

## インターネットチャットを利用した

### 番組メタデータの自動生成システムの実装と評価

大黒 泰平<sup>†</sup> 石島 健一郎<sup>‡</sup> 加藤 友規<sup>‡</sup> 土居 清之<sup>‡</sup> 亀山 渉<sup>†</sup>

† 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

‡ 日本テレビ放送網株式会社

〒105 - 7444 東京都港区東新橋 1 - 6 - 1

E-mail : [taihei\\_d@fuji.waseda.jp](mailto:taihei_d@fuji.waseda.jp)

**あらまし** 筆者らは、番組視聴者が番組コンテンツに対してチャットを用いて入力する発言情報を元に、番組コンテンツに関する番組視聴者の感想情報を記述したメタデータを、番組コンテンツの時系列に沿って生成するための手法を提案している。本報告では、提案手法をまとめると共に、その処理を自動化する為のシステムの実装について報告し、提案システムの評価と今後の課題について述べる。

## Implementation and Experimental Results of Automatic Metadata-Generating System for TV-Programs Using Internet Chat System

Taihei DAIKOKU<sup>†</sup> Kenitiro ISHIZIMA<sup>‡</sup> Tomonori KATO<sup>‡</sup>

Kiyoyuki DOI<sup>‡</sup> Wataru KAMEYAMA<sup>†</sup>

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

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

‡ Nippon Television Network Corporation

1-6-1 Higashi-Shimbashi, Minato-ku, Tokyo 105-7444 JAPAN

E-mail : [taihei\\_d@fuji.waseda.jp](mailto:taihei_d@fuji.waseda.jp)

**Abstract** Nowadays, the convergence of broadcasting and telecommunication worlds is progressing, as well as the amount of digital contents is increasing. In this circumstance, it is required to utilize the metadata to search a program among a great amount of digital contents. In this paper, we propose a metadata-generation method of TV programs from TV viewers' inputs using the Internet chat system. Furthermore its implementation and experimental results are also discussed, and its evaluation is also shown.

## 1. はじめに

2000年のBSデジタル放送の開始、2003年の地上波デジタル放送の開始に代表される様に、TV放送のデジタル化が急速に進んでいる。また、HDDや情報通信網の大容量化により、誰でも簡単に大容量のマルチメディアデータを送受信できる環境が整いつつある<sup>1)</sup>。そのような技術的動向と併せて、通信と放送が融合した「サーバ型視聴形態」が今後は普及してくるものと考えられる。この様な背景で、膨大なマルチメディアデータの中から所望のデータを効率的にユーザに提供できる技術が必要となっている<sup>2)</sup>。

一方、TV-Anytime Forumのような国際標準化団体では、サーバ型視聴形態を実現するための媒体としてインターネットの可能性に着目し、マルチメディアコンテンツの国際的な相互流通システムについて模索している<sup>3)</sup>。また、メタデータを用いてシーン検索やダイジェスト視聴といった、様々なサービスを提供しようという試みが行われている。この様なメタデータサービスが今後高度化することを考えると、マルチメディアデータの検索において、視聴者の感想情報を積極的に利用するということが考えられる。

このような背景から、視聴者の感想情報をメタデータとして利用することを目的に、我々は、番組視聴者が番組コンテンツに対してチャットを用いて発言した内容を元に、番組コンテンツに対する感想情報を定義したメタデータを、番組コンテンツの時系列に沿って生成していく為の手法を提案している。本稿では、その処理を自動化する為のシステムの実装を行った為、実装システムの詳細とシステムの評価について報告する。

## 2. メタデータ生成手法

### 2.1.従来手法

番組視聴者から番組コンテンツに対する情報を得るための手法として、現在までに、「視聴率」が主に用いられてきた。視聴率を用いることの利点としては、「全ての番組コンテンツにおいて、他の番組コンテンツと相

対的に比較可能な形で番組視聴者に関する情報を得られる」、「番組コンテンツを視聴する人であれば、どのような人でも被験者とすることができる」といった点が挙げられる。しかし、このような手法では、「番組制作者が番組視聴者に伝えようとしたメッセージがどの程度伝わっているのかを把握し辛い」、「CMのような広告媒体に対して番組視聴者がどのような感想を持っているのかを把握し図」といった問題点を持っている。

### 2.2.提案手法

これに対して、本研究では、番組視聴者からのチャットでの発言情報を用いて番組コンテンツに対する感想情報を番組コンテンツの時系列に沿って抽出する手法を提案している。本手法を用いることで、番組視聴者が番組コンテンツに対して抱いた感想情報をキーワードとしてメタデータ中に定義することが可能になる。よって、「番組コンテンツの特定の部分に対する感想情報を記述したメタデータを生成することができる」、「番組コンテンツ全体の時系列に沿ってメタデータを付与することができる」といった利点が生まれる。

### 2.3.メタデータ生成モデル

本稿で想定している番組メタデータの生成・循環モデルを図1に示す。図1は、番組コンテンツが制作されてから視聴される過程において、メタデータがどのように生成され、利用されるのかを示したものである。発言内容解析エンジンとメタデータ生成エンジンを経て生成されたメタデータは、番組制作者側にフィードバックされ、より良い番組制作のための精度の高い情報として利用することも可能である。

提案手法では、メタデータを生成するための過程として、1 発言内容の解析、2 キーワード抽出、3 メタデータ生成、という3過程を踏んでいる。以下、1を4節、2を5節、3を6節で詳説する。

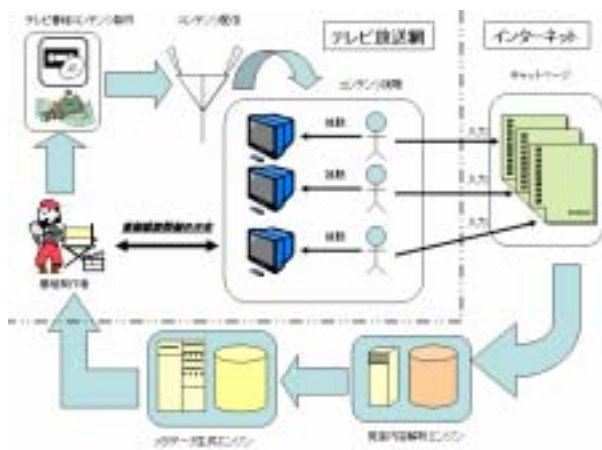


図1 番組メタデータ生成・流通モデル

### 3. 実験について

番組コンテンツに対する発言情報を得るために、研究室にチャットシステムを構築し、この環境で実験を実施した。以下に、その実験手法について述べる。

#### 3.1. 実験環境

実験被験者は、番組コンテンツをPC上で視聴しながら、同一画面上に表示されたチャットページで番組コンテンツに対する議論を行ってもらう。実験被験者は15名（男性8名、女性7名）である。また、実験後、内容ごとにセグメントに分割した番組コンテンツの各部分に対し、被験者が主観的にどの程度面白いと思ったのかを5段階で評定させるアンケートを実施し、1から5ポイントの点数として数値化した。

#### 3.2. 実験番組コンテンツ

本実験では、発言内容の番組ジャンル毎の差異を考慮し、被験者に複数のジャンルの番組コンテンツを視聴させた。実験に使用した番組ジャンルは、代表的な番組ジャンルとして、ニュース番組、バラエティ番組、スポーツ番組、ドラマの4ジャンルとした。

## 4. 発言内容の解析

番組視聴者の発言情報を解析するための分析として、本節では、チャットの盛り上がっていたセグメントを特定するための分析について述べる。

### 4.1. 解析手法

被験者が番組コンテンツのどのセグメントに対して興味を持ったのかを特定するための手がかりを得る為、本稿では以下の仮説を提唱する。

- A) 番組視聴者が番組コンテンツに対して興味を持っている場面では、チャットでの発言回数が増大する。
- B) 番組視聴者が番組コンテンツに対して興味を持っている場面では、発言回数と発言文字列の数が共に増大する。

上記仮説を検証する為、以下の2式を定義した。ここで、Aを各セグメントにおける番組視聴者の発言回数、Bを各セグメントにおける番組視聴者の発言文字数、また、Cを各セグメントの長さを秒で表した値と定義し、を被験者アンケートで得られた各セグメントのポイント被験者の人数分合計した値とする。

$$\alpha = \frac{A}{C}$$

$$\beta = \frac{A \times B}{C}$$

### 4.2. 解析結果

ジャンル毎の分析結果について、ニュース番組とバラエティ番組に関しては、上記の  $\alpha$ 、 $\beta$  の値が共に上昇傾向にあるセグメントが見受けられた。その様な傾向が見られた理由としては、「番組コンテンツのトピックがこまめに化する」、「番組コンテンツの内容が、被験者のチャットでの会話を促すような内容である」といったことが考えられる。一方、スポーツ番組とドラマに関しては、上記の  $\alpha$ 、 $\beta$  の値が共に上昇傾向にあるセグメントが見受けられなかった。その様な傾向が見られた理由として、スポーツ番組に関しては、「場面の展開が早すぎてチャットのスピードがついていけない」という理由が、ドラマに関しては、「ジャンルの性質上、個々のセグメントという短い単位で感想を述べられるものではない」といったことが考えられる。

上記結果より、ここでは、4種類の番組ジャンル内の、ニュース番組とバラエティ番組における分析結果を示す。図2はニュース番組における・・・の値の変化をグラフ化したものであり、図3はバラエティ番組における・・・の値の変化をグラフ化したものである。

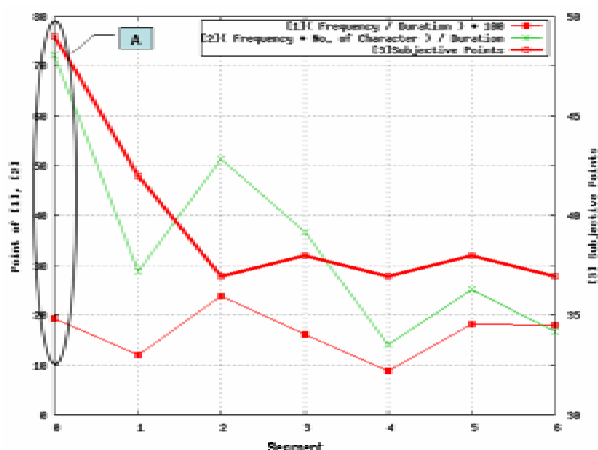


図2 ニュース番組における数値の変化傾向

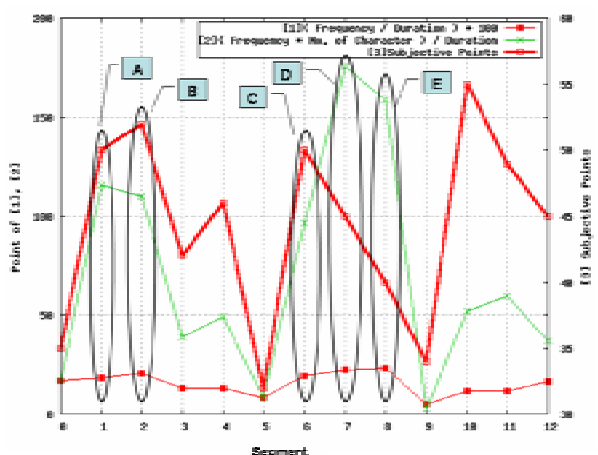


図3 バラエティ番組における数値の変化傾向

図2と図3より、・・・の3値が共に高いセグメントが数箇所見受けられる。即ち、このジャンルの番組コンテンツにおいては、と の値が高ければ、 の値も高いことが高確率で見込まれることから、と の値から番組コンテンツに対する興味の度合いを検出することができる。よって、このようなセグメントに対して、後述するキーワード抽出のための分析をかけていくこととする。

## 5. キーワード抽出

本節では、図2のセグメントAにキーワード抽出の分析を行う例を取り上げ、処理を説明する。

### 5.1.形態素解析

図2のセグメントAに対して、まず、形態素解析を行う。キーワードを含んでいる可能性が高いと考えられる、一般名詞、固有名詞、形容詞、形容動詞、副詞、記号(!,?)、を形態素解析の対象とする。表1から表3に解析結果を示す。

表1 固有名詞の抽出結果

1. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
2. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
3. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
4. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
5. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
6. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
7. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名

表2 形容詞の抽出結果

1. 暑い	アツイ	暑い	形容詞-自立	形容詞-アウオ段	基本形
2. いい	イイ	いい	形容詞-自立	不変化型	基本形
3. ない	ナイ	ない	形容詞-自立	形容詞-アウオ段	基本形
4. 嬉しい	ウレシイ	嬉しい	形容詞-自立	形容詞-イ段	基本形
5. いい	イイ	いい	形容詞-非自立	不変化型	基本形
6. いい	イイ	いい	形容詞-自立	不変化型	基本形
7. すごい	スゴイ	すごい	形容詞-自立	形容詞-アウオ段	基本形
8. すごい	スゴイ	すごい	形容詞-自立	形容詞-アウオ段	基本形
9. カッコいい	カッコイイ	カッコいい	形容詞-自立	不変化型	基本形
10. すごい	スゴイ	すごい	形容詞-自立	形容詞-アウオ段	基本形
11. カッコいい	カッコイイ	カッコいい	形容詞-自立	不変化型	基本形
12. すごい	スゴイ	すごい	形容詞-自立	形容詞-アウオ段	基本形
13. 新しい	アタラシイ	新しい	形容詞-自立	形容詞-イ段	基本形
14. すごい	スゴイ	すごい	形容詞-自立	形容詞-アウオ段	基本形

表3 記号(!)の抽出結果

1.	!	!	!	記号-一般
2.	!	!	!	記号-一般
3.	!	!	!	記号-一般
4.	!	!	!	記号-一般
5.	!	!	!	記号-一般
6.	!	!	!	記号-一般
7.	!	!	!	記号-一般
8.	!	!	!	記号-一般
9.	!	!	!	記号-一般

表1より、野球選手の名前が頻りに発言されている様子が伺える。また表2より、「すごい」、「うれしい」、「カッコいい」といった“賞賛”に関連する内容が交換されているということが伺える。更に、表3より「！」が頻りに発言されていることから、被験者はエキサイティン

グな感情と共に番組コンテンツを視聴しているということが伺える。

## 5.2. キーワードの抽出

5.1 の解析結果を元に、セグメント内の発言からキーワードを抽出する。キーワード抽出に使用する情報を辞書として定義し、辞書の語彙を元にキーワード抽出を行う。表 4 に、本セグメントで定義したキーワード辞書の語彙リストを示す。

表 4 辞書の語彙リスト

キーワード				
イチロー	大リーグ	野球	やきゅう	ベースボール
新記録	しんきろく	凄い	すごい	更新
こうしん	かっこいい	素晴らしい	すばらしい	嬉しい
うれしい	興奮	こうぶん	感動	かんどう
感心	かんしん	感銘	かんめい	！

## 5.3. 雰囲気情報の定義

抽出したキーワードの内、一般名詞と固有名詞に関しては、その語自体が有意な意味を持っていると考えられる。一方、その他の品詞に関しては、語自体が有意な意味を持っているのではなく、語が表す内容が意味を持っている。よって、一般名詞と固有名詞以外の品詞におけるキーワードは、一定の「意味」と対応させて分類するものとする。ここでは、この「意味」を「雰囲気」と呼ぶことにする。雰囲気とキーワードの関係を示したものを、表 5 に示す。

表 5 雰囲気情報とキーワード情報の対応

No.	雰囲気	キーワード
1	Astonishing	すごい、すばらしい
2	Exciting	感動、感心、興奮、感銘、！
3	Fun	かっこいい、楽しい、
4	Happy	嬉しい

## 6. メタデータ生成

図 2 のセグメント A の分析結果を元に、メタデータの作成例を本節では示す。まず、図 4 に作成したメタデータのインスタンスを示す。なお、本稿では、このよう

なメタデータを「キーワード抽出メタデータ」と呼ぶこととする。抽出したキーワードの内、固有名詞と一般名詞については、TV-Anytime Forum が規定するメタデータの内、SegmentInformationTable の中にキーワードとして定義することとした。

また、その他の品詞から抽出した雰囲気情報については、TV-Anytime Forum が規定するメタデータの内、AtmosphereCS を利用し、ProgramInformationTable 内に雰囲気情報として定義することとした。なお、これらの雰囲気情報は、別個のメタデータとして、AtmosphereCS 内のライブラリに記載されており、そのライブラリ内の雰囲気情報を参照するという形で雰囲気情報を定義している。

```

1.  <?xml version="1.0" encoding="UTF-8"?>
2.  <TVAMain xmlns="http://www.tv-anytime.org/2001/08/metadata">
3.    <ProgramDescription>
4.      <ProgramInformationTable>
5.        <ProgramInformation programId="crj://www.km.giti.waseda.ac.jp/news01.rm">
6.          <BasicDescription>
7.            <Genre preferenceValue="40" href="at:8.3">
8.              <Name>Astonishing</Name>
9.            </Genre>
10.           <Genre preferenceValue="40" href="at:8.16">
11.             <Name>Exciting</Name>
12.           </Genre>
13.           <Genre preferenceValue="40" href="at:8.19">
14.             <Name>Fun</Name>
15.           </Genre>
16.           <Genre preferenceValue="40" href="at:8.23">
17.             <Name>Happy</Name>
18.           </Genre>
19.         </BasicDescription>
20.       </ProgramInformation>
21.     </ProgramInformationTable>
22.     <SegmentInformationTable>
23.       <SegmentList>
24.         <SegmentInformation segmentId="1">
25.           <SegmentLocator>20041004115449-20041004115834</SegmentLocator>
26.           <Description>
27.             <Keyword>イチロー</Keyword>
28.           </Description>
29.         </SegmentInformation>
30.       </SegmentList>
31.     </SegmentInformationTable>
32.   </ProgramDescription>
33.   <ClassificationSchemeTable>
34.     <CSAlias alias="at" href="um:tv:metadata:CS:AtmosphereCS:08-2002"/>
35.   </ClassificationSchemeTable>
36. </TVAMain>

```

形容詞・形容動詞・副詞・記号を元に定義される

セグメント内における雰囲気情報の定義

固有名詞・一般名詞から抽出

セグメント内におけるキーワード情報の定義

図 4 キーワード抽出メタデータのインスタンス

## 7. システムの実装

### 7.1. システムの動作フロー

これまでの一連の処理を自動化するためのシステム実装を行った。システムの動作フローを図 5 に示す。使

用言語として、キーワード解析エンジンには Perl・bash を、メタデータ生成エンジンには Java を用いた。また、形態素解析ツールとして Cha-Sen<sup>4)</sup>を用い、XML 文書の操作には Xerces と Xalan を、XML のハンドリングには DOM<sup>5)</sup>を用いた。

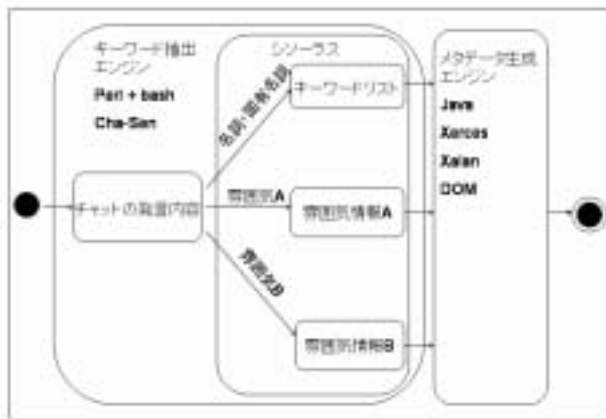


図5 システムの動作フロー

図5に示すように、チャットでの発言内容は、まずキーワード解析エンジンにかけられる。キーワード解析エンジンでは、発言内容の形態素解析、キーワードの抽出、キーワードを雰囲気分類する処理が行われる。次に、抽出されたキーワードをメタデータ生成エンジンにかける。メタデータ生成エンジンでは、抽出したキーワードを TV-Anytime Forum のメタデータスキーマ<sup>6)</sup>に従い、番組コンテンツのメタデータインスタンス中に挿入する処理を行う。

## 7.2. システムの評価

本手法を用いて構築したメタデータインスタンスは TV-Anytime Form のメタデータスキーマに従った検証済みのメタデータインスタンスであるということを確認することができた。

## 8. まとめと今後の課題

本稿では、インターネットチャットを利用して番組コンテンツに関する感想情報を記述したメタデータを番組コンテンツの時系列に沿って生成していく手法を述

べた。結果として、バラエティ番組とニュース番組においては、番組コンテンツの時系列に沿ってメタデータを生成していくことが可能であるということを確認することができた。このような手法を用いることにより、番組視聴者が番組コンテンツのどの部分にどのような感想を抱いたのかという情報を記載したメタデータを生成していくことができるようになるものと考えられる。

今後の課題として、スポーツ番組やドラマといった番組ジャンルに関する同等の手法の確立、音声入力チャットといったチャット以外の番組視聴者インターフェースの検討、等がある。また、形態素解析、キーワード辞書、雰囲気情報辞書など、各種辞書の拡充も重要である。その他、実環境での使用を想定した大規模な実証実験の実施なども検討していく必要がある。

## 参考文献

- 1) 映情学誌, 58, pp647-650 (2004)
- 2) 総務省, “情報通信白書平成 16 年版”,  
<http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h16/index.html>
- 3) TV-Anytime Specification on Metadata:SP003v1.3, TV-Anytime Forum,  
<http://www.tv-anytime.org/>
- 4) Chasen's Wiki,  
<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>
- 5) The World Wide Web Consortium,  
“Document Object Model(DOM)”,  
<http://www.w3.org/DOM/>
- 6) The World Wide Web Consortium, “XML Schema”,  
<http://www.w3.org/TR/2001/REC-xmlschema-0-20010502>,  
<http://www.w3.org/TR/2001/REC-xmlschema-1-20010502>,,  
<http://www.w3.org/TR/2001/REC-xmlschema-2-20010502>