

Web コンテンツの受動的視聴のための 自動変換とスクリプト作成マークアップ言語

灘 本 明 代[†] 服 部 多 栄 子^{††} 近 藤 宏 行^{††}
沢 中 郁 夫^{††} 田 中 克 己[†]

本論文では、膨大な Web 情報を容易に閲覧することを目的とし、チャンネルを選択するように Web ページを選択し、テレビ番組のような番組を作成し、Watch and Listen 型インタフェースを通じて視聴する方式を提案する。受動的視聴とは、これまでテキストや画像で表示されていたコンテンツに代わり、音声やキャラクタアニメーションを用いた番組風コンテンツを作成して見ることである。この番組を作成することを我々は番組化と呼ぶ。これまでの Web ブラウザと異なり、番組化においては、時間軸の同期や演出が問題となる。本論文では、この同期と演出に着目し、Web コンテンツの受動的視聴のための番組化の方法として (1) 現在の HTML ベースの Web コンテンツを自動で受動的視聴向けコンテンツに変換する自動変換方式 (2) 受動的視聴コンテンツ作成のためのスクリプト作成マークアップ言語を提唱する。

Automatic Conversion and a Scripting Markup Language for Passive Watching and Listening of the Web Contents

AKIYO NADAMOTO,[†] TAEKO HATTORI,^{††} HIROYUKI KONDO,^{††}
IKUO SAWANAKA^{††} and KATSUMI TANAKA[†]

This paper proposes a passive Watching-and-Listening approach that makes it easier to view Web information. This approach utilizes a Watching-and-Listening interface that allows the user to choose Web pages in a similar way as we select TV channels, by converting the pages into TV-program-like cartoon animations. During the conversion process, which we refer to as programmization, voice and image-data presentations must be synchronized carefully so that they effectively convey the original information. Moreover, the overall presentation needs to follow a particular presentation structure, such as news and variety shows, in order to improve the familiarity of the program. Our approach solves these issues by developing a novel method of converting HTML documents into programs, and designing a new mark-up language, Scripting-XML (S-XML), which is suitable for describing programs. This paper also addresses the preliminary evaluation results of the current implementation.

1. はじめに

現在使用されている Web の閲覧環境は、利用者に「読む」「スクロールする」「クリックする」などの能動的な操作を要求する Read and Click 型インタフェースである。近年のインターネットの目覚ましい発展により Web 上には種々の情報が氾濫し、膨大な情報量となっている現在、Read and Click 型ユーザインタ

フェースでは、多数クリックして複数の Web ページを取得し、それら Web ページ内の情報を読まなければならない。そのため、利用者の負担が大きく、Web 情報を容易に閲覧できるとはいえない。一方、現在のテレビは「見る」「聞く」の受動的な操作で番組を容易に視聴することができるが、得られる情報はあくまで限られた製作者側の人々から発信された情報である。そこで我々は、一般の人々が発信する膨大な Web 情報をテレビを視聴するように容易に視聴できると望ましいと考えた。

本論文では、膨大な Web 情報を容易に閲覧することを目的とし、チャンネルを選択するように Web ページ(群)を選択し、テレビ番組のような番組風コンテンツを作成し Watch and Listen 型インタフェースで

[†] 神戸大学大学院自然科学研究科情報メディア科学専攻
Division of Information and Media Sciences, Graduate
School and Technology, Kobe University

^{††} 神戸大学大学院自然科学研究科情報知能工学専攻
Division of Computer and Systems Engineering, Grad-
uate School and Technology, Kobe University

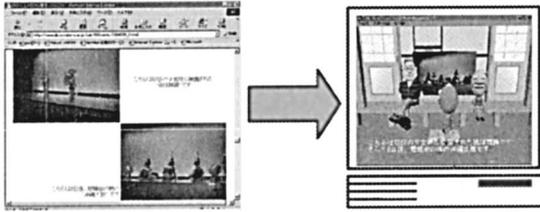


図1 Webの番組化

Fig. 1 TV-Programmization of Web contents.

受動的視聴の可能となる方式を提唱する。受動的視聴とは、これまでテキストや画像で表示されていたコンテンツを、音声合成・読み上げやキャラクターアニメーションを用いて表現した番組を作成して見ることである。以下、受動的視聴のための番組を作成することを番組化と呼ぶ。図1にWebの受動的視聴の例を示す。図1の左半分は現在のWebブラウザによって表示されたコンテンツであり、図1の右半分はそのコンテンツを番組化した画面を示している。

Webの番組化による利点を下記に示す。

- 容易にWebの閲覧が受動的にできる。
- 楽しみながらWebの閲覧ができる。
- コンピュータを技術的に利用できない人もWeb情報を取得できる。

これまでWebの受動的な情報取得にはPointCast¹⁾のようなPUSH型のものがある。これに対し、番組化はWeb情報の音声読み上げを行うため、聞き流したり、別のことをしながら片手間に容易にWebの情報を取得したりできるだけでなく、キャラクターアニメーションを使用しているため、番組として楽しみながら視聴することが可能である。また、これまで「クリック」することすらできない人も、Webの情報の取得が可能となる。

Webの番組化に重要な技術で、現在のWeb閲覧技術と大きく異なる点は

- 時間軸上の同期
- 演出の付加

の2点である。

時間軸上の同期とは、通常Webブラウザは時間軸を気にする必要のない2次元空間での表示であるのに対し、番組化することにより、Web上の画像と文章の同期や文章と文章の同期などの時間軸を考慮する必要が生じる。本論文においては、この画像と文章との同期について述べる。演出の付加とは、キャラクタの台詞や動き、カメラや照明の配置、スタジオセットなどの演出が必要になることである。現在のWeb情報はほとんどがHTMLで作成されているため、これら

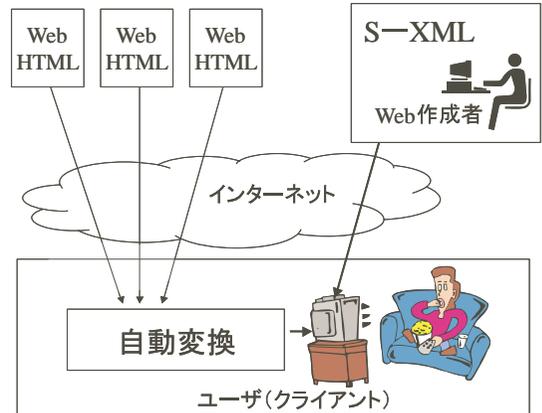


図2 Webの受動的視聴のアプローチ

Fig. 2 Approach of passive viewing of Web contents.

の時間軸上での同期や演出などは考慮されていない。そこで、本研究ではWebの受動的視聴において、同期と演出に着目し図2に示すように、Web利用者側(ユーザ)の立場とWeb作成者の立場から下記の異なる2つのアプローチを採用した。

● 自動変換

現在のHTMLで書かれたWebコンテンツをWeb利用者(クライアント)側で自動変換することにより番組化する技法である。HTMLはタグ構造により構成されており、このタグ構造を解析することにより、HTML文書を自動で番組風コンテンツに自動変換する。このとき、先に述べた時間軸との同期化と演出が問題になる。本研究では同期化問題は、HTMLのタグ解析により解決し、演出はニュースやバラエティなどの番組メタファに注目してあらかじめプリミティブな演出を用意して解決した。

● スクリプト作成マークアップ言語

Web作成者が番組風コンテンツを作成したり、コンテンツ作成時に同時に番組風コンテンツを作成することを可能とするスクリプト作成マークアップ言語を提唱する。そのスクリプト作成マークアップ言語をScripting-XML(S-XML)と呼ぶ。S-XMLは番組化用XML拡張タグ集合であり、演出や同期化の問題を解決すべく、新たなタグ構成を付加した。

以下、2章では関連研究を、3章ではWeb情報の番組化のアプローチの1つである自動変換を、4章ではスクリプト作成マークアップ言語であるScripting-XML(S-XML)を述べる。

2. 関連研究

2.1 番組メタファおよび番組化

情報を番組として呈示する基本概念は、Nonogaki ら²⁾により FRIEND21 で TV 番組メタファとして提唱されている。本研究で述べる番組メタファの基本概念はこの FRIEND21 で提唱された TV 番組メタファを基としている。

また、山口らの提唱する“WebStage”³⁾は、FRIEND 21 で提唱された TV 番組メタファを用いて Web 情報を音声やキャラクタにより表現・呈示する方式を提案している。“WebStage”は本研究とコンセプトが似ているが、Web ページをセグメントに分け、そのセグメントごとに番組素材として提供するものである。それゆえに、タグ解析を行い、同期化領域の発見や演出をふまえて番組化する点や、Web 作成者が Web 作成時に番組を作成できるスクリプト作成マークアップ言語を提唱している本研究とは異なるものである。

さらに、Microsoft 社の提供する WebTV⁴⁾はテレビ端末上に Web 情報を重畳表示するものであり、Web 情報を番組風コンテンツに変換し受動的視聴を行う本研究とはまったく異なるものである。

その他、矢部らの研究⁵⁾ではネットニュースの議論から脚本を作成し、台詞を CG キャラクタに割り当て、議論をテレビ番組のように見せる手法を提案している。この手法の対象はネットニュースの議論に限られていて、Web 情報全般を対象としているわけではない。また、演出方法もパターン化されている。植田らの研究⁶⁾ではスポーツ中継のテレビ映像からシナリオテンプレートを使用してストーリー性を重視したダイジェストの自動生成を提案している。テンプレートは自動的に決定されるため、ユーザの意見や好みを反映させることはできない。

2.2 関連するコンテンツ記述言語

番組記述言語 TVML (TV Program Making Language)^{7),8)}は、リアルタイムに CG、音声を合成して、1 本の TV 番組をすべてデスクトップ上で生成することができる言語である。本研究では Web 情報を呈示する TV 番組形式の実現方法の 1 つとして TVML を用いている。

また、デジタル放送の標準化にともない、XML に基づくデータ放送用の BML (Broadcast Markup Language)、さらにコンテンツごとにタグを設定でき、XSLT によって BML に変換される BXML (Broadcasting XML) が標準化されている⁹⁾。本研究では、Web 情報を利用して番組化を行うことを目

表 1 コンテンツ記述言語の比較

Table 1 The comparison of related content markup languages.

	キャラクタ	演出	同期	Web との関連度
TVML				×
SMIL	×	×		
PPML				×
S-XML				

的とし、XML に番組化のための演出要素を付加する拡張をした S-XML を提唱するものであり、これらの研究とは一線を画するものである。

SMIL (Synchronized Multimedia Integration Language)¹⁰⁾は、HTML のリンク機能に加え、時間軸に沿ったメディア制御ができる。テキストや画像、動画などの各メディアの同期のほか、音声トラックでの言語の選択や、回線にあわせた音声やビデオのバージョンを選ぶ機能などが提供されている。さらに SMIL Boston¹¹⁾では、TV 放送との統合、アニメーション機能が追加され、ニュース番組のようなものを作成することも可能であるが、本研究や TVML のようにキャラクタに台詞をしゃべらせることはできない。

PPML (Personalized Program Markup Language)¹²⁾は、ダイジェスト映像間の接続関係を推定し、その接続部分やシーン切替のタイミングを自動生成する機能を持ち、SMIL と TVML プレイヤの連動を考慮に入れた番組記述言語である。しかし、PPML はダイジェスト映像を対象とした番組用の XML であり、本論文で提唱している S-XML は Web コンテンツを対象としたものである。以上で述べたコンテンツ記述言語と本論文で提唱している S-XML の比較を表 1 に示す。

3. 自動変換

3.1 概要

現在の Web はほとんどが HTML で書かれているが、この HTML はタグ構造により構成されている。このタグ構造を解析することにより、自動で番組記述言語である TVML へ変換し Web の番組化を行う。通常、現在の HTML ソースは上から順番に 2 次元表示の Web ブラウザ上にテキストや画像として表示される。番組化は基本的に HTML タグ構造を解析し、そのタグ構造の論理構造の縦型探索を行い、シーケンシャルに番組化していくが、さらに時間軸上にコンテンツを音声と画像やキャラクタアニメーションなどを用いて表示するため、表示方法が異なってくる。そのために、時間軸上の同期をとる必要があるとともに、

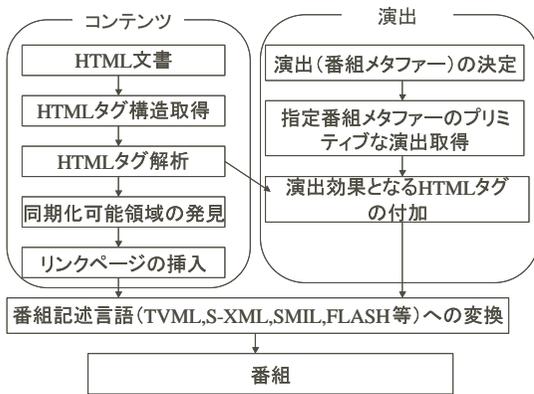


図3 番組化自動変換手順

Fig. 3 Procedure of the automatic conversion into TV-program presentation.

HTML タグの中にある `<CENTER>` タグのように Web コンテンツのレイアウトを示すようなタグは表示不要となるなど、番組化のための種々の処理が必要となる。自動変換の手順を図3に示す。

自動変換の機能の特徴を下記に示す。

- HTML 文書のタグ解析を行う。
- HTML 文書の論理構造の縦型探索のシリアライゼーションを行う。
- 画像と文章、文章と文章の時間軸同期を行う。
- リンク先ページの情報に要約してリンク元ページの番組にマージする。
- 番組メタファによるプリミティブな演出を用意する。

先に述べたように、HTML で書かれた Web ページを番組変換する際、時間軸が非常に重要な問題となってくる。画像とそれに関する文章の同期がとれていなければ、その Web 情報は誤って伝わることになる。そこで本研究では、画像と文章との同期化可能領域の発見方法を提案する。

また、番組化にはキャラクタの台詞や動き、カメラや照明の配置、スタジオセットなどの演出が必要になってくる。ここで実際の TV 番組のようにニュースやドラマ、ドキュメントなどの種々のメタファに着目し、メタファごとの演出を設定し、ユーザが番組を見る前にこれを指定することとする。

3.2 HTML タグ解析

現在の Web ページはテキストを主にして情報を表示するため、HTML タグはテキスト表示のためのタグ構成になっている。これに対し、番組化では、テキスト表示をキャラクタ間の対話形式の音声に変え、時間軸の同期をとり、そして演出を加えて番組として Web

情報を呈示する。HTML 内の文章は台詞として変換するが、番組化を行うために下記の4種類のタグ解析が必要となる。

- 演出効果となる HTML タグ
- 表示不要となる HTML タグ
- 画像の HTML タグ
- 台詞となる HTML タグ

(1) 演出効果となる HTML タグ

HTML タグには、テキスト表示のため、`<BIG>` や `<I>`、`` に代表されるフォントに関するタグがある。HTML 文書内では、これらの開始タグには、各種パラメータの指定が含まれている。現在のテキスト表示ベースの Web ページ上ではこれらのタグを用いて文章や単語を強調している。強調した文章や単語は Web ページ内のキーワードである可能性が高い。したがって、番組化においても強調して利用者に呈示しなければならない。そこで、これらのタグはパラメータごとに、音声や音量を設定する。各タグの設定パラメータが変わることで、音声を変えたり、音量を大きくしたりして強調してユーザに知らせようになる。

(2) 表示不要となる HTML タグ

`<P>` や `<CENTER>` などの Web ページ上のテキスト整形に用いられるタグは音声によって情報を呈示する番組化には必要ない。これらのタグは HTML から番組化への変換時に不要となるため削除する。

(3) 画像の HTML タグ

`` タグは画像ファイルをリンクしている。この画像と同期化可能な文章領域を発見するために画像タグを取得する。

(4) 台詞となる HTML タグ

上記で示したタグ群以外のタグは台詞となるタグ領域として取得する。

上記の演出効果となるタグと表示不要となるタグの例を表2に示す。

3.3 同期化可能領域の発見

上記で述べたように、HTML のタグ構造を解析し、そのタグ構造で表される論理構造の縦型探索を行い番組化するが、時間軸上に情報を呈示する番組化において、時間軸の同期が非常に重要な問題となってくる。画像とそれに関する文章の同期がとれていなければ、その Web 情報は誤って伝わることになる。そこで本節では、画像と文章との同期化可能領域の発見方法を提案する。

表 2 HTML タグ変換
Table 2 HTML tags and their conversion.

タグ	HTMLにおける特徴	番組化への操作
演出効果となるHTMLタグ群		
<H☆>	見出し文字サイズを指定するものであり、パラグラフのキーワードとして利用可能。	見出し語は各ページの特徴を表わすものとして、タイトルと同程度に重要であると考えられる為、音量を大きくする。
	SIZEやCOLORの指定ができるが、これはHTMLソースの作成者が意図して行っていることで、文書全体における重要度は高い。	SIZEやCOLORの指定が行われている部分は、「」(括弧)をつけた言葉に置き換え、番組上でキャラクタの台詞が表示された時に、タグで囲まれていた部分を強調するため、音声や音量を変えて演出を行う。
<D><BIG>等	これらは全て語句を強調する為に用いるタグ。	これらをユーザーが意図する強調レベルに応じて抜き出し、音量を変えて番組に変化を与えていく。
<CITE>	語句の引用を行うタグ。	ユーザーがこのタグを利用する事を明示した場合には引用文を番組に導入する。
 	これらは直前の文書の説明を箇条書きにしたものが多い。	直前部分の文章の内容に応じて表現するメタファーを変えることが可能。また、箇条書きということから、内容的に対等なもの列挙であることが多いので、これに適したメタファーの適用が考えられる。
不要となるHTMLタグ群		
<U><SUP><TT>等	アンダーライン等	番組呈示には不要となるタグ群である
<P><HR><CENTER> <PRE>等	これらはWebページ内のテキスト整形を行う。	番組呈示にはテキスト整形は必要ないため不要となるタグ群である。
<SCRIPT><FORM><FRAMESET><ADDRESS>等	これらはWebページ内各種設定を行う。	番組呈示には不要となるタグ群である。

同期が必要である画像と文章は、HTML 文書において互いに近い距離にあると考えられる。そして、HTML タグは階層構造を持っており、文書や画像をある程度グループ分けすることが可能である。これらのことに注目し HTML タグ構造を解析すると、画像と文章を含む極小なタグ領域を発見することにより、画像と文書の同期化可能領域を発見できる。

画像 i に対する、極小タグ領域 $R(i)$ を下記に示す条件 (1) から (4) をすべて満たすものと定義する。

- (1) $R(i)$ は任意のタグ $\langle X \rangle$ と $\langle /X \rangle$ で囲まれた領域である。すなわち $R(i) = \langle X \rangle, \dots, \langle /X \rangle$
- (2) $R(i)$ は画像 i を含む。すなわち $R(i) \ni i$
- (3) $R(i)$ は少なくとも 1 つの文章 s を含む。すなわち $R(i) \ni s$
- (4) $R(i)$ のいかなる部分タグ領域 $\langle Y \rangle, \dots, \langle /Y \rangle$ も (1) ~ (3) を満たさない。

図 4 に文章 s と画像 i を含む極小タグ領域の例を示す。

画像 i に対する極小タグ領域 $R(i)$ を用いて、画像 i と同期すべき文章を含む同期化可能領域 $S(i)$ を次のように定義する。

- (a) $S(i)$ は $R(i)$ に含まれる。

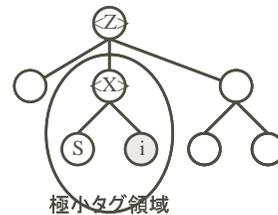


図 4 極小タグ領域
Fig. 4 Minimal tag region.

- (b) $S(i)$ は i と少なくとも 1 つの文章 s を含む。
 $S(i) \ni i \wedge (\exists s)(S(i) \ni s)$
- (c) $S(i)$ は (a), (b) をともに満たす極小な部分領域である。

ここで (c) で述べた「部分領域」とは任意の連続部分領域を意味しており、先に定義した極小タグ領域とは異なり、先頭と末尾が $\langle X \rangle, \dots, \langle /X \rangle$ のように対応するタグになっている必要はない。図 4 の場合は $R(i) = \langle X \rangle, s, i, \langle /X \rangle$ なので、 $S(i) = s, i$ となる。図 5 の構造を持つ Web ページの場合、画像 i に対する極小タグ領域 $R(i)$ は、 $\langle X \rangle, s1, s2, i, \langle Y \rangle, s3, i', \langle /Y \rangle, \langle /X \rangle$ となる。このような場合、 $s1, s2, s3, i, i'$ は $R(i)$ の要素であるが、画像 i の同期

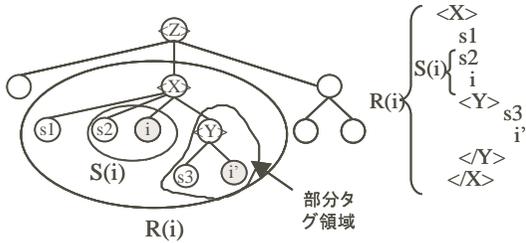


図5 同期化可能領域

Fig. 5 Synchronizable region.

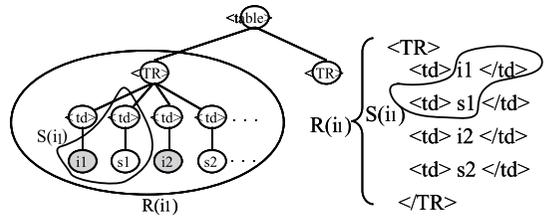


図7 テーブル行タグ解析による同期化可能領域

Fig. 7 Synchronizable region by table-row tag analysis.

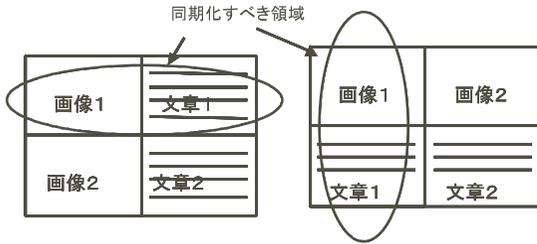


図6 テーブルタグによる同期化

Fig. 6 Implicit synchronization by table tags.

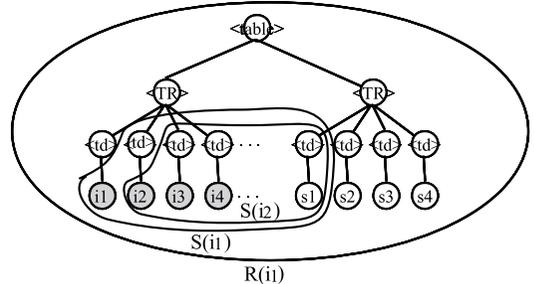


図8 テーブル列タグ解析による同期化可能領域

Fig. 8 Synchronizable region by table-column tag analysis.

化可能領域 $S(i)$ は $S(i) = s2, i$ となり, $s1, s3, i'$ は i の同期化候補とはならないことに注意を要する.

また, Web 上に画像を表示する場合, 通常, テーブルタグを用いる方法がよく使われているが, テーブルタグを用いた場合, Web ページの表示では図 6 に示すように行で画像と文章の同期をとっている場合と列で画像と文章の同期をとっている場合がある. 行で同期をとっている場合, 先に述べた極小タグ領域 $R(i)$ を発見し, 同期化可能領域 $S(i)$ を決定する手法を用いると, 図 7 に示すように, $S(i1) = i1, </td>, <td>, s1$ となり, 画像 $i1$ に同期している文章の候補は $s1$ となり同期化すべき要素を発見することが可能となる. しかし, 画像 $i2$ に同期する文章を発見する場合, 我々の定義では, $S(i2) = s1, </td>, <td>, i2$ または $S(i2) = i2, </td>, <td>, s2$ となり, どちらか特定できないという問題が生じる.

これに対し, 列で画像と文章の同期をとっている場合, 図 8 に示すように, 画像 $i1$ に同期している文章の候補は文章 $s1$ である. (c) で定義したように同期化可能領域 $S(i1)$ は極小な連続部分領域であり, $S(i1) = i1, </td>, <td>, i2, </td>, <td>, \dots, </td>, <td>, s1$ となる. よって, 画像 $i1$ に同期する文章として $s1$ を発見できる. 同様に画像 $i2$ の同期化可能領域 $S(i2)$ は $S(i2) = i2, </td>, <td>, \dots, </td>, <td>, s1$ となり, 画像 $i2$ に同期している文章の候補も文章 $s1$ になる. つまりは, 画像 in に同期している文章の候

補はすべて文章 $s1$ になってしまう問題が生じる. よって, 列で画像と文章の同期をとっている場合は別の方法を用いる必要があり, テーブルタグの列構造をさらに解析する必要がある.

3.4 リンク先ページを考慮したリンク元ページの番組化

これまでは, ページ単位での番組変換について述べてきたが, Web ページの特徴の 1 つである, リンクを考慮した Web ページの番組化について述べる. 本研究では画像と文章の同期からリンク先ページを考慮したリンク元ページの番組化を行った. リンク元ページとリンク先ページを対象として同期化可能な画像と文章を発見することにより, リンク先ページをリンク元ページの番組化にマージして 1 つの番組として表示することが可能となる.

たとえば, ポータルサイトやディレクトリサービスなどに代表されるように, リンク先リストのみで Web ページを構成している場合, このリンク先を示すアンカーのみを番組化したのでは意味がない. そのため, 各々のアンカーのリンク先のトップページを発見し, そのページごとに同期化領域を決定し番組化を行う. その結果得られた番組を Web ページのアンカーの順番でシリアライズし, 1 つの番組を構成する. 図 9 に概念図を示す. また, このときアンカーのリンク先のトップページがさらにアンカーのみで構成されている

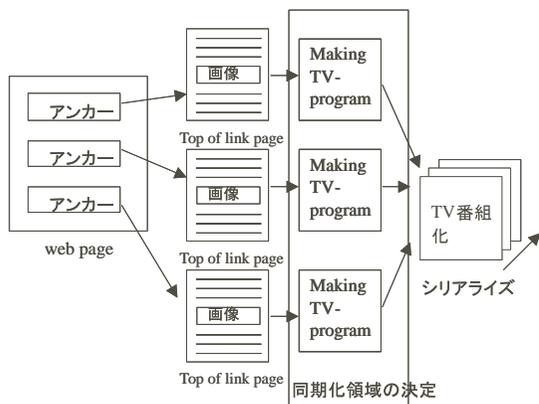


図9 リンクページの番組変換概念図

Fig. 9 TV-programmization of link destination pages.

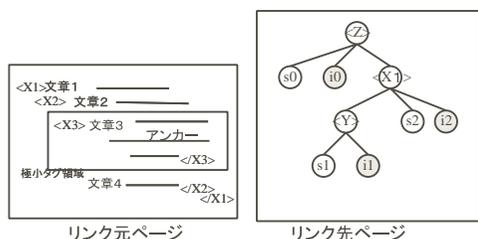


図10 リンクページの極小タグ領域の発見

Fig. 10 Discovery of a representative image in the link destination page.

の場合、さらに各々次のリンク先ページのトップページを番組化すると、大本のページと内容がかけ離れてしまう危険性がある。そのため、リンク先の番組化を行うページはあくまで1階層のページのみと限定している。

図10に示すWebページの文書内にアンカー文字列があり、リンク先ページが存在する場合のリンク元ページの番組化を考える。このリンク元ページ内には画像はなく、アンカー文字列からリンクが図10右のページに設定されているとする。リンク元ページを番組化する際、アンカー文字列と少なくとも1つの文章を含む極小タグ領域を音声読み上げする際にリンク先ページの「代表」となる画像を同期表示させたい。そのリンク先ページ中の「代表」とされる画像として、ページの上に配置される画像とHTMLタグの木構造の上位レベルにある画像とが考えられる。図10の場合、 i_0 は上記2つの条件を満たしている画像となり、リンク先ページ内の「代表」と推定される画像となる。しかしながら、リンク先ページに i_0 のような画像がないとき、ページの上に配置される画像 i_1 とHTMLタグの木構造の上位レベルにある画像 i_2 が同期化可能領域を決定する画像の候補となる。これら

のことより、リンク元ページの番組化に必要となるリンク先ページの画像の抽出はすべてのWebページに対応していないため、今後の課題となっている。

3.5 演出

Webコンテンツを番組として表示するには、キャラクタの動きやカメラ情報、照明の配置、スタジオセットなどの演出が必要である。HTMLではこのような演出を行うタグがないため、HTML文書を番組化に変換しただけではこれらの演出を行うことができない。それゆえに演出を付加しなければならない。しかしながら、番組の視聴者(ユーザ)に演出すべての詳細な設定を行わせることは困難である。そこで本研究では、ニュースやバラエティなどの番組メタファに注目し、番組メタファごとにあらかじめ標準的な演出設定を用意し、ユーザがこの番組メタファを選択して視聴を行うことにした。番組メタファは下記のとおりである。

- ニュース番組
- おとぎ話
- 学校
- バラエティ

これら各々の番組メタファにあらかじめカメラ設定、照明配置、スタジオセットを用意し、ユーザが選んだ番組メタファにより番組化を行う。なお、ユーザが番組化を楽しむために、登場するキャラクタはユーザが指定できるようにしている。

3.6 同期化可能領域の発見の実験

3.6.1 システムの実装

上記に述べた同期化領域の発見、HTMLタグ構造の解析、番組メタファによる演出を基に図3に示した流れを用いてプロトタイプシステムを作成し、通常のHTMLページを番組記述言語であるTVML形式に自動変換して番組化を行った。本プロトタイプシステムのインタフェース部を図11に示す。

3.6.2 評価

プロトタイプシステムを用いて、研究室内の画像を含む15のWebページの番組変換を行い、同期化可能領域の発見の実験を行った。その結果を表3に示す。これらの結果より、テーブルタグを用いておらず、かつ画像と文章が1対1で対応付けられているWebページでは、同期化可能領域の発見は有効であるといえるが、1つの文章に対し対応画像が複数ある場合では、同期化可能領域が発見しにくいことが判明した。また、画像とその説明の文章を記述する場合、TABLEタグを使用しているWebページが多いことが分かる。TABLEタグを用いた場合、画像と文章との同期が1対1で表されている場合は、その同期が行って表現され



図 11 プロトタイプシステム

Fig. 11 The prototype system.

表 3 実験結果

Table 3 Experimental result.

Webページ	TABLEタグの使用	画像と文章が1対1	文章1に対し画像多	同期化成功
1	x	○		○
2	x	○		○
3	x		○	△
4	○	○行	○	△
5	○	○列		○
6	○	○行		○
7	○	○行		○
8	○	○列	○	x
9	○	○行	○	△
10	○	○行		○
11	○	○行	○	△
12	○	○列		○
13	○	○行・列	○	x
14	○	○行・列		○
15	○	○行	○	△

た場合も列で表現された場合も、本研究で呈示した方法が有効であることが分かる。しかしながら、1つの文章に対し対応画像が複数表示されている場合、行で同期をとっている場合は同期化可能領域の発見ができる場合とできない場合とがある。このとき、正しく同期化可能領域の発見ができなパターンは、1行に行中の文章と同期していない画像が含まれる場合、その画像も同期化可能領域に含まれてしまう問題が生じた。また、列で同期をとっているときは、同期化領域の発見ができない場合があることが分かった。よって、本実験により、1つの文章に対し画像が複数同期している場合の同期化可能領域の発見方法を再度検討する必要があることが判明し今後の課題となった。

4. スクリプト作成マークアップ言語

4.1 定義

3章では、既存の Web ページを自動変換により番組化を行うことについて述べた。しかしながら、現在の HTML では同期化や演出などの番組化呈示のためのタグ要素がないため、すべての Web ページに対して十分な番組化が行えるとはかぎらない。そこで、本章では、Web 作成者が Web 作成時に自ら番組を意識した Web ページを作成することが望ましいと考え、番組化の台本作成のための、スクリプト作成マークアップ言語である Scripting-XML (S-XML) を提唱する。S-XML は画像や動画と台詞の同期化や演出的要素の設定などをふまえた、番組化を行うことを目的とした言語であり、以下にその利点を記す。

- Web 作成者が番組化を意識したうえで文書を作成することが可能である。番組化において重要なポイントである同期と演出を設定するタグセットを用意しているため、Web 作成時に番組化の設定を同時に行うことが可能である。
- TVML や SMIL, FLASH¹³⁾などの種々の同期化コンテンツ再生環境に対応可能である。画像と文章との同期タグセットを用意しているため、S-XML のタグを種々の同期化コンテンツ再

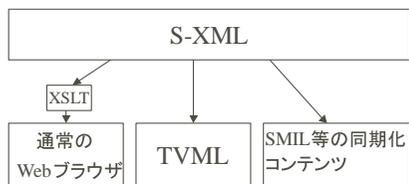


図 12 S-XML コンテンツの表示
Fig. 12 Presentation of S-XML content.

生環境のタグに変換するだけでコンテンツを再生することが可能である。

- XML をベースにしているため、種々の文書とのリンクが可能である。

S-XML は XML の拡張タグ集合であるため、すでに XML にて書かれた文書とのリンクが容易にできる。このため、S-XML は Web 上にある文書のみならず XML で書かれたもしくは変換されたプレゼンテーション用資料やデータベースのデータなどを番組化することが可能である。

- 将来的にデジタル放送を意識した様々な番組化が可能でアプリケーションに適用させることができる。

すでに標準化されているデジタル放送用の拡張制限を入れた BML や BXML, またダイジェスト映像を対象として研究されている PPML との融合が可能である。

S-XML の位置付けを図 12 に示す。現在の S-XML は通常の Web ブラウザと TVML により表示することができるが、将来的には、SMIL や FLASH などの同期化コンテンツを用いて表示することができるようにする予定である。

S-XML は、演出を容易に設定すること、XML で書かれた他の文書を容易に番組化することを可能にすること、および番組化を目的とした文書を他のアプリケーションでも使用できることの 3 点の理由から、番組化に必要であると考えられる要素を下記 2 つの構成で分別した。

- 番組化コンテンツ識別系タグ
XML 文書の文章や画像、動画などのコンテンツを番組化へ利用する方法を定義する。
- 番組化スタイル系タグ
番組化の演出方法を定義する。

4.2 番組化コンテンツ識別系タグ

番組化コンテンツ識別系タグとは、番組のタイトルと問合せを行う際のキーワードの設定、XML 文書内の文章や画像、動画などのコンテンツ中から番組で使用するコンテンツの選択や範囲選択、選択されたコン

表 4 番組呈示コンテンツ識別系タグ一覧
Table 4 TV-program presentation contents tags.

タグ名	特徴
title	タイトルを囲み、キーワードを属性として持つ。キーワードは検索に利用する。
word	ある特定の単語に対応させて表示する画像を指定する。その指定したファイルを同時に提示させたい単語を checkword タグで囲む。
line	台詞としてキャラクタにしゃべらせたい部分を囲む。
img	ファイル名、同期を取るための番号、複数画像の提示方法、配置位置(中央、左方、右方、上方、下方)を指定する。複数画像の提示方法は以下の中から選択する。
movie	ファイル名、同期を取るための番号、動画の開始、終了、再生時間を指定する。
insert	挿入したいファイルを指定する。指定したファイルが複数ある場合は、指定した順に挿入する。
hide	XML 文書としては使用したいが番組では表示したくない部分を囲む。
hlink	リンクのアンカー文字をタグで囲み、番組ではタイトル、Web ページではリンク先を表すアンカー文字として扱う。

テンツどうしの同期対応付けを行うためのタグである。コンテンツ識別系タグの一覧を表 4 に示す。

番組化コンテンツ識別系タグのうち台詞 (line), 同期 (img), リンク (hlink) のタグの特徴を以下に列挙する。

line タグ 台詞としてキャラクタにしゃべらせたい部分を囲む。属性値として以下の事柄を指定することも可能である。

- 台詞と指定した番号の画像や動画との同期化。
- 台詞と異なる字幕を表示。
- リンク先ページの予告紹介のときに元の文書スタイルによっては省略する部分を指定。

img タグ ファイル名、同期をとるための番号、複数画像の呈示方法、配置位置(中央、左方、右方、上方、下方)を指定する。複数画像の呈示方法は以下の中から選択する。

- すべての画像を空間内に配置し、カメラを引いてすべてを同時に映す。
- すべての画像を空間内に配置し、キャラクタが徘徊しながら紹介していく。
- 紙芝居のように 1 つ 1 つを順番に紹介しながら入れ替えていく。

line タグと img タグを使用して画像と文章との同期をとりながら番組コンテンツを作成した例を図 13 に示す。

hlink タグ リンク先を指定する。属性値として以下のようなリンク先の紹介方法を指定することも可能である。

- リンク先ページに含まれている画像のみを紹介する、もしくは <line> タグを使って台詞

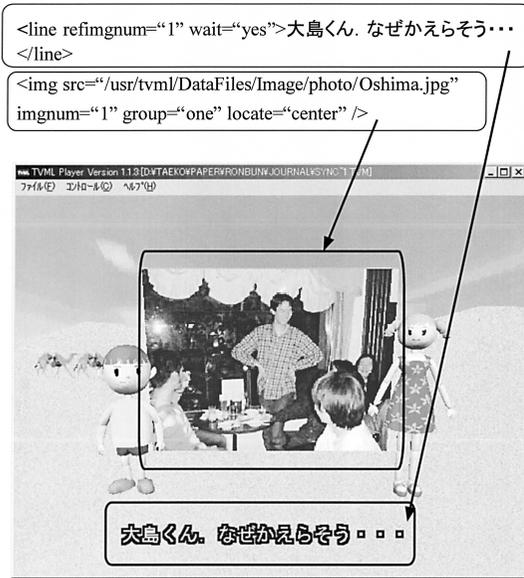


図 13 S-XML 番組化例

Fig. 13 Example of S-XML TV-program presentation.

と画像や動画の同期がとられている部分を取り出してきてリンク先ページの予告編的な番組を提供する。

ここで、自動的に同期部分を抽出する技法は、3章で述べた同期化可能領域の発見手法を用いている。

- 紹介する画像や予告編の作り方として、はじめまたは最後の1つだけを見せる、もしくは上または下からあてはまるものすべてを見せる。
- `<hlink>` タグが出てきたところでリンク先の紹介を行い、すぐさまリンク先へ飛ぶ、もしくはリンク元の番組の再生がすべて終了してからリンク先へ飛ぶ。

リンク先の紹介をするときのスタジオセットやキャラクターなどの演出方法はリンク元ページのスタイルを継承することを基本としたうえで、各々のページ特有のスタイルを反映させる。

図 14 上部のように、XML 文書の本編として番組化するときには、`<variety>` タグで囲まれた部分は 4.3.1 項で述べる演出方法に従って番組化を行い、図 14 下部のように親文書のリンク先の予告編として番組化するときには、親文書のスタイルが `<news>` タグで指定されているので、ニューススタイルのスタジオの中でリンク先で登場するキャラクターがバラエティスタイルを意識した番組展開を行う。

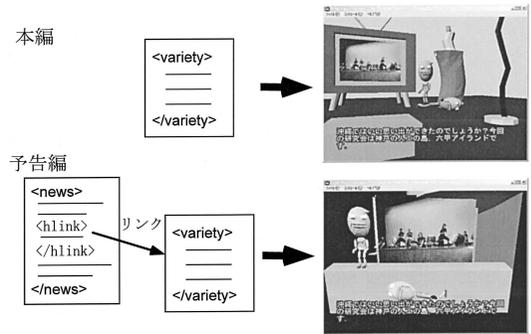


図 14 リンク先の予告編における親スタイルの継承

Fig. 14 The inheritance of the parent style in the presentation of link destination pages.

4.3 番組化スタイル系

番組化スタイル系タグとは、XML 文書に番組化のための演出を付加するためのタグ群である。スタイル系タグはテンプレート系タグとプリミティブ系タグに分けられる。

● テンプレート系タグ

演出を一から詳細に記述しなくても、最低限これらを使用すれば番組化することができるタグ

● プリミティブ系タグ

カメラワーク、登場キャラクターの動きなどの演出を一から自分で記述したいという XML 文書作成者のためのタグ

4.3.1 テンプレート系タグ

S-XML は番組化することを目的とした言語であるため、演出を付加しつつ記述していくことができる。しかしながら、番組コンテンツ作成者が一からカメラ操作やキャラクターの動き、タイミングなど 3 次元空間における番組進行を十分に把握しつつ、テキストで記述していくということは難しい。そこで、作成者がこんな雰囲気、ここでストーリーの転換がきて、といったあいまいな演出イメージからでも番組を作成することができるような、番組作成支援タグが必要であると考え、テンプレート系タグを用意している。

コンテンツ作成者は、`Character` タグで番組内で登場するキャラクターのモデル、声、日本語の場合は方言の初期設定を行う。そして、以下にあげる `news`、`discussion` などのテンプレートタグで、それぞれのテンプレートにあてはめた演出を行いたいコンテンツ部分を囲む。さらに、各テンプレートタグ固有の子要素タグを利用すれば、ストーリー展開やキャラクターの台詞の割振りを指定することができ、文書作成者の意図を番組により反映させることができる。番組形式によって演出や進行方法は異なると考えられるため、テン

レートタグで指定した範囲のコンテンツに適用する。スタジオセットやカメラワーク、キャラクタの台詞の割振りなどの演出の自動生成は、それぞれ異なるアルゴリズムに従って行われる。たとえば、variety タグの子要素として、起承転結タグがある。このタグはストーリー展開を指定するタグであり、このタグに囲まれたコンテンツは、はじめは落ち着いていたキャラクタがそわそわし始めて、騒ぎだし、最後には力ずくでおとなしくさせられてしまうような演出に変換される。

以下に各プレートタグの基本コンセプトとそれぞれに固有の子要素タグの説明を加える。

news メインキャラクタとサブキャラクタに分け、話す量に違いを与える。堅い雰囲気。カメラは基本的に話し手のキャラクタをズームアップ。

topic 主題

detail 主題の詳細

discussion 司会進行役と発言者、反対論者がいる。人数が多数の場合は観客もあり。堅い雰囲気。

issue 議題、論争点

support 支持意見

object 反対意見

school 先生と生徒がいる。先生と生徒の間で質疑応答を繰り返す。

question 質問

answer 回答

variety つねに相槌を入れ、落ち着かないキャラクタがいて、動き回ったりそわそわしたりする。

intro 起。まだ皆落ち着いている。

development 承。少しそわそわし始める。

turnstory 転。さらに動きが激しくなる。

conclusion 結。強引におとなしくさせる。

ghoststory 照明を暗くして下からの青白いスポットライトを当てる。怪談風にゆっくり話す。ストーリー展開は variety と同じ。

dancehall BackMusic が流れ、キャラクタたちは踊り、動き回る。楽しい雰囲気。ストーリー展開は variety と同じ。

4.3.2 プリミティブ系タグ

スタイル系タグのうち、少なくとも上記のプレートタグを使用すれば番組を自動的に生成することはできる。しかし、文書作成者が指定したプレートに基づいた変換によって設定されたスタジオセットやカメラワーク、キャラクタの台詞の割振りとは異なった設定、もしくはさらなる詳細設定を行いたい場合もあると考えられる。そこで、コンテンツの役割の詳細設定やキャラクタの台詞の割振り、パフォーマンスな

```
<xsl:template match="conclusion">
  <xsl:apply-templates select="subject"/>
  <P><TABLE style="background-color:lightgrey; width:600">
    <xsl:apply-templates select="line"/>
  </TABLE></P>
</xsl:template>

<xsl:template match="subject">
  <H3>「<xsl:value-of select="."/>」</H3>
</xsl:template>

<xsl:template match="line">
  <xsl:choose>
    <xsl:when test="@refimgnum">
      <P><TR ALIGN="center"><xsl:value-of select="."/></TR>
      <TR ALIGN="center"><xsl:apply-templates select="."/img"/></TR>
    </xsl:when>
    <xsl:otherwise>
      <TABLE style="background-color:lightgrey; width:600">
        <P><TR><xsl:value-of select="."/></TR></P>
      </TABLE>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>

<xsl:template match="img">
  <TD><xsl:element name="img">
    <xsl:attribute name="src">
      <xsl:value-of select="@src"/>
    </xsl:attribute>
  </xsl:element></TD>
</xsl:template>

<xsl:template match="hlink">
  <xsl:element name="a">
```

図 15 S-XML のための XSLT
Fig. 15 XSLT for S-XML.

どの演出を文書作成者の意図するように一から記述するためのプリミティブ系タグを用意している。プリミティブ系タグを用いて記述すれば、あらかじめ用意されているプレートタグとは異なる独自の番組演出を行うことができる。

4.4 XSLT

S-XML を用いて記述されたページを通常の Web ブラウザ上で表示するためのスタイルシートである XSLT について述べる。以下に XSLT による変換規則を列挙する。

- title タグで囲まれた文字はそのページのタイトルとする。
- キャラクタにしゃべらせる台詞とそれに同期させる画像が指定されていれば、<TABLE> タグを用いて 1 行に台詞、それに対応する画像をその下の 1 行に並べる。
- empha タグで囲まれた文字は「 」をつけて表示し、属性で size や color が指定されていればそれに応じた表示をする。
- hlink タグで囲まれた文字をアンカー文字としてハイパーリンクをつける。

XSLT を図 15 に示す。

4.5 システムの実装

Microsoft Visual Basic を用いて、S-XML 処理系のプロトタイプシステムのユーザインタフェースを作成した。本システムでは、XML DOM をサポートしている Microsoft XML パーサーを使用し、S-XML

```

<?xml version="1.0" ?>
<!DOCTYPE tvprogram SYSTEM "S-xml1.dtd">
<tvprogram>
  <title>田中先生の目から見た DEWS'99</title>
  <hide>田中研究室のホームページの写真館から抜粋</hide>
  <word src="/usr/tvml/DataFiles/Image/photo/sumiya_akama.jpg" wordnum="1"/>
  <style>
    <character model="AHO" name="AHO" voice="e_youngman"/>
    <character model="MARY" name="MARY" voice="j_woman"/>
  </style>
  <variety>
    <visiblechara refnames="AHO"/>
    <visiblechara refnames="MARY"/>
  </variety>
  <intro>
    <line refimenum="1">こちらは今回のワークショップにて、田中先生がデジカメで撮影された画像を編集したページです。田中先生曰わく、「日本のDB界の有名な先生が並んでいますので、学生諸君は顔を見てくださるとのことでした。」</line>
    
    
    
  </intro>
  <development>
    <subject>2日目の夜に開かれた懇親会にて</subject>
    <line refimenum="2">お茶の水女子大学の増永良文先生 DBS研の前々主査でいらっしゃいます</line>
    
    <line>5月はお茶の水女子大学にて、第121回データベースシステム研究会が行われます。</line>
    <hlink address="DBS2000-5.xml" info="line" control="fromtop" jump="last"/>
  </development>
  <turnstory>
    <subject>夕食時のアトラクション(?)</subject>
    <line refmovienum="1">こちらは初日の夕食時に披露された琉球舞踊ですこちらは2日目、懇親会の時の沖縄太鼓です</line>
    <movie src="/usr/tvml/DataFiles/Movie/okinawa.mpg" movienum="1"/>
  </turnstory>
  <conclusion>
    <line>沖縄ではいい思い出ができたのでしょうか？今回の研究会は神戸の人工の島、六甲アイランドです。
    <checkword refwordnum="1">赤間さん</checkword>幹事お疲れ様でした。
  </conclusion>
  </variety>
</style>
</tvprogram>

```

図 16 S-XML ファイル例

Fig. 16 Example of S-XML.

ファイルから TVML スクリプトへの変換を行っている。その結果、生成したスクリプトを TVML プレーヤで再生し、視聴者に呈示する。

S-XML ファイルの一例とそのファイルを番組として再生した結果の一場面と Web ページを図 16, 図 17, 図 18 に示す。図 16 の S-XML ファイルより、図 17 の番組化によるコンテンツの呈示と、図 18 の通常の Web ページによるコンテンツ呈示の 2 つの呈示方法ができる。これらどちらの呈示方法でそのコンテンツを見るかは利用者の選択次第である。

5. おわりに

本論文では、膨大な Web 情報を容易にかつ親しみやすい形で視聴することを目的とし、チャンネルを選択するように Web ページを選択し、テレビ番組のような番組コンテンツを作成して視聴する受動的視聴を提案した。特に、Web コンテンツの番組化は、時間軸の同期や演出が問題となり、これを解決するために、

- 現在の HTML ベースの Web コンテンツを自動



図 17 S-XML による番組化実行結果

Fig. 17 Presentation of S-XML TV-program.



図 18 S-XML による Web ページ表現

Fig. 18 Web page presentation of S-XML content.

で番組型コンテンツに変換する自動変換

- Web コンテンツ作成時に番組化を同時に行える言語であるスクリプト作成マークアップ言語を導入した。

自動変換においては、HTML 文書を自動で番組化コンテンツに変換することを目的とし、HTML 文書の持つ木構造を利用して、画像と文章との HTML 同期化可能領域の発見方法、HTML タグ解析、キャラクタの動きやカメラ情報、照明の配置、スタジオセットなどの演出の付加を行った。同期化可能領域の発見は、画像を中心とした極小タグ領域を発見し、その極小タグ領域を用いて実験を行い、ほぼ同期化可能領域を発見することができた。と同時に、複数画像に対し 1 つの文章のみが対応している Web ページの場合では、同期化可能領域が発見しにくいことが判明し、今後の課題となった。

スクリプト作成マークアップ言語である S-XML の

提唱では、Web 作成者が Web 作成時に番組化を意識しながら Web コンテンツを作成できること、また、既存の XML 文書に演出用タグを付加するだけで番組化ができることの 2 点を目的とし、S-XML の設計・実装を行った。S-XML の実装においては、台詞と画像の同期化、演出の付加を行った表示を行うことができた。しかし、表示された番組の演出は実際の TV 番組に比べればまだまだ十分ではなく、カメラワークなどは実際の TV 番組のレベルに達していない。今後の課題としては、タグセットに応じた番組の演出、リンクで結合された複数の Web ページ群からの番組生成、S-XML ファイルのみならず、HTML ファイルをはじめとした Web コンテンツの利用について考察する必要がある。

謝辞 本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501)による。また、本研究の一部は、次世代情報システム研究所(株)の委託研究の援助を受けている。ここに記して謝意を表す。また、研究に協力していただいた神戸大学大学院自然科学研究科情報メディア科学専攻の草原真知子助教授と神戸大学都市安全研究センター都市情報システム研究分野の角谷和俊講師にも謝意を表す。

参 考 文 献

- 1) PointCast ホームページ,
<http://www.pointcast.com/>
- 2) Nonogaki, H. and Ueda, H.: FRIEND21 project: A Construction of 21st Century Human Interface, *CHI '91 ACM Press*, pp.407-414 (1991).
- 3) Yamaguchi, T., Hosomi, I. and Miyashita, T.: WebStage: An Active Media Enhanced World Wide Web Browser, *CHI '97 ACM Press* (1997).
- 4) Web-TV ホームページ,
<http://www.backweb.com/>
- 5) Yabe, J., Takahashi, S. and Shibayama, E.: Automatic animation of discussions in USENET, *Proc. Advanced Visual Interface 2000 (AVI2000)*, pp.84-91 (2000).
- 6) 植田和憲, 鎌原淳三, 下條真司, 宮原秀夫: シナリオテンプレートによるストーリー性を重視したダイジェスト生成機構, 情報処理学会研究報告, Vol.99, No.61, 99-DBS-119-24 (1999).
- 7) NHK 放送技術研究所: TVML ホームページ,
<http://www.strl.nhk.or.jp/TVML/indexj.html>
- 8) 林 正樹: 番組記述言語 TVML を使った情報の番組化, 情報処理学会 DBS 研究会技術報告,

Vol.2000, No.10, 2000-DBS-120-13, pp.91-98, (2000).

- 9) 久保木準一: デジタル放送と放送業界の動向, 情報処理学会連続セミナー (Nov. 1999).
- 10) SMIL ホームページ,
<http://www.w3.org/AudioVideo/>
- 11) SMIL Boston ホームページ,
<http://www.w3.mag.keio.ac.jp/TR/smil-boston/cover.html>
- 12) 白田由香利, 橋本隆子, 灘本明代, 服部多栄子, 飯沢篤志, 田中克己, 角谷和俊: ダイジェスト映像シーンとマークアップ言語に基づく TV 番組生成システム, 電子情報通信学会データ工学研究会技術報告 (May 2000).
- 13) マクロメディアホームページ,
<http://www.macromedia.com/jp/>

(平成 12 年 6 月 20 日受付)

(平成 12 年 9 月 27 日採録)

(担当編集委員 西田 正吾)



灘本 明代(正会員)

1987 年東京理科大学理工学部電気工学科卒業, 同年(株)TL ヤマギワ研究所入社(株)計算流体力学研究所, クボタシステム開発(株)を経て 1995 年(株)関西新技術研究所入社, 現在に至る。ソフトウェア開発, 可視化技術の研究開発に従事。1998 年より神戸大学大学院自然科学研究科情報メディア科学専攻博士課程に在籍。



服部多栄子(学生会員)

1999 年神戸大学工学部情報知能工学科卒業。現在, 同大学大学院自然科学研究科修士課程に在籍。データベース, ハイパーメディアに興味を持つ。



近藤 宏行(学生会員)

1999 年神戸大学工学部情報知能工学科卒業。現在, 同大学大学院自然科学研究科修士課程に在籍。データベース, ハイパーメディアに興味を持つ。



沢中 郁夫

2000年神戸大学工学部情報知能工学科卒業。現在、同大学大学院自然科学研究科修士課程に在籍。データベース、Webコンテンツの放送配信への利用に興味を持つ。



田中 克己(正会員)

1974年京都大学工学部情報工学科卒業。1976年同大学大学院修士課程修了。1979年神戸大学教養部助手、1986年同大学工学部助教授。1994年同大学工学部教授(情報知能工学科)。1995年同大学大学院自然科学研究科(現在、情報メディア科学専攻)専任教授、現在に至る。工学博士。主にデータベースの研究に従事。人工知能学会、日本ソフトウェア科学会、IEEE Computer Society、ACM等会員。
