

Polyphonet 常時運用の試み テスト運用からの知見と今後の展開

Development of a polyphonet for daily use, trial and future work

杉山達彦*¹ 松尾豊*² 石田啓介*² 濱崎 雅弘*² 森純一郎*^{3*4} 西村拓一*²
 Tatsuhiko Sugiyama Yutaka Matsuo Keisuke Ishida Masahiro Hamasaki Junichiro Mori Takuichi Nishimura

武田英明*^{5*6*3}
 Hideaki Takeda

*¹ (株)ユニクス
 UNICUS Co., Ltd.

*²産業技術総合研究所
 National Institute of Advanced Industrial Science and Technology

*³東京大学
 University of Tokyo

*⁴ドイツ人工知能研究所
 Deutsches Forschungszentrum für Künstliche Intelligenz

*⁵国立情報学研究所
 National Institute of Informatics

*⁶総合研究大学院大学
 The Graduate University for Advanced Studies

A social networking system can become the basis for the information infrastructure of the future. For that purpose, it is important to extract social networks that reflect actual social networks which users have already had. In this paper, we introduce our new web mining system based on social network called Polyphonet Academia, which mines an arbitrary researcher's social network and papers according to registered user's operation. The main achievement of our system is to assist survey, which is researches' daily effort.

1. はじめに

我々は Web マイニング技術や実世界インタラクション技術によるソーシャルネットワークの抽出と、ソーシャルネットワークを用いた利用者の活動支援について研究を進めており、その成果として Polyphonet を開発した [1] .

一般にコミュニティウェアでは、ブートストラップとなるデータがなければ初期におけるユーザは利用メリットが少ないと言われているが [2] , Polyphonet は登録時にソーシャルネットワークを自動抽出することで、使い始めのユーザに対しても有益なソーシャルネットワークデータを提供できるソーシャルネットワーキングシステムである . Polyphonet はコミュニティの参加者を支援するコミュニティウェアとして、また、コミュニティを概観するためのソーシャルネットワーク可視化ツールとして機能する .

本論文では、様々な実装を持つ Polyphonet の中でも、研究者の日常的な活動である情報収集の支援を目的とした Polyphonet Academia (*¹) について、システムの概要とそのテスト運用の結果について述べる .

2. Polyphonet Academia

Polyphonet Academia は日常的に情報収集を行う研究者を対象とした Polyphonet である . 様々な情報収集の方法が考えられるが、起点として想定するのは研究者の氏名である . 学会で講演を聞いた、論文誌で論文を読んだ、知人に紹介されたなどの形で、研究者は情報収集の対象となる研究者の氏名を得ることが多い . Polyphonet Academia では、登録ユーザとなった研究者が、システムに対してマイニングを実行させて、興味のある研究者の情報を、研究者の氏名をキーにして、主体

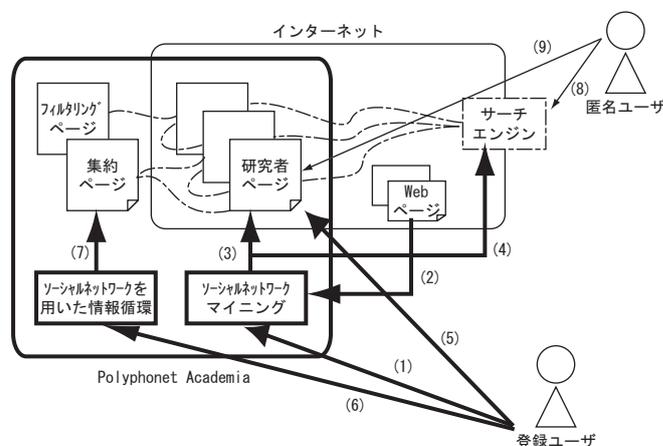


図 1: 提案システムの概要

的に収集していけるように開発した .

Polyphonet Academia では、図 1 の (1) ~ (3) の矢印にて示すように、Web マイニングにより Web 上の情報から研究者間関係を抽出して、ソーシャルネットワークを提示することを行っている . Web マイニングにおいては、研究者の氏名と所属を元に、2 人の研究者が共に出現するページを検索エンジンから収集して分類することで、関係性を推定している . また、関係性に加えてタイトルと著者・共著者の氏名からなる論文情報も Web マイニングしている .

これらデータが、利用者がシステムにアクセスした時点で既に蓄積されている初期データとなる . これは研究者ページ (図 2) として表示されており、利用者にとって自分の知り合いがどこにいるかを示すナビゲート役を果たす . このネットワークはグラフとしてみることができ、探索もできる .

通常の Polyphonet では、これら研究者ページは、登録したユーザの氏名を対象に生成していた . つまり、研究者ページの

連絡先: 杉山達彦, (株)ユニクス, 東京都千代田区内神田 3-17-8 小山ビル 501, 03-6804-0808, 03-5256-1312, sugiyama@unicus.jp

*1 <http://ac.polypho.net/>



図 2: 研究者ページ

数は、登録ユーザの数と一致していた。Polyphonet Academia では、ユーザ登録時に加え、登録ユーザからのマイニングリクエストに基づいても生成することができる(図 3)。登録ユーザは、任意の研究者の情報をマイニングさせて、研究者ページをシステムに登録させられるのである。自分の共同研究者や、自分が学会で知り合った研究者、興味のある研究者、影響力のある研究者など、情報収集の対象としたい研究者のページを生成させることができる。

また、Polyphonet Academia では、マイニングリクエストを入力する方法を使い勝手に考慮して設計した。具体的には、図 3 に示すように、利用者が氏名で Polyphonet Academia のサイト内検索を行った後に、言い換えると登録ユーザが研究者の情報を知りたいと思ってアクションを起こした時に、検索結果のページからマイニングリクエストを呼び出せるようにした。マイニングリクエストの呼び出しの際には、利用者が入力した人名を引き継ぐようにして、利便性を向上させている。

もう一つのマイニングリクエストの呼び出し方法は、共著者に関する情報収集に特化した方法である。Polyphonet Academia が Web マイニングする論文情報には、研究者の共著者の氏名が含まれる。これら共著者の氏名からマイニングリクエストを呼び出せるようにした。マイニングリクエストの呼び出しの際には、利用者が入力した人名を引き継ぐようにして、利便性を向上させている。

このようにして、登録ユーザは、自分が興味をもっている研究者の氏名や、その共著者らの氏名を次々とマイニングして、情報収集の対象となる研究者のページを作れるようになっている。つまり、登録ユーザは、非常に簡単な操作により、自分の興味に従って主体的に、豊富な情報を含む研究者のページを得られる。

また、Polyphonet Academia では、これら Web マイニン



図 3: マイニングリクエストページ

グした研究者ページには、余分な情報が含まれる可能性もあり、登録ユーザが変更できるようになっている(図 1 の (5))。

さらに、Polyphonet Academia において登録ユーザは、Web マイニングした研究者ページを、ソーシャルネットワークに基づいて、集約したり、フィルタリングしたりする機能を備えている(図 1 の (7))。具体的には、特定の関係性を有する研究者らに関する論文情報を集約することができる。また、自分のプロフィールを特定の関係性を有する研究者らだけが見えるようにフィルタリングすることもできる。

Polyphonet Academia の 2 つ目の大きな機能として、マイニングした研究者のページを一般的な検索エンジン (*2) に登録させる機能がある(図 1 の (4))。これにより、図 1 の (8)、(9) に示すように、研究者の氏名で検索エンジンを検索した匿名ユーザを、Polyphonet Academia の研究者ページに誘導することができる。匿名ユーザに公開される研究者ページは、登録ユーザが見るページに比べ限定的な内容となっている。これら、匿名ユーザに公開される研究者ページでは、登録ユーザとなることにより、さらに詳しく研究者に関する情報が見られる旨のメッセージを掲載することで、ユーザ登録を促すようになっている。

多くの登録ユーザを獲得し、多くの研究者ページを提供できるようになることは、あらゆるユーザにとってメリットがある。まず、多くの登録ユーザを獲得し、これら登録ユーザが自らの興味に応じて研究者ページを増やしていくことは、同じ研究者に興味をもつ他の登録ユーザが Web マイニングする時間を節約できる。また、研究者ページは匿名ユーザに限定的にはあるが公開されるので、登録ユーザとならなくても、あらゆるユーザが研究者に関する情報を得ることができる。

Polyphonet Academia は MySQL データベースと PHP および Perl で記述されたプログラムにより構成される。研究者の氏名、所属、写真などのプロフィール情報、Web マイニングで抽出したタイトル、著者リストなどの論文情報、Web マイニングで抽出したものや利用者から追加されたソーシャルネットワークデータとをそれぞれ DB に格納している。ユーザは Web ブラウザを通じてシステムを利用する。

*2 <http://www.google.co.jp/>

3. テスト運用からの知見と今後の展開

Polyphonet Academia は、現在オープンなテスト運用を行っており、本運用に向けた最終的な調整を行っている。

まだ運用規模が小さく、有益なデータが得られたとは言えないが、現在のデータをまとめる。運用期間は 2006 年 12 月 30 日～2007 年 4 月 16 日、平均ページビューは 26.59、ユニークな訪問者数は 1448、総ページビューは 38538 である。登録ユーザ数は 77 である。また、マイニングされた研究者ページ数は 1117、キーワード数は 48416、論文数は 30321 で、検索エンジンに登録された研究者ページ数は 81 である。

ユーザから種々の要望が寄せられているが、研究者の日本語氏名で Web マイニングをしていることから、主に英語で論文投稿をしている研究者から、英語氏名で Web マイニングしたいというものが多かった。また、使い勝手に関する要望や、応答速度に関する要望も多かった。提供している機能、特に、ユーザが主体的に Web マイニングできるという柱となるコンセプトについては好意的なコメントが寄せられていた。

今後も、手動では容易に得ることができない機能を、簡単な操作で得られるというコンセプトで開発を進めていきたい。

4. おわりに

Polyphonet Academia は常時運用版 Polyphonet である。登録ユーザの要望に基づいて、Web マイニングにより、情報収集の対象となる研究者のソーシャルネットワークや論文情報を含む研究者ページを作る機能を実現した。また、研究者の氏名で検索エンジンを検索しており、Polyphonet Academia の潜在的なユーザである匿名ユーザに対して、システムを効果的にプロモーションする機能を実現した。

Polyphonet Academia は、現在オープンなテスト運用を行っており、本運用に向けた最終的な調整を行っている。調整が完了し、かつ、研究者ページが一定数に到達したときに、本運用を開始する計画である。

参考文献

- [1] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満: Web上の情報からの人間関係ネットワークの抽出, 人工知能学会誌, Vol.20, No.1E, pp. 46-56, 2005.
- [2] Takeda, H. and Ohmukai, I.: Building semantic web applications as information/knowledge sharing systems, in Proceedings of End User Aspects of the Semantic Web, 2005.