

キーワードの頻度推移と文献の被引用数を視覚化した 文献検索ツール

謝 英双[†] 三末 和男[†] 田中 二郎[†]

[†]筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

1 はじめに

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。その研究分野にはどのような研究テーマがあるのか、どれが主流の研究テーマな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。さらに、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、特定の研究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はくれない。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。

本研究は、新しい研究分野に取り組もうとしている研究者を対象として、より効率的な文献調査を行うことを支援することを目的とする。

2 文献調査ツールへの要望

既存の文献調査システムへの要望を、以下のような三つの項目に整理した。

- 対象とする研究分野の全体像をみたい
研究者はまず最初に、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ような機能が必要である。
- 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。

- 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。

3 要望に応える機能の実現

上記要望に応える機能を実現するために、我々は著者キーワードの頻度推移と文献の被引用数に着目し、さらに、視覚的表現の直感性と分かりやすさなどの利点[1]を利用することとして、それぞれに対応した視覚表現を設計した。また、これら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う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。

4 ツールの開発

4.1 ツールの概観

図1に、本研究で開発した文献調査ツールの概観を示す。画面中央は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の特徴を示す。画面右側は、重要な論文を見つけるための散布図である。また、検索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

A visual literature-survey tool using citation count and keyword frequency

[†]Yingshuang Xie, Kazuo Misue and Jiro Tanaka

[†]Department of Computer Science, University of Tsukuba

ースの右上に配置される。

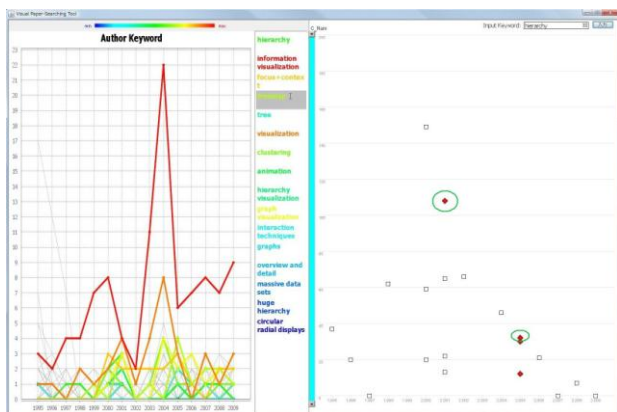


図1: ツールの概観

4.2 インタフェースの設計

研究分野の全体像を提示するために、当該分野における著者キーワード全体を提示したリストを設計した。表示された著者キーワードは着目している研究分野における出現頻度の高さによって上から下に並べる。また、表示された著者キーワードは分野を限定しない全部の著者キーワード集合における出現頻度の順位によって赤色から青色までの色で表示する。

研究テーマの特徴を俯瞰できるように、x軸が発表年を、y軸が著者キーワードの出現頻度を表した折れ線グラフを採用した。一つの著者キーワードを一つの折れ線で表す。折れ線で表したキーワードは、リストで表示した同じキーワードと同じ色で表示される。また、個々の研究テーマ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。

重要な論文を探しやすくするために、x軸が発表年を、y軸が文献の被引用数を表した散布図を採用した。論文の発表年と被引用数は論文の重要性を測る指標として使われる。論文は、その被引用数と発表年によって座標上にプロットされる。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、リンキングとブラッシングによって、一つのツールに統合した。

4.3 ツールの操作

ツールの操作を具体的例を用いて説明する。ツールの初期画面において、研究者はサーチボックスに「hierarchy」というキーワードを入力する。すると、ツールの画面は図1のようにな

る。この時、リストには、「focus+context」や、「animation」、「graph」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」など主流の研究テーマを効率的に見つけることができる。この後、リスト上のキーワードにマウスカーソルを重ねることで、それに関連する論文が散布図上に色付けされて表示される。図1のように、主流の研究テーマ

「treemap」にマウスオーバーすることで、それに関連する一番有名な論文と、2005年に発表されたより有名な論文(円で表した論文)をすばやく見つけることができる。

5 関連研究

検索キーワードと検索結果の両方に着目し、それらを選択するための手掛かりを提供する視覚的情報検索に関する研究は多々ある。たとえば、Marianら[2]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視覚的情報検索インタフェースを開発した。

6 まとめ

研究者が新規参入分野における文献調査を支援する視覚的文献調査ツールを開発した。我々は、キーワードの頻度推移と文献の被引用数に着目し、それらを視覚的に表示することで、三つの文献調査に役に立つ機能を実現した。このような三つの機能を備えたツールによって、研究者が文献調査を行う際の効率を向上させることが可能になった。

参考文献

[1] SK Card, JD Mackinlay, B Shneiderman. Readings in information visualization: using vision to think. 1999.

[2] Marian Dörk, Sheelagh Carpendale, Christopher Collins, Carey Williamson. VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1205-1212, 2008.