

Community Web プラットフォームの展望

Current Status and Challenges of Community Web Platform

大向 一輝*1*2

Ikki Ohmukai

武田 英明*1*3

Hideaki Takeda

*1 国立情報学研究所

National Institute of Informatics

*2 総合研究大学院大学

The Graduate University for Advanced Studies

*3 東京大学

The University of Tokyo

In this paper we summarize our 5 years program called "Community Web Platform" and discuss its future prospect. 55 papers related to the program were submited in years when the Web 2.0 movement has come in the ordinary world. Based on the outcome of these research, we will plan to develop a new research field for "webified" researchers.

1. はじめに

近未来チャレンジ「Community Web プラットフォーム」は最終年度を迎える。この5年間に於いて、ウェブの社会的影響は飛躍的に大きくなり、またウェブを対象とした研究も大きく進展した。筆者らは、Community Web プラットフォームの定義について、本チャレンジの提案時に以下のように述べた[大向 06]。

Weblog や SNS など、現状の Web において個人間のコミュニケーション支援機能が重要視されている。本研究では、Web のこの側面を強調するため、Community Web という造語を用い、この Community Web 環境における個人の活動のモデル化およびその支援を目指す。

また、Community Web の概念を明確にするために、図 1 に示すモデルを提示した。ここでは、ウェブ上で行われている活動を情報のレイヤと人間関係のレイヤに分け、それぞれに対して「収集する」「関連づける」「発信する」といった要素を対応させている。情報のレイヤでは、これらを「Collect」、「Create」、「Donate」と呼び、人間関係のレイヤでは「Relate」、「Collaborate」、「Present」と呼ぶこととした。本チャレンジの当初の目的は、この6種のカテゴリによってウェブ上の情報流通およびコミュニケーションのプロセスを記述し、既存の技術やサービスを含めてすべてのカテゴリをカバーするような基盤を構築することであった。

一方、時期を同じくして Web 2.0 という言葉に代表されるユーザ参加型サービスが急速に立ち上がった。本チャレンジの提案時にすでに存在していたブログや SNS は、この5年間に国内で1000万ユーザ、全世界的には1億を超えるユーザを集めるようになった。その後も Wikipedia やソーシャルブックマーク(SBM)のように、ユーザ同士のコミュニケーションが新たな情報を生み出す例は枚挙に暇がない。以前は個人的な営みであった情報収集プロセスも、それがウェブで公開された状態で行われることによって発信手段となり、他のユーザの情報収集支援や新たな知人関係の発見につながるなど、Community Web モデルにおけるカテゴリ間の境界も曖昧になっている。

このような状況の中で、本チャレンジの位置づけや意義は大きく変わることになった。ユーザ参加型サービスの普及によって、膨大なデータを用いた大規模な分析ができるようになるとともに、それらのサービスの存在を踏まえた新たな仕組みの提案が可能となった。本チャレンジは、多様なバックグラウンドを持つ研究者による議論の場として一定の役割を果たすことができたと考える。本論文では、5年間継続された本チャレンジの総括を行い、得られた成果を今後どのように発展させるかについて議論する。

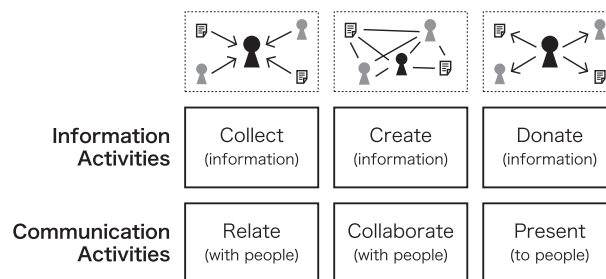


図 1: Community Web のモデル

2. Community Web プラットフォームのこれまで

本チャレンジでは、特定の研究グループを想定せず、ウェブとコミュニティに関わる研究を広く募ってきた。その結果、初年度(JSAI2005)を除く4年間で55件の論文が投稿された*1。また、本チャレンジとは関係なく同様のテーマについて発表された論文も年々増加している。ここでは、55件の中からJSAI2006からJSAI2008の3年間に発表された43件について、その主要なテーマについて述べる*2。

2.1 JSAI2006 の特色

国内最大規模の SNS である mixi のユーザ数が1年間で75万人から490万人まで増加した2006年度のJSAIでは、ソー

*1 セッションごとの発表数の制約(各年度によって異なる)のため、実際にはこれ以上の関連研究が存在する。

*2 本チャレンジに投稿されたすべての論文のリストは <http://sites.google.com/site/communitywebplatform/cfp09> に掲載し、本論文では個別の文献の引用は行わない。

連絡先: 大向 一輝, 国立情報学研究所, 〒101-8430 東京都千代田区一ツ橋 2-1-2, Tel: 03-4212-2585, Fax: 03-3556-1916, i2k@nii.ac.jp

シャルネットワークに関する研究が大半を占めた。

SNS では人間関係を表現するリンクに「知人」という単一のラベルしか存在せず、またそのリンクも経済合理性とは関係なく張られるために、現実のコミュニティの姿を反映していないという批判から [松尾 07]、人間関係の多重性を考慮したネットワークを導入し、情報のやりとりの範囲を必要に応じて限定することができるシステムが 5 件提案された。一方で、SNS の手軽さを利用したコンテンツ作成支援や、複数の SNS の統合システムなど、ソーシャルネットワークの特性を積極的に利用する研究もあり、様々な立場の参加者によって活発な議論がなされた。

また、現実存在する暗黙的なソーシャルネットワークを可視化する研究については、企業内で送受信されたメールの履歴から得られるネットワークと実際の組織構造を比較する研究や、ウェブマイニングによって特定のドメインのネットワークを選択的に抽出する研究があった。

ブログを対象とした研究では、エントリー間のリンクやトラックバックといったブログ特有のコミュニケーションから得られるリンク構造に注目したトピック抽出や、自然言語処理による書き手の社会的立場の推定など、構造だけでなくコミュニケーションの内容に踏み込んだ研究が見られた。

2.2 JSAI2007 の特色

2007 年度にはユーザ参加型サービスの普及は決定的なものとなり、大規模化する研究対象とどのように向き合うか、また日常生活の一部となったウェブにどのように研究成果を組み込んでいくかが主要なテーマとなった。ここでは、「ユーザエクスペリエンス」をキーワードとして、発表された論文をユーザ体験の理解、体験結果の応用、新たなユーザ体験の提案という 3 つに分類した。

ユーザ体験の理解では、多言語間ならびに異なるユーザ属性間のブログ分析によって、それぞれが属するコミュニティの興味・関心の極性を発見する研究が目立った。また、体験結果の応用ではブログないし SNS のネットワーク分析から重要語や重要な関係を特定する手法が提案された。新たなユーザ体験の提案としては、検索エンジンの逆引き辞書を作成し、コミュニティに応じたキーワード推薦を行うもの、ブログの画像つきエントリーに特化した要約手法の提案、研究室のレポート群をクラスタリングして新たな研究テーマを推薦するものなど、実験的な試みが行われている。また、動画コミュニティ「ニコニコ動画」の原型となるシステムにおけるコミュニケーション分析もなされた。

2.3 JSAI2008 の特色

2008 年度のセッションでは、ソーシャルネットワークをスパム排除やイベント情報の推薦、情報アーカイブの組織化など、コミュニケーション以外の用途に利用する提案があった。Wikipedia を対象とした研究も増加し、多言語 Wikipedia を対訳辞書として利用する手法、記事の時系列的な成長パターンによってコンテンツの内容を推定する手法についての発表が行われた。また、コミュニケーションの定量化についてはリンク構造分析と自然言語処理の複合的に利用した手法が見られるようになった。他にも、ソフトウェアやサービスによるコミュニケーション環境が十分に整備されたことで、実際にウェブ上で地域コミュニティを主催し、その知見を報告した研究もあった。

この 3 年間を通じて一貫して知識共有コミュニティの調査を行っている研究もある。質問応答コミュニティを対象とした長期的な調査から、ユーザが年を追うごとにウェブ上のコミュ

ニケーションに慣れ、独自のコミュニティを形成していることが報告されている。

3. ウェブ研究の展望

3.1 データのオープン化

Community Web プラットフォームが対象としてきたユーザ参加型サービスの世界では、データの所有者がサービス運営者ではなくユーザ自身であるとの認識が広まりつつある。これには、情報は生産者の下に帰属すべきであるという原則論と、多種多様なサービスが生まれる中でソーシャルネットワークなどの情報を作り直したり個別に整形するためのコストを低減したいというユーザのニーズが背景にある。とくに、後者の要求に応えるためには、データを相互運用可能な形で流通させる必要があり、定型的なフォーマットや標準規格が求められる。

これに対して、先進的なサービスでは所蔵のコンテンツを XML や RDF でマークアップし、一般に提供している。また、ソーシャルネットワークについては FOAF や XFN を用いた人間関係データの公開を行っている。こういったデータは容易に収集し、機械処理を施すことができる。また、比較的オープンではないサービスでも、ウェブ API を通じて定型化されたデータを入手することができる。

当初はウェブサービス開発者向けに提供されてきたこれらのオープンデータであるが、現在は研究のためのデータセットとしても用いられている。クローリングによるデータ収集と比較して、データセットに基づく研究は、データの取りこぼし等の不確実性が排除されることで追試可能性が高まるとともに、複数の研究チームによってデータの多様な解釈がなされることが期待できる。

Wikipedia を対象とした研究が活発な理由は、その構築プロセスの特異性やデータ量の多さもさることながら、定期的にアーカイブされたデータベースを自由に利用できるからという側面がある。また、Wikipedia のデータを RDF で入手可能な DBpedia^{*3} というサービスもある。こういった研究のインフラに基づいて、国内でも Wikipedia に関する研究を集めたワークショップが開催されている^{*4}。

他にも、完全なオープンデータではないものの、Yahoo! 知恵袋のデータを分析対象とする知識共有コミュニティワークショップ^{*5} や、科研費「情報爆発プロジェクト」向けに Yahoo! 検索の検索語データが提供されるなど、データセットの整備が進んでいる。

3.2 ウェブ化する研究プロセス

ウェブは分散型のアーキテクチャを持ち、その進歩はボトムアップである。このウェブを対象とした研究のプロセスもまた、分散的かつボトムアップの方向に進みつつある。とくに、ユーザ参加型サービスに関する研究では、データが存在する場所を中心にコミュニティが発生し、その中で開発者と研究者による議論が行われている。

昨年開催された SBM 研究会^{*6} では、サービスを提供する事業者と SBM のデータを利用する研究者が一堂に会し、データセットの必要性や詳細なフォーマットに関する議論が行われた。この結果を踏まえて、ライブドアは自社が運営する SBM である livedoor クリップのデータを研究用に公開した^{*7}。

*3 <http://dbpedia.org/>

*4 http://sigsw.org/A803_program.html

*5 http://www.infosocio.org/cfp_workshop_a2009.html

*6 <http://homepage3.nifty.com/toremoro/study/SBM2.html>

*7 <http://labs.edge.jp/datasets/>

また、ニコニコ動画のデータをさまざまな側面から分析する研究会も行われ*8、開発者、研究者がそれぞれの観点から分析結果を発表した。

これらの研究会は Ustream や Twitter といったウェブサービスを通じて中継されたため、大きな注目を集めるとともに、リアルタイムの反応やコメントが発表者に返ってくる点で通常の研究会とは異なる雰囲気の中で行われた。

このように、研究プロセスそのもののウェブ化は着実に進行している。Nature が通常の査読つき論文誌とは別に、事前のピアレビューがなく一般のユーザからコメントを受け付ける Nature Precedings*9を開始するなど、オープン化の試みが各所で行われている。Wikipedia と信頼性を比較された Encyclopedia Britannica においても、一定の条件の下でユーザの参加によるコンテンツの登録・修正が可能となった。

筆者らも、国立情報学研究所が運営する論文情報ナビゲータ CiNii*10において 2000 万件以上の書誌情報をオープンデータとして公開しており [大向 08]、今後は論文情報から得られる研究者のソーシャルネットワークを提供する予定である。

これらのインフラを活用することで、研究対象としてのウェブを捉えるのではなく、研究者自らが利用し、新たな可能性を模索できる環境を構築するのが今後のテーマである。

4. おわりに

本論文では、5 年間のチャレンジの成果について概説し、今後の展望について議論を行った。本チャレンジでは、当初の想定よりも早くウェブとコミュニティの融合がなされたことで、大規模な分析や新たなサービスの提案に関する研究発表が多数行われた。今後は、研究者自身による研究環境の構築を目的として活動を継続する所存である。また、今年度のセッションで発表される 12 件の論文を加えたうえで、最終的な分析と総括を行う予定である。

参考文献

- [松尾 07] 松尾 豊, 安田 雪: SNS における関係形成原理 -mixi のデータ分析-, 人工知能学会論文集, Vol. 22, No. 5, pp. 531-541 (2007)
- [大向 06] 大向 一輝, 松尾 豊, 松村 真宏, 武田 英明: Community Web プラットフォーム, 人工知能学会論文集, Vol. 21, No. 3, pp. 251-256 (2006)
- [大向 08] 大向 一輝: 学術情報サービスのユーザモデルとファインダビリティ, 情報の科学と技術, Vol. 58, No. 12, pp. 595-601 (2008)

*8 <http://nicovideo.g.hatena.ne.jp/Yoshikawa/20081219/1229678324>

*9 <http://precedings.nature.com/>

*10 <http://ci.nii.ac.jp/>