

# 引用論文に着目した研究動向把握支援 インタフェースの基礎検討

## Basic Study on Interface to Grasp Research Trends Based on Cited Articles

程岡 晃一<sup>1\*</sup> 大杉 隆文<sup>2</sup> 松下 光範<sup>1</sup>  
Koichi Hodooka<sup>1</sup> Takafimi Osugi<sup>1</sup> Mitsunori Matsushita<sup>1</sup>

<sup>1</sup> 関西大学総合情報学部

Faculty of Informatics, Kansai University

<sup>2</sup> 関西大学大学院総合情報学研究科

Graduate School of Informatics, Kansai University

**Abstract:** This goal of this study is to develop a system that supports grasping trends of a research topic. In conducting research, it is an important activity to grasp research trends by investigating prior researches in that field. Search engines are often used in such cases, however, it is difficult to efficiently gather papers in the intended research field because search by keywords or authors is not sufficient. In order to solve this problem, this paper proposes an interface to support comparing the articles cited by research papers. Differences in the cited articles can be assumed to reflect the primary concerns of the paper. Our conducted user observation suggested that the proposed interface contributes to the understanding of the differences between papers.

### 1 はじめに

近年、様々な分野の論文が執筆され、膨大な数の論文が世の中に存在している [10]. 研究者は研究活動を行う際、膨大な数の論文から過去に行われている研究を把握し、それらを纏めることで考察を行ったり、今までに行われていない新しい研究を行ったりする。そのため、研究者は自身と同じ研究分野の論文を整理し、その分野の研究動向を把握することが重要である。そこで、研究者は過去の研究を把握するために、論文検索のための検索エンジン (e.g., Google Scholar, CiNii) を利用する。その際にキーワードをクエリとして与え論文を検索する。しかし、クエリに用いるキーワードが適切であるとは限らない。そのため、そのクエリを用いた検索で見つかった論文全てが、必ずしも研究者が読みたい論文であるとは限らない。

この問題の解決策として、キーワード検索だけでなく、ある読みたい論文 (以下、起点論文と記す) が引用している論文から、論文を検索するという方法がある。その引用された論文は起点論文に関連した論文である。また、起点論文を執筆した著者 (以下、起点著者

と記す) と同じ著者の論文から読みたい論文を検索するという方法もある。起点著者が執筆した論文が複数存在する場合、起点著者は同じ研究分野に対し、研究を積み重ねていることがある。そのため、起点著者と同じ著者の論文を検索することは、起点論文と同じ研究分野の論文の収集に繋がる。つまり、自身と同じ研究分野の論文を1つ見つけることができれば、その論文に関連する論文を効率的に収集することが期待される。研究者はそれらの収集した論文を読むことで、研究者自身の研究分野や、関連する分野、取り扱われてきた技術、研究動向などを理解する。このような行為は「サーベイ」と呼ばれ、研究活動において、研究者自身の研究の位置付けを明確にするために重要な行為である。従って、サーベイは膨大な数の論文から、研究者が企図した研究分野の論文を探し出し、多くの論文を読むことが望ましい。

サーベイにおいて、研究動向を把握する際は、研究者が企図した研究分野の論文を把握するだけではなく、ある論文の比較対象になる論文や技術的基盤となる論文がどの論文であるかの関係性を理解することが望ましい。しかし、キーワードを用いた検索では、同じ研究分野の論文を全て収集できるとは限らず、著者名を用いた検索では、企図した研究分野の論文のみを収集

\*連絡先: 関西大学総合情報学部  
〒569-1095 大阪府高槻市霊山寺町 2-1-1  
E-mail: mat@res.kutc.kansai-u.ac.jp

できるとは限らない。そのため、研究動向の把握が困難である。

そこで、本研究では論文を特徴づける要素の1つである引用論文に着目し、複数の論文間の引用論文の一致や差分を可視化することで、研究動向の把握を支援するシステムの実現を目指す。本稿ではその端緒として、引用論文の比較を支援するインタフェースについて検討する。

## 2 関連研究

サーベイを効率化するために、論文の引用・被引用関係を可視化する手法が提案されている。井上らは各論文をノードとし、論文間の引用・被引用関係をエッジとするグラフの可視化システムを提案した [5]。可視化システムにおいて、引用・被引用関係にある論文を全て表示すると、論文の数が多く見た目が乱雑になってしまう。そこで井上らは、その問題を解決するために、引用理由ごとに論文の表示をフィルタリングできる機能を構築した。実験結果から、フィルタリング機能を使うことでユーザのサーベイの検索効率が向上することが示唆された。しかし、可視化画面と引用理由などの情報が別画面で表示されているため、引用理由を1つにフィルタリングしていない場合、可視化画面のみで論文間の関係性を把握することが負担になる。

Nakano らは、論文の関係性をユーザに理解しやすい形で提示するため、引用・被引用関係に加え、共引用関係を用いた論文関係グラフを提案した [2]。共引用関係とは同一の論文に引用されている論文同士の関係のことである。Nakano らはこれらの論文間の関係を全て提示するのではなく、関係の強さを考慮して提示を行った。2つの論文が引用・被引用関係にある場合の関係性については、1つの論文の文章内で何度も引き合いに出されて引用されている論文ほど関係が強いとした。また、2論文間が共引用関係にある場合の関係性については、共引用されている回数をもとに類似度を計算し、その値が高いもののみを提示した。しかし、共引用されるには時間がかかるため、発行されてもいない論文に共引用関係を利用するのは困難である [3]。そのため、共引用の回数が多い論文同士の類似度が必ずしも高いとは限らない。

榎らは、自然言語には表記揺れが存在し、これが検索や分析の精度低下を招いていると示唆している [6]。この問題を解決するために、専門的な技術用語を含む電子情報通信学会の文献検索システム I-Discover と一般的な日常用語を多く含む Wikipedia のデータベースを用いた表記揺れを補償する辞書 Broadic を提案した。Broadic の有効性を確認するため、I-Discover の文献データを用いた時系列・技術要因分析システムを提案し、

Broadic の有無による影響を評価した。その結果、表記揺れを補償することで、より正確な研究トレンドを可視化することができたと示唆された。しかし、トレンドにある研究がユーザ自身の研究分野に対し、どのような研究動向であるか把握するには、論文を複数収集しなければならない。

共著ネットワークを用いて、萌芽領域の中心研究者を予測するための研究が行われている。森らは著者ごとの発表論文数、被引用数、ダウンロード数などの計量書誌データを各著者のパフォーマンスとみなし、ある研究領域において被引用数が高い著者を領域の中心研究者とした [7]。そして、研究領域を特定するために、シードとなる研究者から共著ネットワークを取得した。そのネットワークの構造に基づく特徴量をもとに、ある研究領域における研究者の被引用数を予測する学習モデルを構築した。これにより、中心研究者の予測を行った。しかし、中心研究者の中でどの論文がユーザが企図した研究分野の論文であるか把握し、その中でも論文同士の差異を把握することは困難である。

## 3 デザイン指針

前述した関連研究から、論文情報の関係性を可視化することでサーベイを支援できる。論文の引用・被引用関係の可視化を行う研究では、引用・被引用関係にある論文同士の関係性理解に繋がる。しかし、引用・被引用関係にない論文同士の関係性を理解することは困難である。加えて、共引用関係の可視化は、発行されてもいない論文には適さないため、最新の研究動向を掴むことが困難である。

研究動向の把握を支援する研究では、各分野のトレンドの分析を可能にしている。しかし、最も行われている研究の中でも、研究内容の差異を理解するのは困難である。そのため、同じ研究分野の中でどのような研究の動きがあるか理解することは困難である。

著者情報の分析を行う研究では、共著関係を用いることで、中心性の高い著者の把握に繋がると明らかにされている。しかし、その中心性の高い著者が、どのような研究を行っているかまで理解しなければ、研究動向の把握は困難である。加えて、その中心性の高い著者がユーザの研究分野にとって重要な著者とは限らない。そのため、ユーザ自身が選択した論文を執筆した著者の研究動向を可視化することで、ユーザ自身と同じ研究分野の動向把握支援に繋がると考えられる。

そこで本研究では、ユーザが起点とする論文を執筆した著者の研究動向を可視化する。これにより、引用・被引用関係にない論文も把握できる。しかし、その著者の論文の中でもユーザが企図した研究分野以外の論文も存在する。そのため本稿では、引用論文を比較で

きるようにすることで、ユーザが企図した研究分野の論文を把握できるようにする。引用論文は、引用元の論文の根拠や位置付けを示すために用いられる [8]。そのため、複数の論文を比較する際、引用論文は同じ分野の論文の同一性を判定したり論文同士の関係性を理解したりする際の手がかりになる。例えば、起点論文と比較対象論文（以下、比較論文と記す）の共通する引用論文に着目することで、研究分野の同一性や関連性を把握できる。また、起点論文と比較論文の引用論文の差分から、それらの論文間の着眼点や力点の相違といった論文同士の研究の位置付けを把握できる。

また、同一著者の論文は同じ分野の論文である場合が多いため、その著者の論文を継時的に閲覧することで研究動向の把握の一助になる。しかし、論文には複数の著者が関与する場合が多いため、著者ごとに執筆論文を閲覧できるようにする必要がある。

これらを勘案し (1) 引用論文の一致度を把握できるようにすること、(2) 引用論文の差異を判断可能にすること、(3) 著者ごとに論文を閲覧できること、の3つをデザイン指針とする。

## 4 提案インタフェースの実装

### 4.1 実装の概要

本稿ではコミック工学分野の論文 350 本を対象として、論文表題、著者名、発行年、引用論文から構成されるデータベースを構築した。そのデータベースを対象として、研究動向把握支援インタフェースを実装した (図 1 参照)。

提案インタフェースは、HTML、CSS、JavaScript、jQuery を用いて実装を行った。可視化部分は、Franzらが提供しているオープンソースの JavaScript ベースのグラフィブラリである Cytoscape.js を用いた。Cytoscape.js は視覚化ソフトウェアコンポーネントであるため、Web ブラウザでインタラクティブなグラフを表現できる [1]。これらを用いて、引用論文の比較を支援するインタフェースを実装した。

### 4.2 可視化デザイン

グラフ描画は、人間関係や論文の引用関係などのデータセットを直感的に表現するのに有用である [4]。そのため、提案インタフェースはノードとエッジを用いて描画した。ノードは論文を表し、その中には論文の表題を提示した。エッジは起点著者の論文の中で引用・被引用関係があるノード同士を結んだ。図 1 の画面はユーザが選択した論文 (起点論文) を執筆した著者 (起点著者) の動向を示した。赤色のノードは起点論文、青

色のノードは起点著者が執筆した別の論文を表した (図 1 の①参照)。それらのノードの下に表示されている橙色と灰色のノード (図 1 の②参照) は各々の引用論文のリストである。

図 1 の③は、起点著者の論文の中で、引用・被引用関係のある論文同士をエッジで示している。三角形の矢印が引用先であり、丸の印が引用元である。エッジのうち黒色のエッジは起点論文に対し、引用・被引用関係のあるものである。また、灰色のエッジは起点論文以外に引用・被引用関係のあるものである。また、各ノードを時系列順に表示しているため、どちらの論文が引用元で、引用先かはノードが表示されている座標で直感的に理解できる。

デザイン指針 (1) を満たすために、起点論文と比較論文の共通する引用論文のノードを橙色で、異なるノードを灰色で各々表示した (図 1 の②参照)。これにより、起点論文の引用論文と比較論文の引用論文との一致度を直感的に理解できる。例えば引用論文のリストのうち、橙色のノードが多い比較論文は起点論文と目的や手法の類似性が高いと推測できる。反対に、灰色のノードが多い比較論文は起点論文と目的や手法の類似性が低いと推測できる。さらに、起点論文の引用論文のリストと一致する比較論文が存在しなければ、起点論文の研究の方向性が変化したことの推測に繋がる。また、起点論文の引用論文のリストと一致する比較論文が複数存在すれば、起点論文と目的や手法の類似性が高い研究が複数行われていると理解できる。

図 1 の④はユーザが引用論文のノードをクリックすることで、その引用論文と同一の引用論文のノードが黄色に変化する。これによりクリックした引用論文がどの程度、どのような論文に引用されたか確認できる。

また、引用論文のノードの横に表示した数字 (図 1 の⑤、⑥参照) は、起点著者がその引用論文を引用して何回目であるかを示している。数字が 1 の場合はその引用論文の引用元の論文より古い年をユーザが遡っても、その引用論文を引用している論文を収集することはできない。そのため、ユーザはその引用論文が引用元の論文より古い年の論文で引用されているか参照する必要がない。また、数字の背景が深緑色 (図 1 の⑥参照) の場合、その引用論文は引用元の論文以降引用されていないことを表している。そのため、ユーザはその論文を引用している論文を探す際、その論文を引用している論文より新しい年を参照する必要がない。加えて数字が 1 で深緑色のノードの場合、その引用論文を引用したのは、その引用元の論文のみであるということがわかる。これらにより、ユーザが画面移動することの負担を軽減した。

次にデザイン指針 (2) を満たすために、2 論文間で共通する論文と、一方の論文でのみ引用されている論文のノードの色を区別した。起点論文以外の論文 (図 1

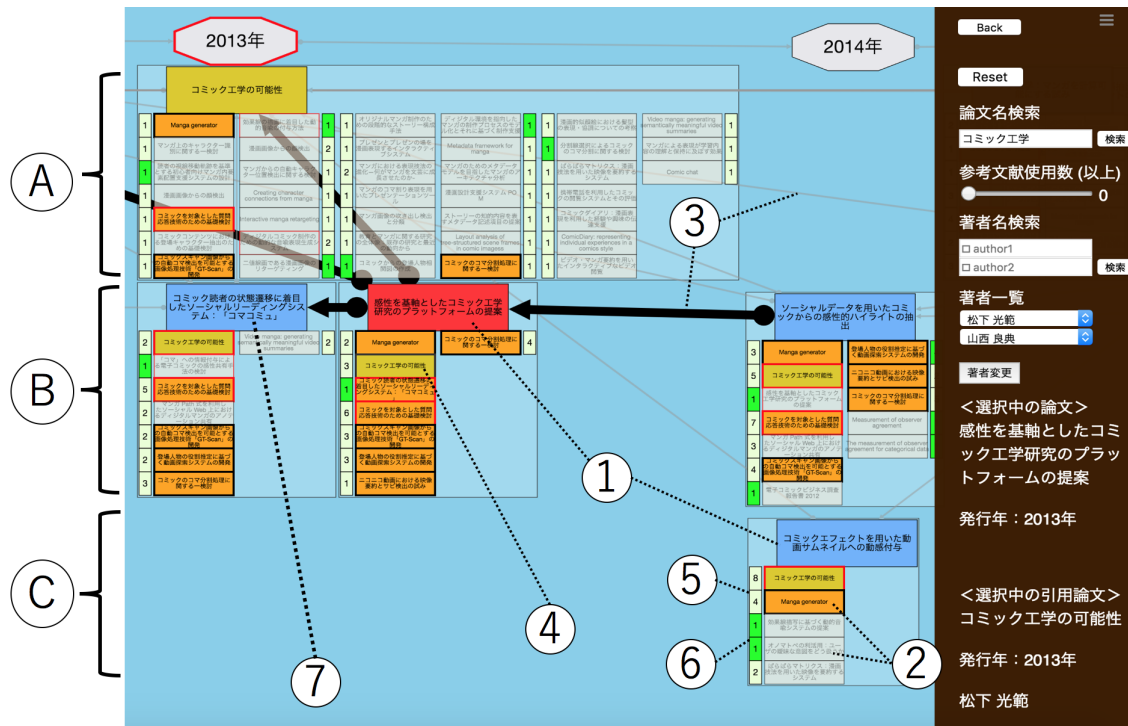


図 1: 提案インターフェース

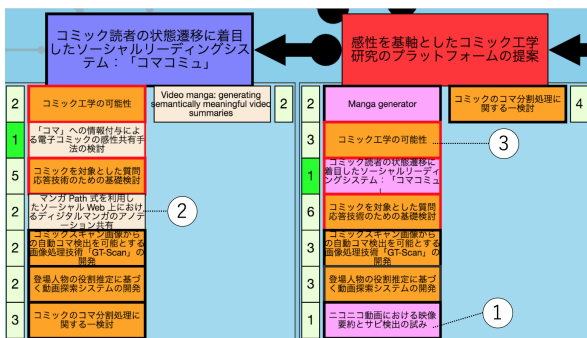


図 2: 引用論文の増減結果

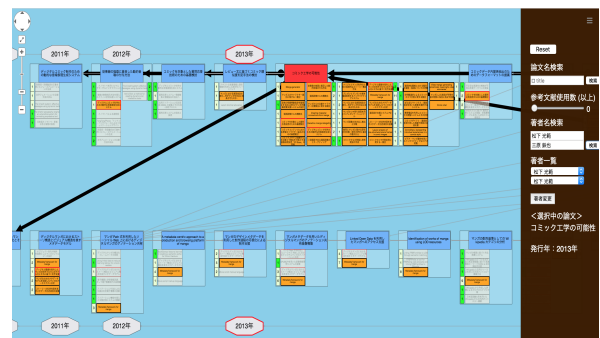


図 3: 共著関係にない著者同士の研究動向

の⑦参照)を右クリックすると、その論文と起点論文の引用論文の色が変わる(図2参照)。桃色のノード(図2の①参照)は起点論文のみで引用された論文であり、黄白色のノード(図2の②参照)は新たに選択された論文にのみ引用された論文である。2つの論文で共通する引用論文は橙色のノードのままである。これにより、直感的に論文同士の差分を理解でき、一方にのみ現れる引用論文がその引用元の論文の特徴を理解する上で手掛かりになる。

また、デザイン指針(3)を満たすために、著者ごとにy座標を統一した。図1では起点著者の1人目が執筆した論文を上段(図1のA参照)、2人目を下段(図1のC参照)に各々表示した。また両者が共同で執筆した論文を中段(図1のB参照)に表示した。これによ

り著者ごとに執筆した論文が容易に理解できる。さらに著者同士の関係性も理解できる。

図3では、ユーザが任意に選択した著者2人の研究動向を表示している。これにより、共著関係にない著者同士の研究動向も比較できる。加えて、両者の研究分野や引用論文の差異の理解に繋がる。

## 5 ユーザ観察

ユーザ観察では、実装したインターフェースが3章のデザイン指針で企図した利用が確認できるか検証する。そこで、ユーザが引用論文を比較することで、論文同士の関係を理解できているかに注目した。

## 5.1 ユーザ観察の手続き

まず、実装したインタフェースの説明を行った。その後、インタフェースの評価を行うため、インタフェースを利用し、機能の妥当性に関するインタビューをリードユーザに行った。なお、実装したインタフェースはコミック工学分野の論文を利用しているため、リードユーザはコミック工学分野の論文を執筆した経験のある者にした。ユーザ観察は、リードユーザが2論文間の関係を理解できているか口頭で適宜確認した。そのために、2つの論文を比較させ、その2つの差異を答えるタスクを課した。比較する論文は起点論文と同じ論文を引用しているものや引用・被引用関係がある論文に設定した。

## 5.2 ユーザ観察の結果

ユーザ観察ではまず2017年の論文Aを起点論文に設定した。その論文Aの引用論文と同じ論文を引用している論文は2016年の論文Bのみであった。そこで論文Bを比較論文に設定し、起点論文との差異を答えるタスクを課した。リードユーザは「論文Aでは、論文Bに扱われていない固有表現の概念を扱うため、その引用論文2本が追加された」と発言した。これは、2つの論文の差分に関する発話であり、デザイン指針(2)で企図した利用が確認された。

次に2017年の論文Cを起点論文に設定した。論文Cの引用論文と同じ論文を多く引用している2015年の論文Dが存在した。そのため、論文Dを比較論文と設定し、起点論文との差異を答えるタスクを課した。リードユーザは「論文Cでは、論文Dで引用されている情報推薦系の論文よりもレビューやコミックを扱う上で必要な論文を引用するようになった」と発言した。これは、2つの論文の差分に関する発話であり、デザイン指針(2)で企図した利用が確認された。

次に2018年の論文Eを起点論文に設定し、他の論文と比較するタスクを課した。論文Eの引用論文と同じ論文を引用している論文は存在しなかった。そのため、リードユーザは「この論文は今までと着眼点が異なるため、他の論文で引用されていたものを引用しなくなった」と発言し、論文Eを新しいテーマの論文であると判断した。これは、起点論文と他の論文の引用論文の一致度に基づく行動であり、デザイン指針(1)で企図した利用が確認された。

一方で、今回の評価ではデザイン指針(3)で企図した利用に関わる行動は観察されなかった。このことから、論文の配置が著者ごとに表示されていることをリードユーザが理解できなかった可能性や、著者ごとに論文を分類することが論文同士の関係の理解に不要である可能性が考えられる。

ユーザ観察終了後に、リードユーザから「情報量が過多であるため、同じジャンルの論文のみを表示してほしい」「広い表示画面を見ることが負担」という意見が得られた。これらのことから、必要な論文を絞り込む機能の必要性が示唆された。

## 6 議論

5章で行ったユーザ観察から、論文同士を比較したことで、各論文同士の一致度や差分を理解できたと示唆された。しかし、ユーザは普段のサーベイ時、引用論文を比較して論文を検索しないため、提案インタフェースの説明なしでインタフェースを利用するユーザは引用論文を比較せず、論文の表題のみしか見ない可能性が考えられる。そのため、今後はインタフェースの説明なしで利用するユーザでも、ノードの色の意味を理解し、引用論文を自然と比較できるインタフェースのデザインにすることを目指す。本章では、そのインタフェースのデザインと、扱った引用論文の情報、の2つの観点から考察を行う。

### 6.1 可視化画面の有用性と問題点

可視化画面では、起点著者のうち2人の研究動向を表示している。その2人のうち、1人のみが執筆した論文を可視化画面の上段と下段に表示した。その表示した論文は、その1人の著者が主著者ではない場合も同じ座標に表示している。そのため、起点論文と関連度が低い可能性のある論文も同じ扱いとして表示しているため、今後は起点著者のうち1人しか執筆していない論文の表示は工夫する必要があると考えられる。例えば、その1人が主著者として執筆していない論文を表示する場合は、y座標や色の変化など区別した表示にすることを検討する。

提案インタフェースは起点著者のコミック工学の論文全てを表示しているため、起点著者が行った研究の網羅的な把握や起点論文と類似度の高い論文などの把握を支援した。しかし、論文数の多い著者の場合、全て把握するのは困難になる。また、著者の中にはコミック工学だけでなく、他の分野の研究を行っていることがある。そのため、起点著者全ての論文を表示するだけでなく、ユーザがピックアップした論文のみ表示する機能を実装する必要があると考えられる。これにより、画面を左右に行き来する手間が省け、情報量過多の解消に繋がる。

また、提案インタフェースではサーベイ論文など引用論文数の多い論文を可視化画面で発見することができる。しかし、引用論文数の多い論文は画面の横幅が長くなってしまおうという問題がある。そのため、引用

論文の表示・非表示の切替機能などが必要だと考えられる。

## 6.2 提示した引用論文の有用性と問題点

実装したインタフェースにおいて、引用論文の表題を各引用元の論文のノードの下に全て表示した。これにより、引用論文から論文を検索する際、引用元の論文を閲覧せず論文を検索することができる。しかし、引用論文の表題全てを表示したため、全体的に文字の情報量が過多になった。そのため、引用論文をジャンルごとに分類するか、キーワードのみを表示し、それらのキーワードをクリックすることで、引用論文の表題が表示されるなどの機能が今後必要だと考えられる。これにより、情報量を減らすことができる。

## 7 おわりに

本稿では、研究動向の把握支援を目的として、引用論文を可視化することで、論文間の関係理解を促すインタフェースを実装し、その有用性を検証した。ユーザ観察の結果から、論文同士を比較したことで、論文同士の一致度や差分を理解できたと示唆された。しかし、各論文の理解や研究動向の理解を支援できたか検証することはできていない。さらに、ユーザ観察では、リードユーザ1人のみを対象としたため、提案インタフェースが他のユーザにも有効に働くか今後検証する必要がある。

また、提案インタフェースのデザインは6章で述べた通り、改善すべき点が存在する。さらに、本稿で構築したデータベースはコミック工学の論文しか扱っていないため、他の分野の論文でも提案インタフェースが有効に働くか検証する必要がある。また、提案インタフェースは論文の学会情報、ページ数などの書誌情報や論文の内容情報を扱うことができていない。これらを扱うことで、論文同士の差分をより理解できると考えられる。そのため、今後は提案インタフェースのデザインの改善と本稿で扱っていない論文情報を扱うことを検討する。

## 参考文献

- [1] Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O. and Bader, G. D.: Cytoscape.js: a graph theory library for visualisation and analysis, *Bioinformatics*, Vol. 32, No. 2, pp. 309–311 (2015).
- [2] Nakano, Y., Shimizu, T. and Yoshikawa, M.: A Visualization of Relationships Among Papers Using Citation and Co-citation Information, *International Conference on Asian Digital Libraries*, Springer, pp. 157–163 (2016).
- [3] Shibata, N., Kajikawa, Y., Takeda, Y. and Matsushima, K.: Comparative study on methods of detecting research fronts using different types of citation, *Journal of the American Society for information Science and Technology*, Vol. 60, No. 3, pp. 571–580 (2009).
- [4] Toeda, N., Nakazawa, R., Itoh, T., Saito, T. and Archambault, D. W.: On Edge Bundling and Node Layout for Mutually Connected Directed Graphs, *Information Visualisation (IV), 2016 20th International Conference*, IEEE, pp. 94–99 (2016).
- [5] 井上絢翔, 韓東力: 参照関係の可視化による論文サーベイの効率化, 第11回インタラクティブ情報アクセスと可視化マイニング研究会, SIG-AM-11-09, pp. 55–61 (2015).
- [6] 槇俊孝, 若原俊彦: I-Discover 文献メタデータを用いた時系列・技術要因分析システム, 電子情報通信学会論文誌 D, Vol. 99, No. 10, pp. 1002–1012 (2016).
- [7] 森純一郎, 原忠義, 榊剛史, 梶川裕矢, 坂田一郎: 大規模学術論文データの共著ネットワーク分析に基づく萌芽領域の中心研究者予測に関する研究, 人工知能学会全国大会論文集, Vol. JSAI2015, pp. 1B21–1B21 (2015).
- [8] 石井仁子, 太田学, 高須淳宏: 引用意図を利用した学術論文閲覧支援のための適切な被引用箇所の特長, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F3-5 (2015).
- [9] 棚瀬達央, 大平茂輝, 長尾確: デジタルコンテンツの部分参照・引用関係に基づく論文作成支援, 第75回全国大会講演論文集 2013, 情報処理学会, Vol. 1, pp. 623–624 (2013).
- [10] 井坂徳恭, 中山泰一: 重要論文検索システム Iask の実装と評価, コンピュータと教育研究会, Vol. 2011-CE-109 (2011).
- [11] 林佑磨, 奥野峻弥, 山名早人: 単語の意味概念行列を用いたキーワード生成による関連論文検索システム, 情報処理学研報, Vol. 2014-DBS-159, No. 10 (2014).