

Semblog:メタデータを用いた Web コンテンツの 再編集・共有プラットフォーム

大向 一輝

国立情報学研究所

[E-mail: i2k@nii.ac.jp](mailto:i2k@nii.ac.jp)

要約：本稿では Weblog (blog) をめぐる言説とテクノロジーについて概説し、メタデータ流通プラットフォームとして我々が提案している Semantic Weblog (Semblog) の取り組みについて述べる。Semblog プロジェクトでは、Weblog によりコンテンツの書き手と受け手、および編集者の距離が縮小していることに注目し、オントロジー等による意味的アプローチとは異なるコミュニティ指向の情報の組織化を目指す。

キーワード：コミュニティ、Weblog、メタデータ

1. はじめに

本研究ではセマンティック Web 技術を利用した個人のためのコンテンツ流通支援システムを提案する。コンテンツの流通プロセスは、その作成および公開だけにとどまらず、前段階における情報収集を含めたサイクルとして捉える必要がある。しかしながら、現在の Web 環境においては流通プロセス全体を統合的に支援する枠組みが用意されていない (Shneiderman, 2002)。

本研究では、上で述べた問題へのアプローチとして Weblog による情報流通について検討を行う。提案システムでは、有効なコンテンツ流通を提供するための基盤として、RDF Site Summary (RSS) を利用する (RDF, 2001)。RSS は Web サイトの概要を記述するために提案されたメタデータ規格である。RSS には Web サイトのタイトルや作成者といった全体的な属性と、サイト内の各コンテンツの概要や更新時間を記述することができる。すでに一部の Web サイトでは RSS の配信が行われており、これを利用して各サイトが配信する RSS を収集し、これを整形することで多くの情報を短時間に閲覧するアグリゲータと呼ばれるアプリケーションやサービスも生まれている。そこで、本研究ではこれらのアプリケーションを拡

張し、情報流通プロセス全体の支援を目指す。

2. Weblog による情報流通

近年の Web において、個人が運営する Weblog サイトが注目されている。Weblog についての明確な定義は存在しないが、一般的には雑記や他サイトへのリンク、それに関するコメントが日々更新されるようなサイトの総称であるとされている (Blood, 2002)。Weblog は情報の受け手であった人々を、再編集という手順を通して情報の送り手に変えるという働きを持っている。

多くの Weblog サイトでは Weblog ツールと呼ばれるコンテンツマネジメントシステム (CMS) が導入されている。Weblog ツールは Web ブラウザ上でのコンテンツ記述・編集を可能にし、その結果は即座に HTML 化されて公開される。これにより、情報公開のためのコストは従来の方法と比較して劇的に低減する。このコストの低減が、スモールコンテンツの生産を可能にしているといえる。

また、Weblog ツールは HTML と同様に RSS を自動生成することが可能である。作成者等の属性はあらかじめユーザに初期設定として Weblog ツールに入力させたものを埋め込み、各コンテンツの概要、更新時間および RSS が指す HTML ファイルの URI 等はコンテンツが入力された際に自動的に記述される。Weblog ツールによって、新たなコストをかけることなく RSS を配信することができるため、一般のユーザにおいてもメタデータの効用が得やすくなっている。

3. Semblog プラットフォーム

本研究ではメタデータと Weblog ツールを用いてユーザの情報収集から生産、公開までを統合的に支援するための「Semblog: Semantic Weblog」システムを提案する。このシステムを利用することで、ユーザは各人の視点に基づく情報収集および情報発信を容易に行うことが可能になる。

本研究では、情報収集および情報発信に際して“Check”、“Clip”、および“Post”という3レベルの興味の度合い (Degree of Interest) を定義し、興味の強さに応じて情報の配信プロセスを変える。

最も弱い“Check”レベルとは、ユーザーが特定の Web サイトや情報ソースに日常的にアクセスすることを意味する。ユーザはその Web サイトのコンテンツ内容をあらかじめ知っているわけではないが、過去の更新履歴からどのような情報が掲載されるかを知っている。本研究では、このような知識が情報流通において重要な働きを持つと考え、ユーザが日常的

に巡回する Web サイトのリストを公開することで、そのユーザがどの分野に興味を持っているかを表明するための支援を行う。

次の“Clip”レベルとは、ユーザが閲覧したコンテンツの中でとくに興味があったものを指定し、保存することで、後日同じコンテンツに再びアクセスしやすいようにすることを意味する。本システムでは、“Check”レベルで登録された Web サイトに含まれるコンテンツの中でユーザがとくに興味があるものを指定することで、指定されたコンテンツのリストを作成、公開する支援を行う。

最後の“Post”レベルとは、非常に強い興味を持つコンテンツに対してそれを引用し、コメントを付加して新しい情報として発信することを意味する。ここでは、単なる興味だけでなく、それに伴う意見の表明がなされる。本システムにおいては、“Post”レベルの情報発信は Weblog ツールが担うものとし、その前段階の情報収集プロセスについての支援を考える。

4. システム構成

本システムの構成を以下に示す。本システムはサービス型とクライアント型の2種のRSSアグリゲータおよび検索用プログラム、そして Weblog ツールから構成される。個々のモジュールはRSSによってデータの交換が行われる。Weblog ツールには MovableType などの既存のシステムを利用する。

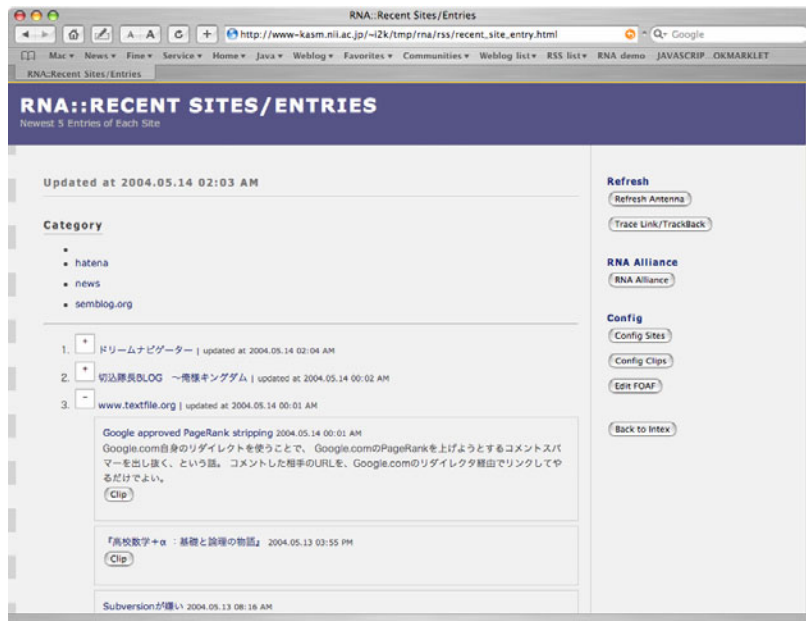
4.1. RNA : RSS 収集・配信サービス

RNA は Perl で記述された CGI プログラムである。ユーザは自身が持つ Web サーバにこれを設置して運用することができる。スクリーンショットを図1に示す。

RNA のユーザは最初に RSS の登録を行う必要がある。他サイトが配信している RSS の URI を設定すると、RNA は HTTP 通信によってファイルを取得する。登録サイトには分類のためにカテゴリを設定することができる。登録サイトのリストは RSS 化され、他のアプリケーションで使用することができる。

RNA は登録された複数の RSS を取得後、サイトごとの最新記事を抽出したもの、サイトにかかわらず更新時間順にコンテンツを並べるものといった3種類のツリーを生成する。生成されたツリーは、そのまま新しい RSS として配信するほか、Web によって可視化することが可能である。RNA で表示するコンテンツのうち、ユーザが興味を持ったものに対しては、1クリックでクリップリストに登録することができる。クリップされたコンテンツは独自の RSS ツリーに格納され、その他の RSS と同様に配信される。通常のツリーは内容が刻々と変化していくが、クリップのツリーからは情報が消されることはない。

図1 RNA



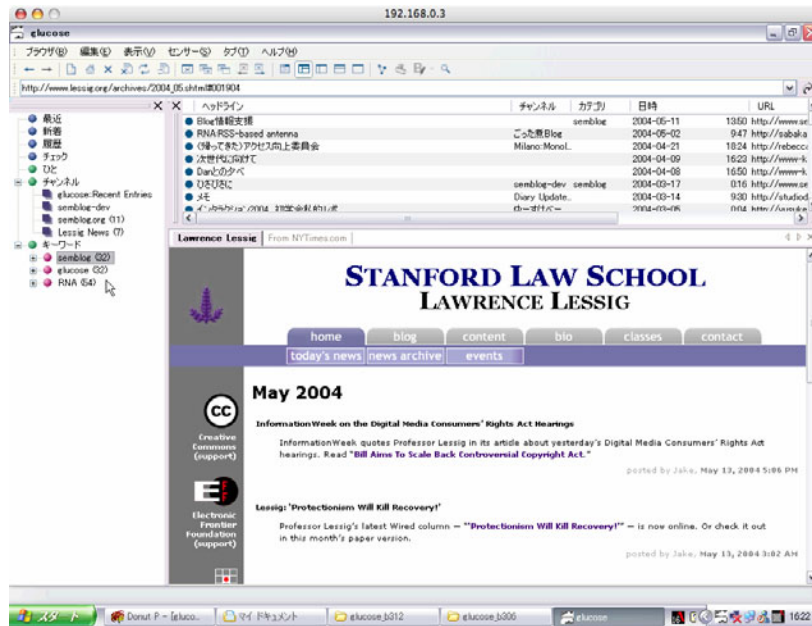
RNA は取得したコンテンツのそれぞれについて TrackBack（逆リンク）の有無をシステムに問い合わせ、存在する場合にはこれを抽出する。抽出されたリンク情報は新たなメタデータとして配信時に追加される。

4.2. glucose : クライアント型 RSS アグリゲータ

glucose は Windows PC 上で動作するクライアント型 RSS アグリゲータである。既存のクライアント型アグリゲータと異なり、glucose では RNA との連携によって情報の流通プロセスを支援することを目指して開発されている。スクリーンショットを図2に示す。

ユーザは RNA と同様に他サイトが配信する RSS の URI を登録する。glucose によって取得された RSS は展開され、3 ペインのインターフェイスによって表示される。左ペインは RSS を配信するサイトのリスト（チャンネル）である。右上のペインには各コンテンツのタイトル、更新日時、サイト名等のリストが表示されており、各項目によってソートすることが可能である。右下のペインには選択されたコンテンツの内容が表示される。また、ティッカー（電光掲示板）機能により、ユーザに対してプッシュ形式で情報を伝えることも可能である。

図2 glucose



RNA と同様に各コンテンツについて TrackBack を抽出することが可能である。抽出されたリンクは右上のペインでメーラの「Re:」表示と同じように表示される。また、リンク先のコンテンツは glucose が先読みすることで、快適に閲覧することができる。

興味のあるコンテンツについてユーザ自身の Weblog に記事を追加する場合には、glucose の Weblog インターフェイスを用いて直接投稿することができる。

5. Semblog プラットフォーム上のアプリケーション

われわれは Weblog ツールと RNA および glucose によって構築される RSS 流通環境を Semblog プラットフォームと呼ぶ。Semblog プラットフォームでは RSS を用いた情報収集から Check 型、Clip 型、Post 型の情報配信を行い、その結果が再び RSS として流通するというプロセスが作られる。このような RSS による情報流通プラットフォーム上での応用例として、複数の RNA を用いた情報推薦手法を提案する。

ここでは、個々の RNA を識別するために、Friend Of A Friend (FOAF) 形式のメタデータ

を用いる (Brickley & Miller, 2002)。FOAFはRDFによって人間関係を記述するためのメタデータフォーマットである。FOAFには本人の名前、メールアドレス、WebページのURIといった基本要素とともに、あるユーザAが別のユーザBを知っている状態をA knows Bという形式で記述する。

このネットワークを利用して、個々のRNAに登録されているサイトもしくはクリップの違いに基づく情報推薦を行う。ユーザは提示されたサイトを1クリックで自身のRNAに登録することができる。また、クリップされたコンテンツのRSSツリーに対しても同様の手法を適用することが可能である。

他にも、ここで構築されたネットワーク上のコンテンツの検索にあたり、ユーザからのホップ数に応じてスコアを変化させるEgocentric Searchといった手法の提案も行っている (Ohmukai, Numa, & Takeda, 2003)。

6. おわりに

本研究では、メタデータとWeblogを利用した情報流通プラットフォームについて提案を行った。提案システムではWeblogツールによってユーザに負担をかけることなくRSSやFOAF情報を配信する。また、Web上のコンテンツを多様化するために、“Check”、“Clip”および“Post”という3種の興味に応じた情報配信や、パーソナルオントロジーの構築を行う。提案システムは2種のRSSアグリゲータとWeblogツール、および応用サービスから構成される。本研究で開発されたアグリゲータは20,000以上のダウンロード数を記録している。今後は、パーソナルオントロジーが情報流通に与える影響について実証実験を行う予定である。

参考文献

- Blood, R. (2002). *The weblog handbook: Practical advice on creating and maintaining your blog*. Cambridge, MA: Perseus.
- Brickley, D., & Miller, L. (2002). FOAF vocabulary specification. <http://xmlns.com/foaf/0.1/>
- Ohmukai, I., Numa, K., & Takeda, H. (2003). *Egocentric search method for authoring support in semantic weblog*. Paper presented at the Workshop on Knowledge Markup and Semantic Annotation (Semannot 2003).
- Ohmukai, I., Takeda, H., Numa, K., Hamasaki, M., & Adachi, S. (2004). *Personal knowledge publishing suite with weblog*. Paper presented at the Workshop on Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Held in conjunction with the Thirteenth International World Wide Web Conference (WWW2004).

RDF Site Summary 1.0 Specification Working Group. (2001). *RDF site summary (RSS) 1.0*.
<http://web.resource.org/rss/1.0/spec>

Shneiderman, B. (2002). *Leonardo's laptop: Human needs and the new computing technologies*. Cambridge, MA: MIT Press.

赤門マネジメント・レビュー編集委員会

編集長 新宅 純二郎

編集委員 阿部 誠 粕谷 誠 片平 秀貴 高橋 伸夫 藤本 隆宏

編集担当 西田 麻希

赤門マネジメント・レビュー 4巻12号 2005年12月25日発行

編集 東京大学大学院経済学研究科 ABAS/AMR 編集委員会

発行 特定非営利活動法人グローバルビジネスリサーチセンター

理事長 高橋 伸夫

東京都文京区本郷

<http://www.gbrc.jp>