に含んだ形態素解析に関する論文が1992年に発表されていることを示している。これらは図中では

「●」として表示されており，ユーザが●上にカーソルを重ねると，その論文の書誌情報がポップアップ表示される．図では，「HMM」にカーソルを重ねた時のポップアップ表示として「村上仁一/嵯峨山茂樹，HMMを用いた形態素解析，1992，情報処理学会第45回全国大会（情報処理学会）」が例示されている．

さて，図4において要素技術として提示されている用語をユーザがクリックすると，その要素技術が他にどのような分野で利用されているのかが一覧表示される．図5は，図4中の“HMM”をクリックした結果を示している．図の2つ表は，左側が学術論文から抽出した分野一覧，右側が特許から抽出した分野一覧を示している．この図から，学術界では1980年代後半には音声認識の分野で，1990年代に入ると手書き文字認識の分野や形態素解析や言語獲得などの自然言語処理の分野でそれぞれ利用されていることがわかる．他方，産業界でも，音声認識や文字認識などの分野でHMMを利用した特許が多く存在するが，自然言語処理の分野では見当たらないことがわかる．

このような技術動向の可視化を行うために，論文では表題に，特許では請求項に，それぞれ着目し，論文と特許から要素技術を抽出している．多くの論文表題には「Aに基づいた」や「Bを用いた」などの表現が含まれる．このAやBは，ある技術を実現するための要素技術を示す用語であると考えられる．そこで，手がかり語を用いて論文表題を解析し，要素技術を示す用語を抽出する¹⁵⁾．他方，特許請求項にも「Aを備えた」や「Bを有する」といった表現が含まれるため，論文表題と同様に手がかり語を用いて請求項から要素技術を示す用語を抽出する¹⁶⁾．

3.2 科学計量学における技術動向分析

「学術研究が産業にどれだけ貢献しているのか」という観点から，学術研究を評価したり，産業と学術研究の関連性を分析したりする試みが，科学計量学の分野を中心に，行われるようになってきている³⁾．その典型的な手法は，Narinらの研究に代表されるように，特許と論文間の引用関係に着目したものである¹⁷⁾．Narinらは，アメリカ，イギリス，旧西ドイツ，日本，フランスの5ヶ国の論文および特許間の引用関係を調査し，各国の科学(Science)の国内および国外の技術(Technology)への影響度について分析している．特許と論文間の引用関係を用いた同様の分析は，例えばレーザー医学¹⁸⁾や宇宙工学¹⁹⁾などの特定の分野を対象にしたり，ある特定の地域からの発表論文や出願特許を対象にしたりするなど²⁰⁾，分析の切り口を変え，様々な側面から行われてきている．なお，この分野の研究動向については，Mayerら²¹⁾に詳しい．

ここで，特許と論文間の引用関係を用いて科学と技術の関連性を分析する上で，特許中でどの程度特許中で妥当な論文が引用されているのかは，分析結果の信頼性に直接影響する大きな問題である．学術論文を対象にした引用分析研究では，論文中の引用を，引用の理由²²⁾や性質²³⁾によって分類するためのカテゴリが複数提案されているが，特許においても同様に，特許中の引用文献を分類するという試みがある．Schmochは，レーザー医学およびポリイミド関連の特許中で引用されている文献1,577件を，引用文献としての質という観点から，以下に述べる3カテゴリに分類している²⁴⁾．

- カテゴリX:直接的に関連する文献．
- カテゴリY:他の引用文献と組み合わせることで直接的に関連する文献
- カテゴリA:技術的な背景として関連のある参考文献

分類の結果，カテゴリX, Y, Zに分類されるものが，それぞれ13.2%，15.7%，71.1%であった．すなわち，カテゴリYを含めても直接的に関連する文献の引用は30%弱にすぎないことを意味する．しかし，Schmochらの別の論文では，宇宙工学の分野の特許では，非常に広い意味で多くの関連文献が引用されているという報告もあるため¹⁹⁾，特許中の引用文献の質に関しては，早急に結論を出すことはできない．いずれにせよ，今後，大域的な調査が必要とされるであろう．

しかしながら，少なくとも技術動向の概要を把握するという目的においては，特許，論文間の引用関係を利用することは有効であると考えられている．以下に，実際に特許，論文間の引用関係を用いた調査結果を示す．

筆者らは，これまでに文献¹³⁾の手法を用い，公開特許公報1993年～2002年の特許3,496,253件から，引用論文86,415件を自動的に抽出している．このデータを用い，特許分類体系のひとつであるIPCコード別に引用論文数を調べた．その結果が表1である．この表から，バイオ，医学，化学，情報関連の分野で引用論文数が多いことが分かる．本稿で述べてきた特許，論文検索システムや技術動向分析システムは，現状では，主に表1に示すような分野で有用であると考えられるが，今後，産業と科学の結びつきが強まるにつれ，より多くの分野での特許，論文検索および技術動向分析システムの利用が活発になると考えられる．

表1 ドメイン(IPC)ごとの引用論文数(上位10件)

IPC	説明	引用論文数
C12N	酵素，遺伝子工学	12,160
H01L	半導体，超伝導体	10,442
A61K	医薬品製剤，化粧品	8,834
C12R	アルコール，微生物	8,656
C12P	微生物による化学物質の製造	8,282
G06F	コンピュータ，メモリ，プログラミング	7,604
C07D	ペニシリン系化合物，インドール系化合物，他	7,434
C07C	有機低分子化合物	6,825
G01N	材料科学	5,553
H04N	テレビ，FAX，ビデオテープ	4,841

4. おわりに

本稿では，特許と論文を対象にした情報検索や技術動向分析に関する諸研究を概観し，著者らが開発中のシステムを含む関連システムやサービスを紹介した．現在，筆者らは，国立情報学研究所主催のNTCIRワークショップにおいて，特許と論文を対象にしたタスクを企画しており，今後，この分野の研究が益々活発になってくると思われる．なお，特許と論文を用いた検索や技術動向分析に興味のある読者は，ntcadm-mining@slis.tsukuba.ac.jp までご連絡いただければ幸いである．

参考文献

1) <http://www.kantei.go.jp/jp/singi/titeki2/keikaku2007.html>
(accessed 2007.07.20)

2) <http://pr.jst.go.jp/patmreg> (accessed 2007.07.20)

- 3) 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(2004)
- 4) <http://www.thomsonscientific.jp/products/dii/index.shtml> (accessed 2007.07.20)
- 5) <http://www.fujixerox.co.jp/product/docupat> (accessed 2007.07.20)
- 6) <http://scholar.google.com/intl/ja> (accessed 2007.07.20)
- 7) <http://www.google.com/patents> (accessed 2007.07.20)
- 8) <http://www.presri.com> (accessed 2007.07.20)
- 9) Nanba, H., Abekawa, T., Okumura, M., and Saito, S. Bilingual PRESRI: Integration of Multiple Research Paper Databases. In Proceedings of RIAO 2004. p.195-211 (2004)
- 10) 釜屋英昭, 難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山. 特許, 論文間の引用情報を用いた論文用語の特許用語への変換. 情報処理学会自然言語処理研究会, NL-178, p.97-102 (2007)
- 11) 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. p.1113-1116 (2007)
- 12) 小栗佑実子, 難波英嗣. 米国特許データベースからの引用文献情報の抽出. 言語処理学会第13回年次大会. p.582-585 (2007)
- 13) 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. 言語処理学会第12回年次大会. p.743-746 (2006)
- 14) <http://www.thomsonscientific.jp/products/aureka> (accessed 2007.07.20)
- 15) 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. p.470-473 (2007)
- 16) 新森昭宏, 奥村学, 丸川雄三, 岩山真. 手がかり句を用いた特許請求項の構造解析. 情報処理学会論文誌, Vol.45, No.3, p.891-905(2004)
- 17) Narin, F., Olivastro, D. and Stevens, K.A. Bibliometrics / Theory, Practice and Problems, Evaluation Review. Vol.18, No.1, p.65-76 (1994)
- 18) Noyons, E.C.M., van Raan, A.F.J., Grupp, H., and Schnoch, U. Exploring the Science and Technology Interface: Inventor-author Relations in Laser Medicine Research. Research Policy. Vol. 23, p.443-457 (1994)
- 19) Schmoch, U., Kirsch, N., Lay, W., Plescher, E., Jung, K.O. Analysis of Technical Spin-off Effects of Space-related R&D by Means of Patent Indicators. Acta Astronautica. Vol. 24, p.353-362 (1991)

- ²⁰⁾ Acosta, M., Coronado, D. The Effects of Regional Scientific Opportunities in Science-Technology Flows: Evidence from Scientific Literature Cited in Firms' Patent Data. <http://ideas.repec.org/p/wiw/wiwrsa/ersa03p321.html> (accessed 2007.07.20)
- ²¹⁾ Mayer, M. Does Science Push Technology? Patents Citing Scientific Literature. *Research Policy*. Vol. 29, 409-434 (2000)
- ²²⁾ Bonzi, S. Characteristics of a Literature as Predictors of Relatedness between Cited and Citing Works. *Journal of American Society for Information Science*, Vol.33, No.4, p.208-216, 1982.
- ²³⁾ Weinstock N. Citation Indexes, in Kent A. (Ed.) *Encyclopedia of Library and Information Science*. New York, Marcel Dekker, Vol.5, p.16-41 (1971)
- ²⁴⁾ Schmoch, U. Tracing the Knowledge Transfer from Science to Technology as Reflected in Patent Indicators. *Scientometrics*, Vol.26, No. 1, p.193-211 (1993)

英文書誌事項

Cross-genre information retrieval and technical trend analysis
by integrating research papers and patents

Hiroshima City University
Hidetsugu Nanba

Abstract

The need for academic researchers to retrieve patents and research papers is increasing, because applying for patents is now considered to be an important research activity. In fact, the development of an IR system of research papers and patents for academic researchers is enshrined in "Intellectual Property Strategic Program 2007" by Intellectual Property Strategy Headquarters, published in May 2007. In this paper, we overview researches, systems and services related to information retrieval and technical trend analysis for research papers and patents.

Keywords:

patent / research paper / technical trend analysis / cross-genre information retrieval / citation relationships