

Community Web プラットフォームのデザイン

Designing Community Web Platform

大向 一輝*1*2

Ikki Ohmukai

武田 英明*1*2

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

In this paper we review a purpose of "Community Web Platform" project proposed in 2005. Recently user participatory services such as Blogs and SNS became mainstream on the Web, and restructuring of their information is required to make for smoother communication among people. We propose a next target of our project, the realization of light-weighted semantic web and social graph, and describe how to operate the project.

1. はじめに

筆者らは、2005年より近未来チャレンジ「Community Web プラットフォーム」を提案し、以後3年間に渡りセッションを運営している。この3年間に、ウェブの世界は大きく変貌を遂げた。筆者は、本チャレンジの提案時に、それまでに普及の兆しを見せていたブログやソーシャルネットワークサービス(SNS)の存在をふまえて、Community Web プラットフォームについて下記のように述べた[大向 06]。

本研究では、個人の存在が Web 上に表出され、そういった個人間でのコミュニケーションがなされている場を「Community Web プラットフォーム」と定義し、プラットフォームの構築や、プラットフォーム上での情報流通の支援のあり方について考察する。

一方、同年9月には、O'Reillyによる「What Is Web 2.0」[O'Reilly 05]をきっかけとした、いわゆる Web 2.0 ブームが到来し、ユーザ参加型のウェブサービスに対する大きな期待が寄せられるとともに、ビジネスとしての成功を狙ったいくつものサービスが立ち上げられることになった。

このような潮流に対して、一部では批判の声があったものの、ユーザ自身がコンテンツの作り手となり、互いにコミュニケーションを行いながら他の参加者を巻き込んでいく User Generated Contents (UGC) の普及は止まることなく現在も続いている。また、こういった活動をどのように支援するかについて多くの議論がなされ、商用サービスとして軌道に乗っている例も見られる。

この点で、当初 Community Web プラットフォームが提案していた理想像の一部は、すでに実現されたと言ってよい。O'Reillyによる Web 2.0 の定義には項目が7つ存在しているが、モバイルコンピューティングやソフトウェア開発手法への言及を除けば、両者の内容は大きく違わない。

上記のような言説を待つまでもなく、コミュニケーションの場としてのウェブの存在は当たり前ものになっており、ブログは国内で数百万サイト、世界では数億サイトが存在している。また、SNS は会員数1千万を超えるものが多数存在している。

こういった状況に対して、人工知能研究者がどのようにウェブに関わり、何を目標として研究を行っていくべきかについて改めて議論を行う必要があると思われる。本論文では、近年のウェブ研究の進展について概説し、新たに環境の変化や課題について述べ、Community Web プラットフォームの目標とタスクを再検討する。

2. ウェブ研究の現状

2.1 ユーザ参加型メディアに関する研究

ブログや SNS といったユーザ参加型メディアがウェブ研究に与えた影響は、個人が持つ知識や他者へ向けたメッセージを入手可能な形で固定化したことにある。

これまで、一般に入手可能な情報は、書籍や雑誌、新聞など、専門家が何らかの目的を持って不特定多数に向けて発信したものに限られていた。情報の内容は、その性質上事実あるいはできる限り客観性が担保できるものでなければならない。

一方で、私的に交わされるメッセージには、そのような要請はない。主観が含まれていたり、感情が吐露されているものも含めて、多岐にわたる膨大な情報が容易に手に入るようになった。また、こういったメッセージは、多くの場合特定少数に向けられている。その際、発信者と受信者の間には何らかの人的関係が存在すると仮定できる。ユーザ参加型メディアの規模が大きくなるにつれて、メッセージが誰から誰に伝達されたのかという関係のネットワークを把握することができるようになる。

ユーザ参加型メディアに対する初期の研究は、各サービスの人口統計や人的ネットワークの全体像を把握するマクロ的視点によるもの、年齢や性別といったユーザ属性の推定、感情的表現の抽出といったミクロ的視点によるものが多い^{*1}。その後、メッセージの伝達関係を利用したオピニオンリーダーの発見や、スパムの抽出など、ユーザ参加型メディアが持つ特徴に注目した研究が進められた。

現在もこの流れは変わっていない。従来のウェブ研究がウェブページ(あるいはページ内の文字列)とハイパーリンクを最小の構成要素としていたのに加えて、ユーザというエンティティを導入し、ユーザ間のネットワークをどのように構築するか、あるいは既存のネットワークをどのように活用するかという問題に主眼が置かれている。

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

*1 ユーザ参加型メディアに関する初の国際会議「WWW2004 Workshop on the Weblogging Ecosystem」におけるサーベイにて

2.2 データの入手

ユーザ参加型メディアを対象とした研究は、その性質上、多数のユーザならびに膨大なデータを必要とする。とはいえ、研究者自身がサービスを運営し、データを得ることは現実的でない。そこで、外部のサービスからデータを入手する方法としては以下の手段が考えられる。

1. プログラムによる自動収集

ブログ等の公開データについては、自動収集プログラムによってある程度の量を集めることが可能である。しかしながら、得られたデータの網羅性については保証されない。また、会員制の SNS では規約によってデータ収集が認められていない場合もある。

2. サービス提供企業との提携

サービス提供企業、研究者の双方にとってメリットがある場合には、個別の提携によってデータの提供を受けられる可能性がある。ただし、得られる情報は提供元企業のポリシーによって異なる。また、他の研究者が同様の条件で追試することは難しい。

3. コンテスト型研究への参加

サービス提供企業と学会、あるいは研究機関との連携によって行われるコンテスト型研究では、参加する複数の研究チームに同じデータが提供される。得られる情報の制約については 2 と同様である。

近年、3 のコンテスト型研究がさまざまな学会で行われており、研究の場として認知されつつある。自然言語処理の分野では TREC 2006^{*2} から始まった Blog Track を先駆けとして、NTCIR-7^{*3} でもブログデータの提供が行われている。また、ICWSM^{*4} ではブログ検索・調査サービスである BlogPulse のデータが公開されており、さまざまな研究が行われている。他にも、情報処理学会数値モデルと問題解決研究会ならびにソネットエンタテインメント株式会社の共催によるブログデータを用いた情報推薦アルゴリズムのコンテスト^{*5}、国立情報学研究所とヤフー株式会社の提携による「Yahoo!知恵袋」コミュニティのデータ公開^{*6} などがある。

一方、1 については後述するデータのオープン化の流れにより、今後プログラムによる自動収集が容易になる可能性がある。ここで得られたデータを研究者間で共有し、コンテスト型研究の基盤とするといった考え方もあり得る。

3. ウェブの構造化

3.1 ライトウェイト・セマンティックウェブ

ウェブを対象とした研究を行う上で、データに含まれるノイズの除去は大きな問題である。これまで、さまざまな手法を用いてウェブから有用な情報を抽出する試みがなされてきた。どのような情報が有用であるかは研究分野によって異なるが、人工知能分野ではコンピュータが理解できるどうかひとつの重要な尺度となる。

Berners-Lee は、ウェブの理想型として、機械可読な情報で構成されたセマンティックウェブのビジョンを提示している [Berners-Lee 00]。セマンティックウェブでは、XML や RDF による情報の構造化、オントロジーに基づく意味論の導入、推論系の導入によって、ウェブ情報の自動収集や自動処理を目指すロードマップが示されている。

このロードマップに沿って多くの研究者が活動している一方で、あまりに壮大な計画であるとの批判もある。セマンティックウェブのアーキテクチャは推論のために設計されており、その他の処理系の存在を念頭に置いていない。ウェブの成功の要因が、簡便な HTML による情報の記述と、検索エンジンなどの多様なアプリケーションの普及にあったと考えると、推論に特化するのではなく、ユーザが理解できる範囲で情報の構造化を行い、既存のアプリケーションを補うようなアプローチが求められる。このような、ユーザ参加型のセマンティックウェブを「ライトウェイト・セマンティックウェブ」と呼ぶことがある。

ライトウェイト・セマンティックウェブの代表的な技術としては、ブログツールが自動生成する RSS や Atom などのフィードがある。フィードはウェブ情報を外形的に構造化するためのシンプルなフォーマットだが、フィードリーダによる情報の効率的な受信やサイトを越えた情報統合に欠かせない手段として定着している。

一方、マイクロフォーマットは情報の内容を簡便に構造化するためのメタデータである [Allsop 07]。マイクロフォーマットでは、カレンダーや住所録などといった用途に応じて少数の語彙を定義し、XHTML に埋め込んで使用する。各フォーマットの策定は標準化団体などを経ずにコミュニティーベースで行われる。

マイクロフォーマットには名前空間がなく、複数のフォーマットが同じ単語を使用した場合に正しい解釈ができなくなる恐れがある。そこで、名前空間を含む RDF を XHTML に埋め込めるようにする RDFa^{*7} と呼ばれるフォーマットが開発されるなど、狭義のセマンティックウェブとライトウェイト・セマンティックウェブの歩み寄りが見られる。また、マイクロフォーマットと XSLT を組み合わせ、ユーザには簡便な記法を提供し、コンピュータには RDF に変換した情報を提示する GRDDL^{*8} といった仕組みの提案が進んでいる。

3.2 ソーシャルグラフ

一般に公開されているウェブページとは異なり、人的関係についての情報はその大半が SNS に格納されているため、自由に取得して活用することが難しい。2.2 節で述べたように、企業との提携によってデータを得ることも不可能ではないが、人的関係は個人情報に属するため、規約によって実現しないことが多い。

オープンな環境で人的関係を記述するためのフォーマットとしては、RDF ベースの FOAF やマイクロフォーマットベースの XFN が提案されている。ただし、両フォーマットともに具体的な用途が想定されていなかったことから、広く普及するには至らなかった。

しかしながら、ユーザ参加型メディアが注目されるようになり、あらゆるサービスで人的関係を構築、管理する機能が導入されつつある。また、既存の SNS においても、その環境の中で第 3 者がアプリケーションを開発し、ユーザに提供するビジネスが始まっている。個人にとって、複数のサービスで同様の

*2 Text REtrieval Conference. <http://trec.nist.gov/>

*3 NII Test Collection for IR Systems. <http://research.nii.ac.jp/ntcir/>

*4 International Conference of Weblogs and Social Media. <http://www.icwsml.org/>

*5 <http://www.ipsj.or.jp/sig/mps/compe08html.release/contest.html>

*6 <http://research.nii.ac.jp/tdc/chiebukuro.html>

*7 <http://www.w3.org/TR/rdfa-syntax/>

*8 <http://www.w3.org/TR/grddl/>

人的ネットワークを維持管理することは非常にコストが高いため、一元的に管理されたネットワーク情報を、個々のサービスが参照して必要な機能を提供するようなモデルが求められる。ここで作られる人々のネットワークは「ソーシャルグラフ」と呼ばれている [Fitzpatrick 07]。

ソーシャルグラフの実現に向けて、SNS におけるデータの相互運用ならびに人的関係データのオープン化という 2 つのアプローチが進められている。

データの相互運用では、Google や MySpace が中心となって策定された OpenSocial^{*9}が注目されている。OpenSocial では、各 SNS に共通した人的関係データへのアクセス方法を定めることで、開発者は複数のサービスで実行可能なアプリケーションを開発することができるようになる。

一方で、オープンな SNS においては、前述の FOAF や XFN を用いて知人関係を記述しているものがある。Google では、ウェブページをクロールする際に、これらのデータを解釈し、人的関係ネットワークを構築するとともに、開発者に対してこのネットワークへのアクセス手段 (Social Graph API^{*10}) を提供している。

両者のアプローチはすぐに統合されるわけではないが、ユーザの要求が高度化するとともにソーシャルグラフの理想へと近づいていくものと予想される。

4. Community Web プラットフォームのデザイン

Community Web プラットフォームのあり方を再検討するにあたって、これまでに述べてきたライトウェイト・セマンティックウェブやソーシャルグラフの概念は必須の要素になると思われる。そこで、これらを人工知能の諸技術によって実現することをプロジェクトの第 1 の目標としたい。

そのためには、ウェブページや人的関係をユーザにどのようにマークアップしてもらおうかというボトムアップのアプローチと、すでに存在するウェブ情報からいかに有用な知識や関係を抽出するかというトップダウンのアプローチがあり得る。前者については、ユーザインターフェイスあるいはユーザエクスペリエンスに関する研究や、コミュニティ支援の方法論を導入し、後者についてはウェブマイニングの技術が必要である。

これらはウェブ上で構造化データを作るための研究だが、一方でこういった環境が普及した時にどのようなアプリケーションが実現できるかについての探求も必要である。セマンティックウェブが目的としてきたオントロジー構築、情報の信頼評価をユーザの参加によって行う手法などが研究対象として考えられる。

5. おわりに

本論文では 3 年目を迎えた Community Web プラットフォームについて、現実のウェブ世界の状況に対応した目標設定を行った。今後は、この目標を達成するための基盤構築や、コンテスト型研究の設定などを進めていく所存である。

参考文献

[Allsop 07] Allsop, J.: *Microformats: Empowering Your Markup for Web 2.0*, Apress (2007)

[Fitzpatrick 07] Fitzpatrick, B. and Recordon, D.: Thoughts on the Social Graph, <http://bradfitz.com/social-graph-problem/> (2007)

[O'Reilly 05] O'Reilly, T.: What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (2005)

[Berners-Lee 00] Berners-Lee, T., Hendler, J., and Lassila, O.: The Semantic Web, *Scientific American* (2000)

[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)

*9 <http://www.opensocial.org/>

*10 <http://code.google.com/apis/socialgraph/>