

Weblog 研究の現状

The State-of-Art of Weblog Research

武田 英明

Hideaki Takeda

国立情報学研究所

National Institute of Informatics

1. はじめに

Weblog(blog)の利用者はこの半年で急増している。また、社会での認知度もそれを上回るほど増えてきている。一方、研究サイドではあまり興味をもたれてこなかった。元々研究的な要素から始まったものではなく、ユーザの間で作られたものであるし、またあまりに急速な発展に研究サイドがついていけなかったという点もある。しかし、この1年ぐらいで研究サイドにも多様なアプローチでweblogに取り組むようになったに思える。本稿ではそのような研究を概観するとともに、特に Semantic Web との関連についても述べる。

2. Weblog とはなにか

漠然として共通理解はあるものの、確固たる定義はない。詳しくは[1]を参照されたい。

多くの人が共通に考える属性としては

- (1) 短い記事の集まっている
 - (2) 時系列表示されている
 - (3) 頻繁に更新されている
 - (4) 同一(あるいは少数)の人物によって書かれている
 - (5) 他の記事やページへのリンクを多くもつ
 - (6) Weblog ツールによって管理されている
- などがあげられる。(1)-(3)までだとニュースサイトも含まれることになるが、(4)によって分かれる。この意味で(4)の性質は重要である。(1)-(5)までではいわゆる Web 日記も含まれることになる。もっとも Weblog と Web 日記は出自が違うものの本質的な違いはない。(6)をいれると循環論法的定義になってしまうが、実際にはこれが最もわかりやすいかもしれない。なお、このようなツール利用を前提とすることで、ツールが共通に提供する機能(RSS 配信や trackback)などを使うことができる。このような機能を共通に持つことが Weblog が発展した理由のひとつであろう。

3. Weblog 研究の会議

Weblog 研究のみをターゲットにした研究会議は少ない。WWW2004 に併設された WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics¹が私の知る範囲での唯一の会議であった。国内では 2004 年 7 月 2 日に当研究会で行われた特集研究会が最初ではないかと思われる。あとは様々な会議の中で関連研究が発表されている。

4. 研究対象としての Weblog と研究のタイプ

研究対象としての Weblog はどのように位置づけられるのであろうか。一つは Weblog を新しいメディアと考え、このメディアと人間あるいは社会との関わりとして Weblog を分析するものである。これは主に心理学、社会学の範疇の研究である。当然メディアの技術を研究するものもある。このメディアにおける技術の評価や新規技術の提案などである。これは工学的研究といえよう。また、これと関連しているが、Weblog がつくる情報空間の分析もある。これは分析やモデル化という意味では理学的研究に近いが、情報科学の分野では上記 2 つはそれほど区別されて研究されているわけではない。

4.1 心理学, 社会学的研究

これは Weblog を新しいメディアと考え、このメディアと人間あるいは社会との関わりとして Weblog を分析するものである。Web においても同様の位置づけができ、実際にそのような研究も行われているが、Weblog と Web の差異は個人性(2章の(4))と個人のつながりの

¹

<http://www.blogpulse.com/www2004-workshop.html>

強調であろう。Web においても個人のホームページという形で個人が前面にくる Web ページは多数存在するが、そのほかのページと交じり合っていて分離は容易ではない。そもそも著者のほうの意識も明瞭に分かれているとはいいがたい。一方、Weblog では当初より個人性が強調されており、著者の意識も高い。さらに Weblog コミュニティと称されるように、Weblog ユーザ間での繋がりに関する意識も強い。一方、Weblog ツールの利用により、個人間のつながりも含めて、他の Web との分離も容易である。このような特徴が Weblog を心理学、社会学研究の対象として独特なものしていると思われる。

4.1.1 個人と Weblog の関係

これは個人がどのように Weblog に接している（書くあるいは読む）かについての研究である。これについては心理学見地から、川浦、山下、川上が Web 日記においてアンケート分析などを行った研究で調べている [2][3]。近年はそれを Weblog へ拡張して行っている [4] [5] [6]。[7] [8]らは同様の議論を事例の分析により行っている。[9]は Weblog のより形式的なデータからジャンル分析を行っている。

4.1.2 社会と Weblog の関係

さらにアメリカで Weblog が興隆した大きなきっかけは 9.11 事件であったことに象徴されるように、アメリカの Weblog は社会との関係が強い。このためジャーナリズム論の立場から論ずるものもある。[10]は事例分析を通じて Weblog のマスコミへの影響の度合いや早さなどを議論している。

4.2 技術的研究

4.2.1 Blog 空間の分析と探索

技術的な文脈でまず注目されたのは膨大な Weblog 情報の分析である。基本的には Web 空間で行った研究の技術が利用できるが、Weblog の特徴的なのは時間情報がある点である。

[11]は既存の Web Community の概念に時間をいれることで community 進化の分析などを行っている (図 1 参照)。日本でも [12]が同様の分析を試みている。

[13]はより詳細に blog 空間を巨視的および微視的に分析している。Weblog の突発性では、“スラッ

シュドット”効果を含む 4 つのタイプに分類できることを示した。また、Weblog では本来参照されるべきだが、リンクとしては張られていない関

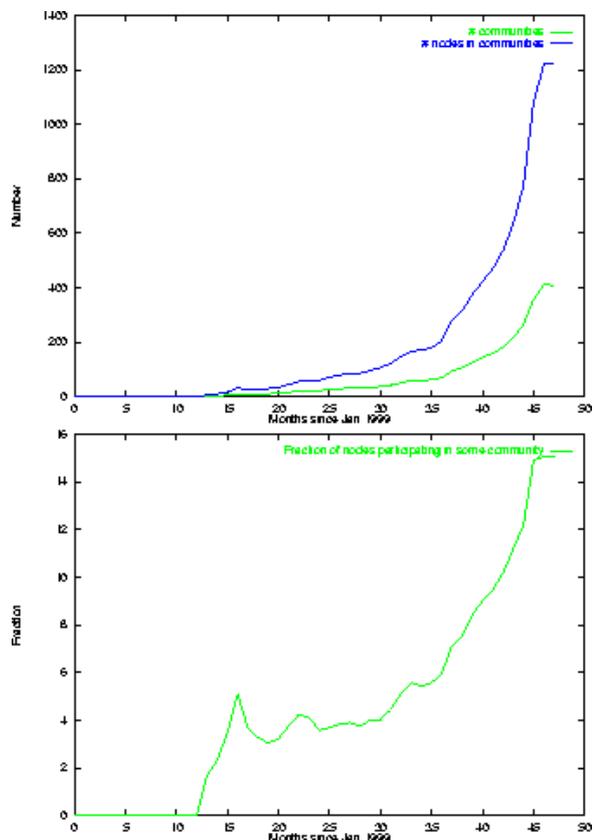


図 1 Community Evolution [11]

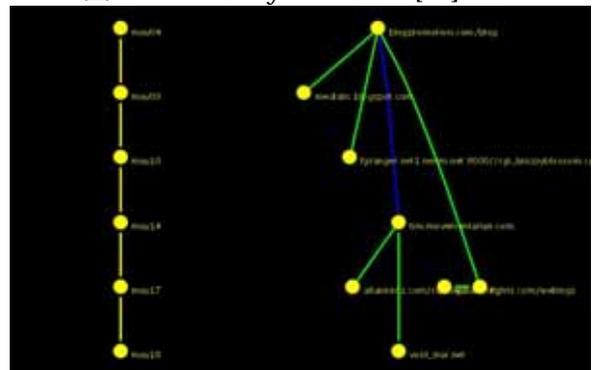


図 2 Epidemic Analyzer の出力例

係というのが多く見られるということを指摘して、その関係を推定するということを行っている。図 2 にその発見した関係を図示したものである。青い線は Blog 同士でリンクがあるもの、緑は推定した関係である。



図3 BlogPulse の出力例

集められた Weblog 情報から何を知るかについてはいろいろなことが行われている。BlogPulse[14] は収集した blog からトレンド解析ができるような仕組みを開発している。図3のその出力例を示す。[15]でも収集した blog から突発性の度合いなどを可視化している。

さらに評判情報だけを取り出す[16]、Web ページの評価に用いる[17]なども行われている。

なお、公開フォーマットが一定しているために、いわゆる Weblog の収集は技術的な困難点はあまりないが、日本で多く見られる Web 日記系まで含めるととたんに難しくなる。奥村ら[18]は多様な日本の Weblog / Web 日記を収集することに成功している。

4.2.2 新しいツール、ソフトウェア

Weblog が普及した要因の一つはツールの整備である。それらのツールは基本的にエンドユーザ側でのニーズにこたえる形で発展してきた。このためツールに関する研究は多くない。

Semblog[19]では個人の情報流通の枠組みとして Weblog に注目している。RSS と FOAF を統合してもちい、その上で多様な機能をもつアグリゲータやアプリケーションを提供している。

Weblog を拡張あるいは別の情報と組み合わせるといった方法も試みられている。「場 log」では GPS 付き携帯電話の写真を利用して、位置データを利用したアグリゲーションを提案している[20]。[21]では Weblog 情報から得られる関係を可視化することで、コンテンツブラウジングの支援を試みている。

5. Semantic Web との関連

Weblog の技術的な側面として重要なのは、メタデータの一つである RSS による情報流通の実現である。多くのツールが RSS に対応することでユーザは意識することなくメタデータ付きの情報

を公開している。RSS はメタデータとして議論の余地があるものではあるが、まずは普及させたという点が重要である。これにより、単に各サイトの情報を必要時に見に行くのではなくて、適宜巡回したり必要なものだけとりだすといったアグリゲーションという利用が実現されている。

RSS そのものは含まれる情報が乏しく、セマンティック Web がめざすメタデータとはいえない。しかし、メタデータ流通の基盤としては利用可能である。

また、セマンティック Web でも重要なテーマとなっている Web の信頼性においても関係がある。Weblog は出自からして個人性が強く、それゆえ Weblog 間の関係を調べることで、個人関係のネットワークを構築することができる。Web の信頼性のひとつの方法はこのような個人関係のネットワークを利用することである。すなわち、Weblog から得られる個人関係を利用することで、信頼性の問題まで含めたセマンティック Web が実現できる可能性がある。

現実的な面でも Weblog はセマンティック Web にとってチャンスであるといえる。セマンティック Web をいまある Web に適用しようとするといういろいろな困難な点が多い。しかし、現在 Weblog による Small contents の流通が普及しつつある。これは新しいフォーマットの普及であり、過去のコンテンツを気にする必要がない。このようなときにこの Small contents の流通にセマンティック Web の技術を埋め込むことができれば、自然と普及することが期待される。

6. おわりに

本稿では Weblog をめぐる研究動向について概観した。Weblog に関する研究は始まったばかりであり、Weblog 自身の発展とともにこれから様々な発展することが期待される。そして、その発展のなかにうまくセマンティック Web の技術や方法論が入れることができれば、Weblog およびセマンティック Web 双方にとって利益になることだと思われる。

参考文献

- [1] 武田 英明, 大向 一輝: Weblog の現在と展望 -セマンティック Web とソーシャルネットワークの基盤として-, 情報処理, Vol.45 No.6, 2004
- [2] Kawaura, Y., Kawakami, Y. and Yamashita, K.: Keeping a diary in cyberspace, *Japanese Psychological Research*, Vol. 40, No. 4, pp.

234 245 (1998).

[3]川浦康至, 山下清美, 川上善郎:人はなぜウェブ日記を書き続けるのか, 社会心理学研究, Vol. 14, pp. 133 143 (1999).

[4] 山下清美:Weblogの心理学, 第6回人工知能学会セマンティックWebとオントロジー研究会, (2004)

[5] 山下清美, 三浦麻子:人はなぜウェブ日記・ウェブログを書き続けるのか(1), 日本社会心理学会第45回大会 (2004).

[6]三浦麻子, 山下清美:人はなぜウェブ日記・ウェブログを書き続けるのか(2), 日本社会心理学会第45回大会, (2004).

[7] Nardi, B. A., Schiano, D. J., Gumbrecht, M. and Swartz, L.: "I m Blogging This A Closer Look at Why People Blog, *Communications of ACM* (2004), (To appear).

[8] Schiano, D. J., Nardi, B. A., Gumbrecht, M. and Swartz, L.: Blogging by the Rest of Us, in *Proceedings of the CHI2004* (2004).

[9] Herring, S. C., Scheidt, L. A. and Sabrina Bonus, E. W.: Bridging the Gap: A Genre Analysis of Weblogs., in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS 04)*, Big Island, Hawaii (2004).

[10] Gill, K.: How can we measure the influence of the blogosphere?, in Glance, N., Adar, E., Hurst, M. and Adamic, L. eds., *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).

[11] Kumar, R., Novak, J., Raghavan, P. and Tomkins, A.: On the bursty evolution of Blogspace, in Hencsey, G., White, B., Chen, Y.-F. R., Kovács, L. and Lawrence, S. eds., *Proceedings of the twelfth international conference on World Wide Web (WWW2003)*, pp. 568 576, ACM Press (2003).

[12] 谷口智哉, 松尾豊, 石塚満:Blogコミュニティの抽出と分析, 第6回人工知能学会セマンティックWebとオントロジー研究会 (2004).

[13] Adar, E., Zhang, L., Adamic, L. A. and Lukose, R. M.: Implicit Structure and the Dynamics of Blogspace, in Glance, N., Adar, E., Hurst, M. and Adamic, L. eds., *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).

[14] Glance, N. S., Hurst, M. and Tomokiyo, T.:

BlogPulse: Automated Trend Discovery for Weblogs, in Glance, N., Adar, E., Hurst, M. and Adamic, L. eds., *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).

[15] 奥村学, 南野朋之, 藤木稔明, 鈴木泰裕:blogページの自動収集と監視に基づくテキストマイニング, 第6回人工知能学会セマンティックWebとオントロジー研究会, (2004)

[16] 鈴木泰裕, 高村大也, 奥村学:WebLogを対象とした評価表現抽出, 第6回人工知能学会セマンティックWebとオントロジー研究会, (2004)

[17] 中島伸介, 竹原幹人, 日野洋一郎, 舘村純一, 原良憲, 田中克己:blog解析に基づくWeb情報検索の信頼性向上技術, 第6回人工知能学会セマンティックWebとオントロジー研究会, (2004)

[18] Nanno, T., Suzuki, Y., Fujiki, T. and Okumura, M.: Automatic Collection and Monitoring of Japanese Weblogs, in Glance, N., Adar, E., Hurst, M. and Adamic, L. eds., *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).

[19] Ohmukai, I., Takeda, H., Numa, K., Hamasaki, M. and Adachi, S.: Personal Knowledge Publishing Suite with Weblog, in *Workshop on Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Held in conjunction with the Thirteenth International World Wide Web Conference (WWW2004)* (2004).

[20] 上松大輝, 沼晃介, 徳永徹郎, 大向一輝, 武田英明:場 log: Weblog 環境における位置情報利用の提案, 第 6 回人工知能学会セマンティックWebとオントロジー研究会, (2004)

[21] 是津耕司, 日野洋一郎, 中島伸介, 門林理恵子, 呉受妍, 林正樹, 田中克己:Weblog 情報を融合したコンテンツブラウジング, 第 6 回人工知能学会セマンティック Web とオントロジー研究会, (2004)