

他者のサーベイ行動を利用した論文検索システムの提案

杉山暢彦* 澤本潤** 杉野栄二** 瀬川典久**

概要. 論文サーベイは研究活動において、研究動向の調査や新規性の確認をするための非常に重要な過程の一つである。論文サーベイには研究分野の動向を把握するためのサーベイと、テーマに関連した研究を把握するためのサーベイといった2つの段階があり、各段階に応じて必要とされる論文は異なる。近年では、Google Scholar や Cinii といった論文検索サービスを利用する事で容易に論文を取得できるようになっているが、これらのサービスでは、キーワードと関連度の高い論文や被引用数の高い論文が上位に表示されるため、ユーザの検索の段階を考慮しているとは言い難い。そこで、本研究ではユーザの論文検索からサーベイまでをシステム上で行えるようにし、蓄積したユーザのデータに基づいてユーザの検索段階に合った論文を提示するシステムを提案する。

1 はじめに

論文サーベイは研究活動において、研究動向の調査や新規性の確認をするための非常に重要な過程の一つである。論文サーベイには研究分野の動向を把握するためのサーベイと、テーマに関連した研究を把握するためのサーベイといった2つの段階があると考えられる。研究分野の動向を把握する段階（浅い検索段階）では、曖昧なキーワードを設定しているため、検索結果の数が膨大になると考えられる。テーマに関連した研究を把握する段階（深い検索段階）では、的確なキーワードが設定されているため、検索結果の数が浅い検索段階よりも少なくなると考えられる。また、各段階に応じて必要とされる論文も異なる。浅い検索段階では分野の把握に役立つ論文、深い検索段階ではテーマに関連の深い論文が必要とされる。

近年では、Google Scholar¹ や Cinii² といった論文検索サービスを利用する事で容易に論文を取得できるようになっている。これらのサービスでは、キーワードと関連度の高い論文や被引用数の高い論文が上位に表示されるため、ユーザとの関連度の高い論文を抽出する事が必要である。また、既存の研究では、論文の参照構造を利用して重要な論文を発見する研究[1]や、論文を外部サービスの情報を利用して分野に分類する研究[2]が行われている。これらの研究では、論文の参照関係や外部情報源を機械的に解析して重要度の高い論文の発見や論文を分野に分類している。重要論文検索の研究では深い検索段階について、論文分類の研究では浅い検索段階について

の支援だと考えられるため、2つの検索段階を考慮した支援はされていない。

そこで、本研究ではユーザの論文検索からサーベイまでをシステム上で行えるようにし、蓄積したユーザのデータに基づいてユーザの検索段階に合った論文を提示するシステムを提案する。

2 提案システム

本システムでは、キーワードによる論文の検索、検索結果のグループ化、検索段階に応じた論文の推薦、取得論文のサーベイを行える。各機能の詳細について以下で説明する。

2.1 キーワードによる論文の検索

本システムでは、キーワードによる論文の検索結果を Cinii から取得してシステム上に表示している。システムで表示する情報は、Cinii から取得可能な論文情報（タイトル、著者、アブストラクト等）を全て検索結果に表示している。これにより、実際に論文詳細ページに行かなくても、検索結果を見るだけで論文詳細ページと同じ情報を得られる。また、検索結果でほとんどの論文情報を得られるため、リンクをクリックするユーザは実際に論文を読んでみたいユーザと考えられるため、リンクをクリックされた論文をユーザの興味のある論文としてシステムに登録する。検索インタフェースの画面を図1に示す。

Copyright is held by the author(s).

* 岩手県立大学大学院ソフトウェア情報学研究所

** 岩手県立大学ソフトウェア情報学部

¹ Google Scholar <http://scholar.google.co.jp/>

² Cinii <http://ci.nii.ac.jp/>



図 1. 検索インターフェース

2.2 検索結果のグループ化

浅い検索段階で利用するキーワードには曖昧なキーワードが多いと推測される。曖昧なキーワードを使った場合、検索結果には様々な分野の論文が含まれる場合がある。そのため、浅い検索段階においては検索結果を分野毎にグループ化する必要がある。

本システムでは、同じキーワードに対して閲覧した論文が似ているユーザ同士をグループ化し、ユーザにはグループ内のユーザが閲覧した論文と使用したキーワードの一覧を表示する。この時、表示するキーワードにはサーベイ機能によって設定したテーマに含まれるもののみとなっている。

グループ化の際に使用した手法について、ユーザ間の距離計算には jaccard 係数を使用し、グループ化には階層的クラスタリングを使用した。この時、グループ化の閾値は試験的に 0.6 として設定している。グループ表示画面の一部を図 2 に示す。

このキーワードに関するグループ

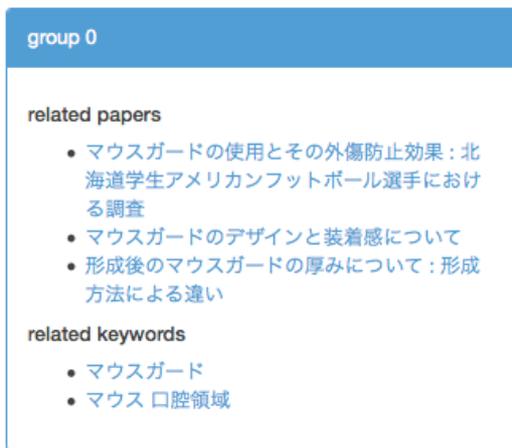


図 2. 検索結果のグループ表示画面

2.3 検索段階に応じた論文の推薦

浅い検索段階では、分野の把握に役立つ論文が必要だと考える。具体的には、「良い参考文献が参照されているとユーザから判断された論文」を浅い検索段階で必要な論文とした。また、深い検索段階では、テーマに関連の深い論文が必要だと考える。具体的には「類似度の高いユーザがテーマに関連が深いと判断した論文」を深い検索段階で必要な論文とした。

抽出方法として、浅い検索段階で必要な論文は、ユーザがサーベイ時に分野の把握に役立つ論文であると評価したものを対象とする。深い検索段階で必要な論文は、対象ユーザと協調フィルタリングを利用する事で取得する。

2.4 論文のサーベイ機能について

本システムを用いた論文サーベイでは、論文に対するテーマの設定と、「分野の把握に役立つ論文であったか」「テーマに関連の深い論文であったか」といった2種類の観点に対する評価が可能となっている。取得した論文に対してテーマを設定する事で、論文だけでなくキーワードにもテーマを設定でき、検索結果のグループ化を行った際に表示するキーワードはこのテーマから選択できるようになるためである。分野とテーマに関する評価は「わからない」「役立つ」「少し役立つ」「あまり役立たない」「役に立たない」といった5段階の評価とした。

3 今後の課題

現在は論文の内容については考慮せず、ユーザの評価のみで各検索段階に必要な論文を抽出している。今後は、論文の内容や参照関係などを利用して各段階に必要な論文を抽出する精度を高くしていく事が課題である

4 参考文献

- [1] 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." 情報処理学会研究報告. コンピュータと教育研究会報告 CE-109: 1-8, 2011
- [2] 福田悟志, CiNii データベースを用いた研究動向分析システムの構築, 言語処理学会第 18 年次大会 539-542, 2012