

筑波大学大学院博士課程

システム情報工学研究科修士論文

論文の被引用数と著者キーワードの頻度推移に
着目した視覚的文献調査ツール

謝 英双

(コンピュータサイエンス専攻)

指導教員 三末 和男

2010年3月

概要

新しい研究分野に参入しようとしている研究者にとって、着目している研究分野に関する情報収集は重要な課題である。文献調査を通して、当該研究分野においてどのような研究テーマがあるのかということや、どれが主流の研究テーマであるのかなどの情報をできるだけ効率的に知りたいと考えられる。また、それぞれの研究テーマに関しては、重要な論文も把握しておきたいと考えられる。

しかし、既存の文献調査システムは、このような研究者に対する支援が十分ではない。たとえば、既存の文献調査システムを利用しても、当該分野における様々な研究テーマを簡単には見つけられない。さらには、どのようなテーマが重要であるかといった情報も提示してくれない。このような不満は、既存の文献調査システムの多くが論文の検索に焦点を合わせており、「研究テーマ」という観点での調査支援が充実していないことに起因する。

本研究では、研究の初心者が文献調査を行う際の効率を向上させる文献調査ツールを開発する。我々は、文献調査ツールへの要望を以下のような三つの項目に整理した。一つ目は、対象とする研究分野の全体像をみたい。二つ目は、個々の研究テーマの特徴を俯瞰したい。三つ目は、重要な論文の探索を助けてほしい。また、三つの要望を満たすために、それぞれに応える機能を開発した。これらの機能を提供することで、既存の文献調査システムの問題点を解決できると考えられる。

この三つの機能を実現するために、我々は論文の被引用数と著者キーワードの頻度推移に着目し、さらに、視覚的表現の直感性と分かりやすさなどの利点を利用することとして、三つの機能のそれぞれに対応した視覚表現を設計した。また、これらの三つの視覚表現を連携させることによって、視覚的文献調査ツールを開発した。本ツールを利用することによって、研究者は着目している研究分野における主要な研究テーマを簡単に見つけることができ、さらに、それに関連する重要な論文も素早く見つけることができる。これにより、新規参入分野における文献調査の効率を向上させることが可能になる。最後に、アンケート形式で評価実験を行い、その結果から本ツールおよび三つの機能の有用性を示した。

目次

第1章 序論	1
1.1 文献調査とその重要性	1
1.2 文献調査により取得したい情報とは	1
1.3 支援対象ユーザ	2
1.4 対象ユーザが文献調査において抱える問題	2
1.5 本研究の目的	2
1.6 本研究の位置付け	2
1.7 本研究の貢献	3
1.8 本論文の構成	4
第2章 既存の文献調査支援システム及び問題点	5
2.1 既存の文献調査支援システム	5
2.2 既存の文献調査支援システムの問題点	7
第3章 問題点を解決するための基本アイデア	8
3.1 文献調査ツールへの要望	8
3.2 本研究の提案手法の概要	9
第4章 論文の被引用数と著者キーワードの頻度推移についての分析	10
4.1 着目した論文データについての分析	10
4.2 なぜ可視化手法を利用するのか	13
第5章 望まれる機能に対応した視覚表現の使用	14
5.1 要望に応える機能に対応した視覚表現	14
5.1.1 研究分野の全体像を提示するための表現	14
5.1.2 研究テーマの特徴を俯瞰できるようにするための表現	15
5.1.3 重要な論文を探しやすくするための表現	18
5.2 リンキングとブラッシング技術の利用	20
第6章 効率的な文献調査を支援する視覚的ツール	22
6.1 視覚的文献調査ツールの概要	22
6.2 インタフェースの設計	23
6.2.1 全体像の提示法	23

6.2.2	研究テーマの特徴の俯瞰法	26
6.2.3	重要な論文の探索法	27
6.2.4	操作体系	29
6.3	ツールの実装	30
6.4	システムの構成	30
6.5	ツールの利用例	32
第 7 章	評価実験と考察	35
7.1	評価実験	35
7.1.1	実験の目的	35
7.1.2	実験の被験者	35
7.1.3	実験の方法	35
7.1.4	実験の結果	37
7.2	考察	37
第 8 章	関連研究	39
8.1	視覚的情報検索に関する研究	39
8.2	研究サーベイの支援に関する研究	40
第 9 章	結論と今後の課題	42
9.1	結論	42
9.2	今後の課題	42
	謝辞	44
	参考文献	45

図目次

1.1	本研究の位置付け	3
2.1	ACM Protal の操作中の画面	6
2.2	Google Scholar の操作中の画面	6
4.1	著者キーワードの説明	11
5.1	リストによる研究分野の全体像の表示	15
5.2	主流の研究テーマの頻度推移の表示	16
5.3	発展しそうな研究テーマの頻度推移の表示	17
5.4	注目されていない研究テーマの頻度推移の表示	17
5.5	研究テーマの特徴を俯瞰する折れ線グラフ	18
5.6	論文を調べる散布図	20
6.1	ツールの概観	23
6.2	著者キーワード全体を表示するリスト	25
6.3	全体的出現頻度に対応する色相の使用	25
6.4	著者キーワードの頻度推移を表す折れ線グラフ	27
6.5	著者キーワードに関連する論文を表示する散布図	28
6.6	論文検索の操作	29
6.7	タイトルを探す操作	30
6.8	文献調査ツールのシステム構成	31
6.9	文献調査ツールの初期画面	33
6.10	「focus+context」に関連する重要な論文	33
6.11	「treemap」に関連する重要な論文	34

表 目 次

4.1	論文要素の例	10
4.2	著者キーワードの頻度推移の例	12
7.1	各被験者のアンケートへの回答	37
7.2	各機能に対する評価	37

第1章 序論

1.1 文献調査とその重要性

研究者が研究をする上で、既存の研究について調べ、論文サーベイを行う必要がある場面よく発生する。文献調査を通して、研究者はいろいろな有益な情報を取得できる。例えば、自分の研究に深く関係している分野について調べることで、当該分野の中で自身の研究がどこに位置づけられるかや、他の研究との差異を知ることで自身の研究の強みを知ることができる。あるいは、初めて触れた研究分野について調べることで、この領域における重要な論文、主な研究テーマなどの有益な情報を知ることができる。これにより、自分のこれから行う研究テーマの選択に非常に役立つ情報を取得できる。

1.2 文献調査により取得したい情報とは

研究者が着目している研究分野の把握の度合によって、文献調査を通して得たい情報が異なる。

新しい研究分野に参入しようとしている研究者にとって、着目している研究分野に関する情報収集は重要な課題である。このような研究者は、文献調査を通して、当該研究分野においてどのような研究テーマがあるかということや、どれが主流の研究テーマであるかなどの情報をできるだけ効率的に知りたいと考えられる。また、それぞれの研究テーマに関して、重要な論文も把握しておきたいと考えられる。このような情報を把握することによって、着目している研究分野の概観を把握でき、これから行う研究テーマの決定が支援される。

一方、自分の当該研究分野についてある程度把握している研究者にとって、自分の研究に深く関係している分野について調べることによって、自分の研究が当該分野においてどこに位置づけられるのかということや、他の研究との差異を知りたいと考えられる。さらに、それらの情報を把握することで、自分の研究の強みを発見できる。

1.3 支援対象ユーザ

本研究では、研究の初心者支援の対象ユーザとする。例えば、自分のやりたい研究テーマはまだ分からない学部4年生、または修士1年生のような人である。このような研究者は、これからある研究分野に新規参入しようとしている研究者であると考えられる。

1.4 対象ユーザが文献調査において抱える問題

本研究の支援対象ユーザは、これから取り組もうとしている研究分野に初めて触れるため、この研究分野をまだ把握していない段階であり、研究分野に関する知識が不足している。そのため、研究者は効率的に文献調査を行うことができない。具体的な例として、研究室に新しく配属された学生の場合を挙げる。1.2節で述べたことによって、このような学生は研究テーマを決める段階にあるので、文献調査を通して、自分の着目している研究分野の研究テーマや、主な研究テーマに関連する重要な論文などの有益な情報を把握しておきたい。しかし、彼らは文献調査をする時、当該分野に関する知識が不足しているため、適切な検索キーワードの選択は難しい。従って、彼らは着目している研究分野を代表とするキーワードを用いて検索するケースが多々ある。これは効率的な文献調査とは言えない。

1.5 本研究の目的

本研究は、研究の初心者を対象として、初心者より効率的な文献調査を可能にすることを目的とする。その達成のために、論文の被引用数と著者キーワードの頻度推移に着目し、さらにそれらを視覚的に表現する文献調査支援ツールを開発する。この視覚的文献調査ツールを使うことによって、文献調査の効率が向上することが期待できる。

1.6 本研究の位置付け

本研究の位置付けを図1.1に示す。本研究で開発したツールは、研究の初心者に向け、検索キーワードと検索結果の両方を選択するための手掛かりを提供する視覚的検索ツールであると考えられる。

我々は、初心者の情報検索に対する三つの要望を整理した。多くの従来の研究は、初心者の情報検索に対する要求を考慮していない。レボウィッツら[1]や、Sarahら[2]が開発したツールは、初心者向けに開発されたが、本研究で整理した初心者向けの三つの要望を満たしていない、また、検索キーワードと検索結果の両方を選択するための手掛かりを提供していない。

一方、検索キーワードの提示、または検索結果の提示に着目する研究が多い。Fernandoら[3]や、Edwardら[4]、Gregら[5]、Antonら[6]、Keithら[7]、Justinら[8]などの研究は、検索結果を選択するための手掛かりを提供することに着目した。一方、Minaら[9]や、Orlandら[10]などの研究は、キーワードを選択するための手掛かりを提供することに着目した。本研

究と検索キーワードの提示と検索結果の提示両方に着目した点が似ている Orland ら [11][12] や、白鳥ら [13]、Marian ら [14] などの研究は、初心者が文献調査を行う際の要望を考慮していない。

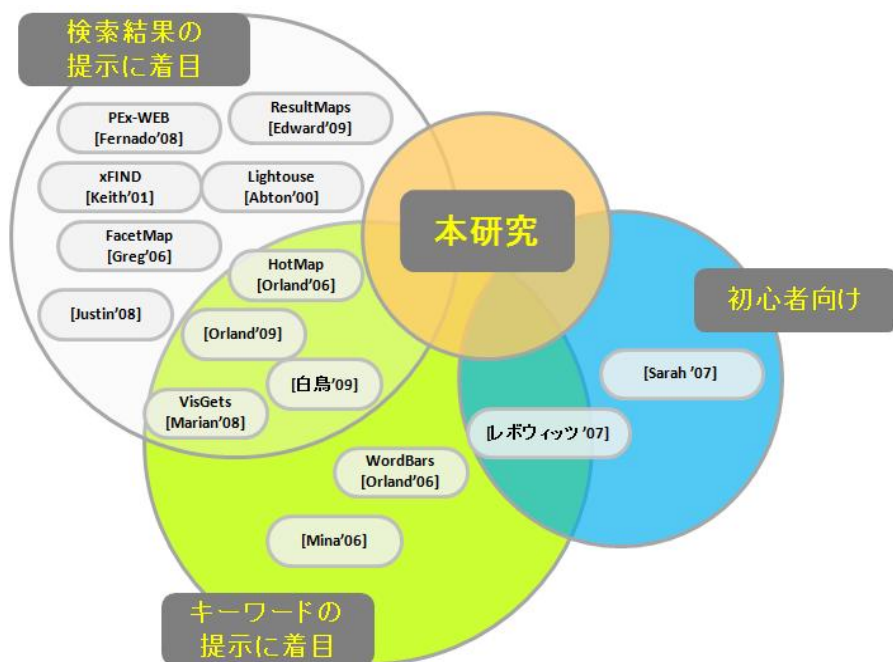


図 1.1: 本研究の位置付け

1.7 本研究の貢献

我々は以下の2点を本研究における貢献と考えている。

一つ目に、本研究は、研究の初心者が文献調査を行う際の要望を整理し、それらの要望を満たすために三つの機能を提案した。また、三つの機能を実現するために、論文の被引用数と著者キーワードの頻度推移に着目し、更にそれらを視覚的に表現する文献調査支援ツールを開発した。このツールを用いることによって、研究者は、着目している研究分野における様々な研究テーマを簡単に見つけることができる。さらに、これらの研究テーマにおける主流の研究テーマや今後発展しそうな研究テーマも効率的に見つけることができる。また、主要な研究テーマに関連する重要な論文を素早く見つけることができる。従って、本ツールによって、研究の初心者は文献調査をより効率的に行えるようになる。

二つ目に、本研究で提案した三つの機能を実現するために、我々は論文の被引用数と著者キーワードの頻度推移を用いることで、検索結果と検索キーワードの両方に着目し、それら

を選択するための手掛かりを提供する検索手法を開発した。この手法を用いることで、初心者が文献調査を行う際の要望を満たすことができる。

1.8 本論文の構成

第2章では、既存の文献調査支援システム及びその問題点を述べる。第3章では、既存の文献調査支援システムの問題点を解決するための基本的なアイデアについて述べる。この後、基本的なアイデアを元に本研究の提案手法の概要について述べる。第4章では、第3章で述べた提案手法を実現するために、論文の被引用数と著者キーワードの頻度推移に着目する理由について述べる。第5章では、第3章で述べた提案手法を実現するため用いられた三つの視覚表現を記述する。第6章では、提案手法を元に第5章で述べた三つの視覚表現を連携した視覚的文献調査ツールの開発について述べる。第7章では、開発した文献調査ツールの有用性と使いやすさを検証するための評価実験について述べる。第8章では、関連研究について述べる。最後に、第9章では、本研究の結論と今後の課題について述べる。

第2章 既存の文献調査支援システム及び問題点

本章では、まず、2.1 節で、既存の文献調査支援システムを簡単に分類して紹介する。この後、2.2 節で、これらの文献調査支援システムの問題点について述べる。

2.1 既存の文献調査支援システム

既存の文献調査支援システムの多くは、テキストのようなウェブサーチインタフェースを利用することで、文献検索を行う [8]。これらのシステムを利用して出てきた検索結果は、一つの観点（例えば、“relevance”、“publication date”などの観点）で、順序付きリストで表示されている。このような表現手法を利用することによって、より良いランクを付けられた検索結果の発見が支援される [3]。ACM Protal[15] や CiNii[16]、Google Scholar[17]、CiteSeer[18] などが有名である。

これらの有名な文献調査システムは、論文の収集方法によって二つに分けることができる [19]。一つは、学会や研究所等が所有している論文を整理し、タイトルや著者名、本文などをインデックスとして検索可能にしたシステムである。ACM Protal や CiNii などがこれにあたる。ACM Protal は主に英語の論文を検索するシステムであり、CiNii は日本語の論文が検索可能なシステムである。

もう一つは、Google Scholar や CiteSeer などのように、研究者や企業が Web 上をクロールすることで論文を収集し、それらを整理し、タイトルや著者名、本文などをインデックスとして検索可能にしたシステムである。

図 2.1 は、ACM Protal というシステムの操作中の画面である。図 2.2 は、Google Scholar というシステムの操作中の画面である。

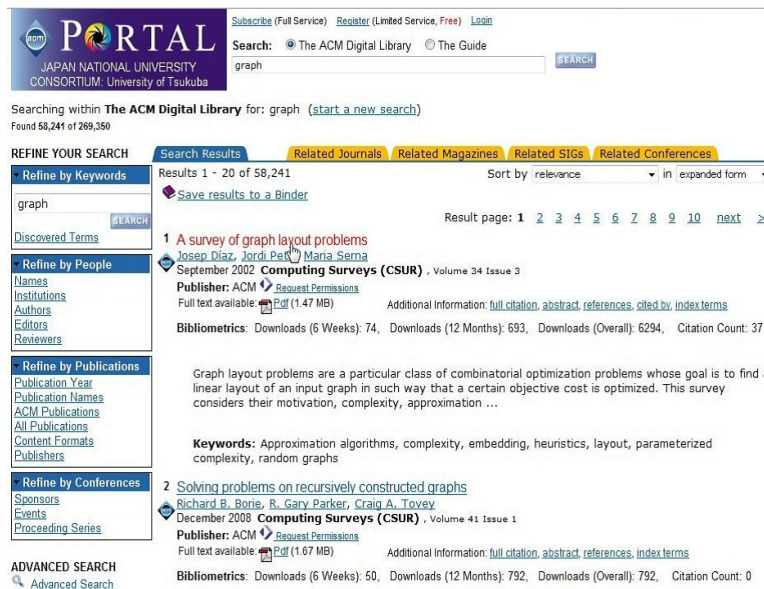


図 2.1: ACM Protal の操作中の画面

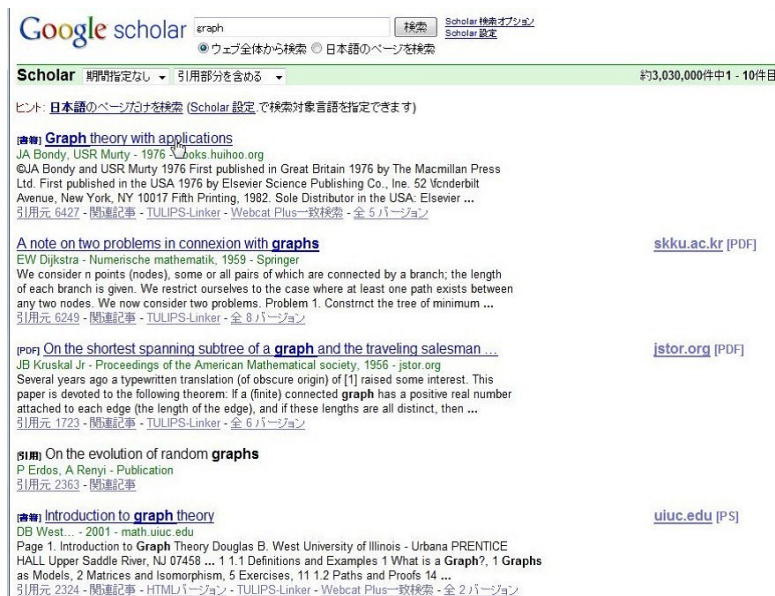


図 2.2: Google Scholar の操作中の画面

2.2 既存の文献調査支援システムの問題点

既存の文献調査支援システムは、文献調査をする上で有用なシステムであるが、完璧なものとは言えない。例えば、ACM Portal や Google Scholar などの文献調査システムを利用すれば、入力するキーワードに関連する大量な論文を迅速に得ることができる。これは、既存の文献調査システムが論文の検索に焦点を合わせるという特徴を反映している。しかし、既存の文献調査システムは、研究の初心者に対する支援が十分ではない。このような研究者は、文献調査を通して、1.2 節で述べた有益情報を取得したい。ところが、既存の文献調査システムでは、それらの有益情報を効率的に獲得できない。たとえば、既存の文献調査システムを利用しても、研究者が着目している研究分野における様々な研究テーマを簡単には見つけることができない。さらには、どのようなテーマが重要であるかといった情報も提示してくれない。これにより、当該分野における主流の研究テーマを見つけにくくなる。このよう不満は、既存の文献調査システムの多くが、「研究テーマ」という観点での調査支援が充実していないことに起因する。

第3章 問題点を解決するための基本アイデア

本章では、既存の文献調査システムの問題点を解決するための基本的なアイデアとして、文献調査ツールへの三つの要望を整理し、さらに、それぞれに応える三つの機能を提起する。この後、このアイデアを基づいて、本研究の提案手法の概要を述べる。

3.1 文献調査ツールへの要望

我々は、文献調査ツールへの要望を以下の三つに整理した。さらに、それぞれに応える三つの機能を備える文献調査ツールを使用することで、文献調査の効率を向上させると考える。

- **対象とする研究分野の全体像をみたい**

研究者はまず初めに、研究分野にどのような研究テーマがあるのかを把握したいと考えられる。従って、様々な研究テーマを一度に眺められる分野の全体像を提示する機能が必要である。この全体像を見ることによって、当該分野における様々な研究テーマを簡単に見つけることができる。

- **個々の研究テーマの特徴を俯瞰したい**

様々な研究テーマを概観したうえで、主流の研究テーマや今後発展しそうな研究テーマを探したいと考えられる。従って、個々の研究テーマのはやりすたりを俯瞰できる機能が必要である。これにより、いろいろな研究テーマの中から、どの研究テーマがずっと注目されているのかや、どの研究テーマが近年から流行し始めたのかなどの有益情報を取得できる。

- **重要な論文の探索を助けてほしい**

主流の研究テーマを見つけた後、それに関連する論文も把握しておきたいと考えられる。従って、研究テーマに関連する重要な論文を簡単に見つけられる機能が必要である。これらの重要な論文を読むことによって、研究テーマに関する研究内容を詳しく理解することができる。

これらの要望に応える三つの機能を提供することで、既存の文献調査システムの問題点を解決できると考えられる。

3.2 本研究の提案手法の概要

3.1 節で述べた三つの要望をそれぞれに応える機能を実現するために、我々は論文の被引用数と著者キーワードの頻度推移に着目し、さらに、視覚的表現の直感性と分かりやすさなどの利点を利用することとして、三つの機能のそれぞれに対応した視覚的表現を設計した。また、これらの三つの視覚表現を連携させることによって、視覚的文献調査ツールを開発した。

研究分野の全体像を提示するために、著者キーワード全体をリストで提示することにした。これにより、研究分野を連想させるキーワードを一度に眺めることができる。

研究テーマの特徴を俯瞰できるようにするために、研究テーマを表す著者キーワードの頻度推移を折れ線グラフで表示することにした。これにより、研究テーマの発展履歴が読み取りやすくなる。

重要な論文を探しやすくするために、文献の被引用数と発表年の二つの観点を用いた散布図で論文全体を表すことにした。これにより、多くの重要な論文を一目で見つけることができる。

さらに、三つの視覚表現を、リンクングとブラッシングによって、一つのツールに統合した。

第4章 論文の被引用数と著者キーワードの頻度推移についての分析

本章では、提案手法を実現するために、論文の被引用数と著者キーワードの頻度推移に着目する理由について述べる。その後、可視化手法を利用する理由について簡単に述べる。

4.1 着目した論文データについての分析

表 4.1¹に示したように、通常、一つの文献は発表年や、被引用数、タイトル、著者、著者キーワード、概要など複数の要素を持っていると考えられる。

表 4.1: 論文要素の例

発表年	被引用数	タイトル	著者	著者キーワード	概要
2009	3	ResultMaps: Visualization for Search Interfaces	E.Clarkson, K Desai, J Foley	treemap, search visualization, and so on ...	We present a treemap-based search ...
2007	84	ManyEyes: a Site for Visualization at Internet Scale	FB Viegas, M Wattenberg, et al.	visualization, social software, and so on ...	We describe the design and development ...
2001	160	Animated Exploration of Dynamic Graphs with Radial Layout	KP Yee, D Fisher, et al.	graph drawing, animation, interaction	We describe a new animation technique ...
...	...	...	...	...	...

我々は表 4.1 に示した要素の中で、文献の**著者キーワード**、**発表年**、**被引用数**と**タイトル**の四つの要素に関する論文データを使用した。

これから、それぞれの要素を使用して表示できる情報について具体的に説明する。

¹このデータは Google Scholar で検索して得たものである。

・ 文献の著者キーワードの使用

図 4.1 に赤い長方形で表した部分は本研究で使った著者キーワードである。このキーワードは著者が論文の要点を表すためにまとめたキーワードであるので、論文の研究テーマを表す的確なキーワードとして使うことができると思う。

着目している研究分野の全体像を表示するために、このような著者キーワードを使用した。つまり、研究者が使用したキーワードを含む全部の論文の著者キーワードを表示することである。これにより、対象とする研究分野における様々な研究テーマを一度に眺めることができる。この全体像を表示することで、研究者は研究分野を代表とするキーワードを使用しても、当該分野における様々な研究テーマを簡単に見つけることができる。

また、個々の研究テーマの特徴を区別するために、それぞれの著者キーワードの年毎の出現頻度を累計して表示させる。この著者キーワードの頻度推移によって、よく注目されている研究テーマや、最近流行し始めた研究テーマなど、主要な研究テーマを発見することができる。

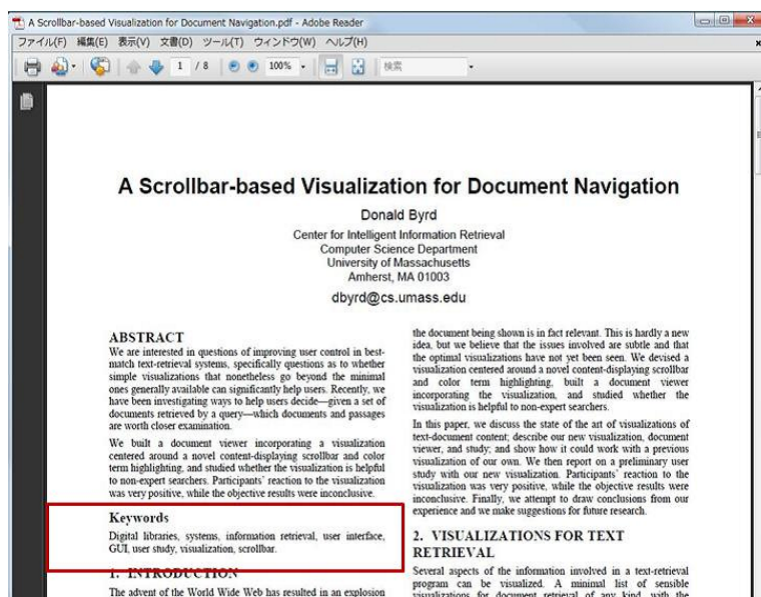


図 4.1: 著者キーワードの説明

・ 文献の発表年の使用

本ツールでは、文献の発表年を二つの重要な目的を達成するために使用する。

一つ目に、文献の発表年は文献の重要度を測る一つの指標として使われる。例えば、近年発表された論文は最新の論文として重要な論文であると思う。あるいは、一番古い論文は当該研究分野の第一者の論文として重要な論文であると思う。

二つ目に、文献の発表年は文献の著者キーワードの頻度推移を表示するために使用される。年毎に同じ著者キーワードの出現する回数を累計すると、この著者キーワードの年より頻度推移を表すことができる。これにより、著者キーワードで表示した研究テーマの特徴が発見できる。例えば、表 4.2 に示したように、10 年間（2000～2009）の三つの著者キーワードの頻度推移の例を挙げる。この表は著者キーワードの年毎の出現回数を表している。この表を見ると、「A」という著者キーワードは近年（2007 年）に初めて出現してきたので、このような研究テーマは今後発展しそうな研究テーマであると考ええる。また、「B」という著者キーワードの頻度推移を見ると、この研究テーマは年毎の出現する回数が 10 年間を通して高いので、このような研究テーマは主流の研究テーマであると考ええる。三つ目の著者キーワード「C」の頻度推移を見ると、この研究テーマは過去に 2 年（2002 年と 2004 年）だけ出現していましたので、このような研究テーマはあまり注目されていない研究テーマであると考ええる。

表 4.2: 著者キーワードの頻度推移の例

発表年 著者 キーワード	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
A	0	0	0	0	0	0	0	2	3	2
B	10	12	7	8	11	15	17	9	8	11
C	0	0	1	0	2	0	0	0	0	0

・ 文献の被引用数の使用

文献の被引用数は文献の重要度を測る指標として研究者によく使われている [20][21]。なぜなら、重要な論文はたくさんの人によって引用されるので、被引用数が多くなると考えられる。本ツールでも、文献の被引用数は文献の重要度を測るもう一つの指標として使用する。文献の被引用数がより高ければ、より多くの研究者がこの文献を知っていると考えられる。従って、このような文献は有名な文献であると考ええる。

・ 文献のタイトルの使用

文献のタイトルを文献を選ぶための補助的手段として使用する。幾つかの文献の重要度がほぼ同じの場合、研究者は文献のタイトルを見ることによって、興味を持っている文献を選んで読むことができる。

4.2 なぜ可視化手法を利用するのか

B Shneiderman ら [22] は、情報可視化について以下のように定義していた。

INFORMATION VISUALIZATION: The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.

この定義によって、情報可視化は大量抽象的なデータを計算機上にインタラクティブな視覚的表現で表示することで、これらのデータが持つ情報の理解を促進することができるということが分かった。言い換えると、可視化手法は直感性や、分かりやすさ、一目で把握できることなどの利点を持っている。これは、本研究で可視化手法を利用する理由である。

第5章 望まれる機能に対応した視覚表現の使用

本章では、3.2 節で述べた提案手法を実現するために、三つの機能にそれぞれ対応した視覚表現の使用について述べる。まず、5.1 節では、本研究で用いたリスト、折れ線グラフと、散布図について詳しく述べる。5.2 節では、これらの視覚表現を連携させるリンキングとブラッシング技術の利用について述べる。

5.1 要望に応える機能に対応した視覚表現

我々は、3.1 節で述べた要望に応える三つの機能を実現するために、それぞれにリスト、折れ線グラフと散布図を利用した。

5.1.1 研究分野の全体像を提示するための表現

我々は、一般的に使用されるリストを利用することで、研究者が着目している研究分野の全体像を俯瞰できるようにする。

リストは、いくつかの情報をエンコードすることができる。最も知られているものは、項目の順番である。通常は、すべてのアイテムは何らかの観点でランキングして表示される。これにより、最も重要な項目はリストの上のところに表示され、ユーザは重要な項目を素早く見つけることができる。

リストの用途はランキングだけに限らない。項目の色を使用することで、役立つ情報を表示できる。本ツールでは、リストを用いることで、研究分野の全体像を俯瞰できる。さらに、リストにおける項目の並び順と色によって、全体像における個々の研究テーマに関する特徴も取得できる。図 5.1 では、本研究で設計したリスト表現を示す。



図 5.1: リストによる研究分野の全体像の表示

5.1.2 研究テーマの特徴を俯瞰できるようにするための表現

研究分野の全体像を概観する上で、主流の研究テーマや今後発展しそうな研究テーマを探したいと考えられる。さらに、このような主要な研究テーマは論文検索のための手掛かりとして使うことができる。我々は、研究テーマのはやりすたりを提示する。このため、研究テーマを表す著者キーワードの頻度推移に着目し、それを視覚的に表す。

本ツールでは、著者キーワードの頻度推移を表現するために、折れ線グラフを用いた。折れ線グラフは、縦軸に数値、横軸に経過時間を取り、経過時間ごとの値を線（折れ線）で結んだグラフ表現のことである（図 5.2, 5.3 と 5.4 参照）。この視覚表現は、時間の経過に伴って変化する量の推移を見るのに利用される。図 5.2、図 5.3 と図 5.4 は一つの研究テーマに対する三つの場合の頻度推移の例を想定して挙げた。図 5.2 で、研究テーマを表す「graph」という著者キーワードの頻度推移によって、この研究テーマが主流の研究テーマであることが分かる。同じように、図 5.3 では、この研究テーマが発展しそうな研究テーマであることや、

図 5.4 で、この研究テーマがあまり注目されていない研究テーマであることが分かる。また、図 5.2、5.3 と 5.4 のように、折れ線の色も役に立つ情報を表示できる。

本ツールでは、個々の研究テーマの特徴を比較しやすくするために、それぞれの研究テーマに対する折れ線は一つのビューで全部表示され、著者キーワードと折れ線の色を対応させることで、複数の著者キーワードの頻度推移を表示する。図 5.5 は、本研究で設計した折れ線グラフ表現である。

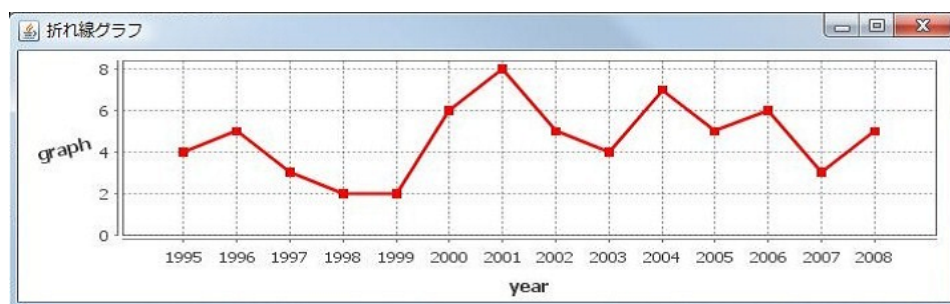


図 5.2: 主流の研究テーマの頻度推移の表示

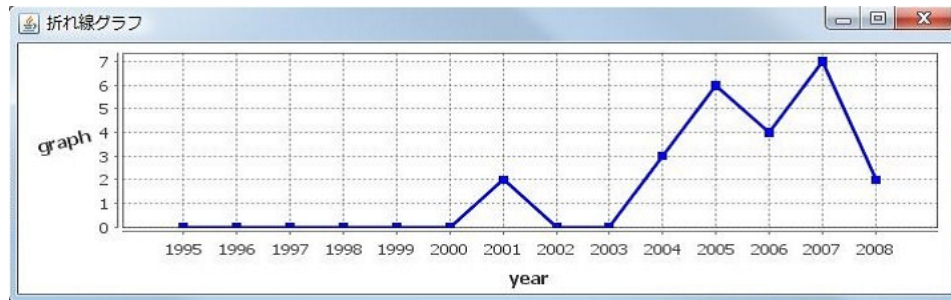


図 5.3: 発展しそうな研究テーマの頻度推移の表示

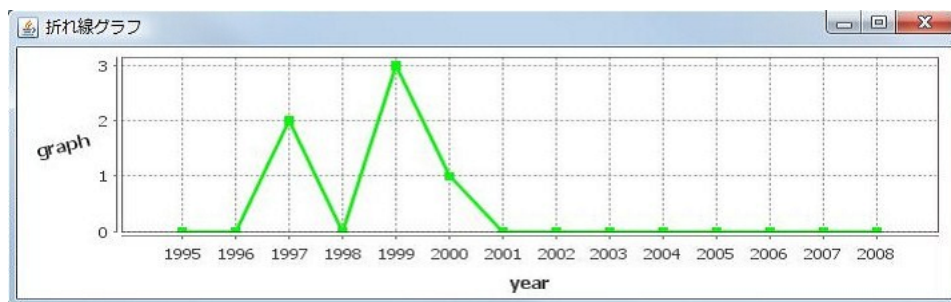


図 5.4: 注目されていない研究テーマの頻度推移の表示

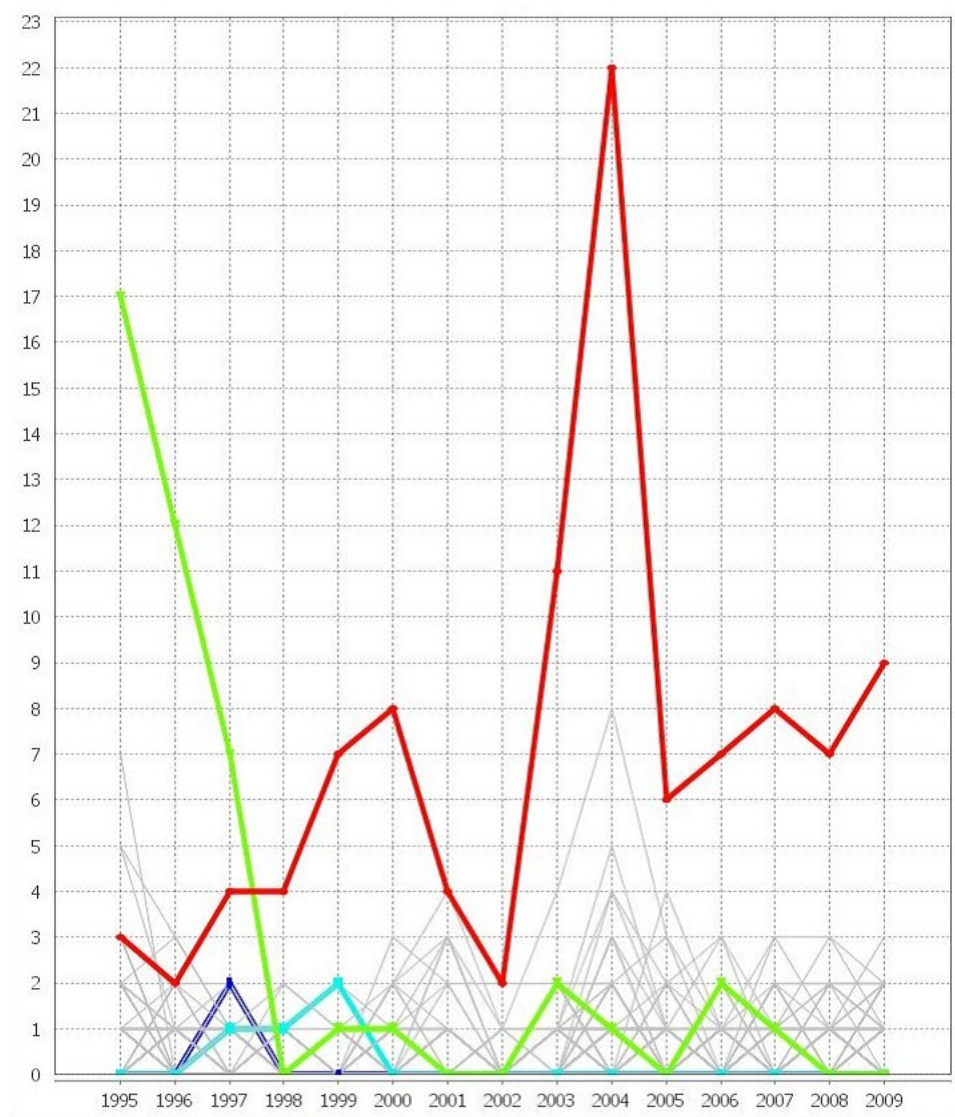


図 5.5: 研究テーマの特徴を俯瞰する折れ線グラフ

5.1.3 重要な論文を探しやすくするための表現

主流の研究テーマなど主要な研究テーマを見つけた後、それに関連する重要な論文を調べることが必要である。我々は以下の二つの要点に基づいて論文を表示することで、重要な論文を探しやすくなると考える。

1. 重要な論文を素早く見つけるために、論文全体を一つの画面で表示する。

2. 多くの重要な論文を簡単に見つけられるように、一つの観点ではなく、二つの観点で論文全体をソートして表示する。

この二つの要点を満たすために、本ツールでは散布図を用いた。

散布図は2種類の項目を縦軸と横軸にとり、データをプロットすることでその分布を視覚的に表現する。散布図において、データはそれぞれの座標に対応した位置にプロットされるので、縦軸と横軸の範囲を適当に設定することで、大量のデータを一つの画面で全部表示することができる。これにより、「要点1」を満たすことができる。そして、散布図において、縦軸と横軸はそれぞれに一つの観点を付けることによって、「要点2」も満たすことができる。これらの2つの要点を満たすことができる表現として我々は散布図を利用した。図5.6は、本研究で設計した散布図表現である。

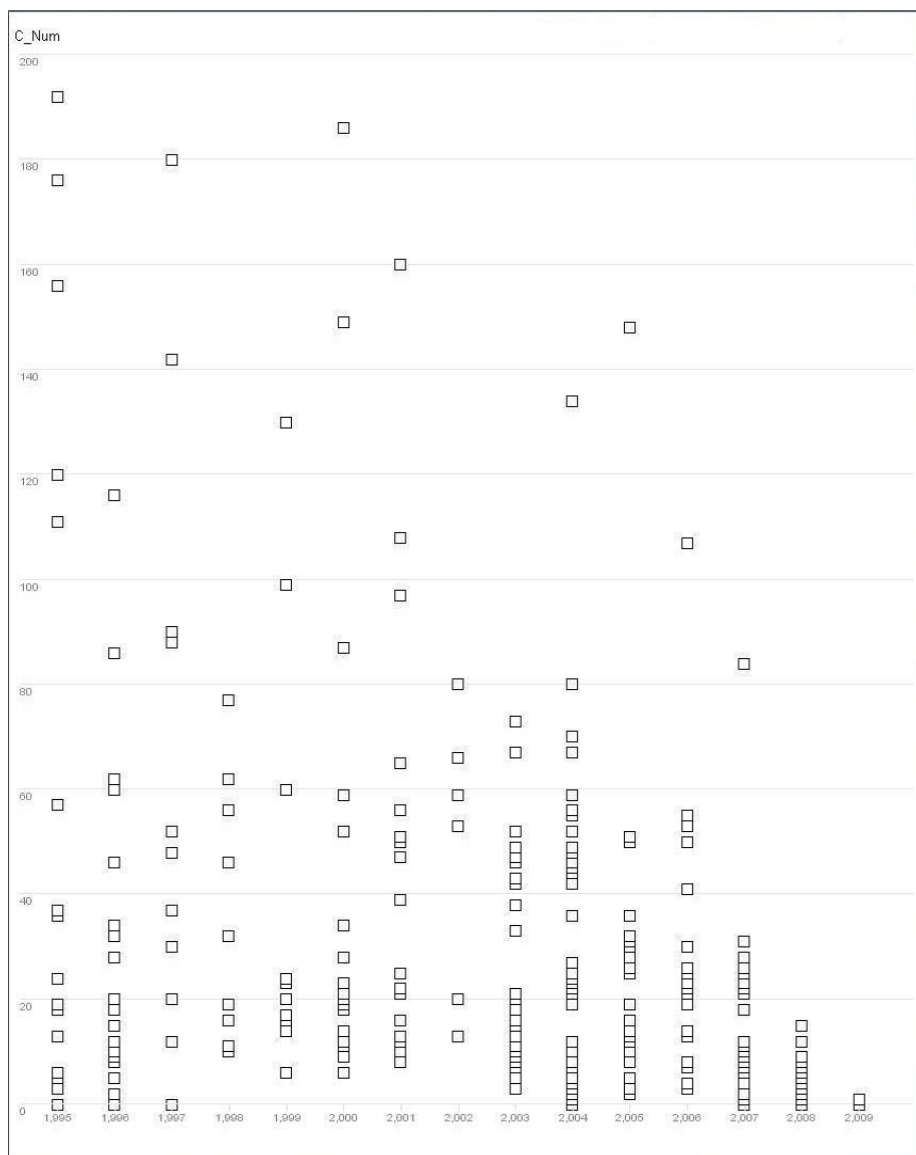


図 5.6: 論文を調べる散布図

5.2 リンキングとブラッシング技術の利用

本研究では、これらの視覚表現を連携させるために、リンクングとブラッシングの技術を利用した。

リンクングは、データを異なるビューで人間が分かる見せ方で提示することを目指す。この技術は、同じデータに関する異なる属性は異なるビューで同時に反映される。これにより、

異なるビューからデータの側面を把握することができる。本研究で、リンキングを用いることによって、リスト、折れ線グラフと散布図を連携して、一つのツールに統合することが可能になる。

ブラッシング[23]は、リンキング技術を用いる上で、一つのビューにおけるアイテムを選択すると、異なるビューにおけるそれに対応したアイテムを同時にハイライトされるという技術である。この技術を使用することによって、著者キーワードからそれに関連する論文の調査をスムーズに進めることができる。

第6章 効率的な文献調査を支援する視覚的ツール

本章では、提案手法を元に開発した、効率的な文献調査を支援する視覚的ツールについて述べる。まず、本研究の視覚的文献調査ツールの概要を説明する。次に、インタフェースの設計について詳しく述べる。その後、ツールの実装とシステムの構成を紹介する。最後に、本ツールの使用例を一つ挙げることによって、本ツールの使用方法について詳しく説明する。

6.1 視覚的文献調査ツールの概要

提案手法に基づいて、本研究の対象ユーザである研究の初心者に向けて、効率的な文献調査を可能にする視覚的文献調査ツールを開発した。図 6.1 に、本研究で開発した文献調査ツールの概観を示す。このツールは主に三つのビューで構成される。画面中央は研究者が着目している研究分野の全体像を表示するリストビューである。画面左側の折れ線グラフは、全体像における個々の研究テーマを表す著者キーワードの頻度推移を示す。画面右側の散布図は、研究テーマに関連する論文全体を一つの画面で表す。研究者がキーワードを入力するための検索ボックスはインタフェースの右上に配置される。

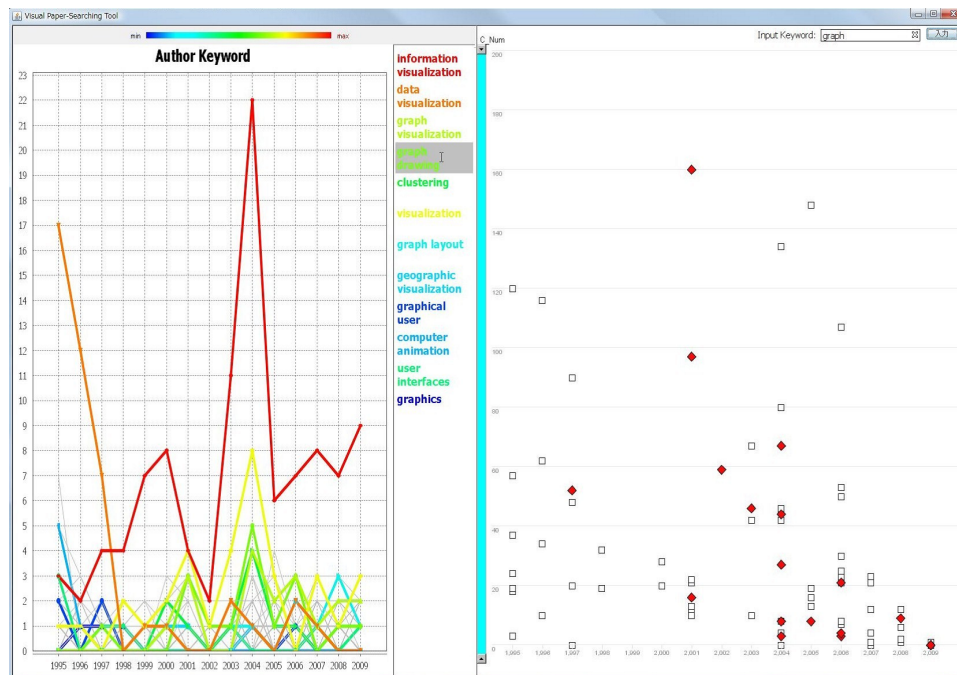


図 6.1: ツールの概観

6.2 インタフェースの設計

6.2.1 全体像の提示法

リストは、研究者が着目している研究分野の全体像を提示するために、研究テーマを表す著者キーワードを全部表示する。図 6.2 のように、一つのリストの項目は一つの著者キーワードを表す。例えば、「clustering」や、「graph visualization」、「hierarchy」などの著者キーワードがある。これにより、研究分野における著者キーワードが表す様々な研究テーマを簡単に見つけることができる。また、一つの著者キーワードにマウスカーソルをオーバーすることで、このキーワードに関連する論文を調べることができる。

そして、個々の著者キーワードの並び順と色の設計を説明する。この前に、二つの用語の定義を説明する。一つは局所的出現頻度、もう一つは全体的出現頻度である。

• 局所的出現頻度

著者キーワードは特定の研究分野における年毎の出現回数である。例えば、グラフという研究分野に着目すると、「graph」というキーワードを使用して文献検索をして、「graph drawing」という著者キーワードを見つけられる。この場合、「graph drawing」という著者キーワードの局所的出現頻度は、「graph」という研究分野における年毎の使われた回数である。

• 全体的出現頻度

著者キーワードは特定の研究分野を限定しない全部の著者キーワード集合における年毎の出現回数である。「graph drawing」という著者キーワードの全体的出現頻度は、「graph」という研究分野に限らない、全部の著者キーワード集合における年毎の使われた回数である。

本ツールでは、色相を利用して、著者キーワードの全体的出現頻度を表す。図 6.3 に示したように、赤色は全体的出現頻度が一番高い著者キーワードを表し、一方、青色は全体的出現頻度が一番低い著者キーワードを表す。例えば、図 6.2 を見ると、「information visualization」という著者キーワードは赤色で表示されているので、この著者キーワードの全体的出現頻度は高いと分かる。一方、「graph navigation」という著者キーワードは青色で表わされているので、この著者キーワードの全体的出現頻度は低いと分かる。これにより、対象とする研究分野における研究テーマが、全部の著者キーワードにおいてどのような位置に置かれているということが把握できる。

しかし、全体的に良く使われている著者キーワードは、局所的にあまり使われていないことがよくある。逆に、全体的に重要ではない著者キーワードは、局所的に重要な著者キーワードになることも良くある。このような情報を提示すると、研究者は対象とする研究分野におけるよく注目されている研究テーマなど、主要な研究テーマは素早く把握できる。従って、本ツールでは、個々の著者キーワードを局所的出現頻度の高さによって、リストの上から下まで並べて表示する。これにより、対象とする研究分野におけるよく使われている著者キーワードはリストの上部に位置するので素早く見つけることができる。

一般的に、対象とする研究分野において、1 回だけ出現する著者キーワードは多くある。このような著者キーワードに関連する論文は一編だけであるので、本研究ではこのような著者キーワードは重要なキーワードではないと考えた。このことから、本ツールでは、局所的出現回数が 2 回以上の著者キーワードだけに着目して表示した。

さらに、著者キーワードの総出現回数だけではなく、年による出現回数の変化も表示することによって、ずっと注目されているということや、近年から流行し始めたということなど、研究テーマに関する特徴を把握できる。従って、我々は著者キーワードの頻度推移に着目し、それを折れ線グラフで表示した。折れ線グラフの設計は次の節で詳しく説明する。



clustering
graph visualization
information visualization
graph drawing
information retrieval
visualization
hierarchy
metrics
small world graphs
hierarchical clustering
graph navigation

図 6.2: 著者キーワード全体を表示するリスト



図 6.3: 全体的出現頻度に対応する色相の使用

6.2.2 研究テーマの特徴の俯瞰法

我々は、研究テーマに関する詳しい特徴を表示するために、研究テーマを表した著者キーワードの年ごとの頻度推移に着目し、それを折れ線グラフで表した。これにより、個々の著者キーワードに関する発展履歴を把握できる。図 6.4 は本ツールで設計した折れ線グラフである。x 軸は年を表し、y 軸は著者キーワードの出現回数を表す。そして、一つの折れ線は一つの著者キーワードを表す。さらに、個々の著者キーワードの特徴を比較しやすくするために、著者キーワード全体が一つの折れ線グラフで表示される。初期画面では、全部の著者キーワードが薄い灰色で表示される。この状態から、対象とする研究分野を選ぶと、当該領域における著者キーワードが全体的出現頻度によって色付けされて表わされる。これにより、着目している研究分野における研究テーマが全体的にどのような位置にいるかということが一目で把握できる。

そして、折れ線の色はリストで使用した色と同じように、著者キーワードの全体的出現頻度の高さを示す。更に、折れ線で表すキーワードは、リストで表示した同じキーワードと同じ色で統一される。これにより、リストで着目した著者キーワードと対応する折れ線を素早く見つけることができる。

図 6.4 のような折れ線グラフを見ると、幾つかの有益情報を発見できる。例えば、赤い折れ線で表した著者キーワードは、ずっと注目されている主流の著者キーワードであると分かる。青い折れ線で表した著者キーワードは、あまり使われていないが、最近から流行し始めた発展しそうな研究テーマであるということも分かる。

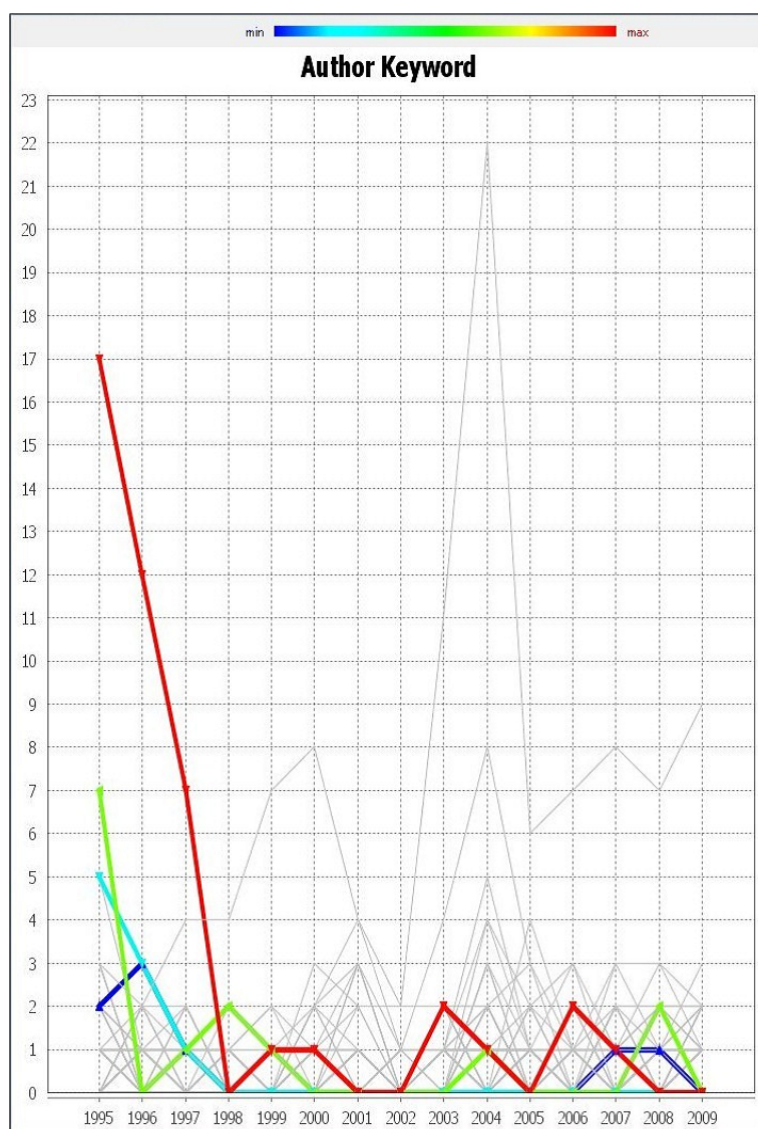


図 6.4: 著者キーワードの頻度推移を表す折れ線グラフ

6.2.3 重要な論文の探索法

我々は、著者キーワードに関連する重要な論文を探しやすくするために、散布図を利用することで、全部の論文を一つの画面で表示した。図 6.5 は本ツールで設計した散布図である。図 6.5 に示したように、一つの長方形は一つの論文を表す。x 軸は論文の発表年を表し、y 軸は論文の被引用数を表す。4.1 節で述べたように、発表年と被引用数は論文の重要度を測る指標としてよく使われている。個々の論文は、発表年と被引用数に基づいてそれぞれの座標に

対応した位置にプロットされる。このことから、論文全体を二つの観点（被引用数と発表年）でソートして表示することができた。これにより、多くの重要な論文は一目で見つけることができる。例えば、「1」が付いた一番有名な論文と「2」が付いた最新の論文は一目で見つけることができる。さらに、任意発表年におけるより有名な論文も素早く見つけることができる。この散布図を利用することによって、多くの重要な論文が一つの画面で同時に表示されるので、著者キーワードに関連する重要な論文を探しやすくなった。

本ツールで設計した散布図の初期画面では、全部の論文を同時に表示する。この状態から、対象とする研究分野を選ぶと、当該分野における論文だけが表示される。そして、一つの長方形にマウスカーソルをオーバーすることで、この長方形で表す論文のタイトルを調べることができる。二つの論文の被引用数が近い時、この二つの論文は近い位置にプロットされる。この場合に対して、左側の青緑色のバーを利用することで、論文間の距離を調整することができる。

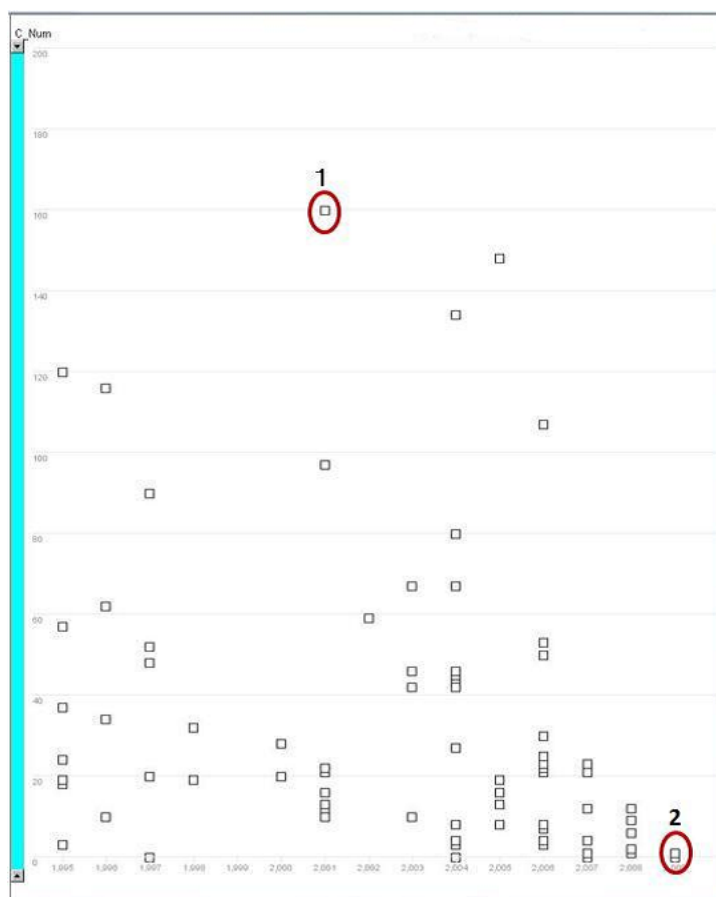


図 6.5: 著者キーワードに関連する論文を表示する散布図

6.2.4 操作体系

本節では、本ツールで実現したインタラクティブな操作を説明する。

一つは著者キーワードから論文への検索をスムーズに行うために使用する。図 6.6 に示したように、リストで表示した著者キーワードにマウスカーソルをオーバーすることで、この著者キーワードに関連する論文がすぐに散布図で異なる形で色付けて表示される。

もう一つは論文に関する詳しい情報（論文のタイトル）を調べるために使用する。図 6.7 のように、論文を表す長方形にマウスカーソルをオーバーすることで、論文のタイトルは長方形の下にすぐに表示され、同時にツールの一番下のところにも青色で表示される。



図 6.6: 論文検索の操作

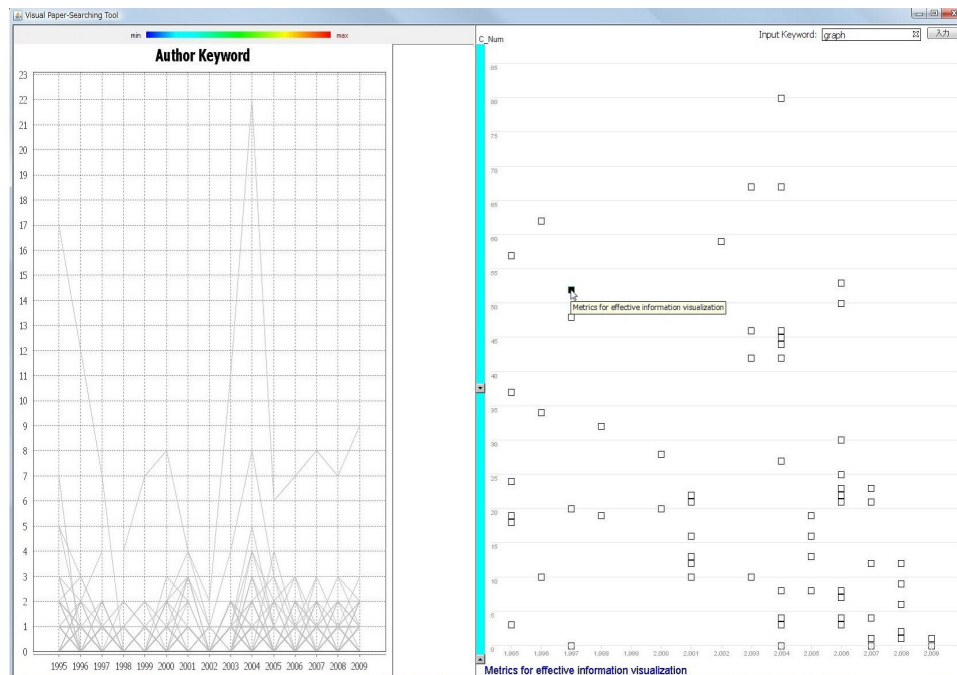


図 6.7: タイトルを探す操作

6.3 ツールの実装

本ツールは、JavaTM6.0(JavaTMPlatform, Standard Edition 6 Development Kit¹)を用いて実装した。折れ線グラフはJFreeChart²ライブラリを用いて実装し、散布図はprefuse³ライブラリを用いて実装した。JFreeChartとprefuseはJavaTMPlatformに基づいて書かれたJavaライブラリである[24][25]。本ツールの入力データはテキストで保存されたデータである。

6.4 システムの構成

図 6.8 は、本研究で開発した文献調査ツールのシステム構成を示す。

文献の著者キーワードや、発表年、被引用数、タイトルなどの論文データ全体は論文データデータベースに保存されている。このデータベースから、本研究で着目した論文の著者キーワード、発表年、被引用数とタイトルを抽出し、三つの視覚表現に対応したテキストデータを作った。この後、折れ線グラフ、リストと散布図はそれぞれに対応したテキストからデータを取得して処理する。最後に、インタフェースの部分から、三つの視覚表現が表した情報

¹<http://java.sun.com/>

²<http://sourceforge.net/projects/jfreechart/>

³<http://prefuse.org/download/>

を得ることができる。

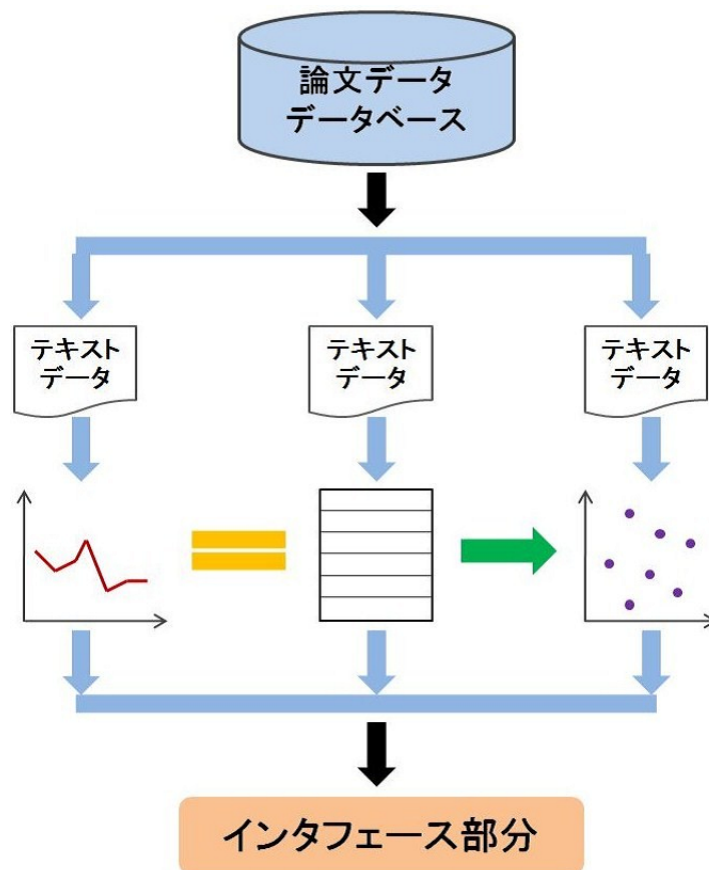


図 6.8: 文献調査ツールのシステム構成

6.5 ツールの利用例

本節では、本研究で開発した文献調査ツールに関する一つの具体的使用例を挙げる。INFOVIS⁴ の1995～2009の15年間の論文データを使用した。

我々は、可視化研究を行う研究室に新しく配属された4年生を使用ユーザと想定した。彼は、階層的なデータに関する可視化研究分野に興味を持っているので、それに関する文献調査を行う場面を想定した。しかし、彼は初めてこの分野に触れるので、どのような研究テーマがあるかをまだ把握していない。従って、彼は「hierarchy」という研究分野を代表とする検索キーワードを使用して文献検索を始める。

図6.9は文献調査ツールの初期画面を示した。左側の折れ線グラフはINFOVISの15年間の全部の著者キーワードの頻度推移を薄色線で表示している。リストには何も表示していない。右側の散布図は15年間の論文全体を表示している。この後、ユーザは右上のサーチバーのところで「hierarchy」という検索キーワードを入力して隣の「入力」ボタンを押す。すると、ツールは図6.10のようになる。この時、リストには「hierarchy」というキーワードを含む論文の著者キーワードを全部表示した。これらの著者キーワードは局所的出現頻度の高さによって上から下まで並び、全体的出現頻度の高さによって赤色から青色までの色で表示した。左側の折れ線グラフはこれらの著者キーワードの頻度推移を示し、右側の散布図は「hierarchy」に関連する論文全体を同時に表示した。この時、リストには、「focus+context」や、「animation」、「graphs」などのキーワードが表示される。これらを見ることで、様々な研究テーマを簡単に見つけることができる。さらに、著者キーワードの並び順とそれに対応した折れ線を見ることで、「focus+context」や「treemap」など主流の著者キーワードを効率的に見つけることができる。主流の著者キーワードを見つけた後、それに関連する重要な論文を調べる。図6.10は、「focus+context」に関連する重要な論文を調べることを示した。「focus+context」にマウスオーバーすることで、それに関連する論文が散布図上に赤色で全部表示される。これらの論文において、緑色の円で表した一番有名な論文と最新の論文をすばやく見つけることができる。また、図6.7に示したように論文を表した長方形にマウスオーバーにすることで、論文のタイトルをすぐに見ることができる。

図6.11は「treemap」に関連する重要な論文を調べることを示した。図6.10に示したような同じ操作をすることで、緑色の円で表した一番有名な論文と、2005年に発表されたより有名な論文を素早く見つけることができる。また、同じ操作で次の研究テーマに関連する重要な論文を調べることができる。

本ツールを使用することによって、着目している研究分野における様々な研究テーマを簡単に見つけることができ、更に、これらの研究テーマにおける主流の研究テーマも効率的に見つけることができる。また、研究テーマに関連する重要な論文も素早く見つけることができる。

⁴可視化分野の国際会議

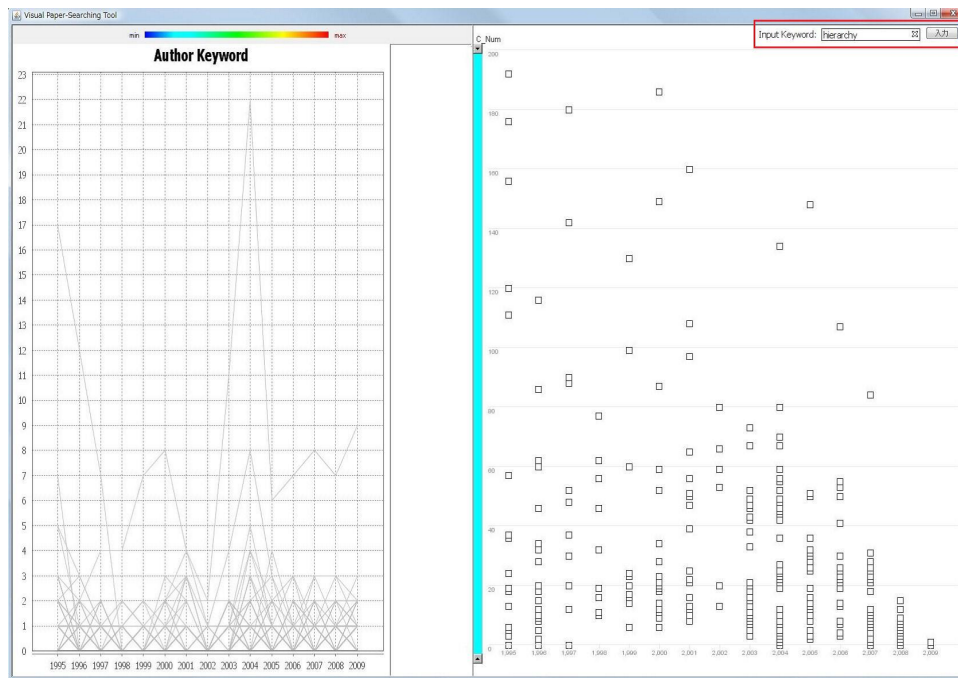


図 6.9: 文献調査ツールの初期画面

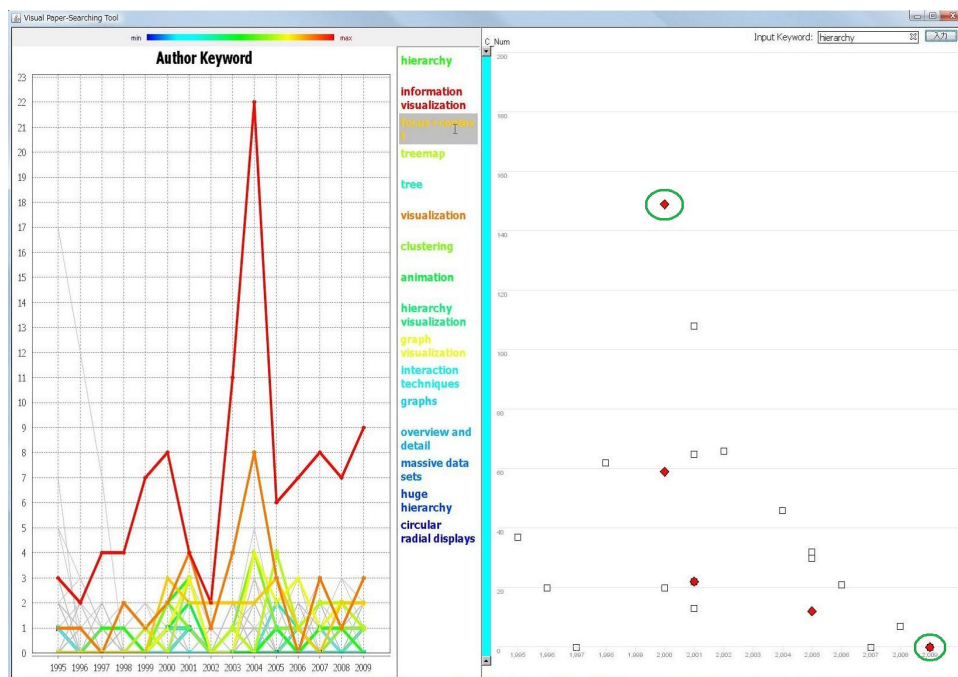


図 6.10: 「focus+context」に関連する重要な論文

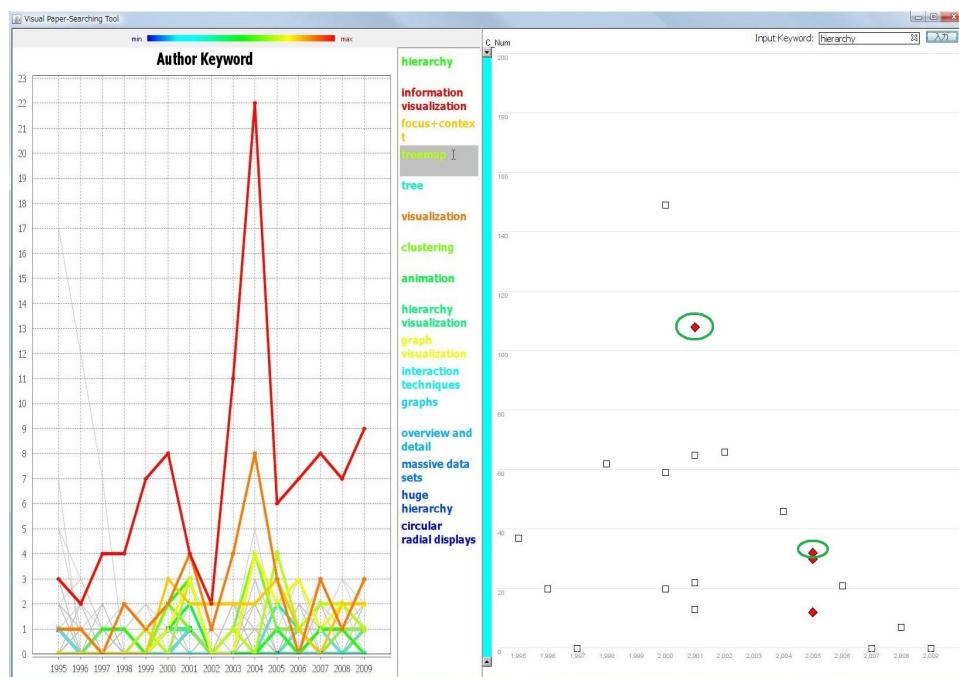


図 6.11: 「treemap」に関連する重要な論文

第7章 評価実験と考察

本章では、本研究で行った評価実験とその結果に対する考察を述べる。

7.1 評価実験

7.1.1 実験の目的

今回の評価実験では、主に以下のことを目的とする。

- 「本ツールが、「研究テーマ」という観点で文献調査を行う上で有用である」という事柄の検証
- 「ツールにおける三つの機能にそれぞれの有用性」についての調査

7.1.2 実験の被験者

コンピュータサイエンスを専攻する学部4年生3名及び修士1年生3名、合計6人を対象に実験を行った。被験者はいずれもこれから新しい研究分野に参入しようとしており、自分のやりたい研究テーマを決める必要がある学生である。被験者は興味を持っている研究分野はあるが、それに関する知識が浅い。

7.1.3 実験の方法

実験用データは第6章のツールの使用例に用いたデータ（INFOVISの15年の論文データ）と同じである。実験が始まる前に、一つの具体的な例を挙げることによって、以下の内容について、被験者に10分程度で説明した。

- 本ツールにおける三つの機能
- それぞれの機能に用いた論文データ及び視覚表現で表示した情報
- ツールのインタラクティブな操作

この後、以下のような二つのタスクを一定の時間内で行った。個々のタスクは5分半程度で行った。全二つのタスクにおいて平均して10分ぐらいかかった。

- タスク 1: 階層的なデータという研究分野に関する文献調査を行う。具体的には、この研究分野における主流の研究テーマとそれに関連する重要な論文を見つける。「hierarchy」という研究分野を代表とする検索キーワードを使用する。
- タスク 2: 被験者は自分が興味を持っている研究分野に関する文献調査を行う。タスク 1 と同じように、研究分野における主流の研究テーマとそれに関連する重要な論文を見つける。被験者は自由に検索キーワードを使う。

二つのタスク終了後、以下の項目の達成度についてアンケート調査を行った。回答はそれぞれ 5 段階評価（5：とても高い、…、1：とても低い）と自由記述によって行ってもらった [26]。

- タスク 1 に関する項目：
 - 項目 (1): 「hierarchy」が表す研究分野における様々な研究テーマを簡単に見つけたか？
 - 項目 (2): これらの研究テーマにおける主流の研究テーマを二つ以上見つけたか？
 - 項目 (3): 主流の研究テーマに関連する二つ以上の重要な論文を素早く見つけたか？
- タスク 2 に関する項目：
 - 項目 (4): 着目している研究分野において、自分が知っている研究テーマ以外の主流の研究テーマを見つけたか？
 - 項目 (5): 知っている研究テーマに関連するすでに読んだ論文以外の重要な論文を見つけたか？

また、本ツール及びそれぞれの機能の有用性に対する評価もアンケートで行った。アンケートの項目は以下の四つである。評価は 5 段階評価（5：とても有用、…、1：全く有用でない）によって行ってもらった [27]。

- 項目 (6): 本ツールは文献調査に対する有用であるか？
- 項目 (7): リストは様々な研究テーマを見つけることに対する有用であるか？
- 項目 (8): 折れ線グラフは主流の研究テーマを見つけることに対する有用であるか？
- 項目 (9): 散布図は重要な論文を探すことに対する有用であるか？

7.1.4 実験の結果

表 7.1、表 7.2 は各項目に対するそれぞれの被験者（A～F）の回答をまとめたものである。

表 7.1: 各被験者のアンケートへの回答

評価項目	A	B	C	D	E	F	平均
項目 (1) の達成度	5	5	4	4	4	5	4.50
項目 (2) の達成度	4	5	5	4	3	4	4.17
項目 (3) の達成度	5	5	5	4	4	5	4.67
項目 (4) の達成度	5	4	4	5	3	4	4.17
項目 (5) の達成度	5	5	4	4	3	3	4.00

表 7.2: 各機能に対する評価

評価項目	A	B	C	D	E	F	平均
項目 (6) の評価	5	4	5	4	4	5	4.50
項目 (7) の評価	5	4	5	4	5	5	4.67
項目 (8) の評価	4	4	4	3	3	4	3.67
項目 (9) の評価	5	5	5	4	4	5	4.67

7.2 考察

表 7.1 のように、タスク 1 とタスク 2 に関する項目の達成度に関する評価を得た。五つの項目の達成度はそれぞれ平均 4.0 以上と良い評価を得た。特に、「様々な研究テーマを簡単に見つけたか」と「研究テーマに関する二つ以上の重要な論文を素早く見つけたか」に関する達成度についてはそれぞれ平均 4.5、平均 4.67 と高い評価を得た。これにより、本ツールは、様々な研究テーマを把握することと、重要な論文を探索することに非常に役に立つと考えられる。また、「主流の研究テーマを二つ以上を見つけたか」に関する達成度はより低い平均を得た。この項目に関する評価は、六つの中に、「4」は三つあり、「3」が一つあった。これに関する原因は、研究分野における研究テーマが多くなると、近い色で表した研究テーマも多くなった。これにより、折れ線グラフにおける近い色で表示した研究テーマの特徴を区別しにくくなるためであると考えた。また、研究テーマが多くなると、折れ線グラフにおける異なる研究テーマの折れ線の間に重なる可能性が高くなる。これは、表 7.2 の項目 (8)（「折れ線グラフは主流の研究テーマを見つけることに対する有用であるか」）に関する評価が低い理由と考えた。しかし、タスク 2 に関する項目 (4)（「研究者が知っている研究テーマ以外の主流の研究テーマを見つけたか」）に関する評価の中に、二つの「5」を得た。この二人の被験者はタスク 2 を行う時、着目している研究分野における研究テーマが少ない（10 個以下）の

で、主流の研究テーマを見つけやすかったと考えた。これにより、本ツールの折れ線グラフは、研究テーマの数が少ない場合に対して有用であると考えた。研究テーマの数が多い場合には、折れ線グラフの上にインタラクション手法を利用することで、研究テーマの特徴を俯瞰しにくいという問題を解決できると考える。

表 7.2 は本ツールおよび三つの機能の有用性に関する評価を示した。「本ツールは文献調査に対する有用であるか」（項目 (6)）に対しては、いずれも 4 以上と高い評価を得た。このことから、本研究で開発したツールは文献調査に対する有用であると言える。また、三つの機能に対して、「全体像を提示する機能」と「重要な論文を見つけやすくする機能」の両方は高い評価を得た。つまり、本ツールのリストを利用することで、研究分野における様々な研究テーマを簡単に見つけることができると言える。また、散布図を利用することで、重要な論文を素早く見つけることができると言える。

第8章 関連研究

本章では、本研究に関連する研究について紹介する。8.1 節では視覚的情報検索に関する研究について述べる。8.2 節で研究サーベイの支援に関する研究について述べる。

8.1 視覚的情報検索に関する研究

視覚的情報検索に関する研究は、今まで多く行われている。

Fernando ら [3] は、大量な検索結果において、関係が近い結果を見つけやすくする PEx-WEB システムを開発した。著者は、ネットワーク図を用いることで、検索結果を一つの画面上に表示した。検索結果をそれが含むトピックによってクラスタリングし、同じトピックに関する結果は同じ色で表示され、近いところに置かれる。

Edward ら [4] の研究は論文を検索することを支援する ResultMaps システムを開発した。この研究も、検索結果をそれが含む研究トピックによってクラスタリングした。そして、これらの検索結果は空間効率が良い階層的表現「treemap」を利用して表示する。

Greg ら [5] は、大量の個人データのサーチとブラウズを行いやすいために、テキストベースの方法の代わりに、FacetMap という、グラフィカルインタフェースを開発した。このインタフェースにおいて、大量のデータがその特徴によってフィルターされる。また、気泡のような視覚表現でそれらのフィルターを表わした。

Justin ら [8] は、Social bookmarking ウェブサイトにおける情報を検索するための視覚的情報検索システムを開発した。著者は、social bookmarking ウェブサイトにおける情報が特定なタグ、カテゴリーとタイトルを付けられたという共同な特徴を利用した。著者は、ネットワーク図を用いることで、キーワードクエリーに関する検索結果を表した。カテゴリーとタイトルにおけるクエリーを含む数の多さによって色相を利用してノードに色を付けた。また、ノードの大きさで検索結果とクエリーの内容的な関係を表した。より小さいノードで表した結果はクエリーに関するより明確な結果である。

ここまで紹介した研究は、全部検索結果に着目し、それを視覚的表現で表示した研究である。これから、検索キーワードに着目し、それを視覚的表現で表した研究を紹介する。

レボウィッツら [1] は、研究者が情報要求に曖昧性が生じるため、抽象度の高いキーワードを用いて効率的な文献調査を行えないという問題点を解決するために、著者と著述文献キーワードの関連性を利用して、各著者の研究内容の傾向を読み取り、それを研究テーマとして解釈し、研究領域をブラウジングするシステムを開発した。

Orland ら [10] は、ユーザが曖昧な検索キーワードを使う問題点を解決するために、検索キー

ワードを洗練して、洗練されたキーワードを自動的に提示する視覚的情報検索システムを開発した。著者は、ユーザが使用した検索キーワードに関するトップ 100 の検索結果のタイトルと断片における全部のタームの出現頻度を計算し、そして、それらのタームは出現頻度の高さによって上から下まで並べられた。また、これらのタームをクリックすることで、検索結果は再分類されて表示される。

また、検索結果と検索キーワード両方ともに着目し、それを視覚的に表示した研究も行われている。

Marian ら [14] は、RSS feeds からのオンライン情報の三つの特徴（時間、空間、話題）に対して、それぞれの検索クエリーをインタラクティブに提示できる視覚的ウィジェットを設計し、これらの三つのウィジェットを組み合わせる RSS feeds からのオンライン情報を検索するための視覚的インタフェースを開発した。

Orland ら [11] は、幾つかのクエリーを併せて検索して出てきた検索結果を比べて関連結果を見つけるための HotMap システムを開発した。一つの検索結果に対して、個々のクエリーがそれにおける出現頻度を計算し、出現頻度の高さによって色相をクエリーに付いた。更に、色を付いたクエリーは全部検索結果の前に表れて検索結果の特徴を表示した。これによって、関係が近い検索結果は簡単に見つけることができる。

8.2 研究サーベイの支援に関する研究

今まで、研究サーベイを支援する研究も多く行われている。例えば、8.1 節で紹介した Edward らとレボウィッツらの研究は、効率的な文献検索を行うことを支援する研究である。このような文献検索を支援する研究だけではなく、論文データ間の関係を分析することで、研究分野における研究トピックの発展動向や、文献著者間の共著関係、文献間の共引用関係など有益情報を発見することを支援する研究も多々ある。これから、代表として幾つかの論文を紹介する。

Ryutaro ら [28] は、最近の研究動向を把握するために、author-topic モデルで確率により著者、トピックと論文の三つの情報間の関係を作り、更に、自己組織化マップを利用することでそれらの関係情報を視覚的に表した。

また、Ryutaro ら [29] は、著者間の共著関係に着目し、それにより三つの視覚的コンポーネントを作って組み合わせた研究コミュニティの発見を支援するシステムを開発した。

Chaomei Chen [30] は、科学文献における研究テーマの発展動向を把握するために、CiteSpace という統合環境を開発した。この環境において、著者は、論文間の引用関係に着目し、ネットワーク表現を用いることで、論文と論文間の引用関係を表した。また、論文を表したノードのリングに付いた色の变化によって、年による論文の被引用情報を表示した。これにより、よく引用された科学文献クラスターを発見できる。また、近年よく引用された科学文献によって、研究テーマの発展動向も把握できる。

Sarah ら [2] は、専門家の特定の目的を満たすためではなく、たくさんの研究者が研究サーベイの経験による研究サーベイに対する要求を集めた。この後、これらの要求を満たすため

に、著者は、「著者」、「共著関係」、「論文」及び「共引用関係」の四つの論文データに着目してそれぞれを視覚的に表し、さらに、四つの視覚的表現を組み合わせたインタフェースを開発した。

全部の著者と著者間の共著関係の概観に着目することではなく、Tze-Haw Huang ら [31] は、特定の著者に着目し、この著者と他の著者間の共著関係を視覚的に表した InterRing というインタフェースを開発した。着目した著者は中心とし、中心から一つのリングは一つの特定期間を表した。特定期間における着目した著者との共著関係がある著者は、特定期間に対応したリングの上に表示された。また、異なる著者は異なる色を付け、着目した著者との共著文献の数によってリングでの割合を配分して表示された。これにより、過去の期間において着目した著者の共著関係を簡単に把握できる。

第9章 結論と今後の課題

9.1 結論

本研究では、研究の初心者を対象として、「研究テーマ」という観点での文献調査を行える視覚的文献調査ツールを開発した。最初に、既存の文献調査システムの問題点を挙げた。問題点を解決するための基本的なアイデアとして、文献調査ツールへの三つの要望を整理し、それぞれに応える以下のような三つの機能を提案した。一つ目は、様々な研究テーマを一度に眺められる分野の全体像を提示する機能である。二つ目は、個々の研究テーマのはやりすたりを俯瞰できる機能である。三つ目は、研究テーマに関連する重要な論文を簡単に見つけられる機能である。この三つの機能を実現するために、論文の被引用数と著者キーワードの頻度推移に着目し、さらに、三つの機能のそれぞれに対応した視覚的表現を設計した。また、これらの三つの視覚表現を連携させることによって、視覚的文献調査ツールを開発した。本ツールを利用することによって、研究者は着目している研究分野における様々な研究テーマを簡単に見つけることができ、さらに、これらの研究テーマにおける主流の研究テーマや発展しそうな研究テーマも効率的に見つけることができる。また、研究テーマに関連する重要な論文も素早く見つけることができる。最後に、アンケート形式で行った評価実験により、本研究で開発したツールおよび三つの機能の有用性を示した。特に、「研究分野の全体像を提示する機能」と「重要な論文を見つけやすくする機能」の二つの機能は非常に有用であるという評価を得た。

9.2 今後の課題

本研究での文献調査支援ツールは、研究の初心者を対象として開発された。本ツールは、このような研究者が着目している研究分野における主流の研究テーマや、それに関連する重要な論文を効率的に把握したいなどの要求を満たした。今後は、研究分野の専門家を含んで対象とし、彼らが当該分野の中で自分の研究の位置付けを把握したいなどの要求も満たすために、ツールを改善しようとする。

また、本ツールは論文の被引用数と著者キーワードの頻度推移に着目し、それらを活用して視覚的に表現した。それら以外にも活用できる論文データがある。例えば、文献間の共引用関係、文献と著者の関係などが挙げられる。これらの論文データも視覚的に表現し、さらに、これらの視覚的な表現機能を導入することで、本ツールはさらに多くの情報探索要求を満たすことができると考えられる。

ツールの機能として更なる改善の余地がある。例えば、現在はタイトルだけを提示しているが、より詳細な情報の提示が考えられる。今後は、論文著者や、概要、論文のリンクなどさらに多くの論文情報を提供する必要がある。また、評価実験により、全体像における研究テーマを表すキーワードが多くなると、キーワード間の色相が近くなり、キーワードに対応した頻度推移を見つけにくいという指摘を受けた。この問題を解決するために、インタラクション手法を利用できると考える。

謝辞

本研究を行うにあたり、指導教員である筑波大学コンピュータサイエンス専攻三末和男准教授には、研究の進め方から論文執筆に至るまで、大変貴重なご助言、ご指導を頂きました。また、研究に取り組む姿勢についても親身にご指導いただき、その中で多くの事を学ばせていただきました。この修士論文を書き上げることができたのはもちろん、人間的にも大きく成長したと実感できるのは、ひとえに先生のご指導のおかげであると思っています。この場を借りて厚く御礼を申し上げます。

私に日本で留学・研究する大変貴重なチャンスを与えて頂きました筑波大学コンピュータサイエンス専攻田中二郎教授には、この研究の遂行にあたり終始適切な助言を頂きました。心より感謝いたします。

専攻の研究発表にあたり、筑波大学コンピュータサイエンス専攻志築文太郎講師と高橋伸講師に的確なアドバイスを頂きまして、ここに感謝いたします。

筑波大学大学院コンピュータサイエンス専攻インタラクティブプログラミング研究室のメンバーの方々にも、特に NAIS チームの皆様には、研究活動に留まらず私生活にもわたって大変お世話になりました。この場を借りて厚く御礼申し上げます。

最後に、物心両面にわたり支えてくれた家族、学生生活の中で関わることでできた友人や全ての方々にも心より感謝致します。本当にありがとうございました。

参考文献

- [1] レボウィッツ紀子, 松村敦, 宇陀則彦. 著者とキーワードの関連性に着目した研究領域ブラウジングシステムの試作. *情報知識学会誌*, Vol. 17, No. 2, pp. 707-712, 2007.
- [2] Sarah Faisal, Paul Cairns, and Ann Blandford. Building for Users not for Experts: Designing a Visualization of the Literature Domain. *11th International Conference Information Visualization (IV '07)*, pp. 75-80, 2007.
- [3] Fernando V. Paulovich, Roberto Pinho, Charl P. Botha, Anton Heijs, and Rosane Minghim. PEx-WEB: Content-based Visualization of Web Search Results. *2008 12th International Conference Information Visualisation*, pp. 208-214, 2008.
- [4] Edward Clarkson, Krishna Desai, and James Foley. ResultMaps: Visualization for Search Interfaces. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1057-1064, 2009.
- [5] Greg Smith, Mary Czerwinski, Brian Meyers, Daniel Robbins, George Robertson, and Desney S. Tan. FacetMap: A Scalable Search and Browse Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 797-804, 2006.
- [6] Anton Leuski and James Allan. Lighthouse: Showing the Way to Relevant Information. *2000 IEEE Symposium on Information Visualization (InfoVis 2000)*, pp. 125, 2000.
- [7] Keith Andrews, Vedran Sabol, Wilfried Lackner, Christian Gutl, and Josef Moser. Search Result Visualisation with xFIND. *Second International Workshop on User Interfaces to Data Intensive Systems (UIDIS'01)*, pp. 0050, 2001.
- [8] Justin J. Donaldson, Michael Conover, Benjamin Markines, Heather Roinestad, and Filippo Menczer. Visualizing social links in exploratory search. *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pp. 213-218, 2008.
- [9] Mina AKAISHI, Koichi HORI, and Ken SATOH. Topic Tracer: a Visualization Tool for Quick Reference of Stories Embedded in Document Set. *Tenth International Conference on Information Visualisation (IV'06)*, pp. 101-106, 2006.
- [10] Orland Hoeber, and Xue Dong Yang. Interactive Web Information Retrieval Using WordBars. *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06)*, pp. 875-882, 2006.

- [11] Orland Hoeber, and Xue Dong Yang. The Visual Exploration of Web Search Results Using HotMap. *Tenth International Conference on Information Visualisation (IV'06)*, pp. 157-165, 2006.
- [12] Orland Hoeber, Daniel Schroeder, and Michael Brooks. Real-World User Evaluations of a Visual and Interactive Web Search Interface. *2009 13th International Conference Information Visualisation*, pp. 119-126, 2009.
- [13] 白鳥佳奈, 伊藤貴之. 大規模表形式データ可視化手法「左京と右京」を用いた文献データの可視化. **電子情報通信学会**, 2009.
- [14] Marian Dork, Sheelagh Carpendale, Christopher Collins, and Carey Williamson. VisGets: Co-ordinated Visualizations for Web-based Information Exploration and Discovery. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1205-1212, 2008.
- [15] ACM Portal, 2010. <http://portal.acm.org/portal.cfm>.
- [16] CiNii (Citation Information by NII). National Institute of Informatics, 2010. <http://ci.nii.ac.jp/>.
- [17] Google Scholar, 2010. <http://scholar.google.co.jp/schhp?hl=ja>.
- [18] Scientific literature digital library, 2010. <http://citeseer.ist.psu.edu/>.
- [19] 榎 剛史, 石塚 満. Web を活用した論文サーベイ支援システムに関する研究. **修士論文**, 2006.
- [20] ウィキペディア, 2010. <http://ja.wikipedia.org/wiki/>
- [21] 小野寺夏生, 山崎茂明, 芳鐘冬樹, 岩澤まり子, 辻 慶太, 緑川信之, 天野 晃, 大谷 裕, 城山泰彦, 児玉 閱, 角田裕之, 山崎静香. 論文の被引用数に影響する要因に関する統計学的研究. **第 56 回日本図書館情報学会研究大会発表要綱**, pp. 41-44, 2008.
- [22] SK Card, JD Mackinlay, and B Shneiderman. Readings in information visualization: using vision to think. 1999.
- [23] Leandro Hernandez Almeida, Rodrigo A. de M. Lourenco, Bianchi Serique Meiguins, and Aruanda Simoes G. Meiguins. WebPrisma: An Interactive Web-Based Tool for Exploratory Visualization Using Multiple Coordinated Views. *2009 13th International Conference Information Visualisation*, pp. 645-650, 2009.
- [24] JFreeChart, 2010. <http://www.jfree.org/jfreechart/>.
- [25] the prefuse visualization toolkit, 2010. <http://prefuse.org/>.
- [26] 小 池諭, 三末和男. 書誌情報ネットワークのビュー操作に着目した論文データ分析支援ツール. **修士論文**, 2009.

- [27] 三末和男, 杉山公造. 図的発想支援システム D-ABDUCTOR の開発について. **情報処理学会論文誌**, Vol. 35, No. 9, pp. 1739-1749, 1994.
- [28] Ryutaro Ichise, Setsu Fujita, Taichi Muraki, and Hideaki Takeda. Research Mining using the Relationships among Authors, Topics and Papers. *11th International Conference Information Visualization (IV '07)*, pp. 425-430, 2007.
- [29] Ryutaro Ichise, Hideaki Takeda, and Taichi Muraki. Research Community Mining with Topic Identification. *Tenth International Conference on Information Visualisation (IV'06)*, pp. 276-281, 2006.
- [30] Chaomei Chen. The centrality of pivotal points in the evolution of scientific networks. *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pp. 98 - 105, 2005.
- [31] Tze-Haw Huang and Mao Lin Huang. Visual Analysis of a Co-authorship Network and Its Underlying Structure. *International Conference on Computer Graphics, Imaging and Visualization (CGIV'06)*, pp. 18-23, 2006.