

J-021

デジタルコンテンツ流通に向けた批評メタデータの自動生成と流通システムの実装と評価 Implementation and Evaluation of the Automatic Criticism-Metadata Generation and the Distribution System for Digital Contents

大勝 琢巳†
Takumi Ookatsu

亀山 渉‡
Wataru Kameyama

1. まえがき

地上波デジタル放送の開始やブロードバンドインフラの普及により、デジタルコンテンツが増大している。このような環境の中で、ユーザがコンテンツを消費するための判断材料として、コンテンツプロバイダが提供する情報がある。これ以外に、一般ユーザの声である口コミと呼ばれる情報も重要であると考え。そこで、口コミを流通させる方法として、一般ユーザの声を、blog を利用して収集し、収集した情報を TV-Anytime Forum[1]の規格に準拠したメタデータ構造に再構築を行うシステムを筆者らは提案している[2,3]。本稿では、その実装の完了、およびその評価について報告する。

2. 提案システムの概要

本研究のシステム概要を図1に示す。本システムでは、大きく分けて2つの種類のblogサーバで構成されている。1つ目は、口コミ情報の収集・解析、そしてメタデータの生成を行う『メタデータ生成用blogサーバ』である。2つ目は、ユーザがコンテンツの評価情報を投稿する『評価投稿用blogサーバ』と呼ばれるものである。

このメタデータ生成用blogサーバには、評価対象となるコンテンツに関するコンテンツのタイトルと概要、配信元を示すロケーション情報をあらかじめ投稿しておく。

ユーザの評価情報を収集するためには、他のユーザが評価投稿用blogサーバを利用して、評価情報を投稿する。このとき、メタデータ生成用blogサーバに事前に投稿されているコンテンツのエントリに対して、Trackbackを行う。Trackbackを行うときに、記述した評価情報に対して、意味的な情報を付加する“意味付加タグ”を挿入する。意味付加タグについては、4節で後述する。この操作によって、メタデータ生成用blogサーバには、複数のユーザがコンテンツを評価した情報が集まることになる。

これらの操作によって集まった評価情報は、メタデータ生成用blogサーバで情報の解析を行い、TV-Anytime Forumに準拠したメタデータへ再構築される。ここで生成される最終的な成果物であるメタデータのことを“口コミメタデータ”と呼ぶことにする。

今回の実験には含まれていないが、生成された口コミメタデータの最終的な利用方法の想定シナリオとしては、口コミメタデータを理解可能なPDRなどの記憶媒体を持った装置にロードされて、ユーザにコンテンツをリコメンドして、コンテンツ視聴の促進に利用することを想定している。

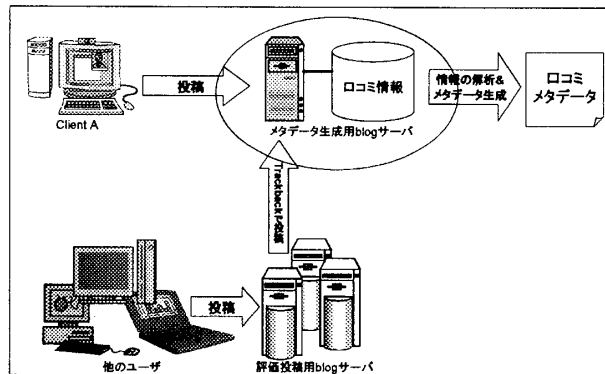


図1 システムの概要図

3. 口コミメタデータの詳細

口コミメタデータを構成するための最低限必要な情報として、以下の1~3のユーザが改変する必要の無い、コンテンツの基本的な情報に関係するものと、4~8のユーザからの評価情報に関係するものの2種類を検討している。そこで、1~3の情報をまとめて『コンテンツ項目』と呼ぶことにし、4~8の情報を『評価項目』と呼ぶことにする。

1. コンテンツのタイトル
2. コンテンツのロケーション情報
3. サービスプロバイダが提供するコンテンツの概要
4. ユーザがコンテンツを評価したときの数値
5. ユーザが評価したテキスト情報
6. 評価を行った人のID
7. Trackbackで評価したときのblogのエントリID
8. 意味付加タグの情報(次節で詳細を述べる)

4. 意味付加タグ

4.1 意味付加タグの詳細

意味付加タグは、投稿されたコンテンツ評価のテキスト情報に、どのような項目に対して評価を行ったのかタグ付けするものであり、独自に定義したネームスペースを利用したXMLのタグセットである。意味付加タグを使用することのメリットは、ユーザの欲しい情報へたどり着くまでの労力の軽減が考えられる。

定義する意味付加タグは、blogを利用して評価情報を投稿する際にユーザに挿入してもらう。今回のシステムでは、口コミメタデータをコンテンツの視聴促進に利用する際に必要であると考え以下の6つの項目を定義する。

- イ. <woms.rating>タグ - コンテンツを10段階(最高10点・最低1点)で評価して、数値化した値を示す
- ロ. <woms.story>タグ - コンテンツのストーリーそのものに関して、評価を行ったことを示す
- ハ. <woms.actor>タグ - 出演している人物・キャラクターなどに関して、評価を行ったことを示す

† 凸版印刷株式会社, Toppan Printing Co., Ltd.

‡ 早稲田大学大学院国際情報通信研究科, GITS Waseda University

ニ. <woms:music>タグ コンテンツで使用されている挿入歌やエンディングテーマなどの音楽に関して、評価を行ったことを示す

ホ. <woms:tempo>タグ コンテンツのストーリーの流れに関して、評価を行ったことを示す

ヘ. <woms:sales>タグ 興行収入などの金銭的な売上げに関して、評価を行ったことを示す

※ 上記で使用されている【woms】は、独自に定義した名前空間である

4.2 意味付加タグの挿入支援 UI

4.1節で意味付加タグの詳細を述べた。実際に、このシステムを利用する場合に、このタグをそのまま記述させるのは、ユーザの負担が大きいのと考える。そこで、ユーザの意味付加タグの挿入に関して、負荷軽減を目的とした手法を提案する。多くのblogのエントリ投稿システムには、リンクの作成や文字を太字にするときに、ユーザがHTMLタグの知識を持っていなくても挿入出来るような、UIが備わっている。本システムでは、従来のUIに対して、違和感のない形でUIを実装することにした。

新規に実装したUIを図2に紹介する。図2の左側の枠で囲われたリンクの作成や文字の太字を行うボタンは標準で実装されているものである。対して、右側のやや標準よりも大きめのボタンが今回実装した意味付加タグ挿入用のボタンである。ボタンの実装方法であるが、標準のボタンがJavaScriptで実装されていたので、今回追加するボタンも標準のボタンと同様にJavaScriptで実装を行った。

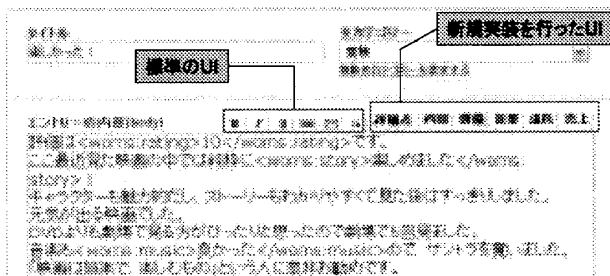


図2 意味付加タグを挿入するインターフェース

5. 評価実験

5.1 評価システムの環境一覧

評価システムでは、OSにLinux(Kernel 2.4系)を利用して、Webサーバにフリーで安定感のあるApache 1.3.31を採用した。メタデータ生成用blogサーバのアプリケーションに、blosxom v2.0を利用して、評価投稿用blogサーバのアプリケーションには、Movable Type 3.151-jaを利用した。開発言語には、blogアプリケーションがperlで実装されているので、perl v5.6.0を採用した。

5.2 評価実験の詳細

実験に参加してもらった被験者は、12名(男性:6名、女性:6名)に協力して頂いた。対象コンテンツは、実験に参加する被験者の多くが知っていると思われる映画コンテンツを13個選別した。そして、被験者にはエントリに登録されている13個の映画コンテンツの中から、評価を行える映画コンテンツを選択して、評価投稿用blogサーバにアクセスして、Trackbackを利用して最低2つの映画コンテンツの評価を行ってもらった。評価終了後に、アンケートを実施する。

5.3 システム評価

被験者の評価情報から、情報を抽出して生成されたロコミメタデータがTV-Anytime Forumに準拠したメタデータであると確認を行った。同時に、メタデータの妥当性の検証を行い、Validなメタデータであることを確認した。

5.4 アンケートの結果と分析

被験者に行った意味付加タグに関するアンケートの結果を示して、意味付加タグの評価を行う。

4.2節で詳説を行った意味付加タグのUIについては、12名中10名の人が特に「違和感を覚えない」と回答した。このことから、意味付加タグを挿入させるUIをプラグインで配布するときは、既存のUIと同様の方法で実装をすることが良いと考えられる。

しかし、意味付加タグを評価情報のテキストに挿入することには、12名中7名の人が「面倒だ」と回答した。そこで、7名の内訳をみると、blog利用経験者が5名であった。これは、blog利用経験者にとって、意味付加タグの挿入という行為が、今までblogを利用するときには無かったので、面倒だと感じたと思われる。逆にいうと、blog未経験者は比較対象が存在しないので、意味付加タグの利用を負担だとあまり感じないと思われる。

そこで、意味付加タグを挿入することを「面倒だ」と回答した被験者には、評価情報をコンテンツ流通に利用することを説明して、再度同様の質問をすると、7名のうち3名が肯定的な意見へと変化した。残りの4名に関しても、「多少面倒だ」という少し肯定的な意見へと変化した。

このアンケートより、意味付加タグを挿入する意味の理由をユーザにあらかじめすることで、より多くの意味付加タグを利用したロコミ情報が得られると考えられる。

6. まとめと今後の課題

本稿では、blogシステムを利用して、ユーザのコンテンツ評価情報を収集・解析して、TV-Anytime Forumに準拠したロコミメタデータの自動生成ができるシステムの実装を行った。それと、メタデータに意味情報を付加する意味付加タグの提案と実装を行い、実装システムを利用して、システムの評価と意味付加タグの評価実験を行った。

評価の結果、実装システムを利用して、意味情報を付加したロコミメタデータの自動生成が可能であったことを確認した。アンケートの結果から、意味付加タグの存在意義を事前にユーザに説明することで、意味付加タグの挿入も半数以上の人が肯定的な意見であったことを確認した。

今後の課題としては、ユーザが意味付加で使用するキーワードとなる品詞を選択しているが、よりユーザの負荷軽減のために、形態要素解析と組み合わせて、キーワードの自動抽出を検討する必要がある。それと、今回は実験段階ということで、定義した意味付加タグの項目の精査をあまり行っていないが、実サービスで使用するときは、意味付加タグのより深い精査が必要だと考える。

参考文献

- [1]TV-Anytime Forum, <http://www.tv-anytime.org/>
- [2]大勝琢己, 亀山渉, “Blogを利用したメタデータ流通に関する考察”, 情報処理学会第67回全国大会 2U-6
- [3]大勝琢己, 亀山渉, “デジタルコンテンツ流通に向けたメタデータ自動生成と流通システムの検討”, 情処研報 2005-AVM-49, Jul.2005