

WA-TV：次世代蓄積型テレビのための 番組コンテンツのウェブ化と情報補強

宮 森 恒[†] 馬 強[†] 田 中 克 己^{†,††}

本稿では、テレビ番組をウェブコンテンツにメディア変換し、その結果を、インターネット上での補完情報検索の結果と統合することで、番組の一覧性と理解性を向上させる方式を提案する。提案方式では、まず、テレビ番組に付随するクローズドキャプションを利用して番組を階層的に分割する。次に、補完情報検索を利用することで、分割された各シーンごとに、関連したウェブページが取得される。関連ウェブページと統合された番組コンテンツは、異なる詳細度の階層間を、ズームメタファを用いてシームレスに切り替えられるインタフェースによって閲覧される。TV 番組からウェブコンテンツへのメディア変換によって、テレビ番組の一覧性が向上し、テレビ番組の概要を早見したり、好きなシーンを効率良く探したりすることができるようになる。また、関連情報との統合によって、付加価値が向上し、キーワード検索の手間をかけることなく関連ウェブページを簡単に閲覧できるようになる。本稿では、番組録画後のオフライン利用、および、放送を受信しながらのオンライン利用の両方に対応できるように改良した、次世代蓄積型テレビを想定したプロトタイプシステム“WA-TV”(Webifying and Augmenting TV-content system)を実装、評価することで、提案方式の有効性を確認した。

WA-TV: Webifying and Augmenting TV Programs for Next-generation Storage TV

HISASHI MIYAMORI,[†] QIANG MA[†] and KATSUMI TANAKA^{†,††}

A method is proposed for enhancing the browsability and the understandability of TV programs that converts TV programs into Web content and integrates the results with complementary information retrieval using the Internet. In the proposed method, a TV program is first hierarchically segmented using information in the program's closed captions into different levels of details such as topics, subtopics, etc. Then, related Web pages are retrieved for each segment using complementary information retrieval. The program content integrated with the related Web pages are viewed using the zooming interface where users can seamlessly move back and forth between the storyboard and the normal playback screens showing different level of details. Converting the programs into Web pages enables the programs to be skimmed over to get an overview and for particular scenes to be easily explored. Integrating complementary information enables the programs to be viewed efficiently with value-added content. Implementing and testing a prototype system for next-generation storage TV, “WA-TV” (Webifying and Augmenting TV-content system), which has been improved for usage of both offline and online browsing, validated the approach taken by the proposed method.

1. はじめに

近年、HD レコーダの性能向上にともない、一般の利用者が録画できるテレビ番組のデータ量も飛躍的に増加している。現在、1TB の容量を持つ HD レコーダが登場し、画質モードによっては、1,700 時間以上の

録画が可能である。しかし、録画した映像が増大しても、人間が 1 日に視聴できる時間には物理的に限界がある。このため、膨大な映像データから興味のある部分を効率良く探したり、内容の概要を手軽に理解したり、限られた時間でハイライトだけを見たりするといった、情報を効率良く取得する仕組みが重要となる。

また、テレビ番組は、従来からプロの番組制作者によって作成された映像コンテンツであり、一般にリアリティや質がきわめて高い。しかし、番組制作のためには公共性や視聴率を確保する必要があるため、一定時間内の番組で提供される情報の詳細さや多様さにつ

[†] 情報通信研究機構
National Institute of Information and Communications
Technology

^{††} 京都大学
Kyoto University

いては限界があり、利用者にとって必ずしも十分とはいえないことがある。番組内で紹介された関連情報を必要に応じて詳しくあるいは幅広くいろいろな視点で見たいという要求は、人々の生活スタイルの多様化にともない、今後もますます増加していくと考えられる。このため、番組中に利用者が興味を持った情報や関連する情報を効率良く見ることのできる仕組みが重要となる。

従来、膨大な映像データから興味のある部分を効率良く探したり、内容の概要を手軽に理解したりするといったさまざまな視聴方法についての研究は活発に行われてきた。たとえば、TV2Web¹⁾では、異なる詳細度の構造化データを用意し、ズームメタファによるスムーズな切替えを行うことで、録画番組の一覧性を向上させる手法が提案されている。しかし、構造化データの生成手法等に課題が残っており、実際のテレビ視聴へ応用するレベルには達していなかった。また、ウェブコンテンツをはじめとする関連情報と関連付けることで、アクセス可能な情報のバラエティを補強するというアプローチはこれまでほとんど行われていなかった。

そこで本稿では、テレビ番組をウェブコンテンツにメディア変換し、その結果を、インターネット上での補完情報検索の結果と統合することで、番組の一覧性と理解性を向上させる方式を提案する。TV番組からウェブコンテンツへのメディア変換によって、テレビ番組の一覧性が向上し、テレビ番組の概要を早見したり、好きなシーンを効率良く探したりすることができるようになると考えられる。また、関連情報との統合によって、付加価値が向上し、キーワード検索の手間をかけることなく関連ウェブページを簡単に閲覧できるようになると考えられる。本稿では、字幕データ分割、異なる詳細度を持つ構造化データの自動生成、各詳細度における補完情報検索とその結果との統合といった一連の手法を、番組録画後のオフライン利用、および、放送を受信しながらのオンライン利用の両方に対応させた、次世代蓄積型テレビを想定したプロトタイプシステムWA-TVを実装、評価することで、提案方式の有効性を示す。

本稿の構成は以下のとおりである。2章では関連研究を整理する。3章では、WA-TVの概要と処理手順について説明し、4章では、実験および考察を示す。最後に5章でまとめを述べる。

2. 関連研究

従来、膨大な映像データから興味のある部分を効率

良く探したり、内容の概要を手軽に理解したり、限られた時間でハイライトだけを見たりするといった情報を効率良く取得する視聴方式については、さまざまな研究が活発に行われてきた。

たとえば、映像コーパスから興味のあるショットを探索するために視覚的要素を利用したインタフェースが提案されている²⁾。映像セグメントを表す複数のキーフレームから構成されるストーリーボードと視覚的特徴によるフィルタリング機能を組み合わせることで効率的な探索が可能だとしている。

また、番組概要の把握やシーン探索を効率良く行うための表示インタフェースとしてTV2Web¹⁾やVideo Manga³⁾が提案されている。TV2Webでは、映像セグメントのサムネイルと字幕データから構成されるストーリーボードを異なる詳細度レベル(セグメント、シーン、ショット)で構造化し、これら構造データをズームメタファによってスムーズにつなぎながら(切り替えながら)表示することで効率的な概要把握やシーン探索が可能だとしている。Video Mangaでは、クラスタリングされた映像セグメントが長くて稀なものの重要度を高く計算し、その値に応じて表示するキーフレームの大小を制御する。異なるサイズのキーフレームをマンガのような表示形態に配置しサマリとして表示することで、より重要な場面を視覚的に把握しやすいインタフェースを実現している。文献4)で提案されたシステムは、番組映像中に配信側で用意された汎用的な番組索引と検索用ルールが含まれることを前提とし、これらを利用して利用者の嗜好を考慮したハイライトをダイジェスト視聴することができる。

これらの手法は、いずれもキーフレームや重要な映像セグメントを空間的あるいは時間的に展開することで、一覧性を向上させたり、短時間で概要把握を可能としたりするものであり、基本的に、利用者に提示する情報量を削減するといったアプローチであった。一方、提案方式は、このような一覧性の向上や短時間で概要把握を可能とするアプローチは踏襲しているものの、ウェブの特徴であるハイパーリンクの利用によって、番組映像以外のより詳しい情報あるいは幅広い関連情報を統合表示し、利用者に提示する情報量を必要に応じて増やすアプローチもとっている点がユニークであると考えられる。

特に、TV2Webでは、異なる詳細度の構造化データに対する、ズームメタファによるスムーズな切替え手法が提案されているが、構造化データの生成手法等には課題が残っていた。本手法では、字幕データ分割、異なる詳細度を持つ構造化データの自動生成、各詳細度

における補完情報検索とその結果との統合といった一連の手法を、録画番組およびオンエア番組に対して提案し、次世代蓄積型テレビのプロトタイプとして実装、評価している点が従来研究とは異なるものである。

3. WA-TV の概要

WA-TV (Webifying and Augmenting TV-content system) は、テレビ番組をウェブコンテンツにメディア変換し、インターネット上での補完情報検索の結果と統合する新しい番組コンテンツの視聴方式を提供する。図 1, 図 2 に、WA-TV の概念を示す。利用者は、通常のテレビ視聴ができるだけでなく、ウェブコンテンツにメディア変換された番組内容一覧を示す画面へ自由に行き来することができる(図 1)。また、変換されたウェブコンテンツは番組内容の一覧を示すだけでなく、番組内容に関連したウェブページへのリンクを統合することで番組内容の情報を補強している(図 2)。

WA-TV は、次世代蓄積型テレビのプロトタイプであり、次の特徴を持つ。

- ウェブコンテンツへのメディア変換
- 補完情報検索による情報補強

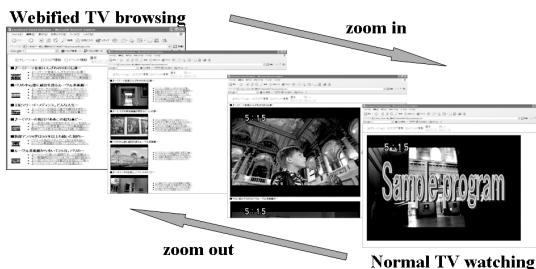


図 1 WA-TV の概念 (1)
Fig.1 Concept of WA-TV (1).

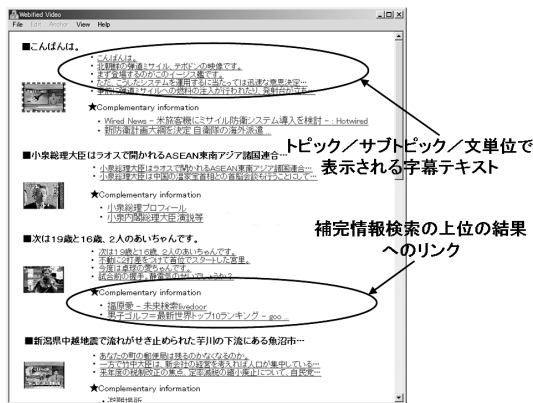


図 2 WA-TV の概念 (2)
Fig.2 Concept of WA-TV (2).

● ズームによる閲覧インタフェース

ウェブコンテンツへのメディア変換により、番組内容の一覧性が向上し、関連情報を補強するための土台が用意される(図 1)。また、補完情報検索により、もとの番組内容に対して、関連情報が新たに追加される(図 2)。最後に、ズームによる閲覧インタフェースにより、表示される内容の詳細度を制御することができ、より直感的で分かりやすい閲覧が可能になると考えられる(図 1)。

WA-TV の処理手順は図 3 のとおりである。まず、テレビ番組を録画し、クローズドキャプション(以後、字幕データと呼ぶ)を用いた階層的なトピック分割を行う。得られた異なる詳細度の字幕データと、対応する部分映像を、ペアとして構造化する。字幕データと部分映像の組は、画面上でストーリーボード形式に表示されるようレイアウトされる。また、補完検索で得られた関連情報は、適宜、ストーリーボード上の対応する位置に統合してレイアウトされる。出力されたウェブコンテンツの表示は、ズーム操作によって、部分映像の大きさが連続的に変化するように、また、異なる詳細度データへの遷移が適宜行われるように制御される。これにより、通常のテレビ視聴画面とストーリーボード閲覧画面をシームレスに移動することができ、番組中の興味のあるシーンを効率良く探索することができる。また、関連情報へのリンクが統合されているため、状況に応じてより詳しい/幅広い情報へ効率良くアクセスすることができる。以下、各手順の処理内容を説明する。

3.1 番組の分割と階層構造化

番組映像の分割および階層構造化は、クローズドキャプションを用いたトピック分割を行うことで実現する。

3.1.1 オフライン分割手法

まず、番組映像の録画が完了した後、オフラインで

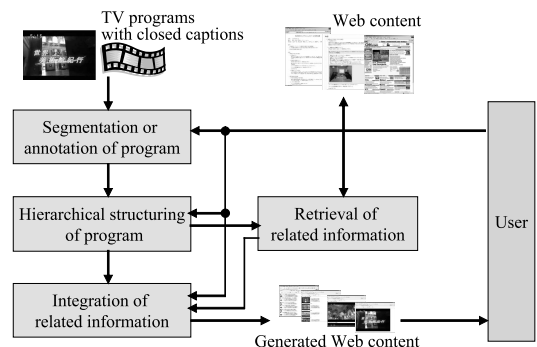


図 3 WA-TV の処理手順
Fig.3 Processing steps of WA-TV.

処理をする場合の分割手法について説明する。

クローズドキャプション（以後、字幕データと呼ぶ）は、番組中の発話がなされた時刻 t と、発話内容を記述したテキスト s のペア (t, s) からなるデータであり、文字多重放送の番組であれば適当なデコーダを使うことで簡単にテキストデータとして記録することができる。

字幕データを用いたトピック分割には大きく分けて2つのアプローチが考えられる。1つは学習に基づくアプローチだが、

- ドメインが限られる可能性があり、汎用的な話題に適用するのが難しい、
- 十分なトレーニングデータが必要、

という問題点がある。もう1つは、統計手法に基づくアプローチである。ここでは、トレーニングデータを特に必要とせず、任意のドメインで利用できるという利点のある、単語分布に基づくトピック分割手法⁵⁾を用いることにした。このトピック分割手法では、以下の条件を前提としている。

- トピックは、各トピックの単語分布によって定義される。
- 異なるトピックは異なる単語分布を持ち、統計的に互いに独立である。
- あるトピックが与えられたとき、そのトピック内の単語は統計的に独立。

この前提条件は、文献 5) において統計的手法による分野非依存のテキスト分割を考えるうえで導入された仮定である。よって、本稿が対象とするさまざまな分野の字幕データの統計的手法によるテキスト分割を考える際においても1つの前提条件として妥当であると考えられる。

いま、 n 個の単語で構成されるテキストを $W = w_1, w_2, \dots, w_n$ とし、 W を m セグメントに分割したテキストを $S = S_1, S_2, \dots, S_m$ とする。分割 S となる確率は、 $Pr(S|W) = Pr(W|S)Pr(S)/Pr(W)$ で表される。与えられた W で $Pr(W)$ は定数であるため、もっともらしい分割 \hat{S} は、 $\hat{S} = \operatorname{argmax}_S Pr(W|S)Pr(S)$ で与えられる。

詳細は文献 5) を参照してほしいが、 $Pr(W|S)$ と $Pr(S)$ は、仮の分割セグメント内の一致単語数やその中の異なる単語数、等で表現することができる。ここで、分割コスト $C(S) = -\log Pr(W|S)Pr(S)$ とし、 $C(S)$ を最小化する S を求める。分割位置とセグメントをそれぞれノード、エッジとするグラフにおいて、コスト最小となるパスを DP アルゴリズムで検出することで、もっともらしい分割 S を得ることがで



図 4 異なる詳細度で分割された字幕および映像の構造化
Fig. 4 Structuring of closed captions and video segmented in different levels of details.

きる。

以上のトピック分割を字幕データ s に適用させることでトピック単位に分割された結果 S を得る。分割された各トピックの字幕データを再び s として同様のトピック分割を行うことで、サブトピック単位に分割された結果 S を得る。映像データは、字幕データの分割結果から得られる時刻に基づいて分割される。

以上より、番組映像は、トピック単位、サブトピック単位、文単位の3つの異なる詳細度データに分割された。これを図4のように階層的に構造化する。

3.1.2 オンライン分割手法

番組を録画後ではなく、放送時に受信しながら視聴したい場合が考えられる。このような場合、字幕データは、番組ごとにまとめたテキストデータの集合ではなく、連続的に受信される断片的なテキストデータとして処理される必要がある。ここでは、受信した字幕データに対する、文 (sentence) の識別を行わず、1回分の受信データをブロック (block) と呼び、分割処理の最小単位とする。

ここでは、同じストーリーや話題について述べている字幕データの類似度と語の共起関係は高いと考え、類似度と共起関係の値およびその変動幅に注目して、セグメントの切れ目を決める手法をとることとした。本手法は、直感的には、受信された字幕データの中でキーワードペア間の共起関係を計算し、共起関係が強ければ、それらの字幕データは1つの話題を述べている可能性が高いと考える。

語 w_i と w_j の共起度 (co-occurrence), $cooc(w_i, w_j)$ を次のように定義する⁶⁾。

$$cooc(w_i, w_j) := \frac{df(\{w_i, w_j\})}{df(\{w_i\}) + df(\{w_j\}) - df(\{w_i, w_j\})}$$

ただし、 $df(\{w_i\})$ は、テキストコレクションにおける、語 w_i を含むテキストの数である。 $df(\{w_i, w_j\})$

は語 w_i と w_j を同時に含むテキストの数である。

分割手順を以下に示す。ここでは、 CT_i を時間 t_i におけるキーワード集合とする。 ST と ET はそれぞれ抽出されるセグメントの開始時間、終了時間、 \emptyset は空集合を表す。

- (1) $CT_0 = \emptyset, ST = 0, i = 1$ とする。
- (2) 字幕データを受信する。データがなければ、終了する。
- (3) 時点 t_i ($i \geq 1$) でデータを受信したら、キーワード集合 K を受信された字幕データから抽出する。
- (4) $CT_i = CT_{i-1} \cup K$ とする。
- (5) CT_i におけるすべてのキーワードペアの中に、共起関係の強いキーワードペアの割合 (co-occurring word frequency), $cwf(t_i)$ を計算する。ここでは、共起関係の強いキーワードペアとは、無向共起度がある閾値 θ より大きい2つのキーワードのことである。 m は CT_i におけるキーワードの数である。また、 $cr(w_i, w_j)$ の cr は co-occurrence relation の略である。

$$cwf(t_i) = \sum_{j=1, k=j+1}^{j=m-1, k=m} cr(w_j, w_k) / \frac{m \cdot (m-1)}{2}$$

$$cr(w_j, w_k) = \begin{cases} 1, & cooc(w_j, w_k) \geq \theta \text{ のとき} \\ 0, & cooc(w_j, w_k) < \theta \text{ のとき} \end{cases}$$

- (6) $cwf(t_i) \geq \Theta$ であれば、(9)へ、でなければ、次へ。ただし、 Θ はあらかじめ定義された閾値である。
- (7) $ET = t_i$ とする。開始時間と終了時間がそれぞれ ST と ET である字幕データのかたまりをセグメント $topic_i$ として出力する。 CT_i は $topic_i$ のキーワード集合として出力され、話題構造の抽出に利用される。
- (8) $CT_i = \emptyset, ST = t_i, i = i + 1$ とする。
- (9) 字幕データを受信する。これ以上のデータがなく、かつ、 $CT_i = \emptyset$ であれば、終了する。これ以上のデータがないが、 $CT_i \neq \emptyset$ であれば、(7)へ。その他の場合、(3)へ。

以上の処理で得られたストーリー (=サブトピック) 単位の分割結果を結合して階層構造化する。結合は、各セグメントのトピック構造の結合に基づいて行う。以下、トピック構造モデルとその結合操作について説明する。

話題構造は主題語と内容語の集合のペアから構成される。主題語は、Web ページやテキストストリームの話題において主題となる語である。本稿では、話題

において、出現頻度の高い、かつ、その他のキーワードとの共起関係の強いキーワードを主題語とする。一方、内容語は、同じページまたはテキストストリームに出現し、主題語との共起関係の強いキーワードである。いい換えれば、主題語はその話題のタイトルを表す役割があり、内容語は話題の本体を表す。

話題構造は2つ以上のノードを持つ、1つの連結成分からなる DAG (Directed Acyclic Graph) を用いて表現できる。つまり、ある話題構造 t は、キーワードを表す頂点の集合 V ($|V| \geq 2$) と、キーワード間の主題語・内容語関係を表すエッジの集合 E ($E \subseteq V \times V, |E| \geq 1$) を用いて表現できる。このような1つの連結成分からなる DAG を話題グラフと呼ぶ。

ある話題構造 t の話題グラフ $G(t)$ は、次のように定義される：

$$G(t) = (V, E)$$

2つの話題構造 t と t' の結合は、この2つの話題構造の話題グラフの和である。ただし、この2つの話題グラフの和は1つの連結成分からなる DAG である必要がある。つまり、2つの話題構造の結合の結果は、話題構造である。

$$t \bowtie t' = \begin{cases} G(t) \cup G(t'), & G(t) \cup G(t') \text{ が 1 つの} \\ & \text{連結成分からなる} \\ & \text{DAG である場合} \\ \phi, & \text{その他} \end{cases}$$

ただし、 ϕ は空を表す。 $t \bowtie \phi = \phi$ とする。

もし、2つのセグメントのトピック構造の結合結果が ϕ でなければ、それらは結合されたと考える。結合可能なすべてのセグメントは1つのトピックとしてまとめられる。すなわち、トピックは、トピック構造の結合に基づいてまとめられた、一連の関連セグメントの集合であるといえる。

$$topic = s_i \bowtie s_{i+1} \bowtie \dots \bowtie s_j$$

$$s_i \bowtie s_{i+1} \bowtie \dots \bowtie s_j \neq \phi$$

$$s_i \bowtie s_{i+1} \bowtie \dots \bowtie s_j \bowtie s_{j+1} = \phi$$

ここで、 s_i はトピックの最初のセグメントであり、 s_{j+1} は次のトピックの最初のセグメントである。

以上より、番組映像は、トピック単位、ストーリー (=サブトピック) 単位、ブロック単位の3つの異なる詳細度データに分割された。

3.2 補完情報検索と統合

補完情報検索は、字幕データから話題構造を抽出し、それに基づく構造化質問を生成することで、従来の類似検索とは異なった、より詳細なあるいは別の観点からのウェブページを検索できる手法である⁶⁾。

話題構造は、タイトルと本体のキーワードを持つため、主題類似・内容類似といった情報検索が可能となるだけでなく、主題は類似しているが内容は異なるといった傾向を持つ、似て非なる情報の検索も可能となる。

ここでは、次に示す 4 種類の質問を生成する。以下、テキストストリームのある話題構造 t を $(\{s_1, s_2, \dots, s_n\}, \{c_1, c_2, \dots, c_m\})$ とする。 $m \geq 1$, $n \geq 1$ である。 s_i と c_i は、それぞれ主題語と内容語を表す。“insubject” と “incontent” に後置される検索文は、それぞれ Web ページの見出しと本文を検索対象とする。“ \wedge ” と “ \vee ” はそれぞれ論理積と論理和を表す。“ \neg ” は、論理否定を表す。たとえば、質問 ($insubject : k_1 \wedge k_2$) \wedge ($\neg(incontent : k_3 \wedge k_4)$) は、 k_1 と k_2 が見出しに含まれ、 k_3 と k_4 が本文に含まれないページを検索する。

- SD (Subject-Deepening) 質問 (Q_{d_s}):

$$Q_{d_s} = (incontent : s_1 \wedge s_2 \wedge \dots \wedge s_n) \wedge (\neg(insubject : c_1 \vee c_2 \vee \dots \vee c_m))$$

- CD (Content-Deepening) 質問 (Q_{d_c}):

$$Q_{d_c} = (insubject : c_1 \wedge c_2 \wedge \dots \wedge c_m) \wedge (\neg(incontent : s_1 \vee s_2 \vee \dots \vee s_n))$$

- SB (Subject-Broadening) 質問 (Q_{b_s}):

$$Q_{b_s} = (incontent : c_1 \wedge c_2 \wedge \dots \wedge c_m) \wedge (\neg(insubject : s_1 \wedge s_2 \wedge \dots \wedge s_n))$$

- CB (Content-Broadening Query) 質問 (Q_{b_c}):

$$Q_{b_c} = (insubject : s_1 \wedge s_2 \wedge \dots \wedge s_n) \wedge (\neg(incontent : c_1 \wedge c_2 \wedge \dots \wedge c_m))$$

以上の構造化質問を生成し、Google API を用いた検索を行うこととした。

3.3 ズームを利用した閲覧インタフェース

階層構造化された番組データと、補完情報検索で得られた関連ウェブページを統合し、ウェブコンテンツとして表示できるようにする。本稿では、字幕データと部分映像の組を、画面上の縦方向に並べるストーリーボード形式を採用することとした。映像と字幕データは、トピック単位、サブトピック単位、文単位といった異なる単位で分割されているが、ここでは説明の都合上、分割された各部分映像のことを単にシーンと呼ぶ。図 5 に示すように、シーンのサムネイルとシーンに対応する字幕データを横に並べ、これを縦方向に時間順に配置する。

図中、シーン 1 がトピック 1 の映像に対応していたとすると、シーン 1 の全テキスト (すなわち、シーン 1 のテキスト 1, 2, 3) がトピック 1 の字幕データに

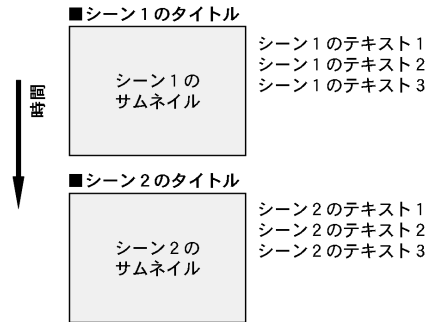


図 5 ストーリーボード表示
Fig. 5 Storyboard display.

対応する。シーン 1 の各テキスト 1, 2, 3 は、シーン 1 のサブトピックの字幕データ 1, 2, 3 にそれぞれ対応する。シーン 1 のサムネイルは、トピック 1 の映像を再生するようにハイパーリンクで関連付けられ、シーン 1 のテキスト 1 は、シーン 1 のサブトピック 1 の映像を再生するように同様にハイパーリンクで関連付けられる。

さらに、補完情報検索で得られた結果を統合表示した例を図 2 に示す。各トピック、各サブトピックごとに、補完情報検索で得られたウェブページへのハイパーリンクが追加されている。これにより、利用者はより詳細な話題、あるいは、その話題を広げるような情報に必要に応じて効率良くアクセスすることができる。

図 1 は、WA-TV の画面遷移の様子である。ズーム操作により、表示映像の大きさが連続的に変化するだけでなく、異なる詳細度データへの遷移が適宜行われる。たとえば、図 4 のトピック 1, トピック 2, ..., トピックを含むストーリーボードは、ズームインの操作によってサムネイルがある大きさに達すると、サブトピック 1, ..., サブトピックを含むストーリーボードに滑らかに表示を切り替える。さらにズームインの操作を続けてサムネイルがある大きさに達すると、サブトピックのストーリーボードは、文レベルのストーリーボードに滑らかに表示を切り替える。さらに、ズームインすると、通常のテレビ視聴画面の表示に滑らかに切り替わる。ズームアウト操作はまったく逆の効果となる。これにより、通常のテレビ視聴画面とストーリーボード閲覧画面をシームレスに横断することができ、番組中の興味のあるシーンや関連情報を効率良く探索することができる。

4. 実験および考察

4.1 オフライン分割手法

ニュース、ドキュメンタリ、ドラマ、スポーツのテ

表 1 オフライン分割結果
Table 1 Result of offline segmentation.

	適合率	再現率	F 値
1 次分割	0.63 (32/51)	0.43 (32/74)	0.51
2 次分割	0.55 (97/177)	0.37 (97/262)	0.44

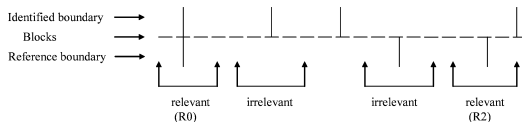


図 6 テキストストリームのセグメンテーションの判定モデル: R_x は, reference boundary と identified boundary の距離 (ブロック数) が x 以下であれば, identified boundary を relevant と判定することを意味する

Fig.6 Model of deciding segment boundary for text stream: R_x means that “identified boundary” is decided as “relevant” if the distance (number of block) between “reference boundary” and “identified boundary” is less than or equal to x .

テレビ番組を対象として, 字幕データを用いたオフライン階層トピック分割の精度について調べた。

その結果, ニュースについては 1 次分割の適合率・再現率がそれぞれ約 0.6, 約 0.4, 2 次分割の適合率・再現率がそれぞれ約 0.5, 約 0.4 であった (表 1)。ただし, これは分割位置の一致を厳密に計算した場合の結果であり, 主観的にはトピック境界でほぼ正確に分割されていることを確認した。正確にならない例としては, 1 トピックが 20 文以上の長い冒頭ニュースの場合, 1 次分割で 2 つ以上のトピックに分割されることがあることがあげられる。しかし, 長いニュースは 2 つ以上に分割されると内容を把握しやすくなるため, 実効的にはあまり問題にはならないことを確認した。しかし, トピック境界があいまいになりがちな, ドキュメンタリ, バラエティ, ドラマ, スポーツについては, 単語分布だけに頼らない改良が必要であることが分かった。文脈の転換点等を示す単語・音響・動画等から境界を推定する分割方法を考える必要がある。

4.2 オンライン分割手法

まず, 28 カ月間の NHK ニュース 7 の字幕データを利用して, 共起度辞書を作成した。この共起度辞書を利用して, 字幕データのセグメンテーションに必要な語の共起度を調べた。共起度辞書に登録されていない語のペアの共起度は 0 とした。なお, ストップワードを省くため, 940 ワード (そのうち 593 個は英単語である) のストップワードリストを作成した。

正解の判断モデルは, 図 6 に示されているように, Informedia⁷⁾ で利用されているモデルを修正したも

表 2 オンライン分割結果 ($\theta = 0.2, \Theta = 0.25$)
Table 2 Result of online segmentation ($\theta = 0.2, \Theta = 0.25$).

	再現率	適合率	F 値
ストーリー分割	0.426	0.330	0.372
トピック分割	0.360	0.387	0.373

表 3 補完検索結果の適合率

Table 3 Precision of complementary retrieval result.

	適合率
詳細な内容の補完	0.63 (32/51)
幅広い内容の補完	0.71 (36/51)

のを利用する。システムが判定したセグメントの境 (identified boundary) と, 人間が判定したセグメントの境 (reference boundary) との距離 (ブロックの数) が 1 以内であれば, システムが正しく判定したという。

実験では, 2 日間の字幕データ (NHK ニュース 7) を用いた (821 ブロック)。人間によって判定されたセグメントは 73 個であった。実験結果を表 2 に示す。これより, Informedia のインクリメンタル・クラスタリング手法の最善の F 値 (0.367⁷⁾ より, 提案手法が良い結果を出していることが分かる。ただし, 表 1 と比較すると, オンライン分割よりもオフライン分割の精度の方が良い結果を示していることが分かる。

4.3 補完情報検索の妥当性

補完情報検索で得られたウェブコンテンツの妥当性に関して, トピック分割がほぼ正確に行えたニュースに関して調べた。関連する補完情報を 2 名の評価者が選択して判断することで調べた結果, 詳細な内容の補完の適合率は約 0.6, 幅広い内容の補完の適合率は約 0.7 程度であった (表 3)。

補完情報検索の結果を複数, 利用者に提示するようにすれば, 補完情報としてある一定の精度で適切な内容を提供できると考えられる。より多くの被験者による評価実験を行う必要があると考えられる。

4.4 ズーミングインタフェースの評価

テレビ番組を早送り・巻戻しする従来のインタフェースと比較する形で, ズーミングインタフェースの使用感に関する評価実験を行った。年齢と性別で分類した計 126 名を対象とし, 実験データは, 30 分のニュース番組 5 本を用いた。

被験者の方には, 番組内容の概要や詳細をたずねる質問にズーミングインタフェースおよび従来インタフェースを用いて回答していただき, 画面表示を見た

表 4 評価に用いた操作画面

Table 4 Operation screen used for evaluation.

操作画面	構成
操作画面 1	階層データ (図 4 で示した 3 階層), ズームあり, 自動トピック分割
操作画面 2	階層データ (図 4 で示した 3 階層), ズームあり, 人手によるトピック分割
操作画面 3	階層データ (図 4 で示した 3 階層), ズームなし
操作画面 4	平坦データ (図 4 で示した文単位のみ), ズームあり
操作画面 5	従来の早送り・巻戻しインタフェース

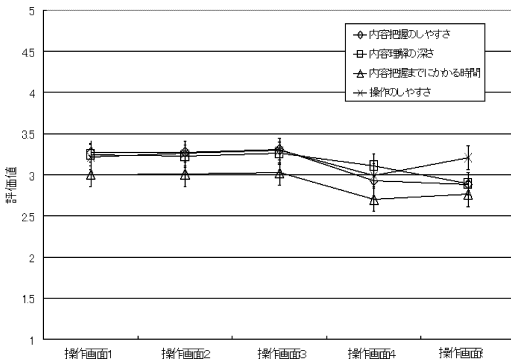


図 7 各操作画面の使用感の比較

ときの内容把握のしやすさ, 内容理解の深さ, 内容把握までにかかる時間, 操作のしやすさといった項目に対するアンケートに 5 段階評価で記入していただいた. 比較に用いた操作画面を表 4 に示す.

図 7 の各操作画面の使用感の結果 (各プロットに 95%信頼区間をあわせて示す) を見ると, 階層データありに相当する操作画面 1~3 の結果はほぼ同等の結果となった. 自動および人手によるトピック分割やズームの有無による違いは, 操作画面の内容把握のしやすさ, 内容理解の深さ, 内容把握までにかかる時間, 操作のしやすさに対して実質的に大きな影響を与えないことが示されていると考えられる. また, 操作画面 1~3 に比べ, 階層なし (操作画面 4) や従来インタフェース (操作画面 5) の結果はいずれも低くなっており, 階層データを構成する有効性が示されている.

一方, 図 8 の今後使っていきたい操作画面のアンケート結果を見ると, ズームありの操作画面 1 or 2 の方がズームなしの操作画面 3 より, より多くの利用者に受け入れられる可能性がある点が示されている. 要因の 1 つとしては, ズームありの方が階層の切替わりをスムーズに表現できるため, 切替わりによる文脈変化を容易に維持でき, より多くの人に分かりやすいイ

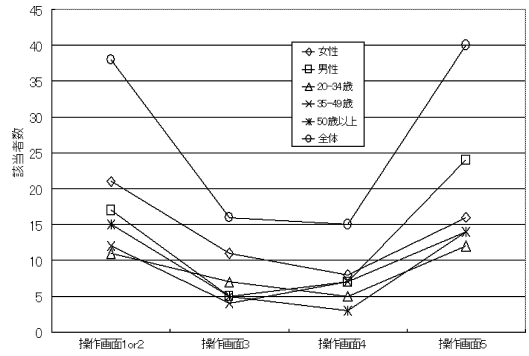


図 8 今後使いたい操作画面

ンタフェースを提供できるためと考えられる.

また, 操作画面 1 or 2 と操作画面 5 の結果を比較すると, ほぼ同程度の数の被験者が従来インタフェースを今後も使いたいと感じていることが分かる. これは, 使用感としてはある程度優れていると感じるズームインタフェースに興味を示す一方で, これまで慣れ親しんでいるインタフェースがいかに根強く利用者に浸透しているかを示す結果となっている. 今後, 新しいインタフェースをデザインする際には, 従来インタフェースをある程度配慮した操作環境を考える必要があることを示している. インタフェースを設計するうえで 1 つの興味深いヒントになると考えられる.

なお, 本評価実験においては, 実験データとしてニュース番組を用いた. 今後の課題として, トピック境界があいまいになりがちなドキュメンタリ, パラエティ, ドラマ, スポーツについての字幕データ分割手法を改良するとともに同様の被験者実験を通して有効性をさらに検証する必要があると考えられる.

4.5 考察

WA-TV の重要な点は, テレビ番組をウェブコンテンツにメディア変換することにより, これまで受動的に視聴するスタイルが一般的だった番組映像を, より積極的に「閲覧」することができるようになった, という点である. 本稿では, 番組内の関連箇所へジャンプするハイパーリンクと, より詳しいあるいは幅広い関連情報へジャンプするハイパーリンクをウェブ上に統合することで, 能動的閲覧が実現されることを示した.

さらに, ウェブコンテンツに変換されていることを利用すると, 複数番組の情報統合も可能になると考えられる. たとえば, 複数の番組を同一の日付でまとめたグループに対する一覧を表示したり, 同一番組・同一トピックを時系列的にまとめたグループに対する一覧を表示したりすること等が考えられる. これらの表

示は、その階層レベルによって、動画像と音響データで再生されたり、テキストデータと音響データのみで再生されたりしてもよい。さまざまな提示方法が提案法の枠組みの中で実現できると考えられる。

番組の分割については、本稿では、字幕データを用いたオフライン分割、および、オンライン分割の2通りの手法を実装した。実験の結果からも分かるように、一般に、オフライン分割の方がオンライン分割に比べて精度の高い結果が得られる。オフライン分割は、番組録画が終了した後に番組視聴をする場合に適用され、見たいシーンを探すためのトピック一覧の正確性を重視した見方を提供する際に有効と考えられる。一方、オンライン分割は、番組放映中に視聴する場合に適用され、より高い頻度で多くの補完ウェブページを提示する見方を提供する際に有効であると考えられる。

WA-TV では、オンエア番組が放送されてから補完検索の結果が提示されるまでにはある程度の時間差が生じる。時間差の主な内訳としては、

- オンエア番組と字幕データの時間差 t_1 (あるシーンが表示されてから、そのシーンに対応する字幕データが提供されるまでの時間)、
- 字幕データの分割および補完情報検索を行い、結果が表示されるまでの時間 t_2 、

があげられる。我々の実験環境では、 t_1 については数秒程度(放送局の字幕化手法に依存する)、 t_2 については3秒程度であった。よって、オンエア番組を録画しながら10秒程度のタイムシフト再生をすることを利用者が許容すれば、これらの時間差を吸収して、オンエア番組とほぼ同期した補完検索の結果提示を実現することも可能であると考えられる。

また、WA-TV では、オンライン分割の精度が共起辞書に依存するため、その辞書に含まれない新語が出てきた場合には適切な補完情報検索の結果が得られる保証はない。しかし、補完情報検索で抽出される話題構造は、キーワードの出現頻度も考慮しているため、字幕データ中に新語が頻りに出現した場合には、話題構造に新語が含まれる可能性があり、そのような場合は新語に対応した適切な補完情報検索の結果が得られることが期待できる。また、たとえば、共起辞書を1日ごとに更新する等の対応をとることにより、ある程度迅速に新語に対応することが可能であると考えられる。

本稿では、オフライン分割として良好な結果が得られたニュース番組を主な実験データとして用いた。オンライン分割および補完情報検索については、ニュース番組以外の他のジャンルの番組を適用した場合、そ

の性能がニュース番組の場合と異なってくる可能性がある。今後、被験者実験とあわせてそれらの有効性を検証する必要があると考えられる。

5. ま と め

本稿では、テレビ番組をウェブコンテンツにメディア変換し、インターネット上での補完情報検索の結果と統合する新しい放送コンテンツの視聴方式を提案した。また、次世代蓄積型テレビを想定したプロトタイプシステム WA-TV を実装した。これにより、TV 番組からウェブコンテンツへのメディア変換によって、テレビ番組の一覧性が向上し、テレビ番組の概要を早見したり、好きなシーンを効率良く探したりすることができるようになることを確認した。また、関連情報との統合によって、付加価値が向上し、キーワード検索の手間をかけることなく関連ウェブページを簡単に閲覧できるようになることを示した。

今後は、分割手法や補完情報検索、インタフェースの改善等プロトタイプシステムの改良をさらに行っていく予定である。WA-TV が手軽に構築でき、まとまった被験者による利用実験を行うことにより、一覧性と理解性の向上がもたらす新しい番組視聴方式の有効性、実用性等をさらに踏み込んだ形で実証することができると期待される。

参 考 文 献

- 1) Sumiya, K., Munisamy, M. and Tanaka, K.: TV2Web: Generating and browsing web with multiple LOD from video streams and their metadata, *ICKS2004*, pp.158-167 (2004).
- 2) Christel, M.G. and Huang, C.: Enhanced access to digital video through visually rich interfaces, *ICME*, MD-L5.1 (2003).
- 3) Uchihashi, S., Foote, J., Girgensohn, A. and Boreczky, J.: Video Manga: Generating semantically meaningful video summaries, *Proc.ACM Multimedia 99* (1999).
- 4) 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: TV 受信端末におけるダイジェスト視聴システム, 情報処理学会論文誌: データベース, Vol.41, No.SIG3(TOD6), pp.71-84 (2000).
- 5) Utiyama, M. and Isahara, H.: A Statistical Model for Domain-Independent Text Segmentation, *ACL/EACL-2001*, pp.491-498 (2001).
- 6) 馬 強, 田中克己: 話題構造に基づく放送と Web コンテンツの統合のための検索機構, 情報処理学会論文誌: データベース, Vol.45, No.SIG10(TOD23), pp.18-36 (2004).
- 7) Hauptmann, A., Chang, J.C., Hu, N.N. and

Wang, Z.R.: Text Segmentation in the Informedia Project. <http://www-2.cs.cmu.edu/~hnn/project/ML-project/ml-report.htm>

(平成 17 年 9 月 19 日受付)

(平成 18 年 3 月 13 日採録)

(担当編集委員 石川 博, 有次 正義, 片山 薫,
木俣 豊, 土田 正士)



宮森 恒 (正会員)

独立行政法人情報通信研究機構 (NICT) メディアインタラクショングループ主任研究員。1997 年早稲田大学大学院後期博士課程修了。1996 ~ 1997 年同大学理工学部助手。

工学博士。主に、マルチメディアコンテンツ処理に関する研究に従事。ACM, 電子情報通信学会, 映像情報メディア学会, 日本データベース学会各会員。



馬 強 (正会員)

1998 年広島県立大学経営学部経営情報学科卒業。2000 年神戸大学大学院自然科学研究科博士前期課程修了。2004 年京都大学大学院情報学研究科博士後期課程修了。同年独立行政法人情報通信研究機構入所。博士 (情報学)。情報検索, マルチメディア情報システム, 情報統合に興味を持つ。IEEE, ACM 等各会員。



田中 克己 (正会員)

京都大学大学院情報学研究科社会情報学専攻教授。1976 年京都大学大学院前期博士課程修了, 工学博士。主にデータベース, マルチメディアコンテンツ処理の研究に従事。IEEE

Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 日本データベース学会各会員。