

Semblog: セマンティックウェブ技術を用いたスモールコンテンツの

再編集・共有プラットフォーム

1. 背景

本プロジェクトではセマンティックウェブ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。とくに、**Web** 上に存在する膨大な情報の中から必要なものを抽出し、新たな情報の生成に寄与させる手法については、大規模な検索エンジンを利用する以外に実用的なものを見いだすことができていない。

このような問題に対し、セマンティックウェブでは、Web 上のコンテンツについて機械可読なメタデータを付加し、エージェントによる情報処理を促進することを目標としている。このビジョンの実現に向け、W3C をはじめとして多くの研究および開発がなされている。一方、現状のセマンティックウェブ技術が抱える問題点として、ユーザに対しどのようにメタデータを記述させるかといういわゆるオーサリング技術の不足が指摘されている。セマンティックウェブの要素技術である Resource Description Framework (RDF) や RDF Schema (RDFS)、あるいは Web Ontology Language (OWL) は、XML から由来する記法の複雑さだけでなく、適切な語彙の選択の難しさを内包している。これらの言語を専門家だけではなく一般のユーザへ普及させるためには、アノテーションやマークアップを容易にすることが重要である。

2. 目的

本プロジェクトではセマンティックウェブ技術と 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」システムを提案する。Semblog では RSS 等の”Light Weight”なメタデータにより情報流通を活性化することを目指し、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。その際には、ユーザの興味を”Check”, ”Clip”, および”Post” という 3 レベルに分類し、興味の強さに応じて情報の配信プロセスを変える。

本システムでは、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。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、これを利用したアプリケ

ーションやサービスも生まれている。アグリゲータと呼ばれるアプリケーションは、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ことを可能にしている。

3. 開発内容

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「RNA」「glucose」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。また、動的に他のモジュールを呼び出す場合には XML-RPC プロトコルによる通信を行う。Weblog ツールには MovableType などの既存のシステムを利用する。これらのツールは RSS ならびに XML-RPC をサポートしている。

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。RNA の基本機能およびインターフェイスを以下に述べる。

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。また、アグリゲータのサイトリストの標準フォーマットである OPML の読み込み、書き出しにも対応している。

RNA は登録された RSS を取得後、パース処理を行い、複数の RSS ツリーから 1 つの「global」RSS ツリーを構築する。global RSS ツリーは取得された全ての情報が格納されている。次に、RNA はコントローラの要求に応じて global ツリーを加工し、部分ツリーを生成する。ここでは、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。また、ユーザはルールを記述したプラグインスクリプトを用意することで自由に部分ツリーを生成することができる。

生成された部分ツリーは、そのまま新しい RSS として配信するほか、XSL スタイルシートを用いて Web ブラウザ側もしくはサーバ側の XSLT エンジンによって可視化することが可能である。また、RNA 内部の HTML 変換エンジンによって、ユーザがテンプレートファイルを用意することで部分ツリーを HTML 化することも可能である。ここで用いられるテンプレートは HTML と類似したものになっており、XSL スタイルシートよりも理解しやすく一般ユーザにもカスタマイズしやすいものになっている。

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。

RNA は取得したコンテンツのそれぞれについて後述の TrackBack 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。また、Description 内に記述されているハイパーリンクを同様に抽出する。抽出されたリンク情報は新たなデータとして配信時に追加される。

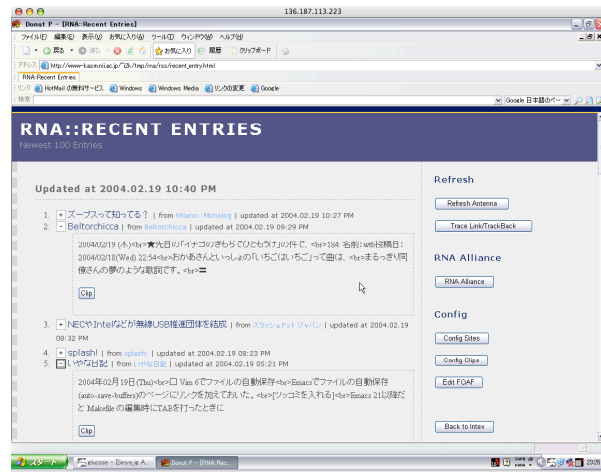


図 1 RNA

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。主な機能、インターフェイスを以下に述べる。

RNA と同様に他サイトが配信する RSS の URI を登録する。OPML の入出力にも対応する。また、RSS を配信していないいくつかのニュースサイトについてはセンサープラグインという Python スクリプトによって記事を切り出し、RSS 化することが可能である。

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。

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ポストすることができる。このインターフェイスには XML-RPC を利用している。

Weblog へのポスト機能と同様に、ユーザの持つ RNA のクリップに情報を追加することができる。

glucose は P2P のサーバントとして動作し、glucose を起動しているユーザ同士でネットワークを構成する。P2P ネットワーク内では閲覧されたコンテンツの URI 情報がランダムに配信され、他のユーザの glucose に対してコンテンツ推薦を行う。その中に興味のあるコンテンツがある場合には、クリップやポストによって反映させるほか、そのサイトの RSS を glucose に登録することができる。

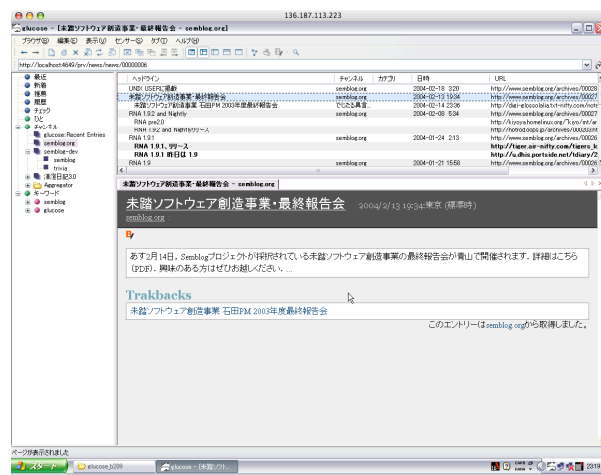


図 2 glucose

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、本プロジェクトでは 2 種のアプリケーションを提案する。1 つは複数の RNA を用いた情報推薦であり、もう 1 つは RSS に記述されたリンク関係を利用した検索である。

4. 開発物の特徴

RNA について

- ・実行環境に対する依存性の低減

RNA の最大の特徴は Pure Perl で実装されたスクリプトであることである。多くの類似ソフトでは、サーバ環境によって必要なライブラリが存在せず、それゆえユーザが利用できないことが多い。RNA では、XML および RSS の処理を行うパーサを Perl 言語で実装し、問題のあるライブラリとの依存関係を解消している。この結果、ラ

イブラリの追加が不可能なレンタルサーバ等の環境での動作を確認することができた。

- テンプレートシステム

個人用の CGI スクリプトにおいては、ユーザの好みに応じてデザインを変更することが可能であることが望ましい。RNA では MVC モデルに基づくテンプレートシステムとして実装することによりカスタマイズ性を向上させている。ユーザは特殊な変数の入った HTML ファイル（テンプレート）を記述しておくことで、システムが RSS の各要素を変数に置き換え、HTML としてファイルを生成する。

- プラグインアーキテクチャ

RNA はプラグインアーキテクチャを採用しており、複数の RSS から新たなツリーを構築するにあたりルールベースのスクリプトを追加することができる。プラグインによって生成された RSS は前述のテンプレートシステムによって自動的に HTML に変換される。これにより、カテゴリ毎にフィルタリングされたコンテンツのリストを作成することや、複数のツリーを一括で表示するような新たなインターフェイスを構築することが可能になる。

- バーチャルコミュニティポータル

多くの類似サービスでは、ユーザが登録した RSS のリストを外部から参照することはできず、ユーザのもつ Weblog サイトに埋め込むためのスクリプトが発行されるだけのものが多い。すなわち、これらのサービスは個人のページ閲覧の効率を高めるか、もしくは個人が発信する情報の内容を増やすために利用されているといえる。一方、RNA ではそれ自身が 1 つのサイトとなり、登録された RSS のリスト等を全てアクセス可能にする。この特徴を生かして、Weblog サイトを持っている個人同士のコミュニティのポータルサイトを容易に構築することができる。このポータルサイトは仮想的なものであるため、各 Weblog が複数のコミュニティに属するものであっても何ら影響を与えない。

glucose について

- プッシュ型情報ブラウジング

glucose では登録されたサイトのニュースをランダムに表示するティッカー機能を搭載している。これにより、能動的な情報取得とは異なるプッシュ型のブラウジングが可能になる。

- 各種サービスとの連携

glucose では RSS を利用した外部サービスとの連携を積極的に進めている。Glucose 単体では外部への情報公開が不可能であるという問題の解決には RNA を利用し、登録した RSS 以外の情報を得るためには Feedback や Bulkfeeds 等の全文検索エンジンを利用することでキーワード検索が可能になっている。ユーザはこの検索結果から RSS を登録することも可能であり、これによって容易にチェックするサイトの数を増加させることができる。

5. 今後の課題、展望

・ユーザビリティの向上

RNA や glucose は Web コンテンツの閲覧を補助するものである以上、既存の Web ブラウザを超えるユーザビリティを提供する必要がある。ユーザビリティについては近年タブブラウザと呼ばれるアドオンが普及しているが、glucose においてもこれらと同等の利便性を提供し、さらなるユーザの増加を目指す。また、今後の FOAF 連携機能を進めるにあたっては、従来の 3 ペインインターフェイスが適さないことも考えられるため、新たなユーザインターフェイスの設計が必要である。

・入力インターフェイス

Weblog ツールへの直接の書き込み機能を提供している RSS リーダは少ない。glucose は MovableType と同一の書き込み機能を提供することで優位性を保っている。しかしながら、これらのインターフェイスは汎用であるがために支援の余地が少ない。今後は用途に応じた複数のインターフェイスを用意し、コンテンツの入力自体を支援することを目指す。

・ P2P

glucose に内蔵された P2P 配信の機能向上として、ユーザのクラスタリングによるコンテンツ推薦手法を考える。

・ 実験

すでに RNA の総ダウンロード数は 1500 以上、glucose は 12000 程度となっているが、今後はこれらのソフトウェアの有効性を検証するために、ユーザの協力を得て実証実験を行う。

6. 開発者

大向 一輝（開発代表者）国立情報学研究所 i2k@grad.nii.ac.jp

濱崎 雅弘（共同開発者）国立情報学研究所 hamasaki@grad.nii.ac.jp

安達 真（共同開発者）早稲田大学 shn@suou.waseda.jp

参考

関連Webサイト：<http://www.semblog.org/>