

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

大勝 琢巳^{†‡} 亀山 渉[†]

[†] 早稲田大学大学院国際情報通信研究科

〒367-0035 埼玉県本庄市西富田大久保山 1011

[‡] 凸版印刷株式会社

〒112-8531 東京都文京区水道 1-3-3

E-mail: t-ohkatsu@asagi.waseda.jp

あらまし: 地上波デジタル放送の開始、ブロードバンドインフラの普及によって、デジタルコンテンツ流通が本格化している。しかしながら、流通するコンテンツが増加する環境で、消費者が所望するコンテンツを効率良く発見させる手法の確立がなされていない。筆者らは、その解決方法として、他のユーザが評価した情報(いわゆる口コミ)を流通させるシステムを提案してきた。本稿では、簡単に情報配信が可能な blog を利用して口コミを流通させるシステムについて報告する。

キーワード: メタデータ, ウェブログ, ブログ, コンテンツ流通

On the Metadata Automatic Generation and the Distribution System for Digital Contents Distribution

Takumi OOKATSU^{†‡} Wataru KAMEYAMA[†]

[†] Graduate School of Global Information and Telecommunication Studies, Waseda University

1011 Ookuboyama, Nishi-Tomida, Honjo-shi, Saitama, 367-0035 JAPAN

[‡] Toppan Printing Co.,Ltd.

1-3-3 Suido, Bunkyo-ku, Tokyo 112-8531 JAPAN

E-mail: t-ohkatsu@asagi.waseda.jp

Abstract: Because of the starting of terrestrial digital broadcast in December 2003 and the rapid wide-spread of broadband infrastructures, digital contents have been distributed and re-distributed among us again and again. Various kinds of researches in this field have being done actively by many academic and industrial researchers. However, until now consumers are still suffering from the lack of efficient methods to acquire their desired contents. In order to solve such a problem, we have proposed a novel information distribution system focused on word-of-mouth communication. Word-of-mouth information can be considered as evaluated information by other users. In this paper, the proposed system utilizing weblog to distribute word-of-mouth information as well as its implementation are introduced.

Keyword: Metadata, Weblog, blog, Contents Distribution

1. はじめに

日本では、CS・BS デジタル放送開始に続いて、2003年12月より一部地域において地上波デジタル放送が開始された。この放送のデジタル化によって、コンテンツもデジタル化され、その恩恵として放送コンテンツの高画質化・多チャンネル化・高機能化が実現されると予想される。加えて、ADSLやFTTHなどの広帯域なネットワークインフラが一般家庭へ普及することによって、動画や音声を中心とした Audio Visual 系の大容量コンテンツへとユーザの要求が高まっている。それに対応するように、ユーザ側では大きなデータを保存可能な大容量 HDD などの記憶媒体が備わった PDR(Personal Digital Recorder)などの機器が普及している。

上記の環境や機器の変化を伴って、放送と通信の融合、すなわち「メディア・コンバージェンス」を引き起こすと予想される。これに対応するべく、TV-Anytime Forum[1]のような国際標準化団体や Association of Radio Industries and Businesses(ARIB)[2]などの団体では、新しいコンテンツ視聴方法である「サーバ型放送」を提案している。サーバ型放送では、デジタルコンテンツを大容量記憶媒体に保存を行う。そして、ユーザごとの複雑な配信経路や、コンテンツの利用環境の違い、膨大なコンテンツからユーザが所望するコンテンツを発見するための検索方法などの問題を、メタデータを利用して解決することを目的としている。

我々は、メタデータの使用方法の中で、今後急速に増大すると予想されるデジタルコンテンツの中から、ユーザが所望するコンテンツを発見する手法が重要だと考える。発見手法について、他のユーザがコンテンツを評価した情報(実社会での口コミに値するもの)を流通させるシステムを提案してきた[3]。

本稿では、誰でも簡単に情報配信が可能な Weblog (以後、blog と呼ぶ)を利用して口コミを流通させるシステムについて、実装が完了したので報告する。

2. コンテンツ流通について

図1に、現在のコンテンツとメタデータの流通環境の概要を示す。まず、コンテンツ本体に関しては、著作権の問題があるため CtoC 間の流通は存在しないが、それ以外の流通はすでに存在している。次にメタデータの流れをみると、主にコンテンツ製作者が、コンテンツ制作段階で使用する BtoB 環境を対象とした AAF[4]、P/Meta[5]や MXF[6]などの規格が存在する。BtoC 間では、コンテンツプロバイダが、コンテンツに関する情報(タイトル・ジャンル等)をユーザに送る規格として、TV-Anytime Forumの規格が存在している。

もちろん、これらのメタデータは重要な情報ではあ

るが、これ以外に消費者にとって有用な情報として、一般の消費者が自由にコンテンツを評価した情報があると考えられる。例えば、インターネットを利用する消費者は、映画などを見に行く時に、インターネットの掲示板などを参考にする。このような CtoC 間の情報の流通環境を整備することで、消費者は多くの口コミ情報を得ることが可能になる。これにより多くのコンテンツが消費されると考えられる。

次節より、CtoC 間の口コミ情報を流通させるシステムを提案する。

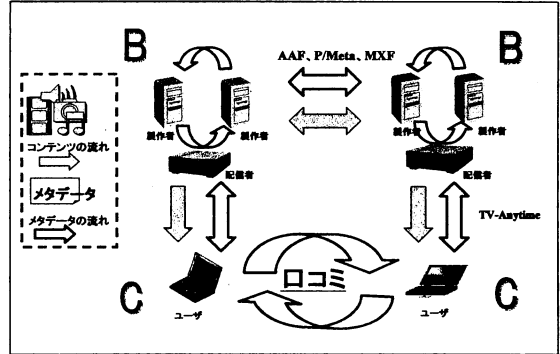


図1 コンテンツとメタデータの流れ

3. 提案手法

3.1. blog の有効性

本研究では、ユーザがコンテンツを評価した情報をメタデータ化して、生成したメタデータをコンテンツ流通の促進に役立てることが目的である。そのためには、まずユーザのロコミ情報を収集する必要がある。

その基盤技術として blog を利用するが、そのメリットを以下にまとめる。

- 誰もが簡単に情報配信可能
- 他の blog へのリンク関係を強化する Trackback の機能がある
- blog は XML との親和性が高い
- 記事の更新情報を記述した RSS[7,8,9]、Atom[10] の自動生成機能

一般ユーザの声を集める点からみると、これまでの Web サイトは、ファイルのアップロードや、HTML タグなどの専門的な知識をユーザは要求されるが、一般的な blog サイトにはエントリ投稿に関して、簡便なユーザ・インターフェース(UI)が備わっている点が優れている。同時に、Trackback の機能により、他ユーザ・コミュニティへのつながりが発生し、より多くのロコミを得られると考える。これに、メタデータ記述フォ

ーマットである XML が blog の要素技術のひとつに採用されていることから、メタデータとの親和性が高く、最終的な目的であるロコミ情報のメタデータ生成、及びその流通が行い易いと考ええる。

3.2. システムの概要

本研究のシステム概要を図 2 に示す。本システムでは、ロコミ情報の収集・解析とメタデータの生成などを担当する blog サーバを用意する。図 2 の中では、メタデータ生成用 blog サーバが、これにあたる。この blog サーバに、事前に評価を行うコンテンツに対する blog のエントリを投稿しておく。図 2 の中では、Client A が事前にエントリの投稿を行う。

他のユーザのロコミを収集するために、Client A が事前に用意したエントリに対して、他のユーザが自分の blog サイトを利用して、コンテンツの評価を投稿すると同時に、Client A のエントリに対して、Trackback を行う。この時、テキスト情報に意味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、後ほど 5 節で詳細を述べる。この操作によって、Client A の blog サーバには、複数のユーザの情報が集まることになる。

この様にして集まったロコミは、Client A の blog サーバのバックグラウンドで情報の解析を行い、TV-Anytime Forum に準拠したメタデータへ変換される。ここで生成されるメタデータを「ロコミメタデータ」と呼ぶことにする。

今回の実験には含まれないが、生成されたロコミメタデータは、最終的な利用方法として、PDR などの記憶媒体を持った装置にロードされ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。

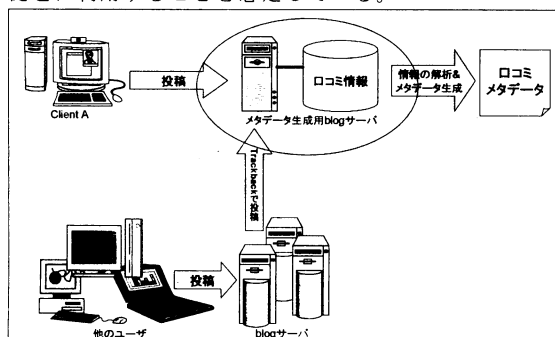


図 2 システムの概要図

4. ロコミメタデータ

4.1. TV-Anytime のメタデータ仕様

3.2 のシステムで生成するロコミメタデータは、サーバ型放送のために標準化を目的としたメタデータの

スキーマを提案している TV-Anytime Forum に準拠したメタデータとする。本来、TV-Anytime Forum は、BtoC 間の利用を想定しているが、CtoC 間での利用で必要と思われる語彙を定義していると考ええる。今回使用する以下の 3 つの項目を説明する。

- (I) コンテンツのタイトルやジャンルなどのコンテンツ固有の情報
- (II) コンテンツのインスタンスの場所を示すロケーションや放送スケジュールなどを示す情報
- (III) コンテンツを評価した情報

(I) は、ProgramInformationTable 内の BasicDescription に記述する。

(II) は、ProgramLocationTable 内に記述する。

上記 I と II の 2 つの情報に関しては、コンテンツ制作会社やコンテンツプロバイダがユーザに提供するメタデータで、基本的にはユーザが変更することが出来ない不変の情報である。この項目に関しては、事前に blog のエントリとして、情報を投稿しておくこととする。

これに対して、III の情報は、ユーザがコンテンツを自由に評価したロコミ情報であり ProgramReviewTable 内に記述する。今回のシステムで抽出するロコミ情報は、このテーブルに格納することとする。次項で詳細を説明する。

4.2. ロコミメタデータの詳細

ロコミ情報に関しては、以下の項目を定義する。

- **Reviewer** : 情報を投稿した人の ID
- **FreeTextReview** : 評価内容のテキスト情報
- **ReviewReference** : 評価を投稿した blog のエントリ ID
- **Rating** : コンテンツを評価したときの数値
- **BasicDescription** : テキスト情報に、どのような項目に関して評価を行ったのか、意味を付加する情報

BasicDescription 以外は、TV-Anytime の標準的な語彙である。しかし、BasicDescription は、スキーマの拡張を行い評価内容のテキスト情報に対して、意味情報を付加するものである。テキスト情報に意味情報を付加することのメリットは、平文のテキスト情報では、多くのテキスト情報が収集された際に、ユーザが目的の情報(例えば、あるコンテンツの挿入歌の情報を欲しいなど)を得るために、多くの労力を要する。しかし、意味情報を付加することで、ユーザが目的の情報を得やすいと思われる。更に、この項目を理解可能な機器

であれば、情報の自動抽出も可能になると思われる。

5. 意味付加タグ

5.1. 意味付加タグの詳細

意味付加タグとは、投稿するテキスト情報に、どのような項目に対して評価したのか、意味的な情報を付加するタグである。これは、blog でエントリを投稿する際にユーザに記述してもらう。今回のシステムでは、ロコモ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思われる以下の項目を定義することとする。

- `<woms:rating>`タグ - コンテンツを評価して、数値化した値を示す
- `<woms:story>`タグ - 内容そのものに関して、記述された事を示す
- `<woms:actor>`タグ - 出演している人物・キャラクターに関して、記述された事を示す
- `<woms:music>`タグ - コンテンツ中で使用されている音楽に関して、記述された事を示す
- `<woms:tempo>`タグ - 物語の流れに関して、記述された事を示す
- `<woms:sales>`タグ - 興行収入などの金銭的な売上げに関して、記述された事を示す

※ 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

タグの挿入に関しては、意味付加タグをキーワードとなるような単語に付加する。例えば、形容詞や名詞などである。具体的な例を以下に示す。

```
話の展開は<woms:tempo>早い</woms:tempo>
音楽は結構<woms:music>よかった</woms:music>です
```

5.2. 意味付加タグの挿入方法

5.1 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、ユーザ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。

現在、blog サービスは増加の傾向にあるが、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。そこで、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。本システムでは、blog 投稿用のblog アプリケーションとして、Movable Type[11]を採用した。そのUIを図3に紹介す

る。

図3の左側の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。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。ボタンの大きさの違いはあるが、違和感無く実装が行えた。

UI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。

ちなみに、意味付加タグは、上記UIを利用して独自のタグセットをblogのエントリに挿入するが、HTMLを表示する一般的なWebブラウザ(Internet ExplorerやNetscapeなど)は、意味付加タグを解釈しないので、ブラウザでエントリを表示するときには問題は無いと考える。

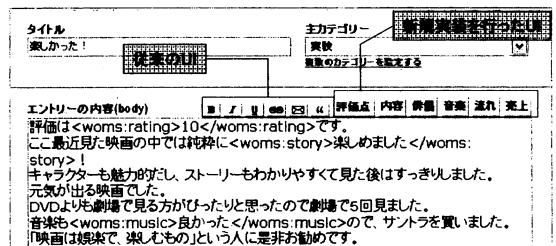


図3 UIの実装例

6. 要素抽出とメタデータ生成手法

ロコモメタデータを生成する際の、メタデータを構成する成分を抽出する手法を述べる。4.1 で説明した項目のユーザが変更することの出来ない項目である(I)と(II)は、あらかじめ投稿されたコンテンツに関するエントリから抽出するものである。抽出には、特定の記述方式を決めて、情報を抽出する。

(III)について、必要な情報は、主にTrackbackを利用して他のユーザが記述した情報である。Trackbackには、米Six Apart社の製品であるMovable Typeで実装されているTrackBack Ping[12]の仕様がある。このTrackBack Pingで取得可能な情報は、以下4つの項目である。

- title - エントリのタイトル
- excerpt - エントリの要約
- url - エントリのpermalink
- blog_name - エントリを投稿したブログの名前

上記のTrackBack Pingで取得可能な情報を、4.2 で示したロコモメタデータの項目に対応させることにする。その対応表を以下に示す。

表1 メタデータ成分の対応表

ロコミメタデータ	TrackBack Ping の情報
Reviewer	blog_name
FreeTextReview	excerpt
ReviewReference	url

Rating と BasicDescription の 2 つは、5 節で述べた意味付加タグから抽出した情報を挿入する。具体的なデータ取得の手順は以下である。

1. TrackBack Ping によって得られる url から、HTML を取得する
2. 取得した HTML の解析を行い、意味付加タグでタグ付けされた情報を取得する
3. 取得した情報のうち、rating タグでタグ付けされた情報は、ロコミメタデータの Rating の項目で利用する
4. それ以外は、コンテンツを評価するテキスト情報の意味を付加する情報として BasicDescription の項目で利用する

注意点として、一般的な blog のエントリ投稿システムにはこの意味付加タグを追加する UI はもちろん実装されていないので、意味付加タグの挿入は強制ではない。そして、意味付加タグを利用しないユーザはテキスト情報だけを有効な情報として、ロコミメタデータに利用する。

最後に、抽出した情報を元に TV-Anytime Forum に準拠したロコミメタデータを生成する。

7. 実験と評価

7.1. 実験環境

実験環境で使用した環境一覧を表 2 に示す。

表 2 実験環境システムの環境一覧

OS	Linux kernel 2.4 系
Web サーバ	Apache 1.3.31
開発言語	perl v5.6.0
メタデータ生成用 blog	blosxom v2.0
評価投稿用 blog	Movable Type 3.151-ja

7.2. 実験方法について

実験は以下の要領で行った。

- 【1】 事前に評価を行うコンテンツに関するエントリの投稿を行い、他のユーザからの Trackback によるロコミ情報の投稿に備える。今回準備したコンテンツは、多くの被験者が知っていると思われる映画を評価の対象として、対象コンテンツの数を 13 個用意した。
- 【2】 被験者 13 名(男性：7 名、女性：6 名)に対して、

なるべく意味付加タグの挿入を行い、Trackback を利用してコンテンツの評価を行ってもらう。

- 【3】 メタデータ生成用 blog サーバで投稿されたロコミ情報をつなぎ合わせて、自動的にロコミメタデータの生成を行う

- 【4】 生成されたロコミメタデータが TV-Anytime Forum 規格に準拠しているのか、Validity Check を行う

7.3. 結果と評価

実験結果から生成したロコミメタデータのインスタンスの一部を図 4 に示す。

図 4 生成したロコミメタデータのインスタンス

blog の Trackback 機能を利用して、一般ユーザが自由にコンテンツを評価したロコミ情報を収集して、収集した情報からロコミメタデータを生成することが可能であることを確認した。

実装を行った意味付加タグを挿入する UI を利用して、意味付加タグの挿入を行い、挿入した意味付加タグを解析して、ロコミメタデータが生成可能であった。

自動で生成したロコミメタデータが TV-Anytime

Forum 規格に準拠したメタデータであると確認を行った。同時に、生成した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。

8. まとめと今後の課題

本論文では、blog システムを利用して、一般ユーザーがコンテンツを評価した情報を収集して、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。実際に、システムを利用してロコミメタデータの生成を行った。ロコミメタデータ生成にあたり、メタデータに意味情報を付加する意味付加タグの定義と実装を行った。

今後の課題としては、意味付加タグをユーザーが積極的に利用してもらうために、実装した UI のユーザー評価を行う。現在は、ユーザーに意味付加で使用するキーワードとなる品詞を選択してもらっているが、形態要素解析などを利用して、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が少ない。今後、サービスへの移行を考えた場合、項目のより深い精査が必要だと考える。

文 献

- [1] TV-Anytime Forum,
<http://www.tv-anytime.org/>
- [2] 社団法人電波産業会 (ARIB : Association of Radio Industries and Business),
<http://www.arib.or.jp/>
- [3] 大勝琢巳、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
- [4] AAF Association,
<http://www.aafassociation.org/>
- [5] Richard Hopper, “P/Meta –Metadata Exchange Scheme, v1.0”, EBU Technical
- [6] Bruce Devlin, “MXF – the Material eXchange Format”, EBU Technical Review, July 2002
- [7] Dan Libby, RDF Site Summary (RSS) 0.9 official DTD, proposed, 1999, Netscape Communications,
<http://my.netscape.com/publish/formats/rss-0.9.dtd>
- [8] Dan Libby, RSS 0.91 Spec, revision 3, 1999-07-10, Netscape Communications,
<http://my.netscape.com/publish/formats/rss-spec-0.91.html>
- [9] Dan Brickley, et al., RDF Site Summary (RSS) 1.0, 2000. RSS-DEV Working Group,
<http://purl.org/rss/1.0/spec>
- [10] M. Nottingham and R. Sayre, eds., The Atom Syndication Format draft-ietf-atompub-format-08, 2005-04-18,
<http://atompub.org/2005/04/18/draft-ietf-atompub-format-08.html>
- [11] Movable Type,
<http://www.sixapart.com/movabletype/>
- [12] Trackback technical specification,
http://www.sixapart.com/pronet/docs/trackback_spec