社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－DBS－130 (6)
2003－ＦＩ－ 71 (6)
2003／5／22

# 文献に基づく生物学研究支援: 生物学者の視点からの支援可能性の検討

丸橋 弘治　　仲尾　由雄
(株)富士通研究所
〒211-8588 川崎市中原区上小田中 4-1-1
{maruhashi.koji,ynakao}@jp.fujitsu.com

分子生物学の進歩に伴い、大量の生物学データ・文献の蓄積が飛躍的に増加している。近年、自然言語処理技術や情報検索技術を利用して、生物学文献に蓄積された知識を抽出・活用しようという試みが盛んになっているが、その多くは、タンパク質相互作用を直接的に記述した部分を処理対象としている。しかし、生物学研究の支援という観点から捉えると、生物学者が多様な生命現象を総合的に理解することをいかに支援するかという別の課題もある。近年の生物学研究は、研究課題の細分化・専門化が進んでおり、個々の文献で扱われている課題は、その領域の専門家以外には理解が困難な場合も多い。本稿では、この問題の解決を目標に、著者情報を手がかりに、論文を自己組織化して提示する手法について検討を行った。シロイヌナズナに関わる１万件の文献を題材に、著者情報に基づき、論文集合を階層的に組織化する実験を行ったところ、様々な粒度の研究テーマに対応する論文集合が自動的に求められる可能性が示された。

# An Empirical Study for Developing Biomedical Research Survey Tools Based on Author Relation Mining

Koji Maruhashi and Yoshio Nakao
Fujitsu Laboratories Ltd.
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-8588, Japan

The biomedical literature is an essential information source to study the mechanisms of biological processes. Many efforts have been conducted to automatically extract biological knowledge, e.g., protein-protein interactions, from machine-readable literature, especially from the MEDLINE database. There is another important issue how to help biomedical researchers to understand life phenomena comprehensively. Recent biological research topics tend to be too specialized so that a biomedical researcher feels difficulty in understanding the content of articles in an unfamiliar biological field. In this paper, we discuss the problems that a biomedical researcher usually faces when to find an appropriate set of articles from an unfamiliar field and report a preliminary experiment where ten thousand articles of molecular biology reporting on Arabidopsis were organized hierarchically according to the structure of author information. The experimental result suggests that author information is a good clue for identifying appropriate research topics and relations among research topics.

## 1. はじめに

近年の分子生物学の飛躍的な進歩に伴い、塩基配列、アミノ酸配列、タンパク質立体構造、遺伝子発現プロファイル、生物学パスウェイといった生物学情報の蓄積は飛躍的に増加している。同時に、過去の生物学関連の文献のデータベース化も進んでいる。

最も著名な医学・生物学関係の文献データベースである MEDLINE は、1971 年より米国 NLM(National Library of Medcine)によりオンラインサービスが行われている。MEDLINE には、現時点で、4,600 を超える医学・生物学関連の論文誌について、約 1,200 万件の論文抄録と書誌情報が蓄積されており、医学・生物学の研究者の貴重な情報源となっている。また、文献で報告された実験データ（ファクトデータ）のデータベース化も進められている。例えば、1988 年に NLM の一部門として設立された NCBI (National Center for Biotechnology Information)により、DNA 塩基配列データベース GenBank などの整備が進められてきた。それらのファクトデータベースと MEDLINE 文献データベースとの間にはリンクが張り巡らされており、文献とファクトとを相互に参照することが可能になっている。

生命現象は、分子、細胞、組織、個体、集団といった様々なレベルの現象がレベルを超えて相互作用しあい、複雑な多様性を作り出している。そのような多様性を一度に理解・説明することは困難であり、そのため、個々の生物学研究は、特定の観点に沿って注目すべき生命現象を切り出し、その範囲の現象に統一的説明を与えることを目標とすることになる。しかし、そのように切り出した生命現象であっても、それをもたらすメカニズムは、分子レベルのミクロな反応から、個体あるいは種の環境適用といったマクロな方向性にまで、幅広く関わっており、やはり、一気に説明しつくすことは困難である。そのため、生物学研究の各ステップにおいては、細分した研究課題を、ひとつひとつ解決していくというアプローチをとることが多い。そして、各ステップの研究成果が、論文の形で文献ＤＢに蓄積されることになる。

このように、生命の多様性に比例して論文数も膨大になり、今後も爆発的に増加することが予想される。個々の論文は、一般に、生命の多様性のある側面、あるいは、特定の側面をもたらす生物学的メカニズムを報告するものである。よって、それぞれが無視できない有意義な情報である一方で、個々の論文で扱われる課題は、細分されすぎているため、その課題に関連する領域の専門家以外には理解が困難な場合も多い。

本研究は、自然言語処理・情報検索などの情報技術に基づき、生物学者の研究の要求に沿える文献活用支援システムの実現を目標に、生物学分野の文献データの実態を生物学者の視点から探るものである。

## 2. 生物学文献の総合的理解の支援

近年、自然言語処理技術や情報検索技術を利用して、生物学文献に蓄積された知識を抽出・活用しようという試みが盛んになっている[1, 2]。その多くは、タンパク質相互作用を直接的に記述した部分を処理対象としている(例えば[3, 4, 5])が、生物学研究の支援という観点から捉えると、生物学者が多様な生命現象を総合的に理解することをいかに支援するかという別の課題もある。

生物学の研究においては、新たな発見により、従来無関係と考えられてきた周辺領域の研究が、自身の研究と深く関係していることが見出されることも多い。そのような場合、生物学研究者は、周辺領域の文献から情報を収集し、仮説を構築し、検証実験を行うことになる。また、新たな研究領域を模索している研究者は、利用可能な研究資源と現状の技術レベルにあった研究課題を見出すために、周辺領域の文献を分析する必要が生ずる。しかし、前述のように、近年の生物学研究は、研究課題の細分化・専門化が進んでおり、個々の文献で扱われている課題は、その領域の専門家以外には理解が困難な場合も多い。そのため、比較的近い研究領域に

関しても、研究の進展状況の把握が困難であり、関連する研究成果をすばやく活用できない場合がある。

例えば、植物の重力屈性・細胞分化という異なる切り口で独立に進められた研究が、相互に関連していたという事例がある。これは、1996年にFukakiら[6]が報告した重力屈性異常の突然変異体sgr1と、1995年にScheresらが報告した内皮細胞層が消失した突然変異体scrが、同一遺伝子の変異によるものだと、1998年に判明した[7]というものである。内皮細胞層が重力屈性と関与していることは古くから指摘されていたにも関わらず、この変異体の関係は、ゲノム中の遺伝子座位を求めるまで気がつかれなかった。

各生物学研究者が取り組んだ研究課題には、その研究者の直観が強く反映されていると考えられる。そして、異なる領域の生物学研究者にとっては、生物学研究者としての直観の方が、個々の細分化された研究課題より理解しやすい場合が考えられる。実際、複数の生物学研究者に、異なる領域の文献を調査する場合にどういう手順をとるかを尋ねたところ、まず、生物学的ストーリーに沿って関連研究を位置づけたレビュー論文を見つけようとするという趣旨の同じ答えが返ってきた。

逆に、生物学の研究者は、自身の着目した生命現象の側面の解明に必要な研究課題に取り組んでいるとも考えられる。よって、一見、無関係に見える研究課題でも、同じ研究者が取り組んだ研究課題であれば、その生物学者の直観においては、矛盾なく結び付けうるということを示唆している可能性がある。言い換えれば、生物学者の直観に基づくストーリーに沿って研究課題を並べることで、個々の文献で取り扱われた研究課題の位置づけが明確になる可能性がある。

以上の考えに基づき、以降では、生物学の研究者の著作をまとめて参照可能とすることで、生物学者の文献活用を支援する可能性を実例に基づき検討する。

## 3. 著者情報を手がかりとした文献の自己組織化とテーマ抽出

### (1) 研究テーマに基づいた文献抽出

生物学者の文献活用のひとつに、予め想定した何らかの研究テーマについて、周辺領域から、関連テーマを扱った論文を収集しようとするケースがある。このとき利用者は、実際の生物学における研究テーマの構造を元に、関連研究テーマの範囲を想定する。そして、その関連研究テーマを扱った論文の集合を、漏れなく、かつ誤りなく抽出することを望む。実際に論文集合を抽出する際には、論文に付随する何らかの情報を抽出し、それを手がかりにして行うことになる。

文献に含まれる単語や、論文に付与されたキーワードなどを手がかりにして、利用者の望む論文集合が抽出できる場合もある。しかし、大抵は必要とする論文が何割か欠落していたり、関係が浅い論文が何割か含まれていたりする。それは手がかりとする情報の構造が生物科学の研究テーマの構造と一致していないことが主要な原因であると考えられる。

そこで、論文の著者情報を手がかりに、より確実かつ正確に利用者の望む研究テーマを扱う論文の集合を抽出することを考えた。前章でも述べたとおり、研究者は自身が着目した生命現象の側面の解明を研究テーマとし、それに沿って研究課題を選んでいる。そのため、ひとりの研究者の取り組んだ研究課題の集合は、生物科学の研究テーマの構造が反映されたものとなっている。従って論文の著者情報の構造は、生物科学の研究テーマの構造とよく一致していることが期待される。

生物学的観点からは関連のない論文に、同じ著者が脈絡もなく関与するようなことがあると、著者情報に基づき抽出した論文集合の構造は、生物学の研究テーマの構造と対応しなくなってしまう。たとえば、測定技術の専門化が、生物学的テーマとしては全く関係のない多数の論文に著者として関与していると、その人物を手がかりに抽出した論文集合からは、測定技

術としての共通性しか見出せないといった事態が考えられる。

そのような事態が実際には稀であることを確かめるために、分子生物学の論文調査を行った。まず、MEDLINE データベースから、分子生物学のモデル生物として利用されている Arabidopsis(シロイヌナズナ)に関する論文を PubMed(http://www.ncbi.nlm.nih.gov/entrez/)の検索機能によって抽出した。具体的には、”Arabidopsis”というキーワードで検索し、検索結果として得られた 10,568 件の論文を分析対象として抽出した。次に、この論文集合において、同じ名前の研究者が著述した論文の集合を抽出し、生物学的テーマと対応しているかを分析した。具体的には、分析対象とした論文のいずれかで著者として登場した研究者(著者名)24,408 名のなかで、50 件以上の論文を著述した研究者 27 名を取り上げ、その著述論文の集合が、何らかの生物学的研究テーマに即したものとなっているかを分析した。

表 1 に、分析対象とした研究者のうち、著述数の上位 10 名の研究者について、その著述論文に振られている主要な MeSH タームを 5 つずつ取り出して示した。ここで MeSH タームは、重要度を tf-idf 法の変種（著者の著述した論文集合において当該タームが付与された論文数を tf として扱い、df は分析対象論文集合の範囲で計算したもの）によって計算し、自動的に抽出した。表 1 で網掛けした部分は、先頭の列に挙げた研究者の研究(論文集合)の生物学的テーマとしての特徴を現していると判定(本論文の著者の一人の主観評価による)された MeSH タームである。これらの MeSH タームに示されるように、今回調査した範囲では、著者を手がかりに抽出した論文集合は、何らかの生物学的研究テーマに適切に対応していた。すなわち、少なくとも多数の論文を著述している研究者は、何らかの研究テーマに即して論文を著述していることが確かめられた。この事実は、生物学の研究テーマの構造に沿って論文集合を抽出する手がかりとして論文の著者情報をもちいるのが妥当であることを示唆している。

### (2) 研究テーマ構造が不明な研究領域からの文献抽出

前述のように、生物学の研究においては、新たな発見により、従来無関係と考えられてきた周辺領域の研究が、自身の研究と深く関係して

(表 1)著述論文数上位の著者と著書の MeSH ターム

| 著者名 | MeSH ターム | | | | |
|---|---|---|---|---|---|
| Van Montagu | Support, Non-U.S. Gov't | Cyclin-Dependent Kinases | Oxidative Stress | Plants, Toxic | Protein p34cdc2 |
| Shinozaki | Water | DNA, Complementary | Desiccation | Base Sequence | Support, Non-U.S. Gov't |
| Meyerowitz | Support, U.S. Gov't, P.H.S. | Support, U.S. Gov't, Non-P.H.S. | Genes, Homeobox | AGAMOUS Protein, Arabidopsis | Homeodomain Proteins |
| Inze | Cyclin-Dependent Kinases | Cell Cycle | Oxidative Stress | Cyclins | Support, Non-U.S. Gov't |
| Chua | Microfilament Proteins | Actins | Support, U.S. Gov't, P.H.S. | Plants, Genetically Modified | Phytochrome |
| Yamaguchi-Shinozaki | Water | Desiccation | Abscisic Acid | Sodium Chloride | DNA, Complementary |
| Chory | Light | Support, U.S. Gov't, Non-P.H.S. | Signal Transduction | Phytochrome | Support, U.S. Gov't, P.H.S. |
| Deng | Support, U.S. Gov't, P.H.S. | Light | Carrier Proteins | Morphogenesis | Support, U.S. Gov't, Non-P.H.S. |
| Ausubel | Plant Diseases | Virulence | Pseudomonas | Support, U.S. Gov't, P.H.S. | Pseudomonas aeruginosa |
| Dean | Chromosome Mapping | Genetic Markers | Chromosomes, Artificial, Yeast | Polymorphism, Restriction Fragment Length | DNA Transposable Elements |

著述論文数の上位 10 名の研究者の著述論文について、重要度の高い 5 つの MeSH タームを抽出した。左から重要度の高い順に並んでいる。実際の研究グループの研究内容をよく表しているものに網掛けを施してある。

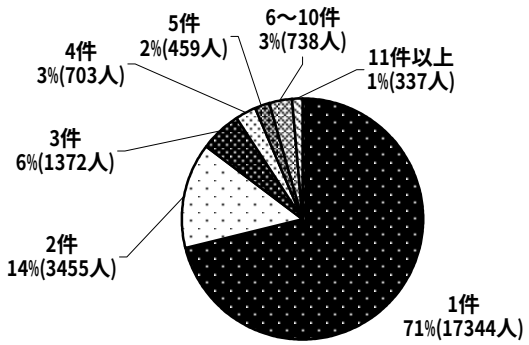
いることが見出されることも多い。そのような場合、利用者は、自身の専門と関連が薄い研究領域など、その領域における研究テーマの構造(その領域でどのようなテーマがあり、また、それぞれのテーマが相互にどのように関連しているか)の知識すらない状態で、文献調査を行う必要に迫られることがある。その場合、まず、その領域の主要研究テーマを把握しないと調査すべき論文の範囲を特定できない。よって、著者情報を利用して、主要な研究テーマが抽出できれば、それを手がかりとして提示することで、利用者は、適度な範囲の論文集合に、分析対象を絞れるようになる可能性がある。また、抽出した主要テーマおよびテーマ間の関連を利用者に示すことで、未知の領域の研究テーマの理解を支援することなども考えられる。そこで、本節では、著者情報を手がかりに、研究領域の主要テーマを抽出しうるかを検討する。

利用者が分析対象とすべき論文集合を絞ることを支援するには、まず、関連テーマの大きく研究をまとめて提示することで、選択すべき対象を少なくすることが考えられる。そこで、著述論文数の多い研究者に関する論文集合をまとめて、利用者に提示することを考えた。この場合、著述論文数の多い研究者は少数であり、かつ、それら少数の研究者による論文が元の論文集合をなるべく広くカバーしていることが望ましい。この観点から、前述のシロイヌナズナに関する論文集合を分析した結果を図１と図２に示す。

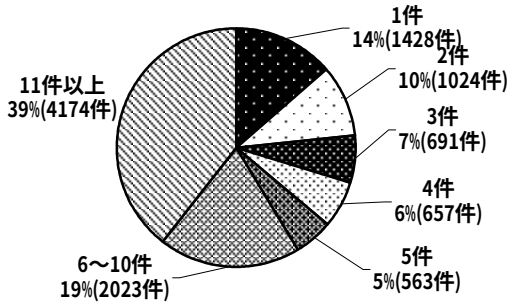
図１は、著述論文数による研究者の構成比を示したグラフである。これによれば、6 本以上の論文を著述した研究者は全体(24,408 人)の 4%に過ぎず(1,075 人)さらに 11 本以上の論文を著述した研究者は全体のわずか 1%に過ぎない(337 人)。この結果から、多数の論文を著述している研究者は研究者全体においてごくわずかであることがわかる。

図２は、著述論文数の多い順に研究者(分析対象論文の著者)を選択し、その研究者による論文集合を利用者に提示した場合に、利用者は

(図 1)著述論文数による研究者の構成比




(図２)最有力著者の著述数による論文の構成比




どれ位効率的に分析対象を絞っていけるかという観点から、分析対象論文集合を分析した結果である。具体的には、以下の手順により、提示論文集合のカバー率に相当する値を求めた。まず、それぞれの論文に対して、共著者中で最大の論文著述数を持つ研究者(以下、最有力著者と呼ぶ)を１名ずつ選択する。次に、各研究者について、その研究者が最有力著者となっている論文数（以下、有効論文数と呼ぶ）を集計する。そして、有効論文数同一の研究者による論文集合について、元の論文集合に対する構成比を集計した結果が図２である。この図によれば、11 本以上の論文を著述した 1%の研究者を著者情報に含む論文は全体(10,568 件)の 39%を占め(4,174 件)、6 本以上の論文を著述した 4%の研究者を著者情報に含む論文は全体の 58%を占める(6197 件)。

この結果は、著述数の多い順に研究者を選択し、その研究者による論文集合を利用者に提示

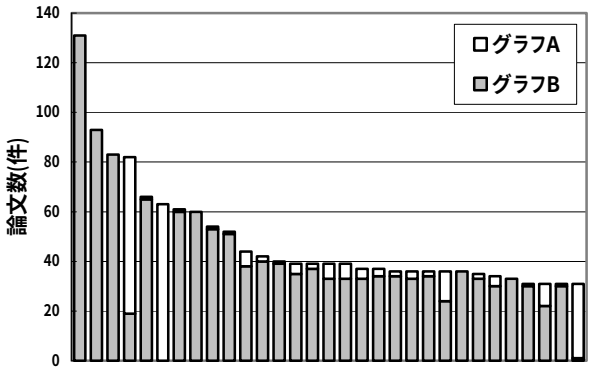
すれば、効率的に分析対象とすべき論文集合を絞ることが支援できることを示唆しているが、論文集合を絞るという意味では、最初の選択肢である提示集合にできるだけ重なりが少ないことが望ましい。すなわち、最初の選択の手がかりとする著述論文数の多い研究者同士は、共著関係が少ないことが望ましいことになる。

図３は、著述論文数の上位 30 人の研究者間の共著関係を調査した結果である。グラフ A は、著述論文数の大きい順に研究者を並べ、その著述論文数をプロットしたグラフである。グラフ B は、各研究者の有効論文数のグラフである。例えば、４番目に多く論文を著述した研究者は、著述論文数は 82 件のうち 19 件で、自分自身が最有力著者であったことを示している。グラフ A の値に対してグラフ B の値が小さい人物はほとんどいないことがわかる。つまり、この範囲のほとんどの人物は、大半の著述論文において、共著者のなかで自分が最も著述論文数が多い。この事実は、著述論文数の多い人物同士は共著関係が少ないということを示している。言い換えれば、バイオ研究においては、研究の中心となる人物同士は共著することが少ない傾向があることを示している。

図３に示した人物のなかには、グラフ A の値に対してグラフ B の値が小さい人物も含まれている。このような人物は、ほとんどの著述論文において、自分よりも多くの著述論文数をもつ人物が共著者となっている。分析対象集合の絞込みを支援する上で、なるべく少ない数の提示論文集合で、なるべく広く元の論文集合をカバーすることを考えると、このような人物による論文の集合は、最初の選択肢としては用いないことが効率的である。このようにすれば、図１で示した著述論文数の多い人物数よりもさらに少ない数の論文集合を提示するだけで、利用者は絞りこみを完了できる可能性はある。ただし、生命現象を捉える観点は、必ずしも背反であるとは限らないので、このように提示することがよいかについては、議論の余地がある。

以上の結果は、未知の分野の研究内容把握の際、多数の論文を著述した研究者の著者情報に基づいて、出発点とする適当な大きさの論文集合を抽出できることを示している。

(図 3)著書数上位 30 名の著述論文数と有効論文数




著述論文数の上位 30 名を抽出し、著述論文数の多い順に並べた。
(グラフ A)著述論文数(グラフ B)有効論文数

## 4. まとめ

今回、生物学者の文献活用の支援について検討した。今回は次の二つのケースについて考え、いずれにおいても文献の著者情報を利用することが妥当であることを示した。第 1 に、利用者が、何らかの研究テーマを想定し、特定の領域から、関連テーマを扱った論文の集合を抽出しようとするケースを取り上げた。そして、著者情報を手がかりに、研究テーマの構造に沿った論文集合が抽出できることを例証し、それにより、利用者の望む論文集合が的確に抽出できるよう支援する可能性を論じた。第２に、特定の領域について、主要な研究テーマを抽出しようとするケースを取り上げた。そして、植物を対象とする分子生物学研究の論文集合を題材に、各研究者の著述論文数と共著関係を分析し、著述論文数の多い研究者は比較的少数で、しかも互いに共著しあうことが少ない傾向があることを示した。このことにより、著者情報が、主要な研究テーマを抽出する手がかりとなる可能性が強く示唆された。

しかし、著者情報に基づき利用者の文献活用を支援する方法を確立するためには、いくつかの解決するべき問題が残されている。また、著

者情報を他の情報と併用することで、より高度な支援機能を実現できないか検討する余地がある。

例えば、利用者の望む研究領域の論文集合を抽出できたとしても、その数が膨大であれば、そのままでは利用者は活用することができない。論文集合を効率よく理解するためには、その論文集合の内部構造（サブテーマとサブテーマ間の関係を反映した構造）を知ることが必要となる。その場合でも、著者情報を利用することが有効であると考えられる。たとえば、著者情報をてがかりに抽出した論文集合を、さらに、共著関係によって分析することで、内部構造を求めることが考えられる。研究者一人一人が何らかの形で生物学的な研究テーマに沿って研究しているなら、共著関係により抽出した内部構造も、生物学的な研究テーマに沿ったものになるはずである。

シロイヌナズナを用いた植物の光形態形成研究の主要な研究者 3 名(Chory、Chua、Deng)が最有力著者である論文集合(以後、主要グループと呼ぶ)を抽出し、共著関係によって分析した。分析の結果、これらの主要グループには、ほとんど内包されてしまうような他の研究者の著述論文の集合が含まれていた。以後、これらの集合の共通部分を、共著グループと呼ぶ。表２に、主要グループと共著グループ（あわせて「研究グループ」と呼ぶ）について、各グループの論文数と、光形態形成の主要キーワードをタイトルに含む論文数を示す。この結果は、共著関係によって主要グループの内部構造が抽出できることを示唆している。例えば、Deng の主要グループに着目すると、COP1，COP9 というキーワードで特徴付けられる２つの主要研究テーマがあり、これらのテーマの両方に関連する共著グループと、一方のみに関連する共著グループがあることがわかる。このように、論文集合を共著関係によって分析し、研究テーマに沿った内部構造を求めれば、利用者に論文集合を理解するヒントを提供できる可能性がある。

(表 2)研究グループと研究テーマとの対応

| 研究グループ | all | Phytochrome | COP1 | COP9 |
|---|---|---|---|---|
| Chory | 64 | 12 | - | - |
| Chory::Chory | 17 | 5 | - | - |
| Chory::Fankhauser | 6 | 2 | - | - |
| Chory::Li | 16 | 1 | - | - |
| Chory::Nagpal | 3 | 1 | - | - |
| Chory::Reed | 4 | 2 | - | - |
| Chua | 59 | 8 | 1 | 1 |
| Chua::Chua | 35 | 7 | 1 | 1 |
| Chua::Kunkel | 4 | 1 | - | - |
| Deng | 51 | 1 | 24 | 12 |
| Deng::Deng | 17 | - | 9 | 1 |
| Deng::Hardtke | 2 | - | 1 | - |
| Deng::Matsui | 5 | 1 | 3 | 2 |
| Deng::McNellis | 3 | - | 3 | - |
| Deng::von Arnim | 6 | - | 4 | - |
| Deng::Wei | 18 | - | - | 9 |

(列の説明)
研究グループ:主要な研究者名と、その共著者名で示す。共著者名なしの行は主要な研究者のグループ全体を示す。
all:その著書グループの総論文数を示す。
Phytochrome,COP1,COP9:論文タイトルに各キーワードを含む論文数を示す。

しかし、分割可能な複数の研究テーマを、同一のグループが行っている場合も考えられ、この場合は著者情報だけでは内部構造を与えることができない。ここに、著者情報以外の情報の併用という課題がある。例えば、あるグループの研究テーマが時間と共に移行したケースが考えられる。このケースでは、内部構造を求めるには、時系列情報の併用が必要となる。また、同一グループが複数の研究テーマを同時に行っている場合もあり得る。その場合はコンテンツベースの自己組織化機能などを併用して論文集合を細分し、内部構造を求める必要がある。あるいは、論文の引用関係によって内部構造を求めることも考えられる。著者情報以外の手がかりを併用する手法の開発、また、引用関係などによる自己組織化と、著者情報による自己組織化の特性の違いを調査することは、今後の課題である。特に、著者情報と別の情報を併用して論文集合を自己組織化することは、文献活用支援の可能性を拡大する上で重要な課題である。

例えば、著者情報と時系列情報を併用して論文集合を解析することで、研究テーマの推移を

検出できる可能性がある。そして、それにより、文献活用支援の可能性が飛躍的に拡大する可能性がある。ある研究者について、研究テーマの変化を捉えることで、その研究者の直観がどう働いたのかを推測できたなら、そこから利用者の研究に有意義な示唆を抽出できるかもしれない。

また、複数の研究テーマについて、研究テーマの関係の推移を分析することで、生物学分野の研究テーマの隠れた方向性を発掘できる可能性もある。例えば、当初関係が見出せなかった研究テーマが関連性を深めていく様を時間と共に観察すると、まずコンテンツベースによる関係が徐々に現れ、続いて論文の引用関係が見られるようになり、最終的に研究者間の共著関係が表れる、といった変化が検出できるかもしれない。このことは仮説に過ぎず、今後検証していくことが必要であるが、もしこのような法則が見出せるならば、研究テーマ、ひいては、生命現象の隠れた関係を発掘する手法や、今後の研究の方向性を予測する手法の開発へと発展する可能性がある。そして、そのような手法により、生物学研究者の研究の方向性に強い影響を与えうる文献活用支援方法が実現できるかもしれない。今後の生物学研究の文献活用方法の発展のためには、著者情報の重要性を充分認識し、活用していくことが重要である。

## 参考文献


[1] Hirschman, L., Park, J., Tsujii, J., Wu, C. and Wong, L.: Literature Data Mining for Biology, in *Proc. of PSB 2002*, pp. 323-325 (2002).

[2] C. Friedman, R. M., L. Hirschman and Wu, C.: LINKING BIOLOGICAL LANGUAGE, INFORMATION AND KNOWLEDGE, in *Proc. of PSB 2003*, pp. 388-390 (2003).

[3] Blaschke, C., Andrade, M. A., Ouzounis, C. and Valencia,~A.: Automatic extraction of biological information from scientific text: protein-protein interactions, in *Proc. ISMB1999*, pp. 60-67 (1999).

[4] Friedman, C., Kra, P., Yu, H., Krauthammer, M. and Rzhetsky, A.: GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles, in *Proc.ISMB2001*, Vol. 17, pp. S74-S82 (2001).

[5] Ding, J., Berleant, D., Nettleton, D. and Wurtele, E.: Mining MEDLINE: Abstracts, Sentences, or Phrases?, in *Proc. of PSB 2002*, pp. 326-337 (2002).

[6] Fukaki, H., Fujisawa, H. and Tasaka, M.: SGR1, SGR2, and SGR3: Novel Genetic Loci Involved in Shoot Gravitropism in Arabidopsis thaliana, *Plant Physiology*, Vol. 110, No.~3, pp. 945--955 (1996).

[7] Fukaki, H., WysockaDiller, J., Kato, T., Fujisawa, H., Benfey, P. N. and Tasaka, M.: Genetic evidence that the endodermis is essential for shoot gravitropism in Arabidopsis thaliana, *Plant Journal*, Vol. 14, No. 4, p. 425 (1998).