Video-Augmented Web : ビデオストリームの Web ページへの動的統合

Video-Augmented Web: Dynamic Integration of Video Stream into Web Pages

湯本 高行♠ 吹野 直紀♠ 馬 強♥
角谷 和俊♦ 田中 克己♦

Takayuki YUMOTO Naoki FUKINO
Qiang MA Kazutoshi SUMIYA
Katsumi TANAKA

本稿では，ビデオストリームの Web ページへの動的統合システム Video-Augmented Web を提案する．Video-Augmented Web は，ユーザが Web ページの任意の個所を選択することにより，その内容を適切な長さのビデオによって補強するものである．また，Web ページの意味構造を利用し，これに対して，統合のためのクエリを記述することによって，統合のためのテンプレートを実現する枠組についても述べる．このテンプレートによってさまざまな分野，用途に応じた統合が可能であるが，その一例として，サッカーの試合のニュース記事の Web ページを試合の映像で補強する例を示す．

In this paper, we propose a dynamic integration of video stream into web pages called Video-Augmented Web. Video-Augmented Web augments sentences selected by users in a web page by video clip. We also explain the method to use semantic structure from a web page, and to formulate and describe a query for content integration by using it. This query description is a basis towards a framework for defining content integration template. We show an example of a web page augmentation about football game by video clip.

1. はじめに

現在，ブロードバンドネットワークの普及により，大量のコンテンツが利用できる．また，従来の文書型のWebコンテンツだけではなく，ストリーミング放送のようなビデオコンテンツも急速に増えている．近年の蓄積型テレビの登場や，2003年12月から開始される地上波ディジタル放送はこの傾向にさらに拍車をかけるものである．しかし，これらのWebページやビデオコンテンツは別々に利用されており，両メディアの特徴を生かして利用されているとは言いがたい．そのため，両メディアの統合は新たな利用形態としての可能性を秘めている．

♠ 学生会員 京都大学大学院情報学研究科修士課程
{yumoto, fukino}@dl.kuis.kyoto-u.ac.jp
♥ 学生会員 京都大学大学院情報学研究科博士後期課程
qiang@dl.kuis.kyoto-u.ac.jp
♦ 正会員 京都大学大学院情報学研究科
{sumiya, tanaka}@ dl.kuis.kyoto-u.ac.jp

そこで，本稿では，Webページとビデオコンテンツの動的な統合システムVideo-Augmented Webを提案する．Video-Augmented WebはWebページをもとに，関連するビデオを動的に検索・統合・呈示するものである．統合の中心となるWebページの内容や統合の目的によって，Video-Augmented Webのアルゴリズムは異なる．本稿では，サッカーのニュース記事と同時に試合のハイライトシーンを見る場合を想定する．その目的に合わせ，ニュース記事の意味構造を利用し，その意味構造に対して，クエリを割り当てることによって動的な統合を実現する．

Video-Augmented Webでは，Webページの長所であるインタラクション性に注目し，ユーザに注目している(複数の)文をマウスなどで選択させることにより，ユーザが欲しているシーンをより明確かつ簡単に指定することを可能にしている．

2. Web ページのビデオによる動的拡張

2.1 関連研究

複数のメディアのコンテンツを統合するアプリケーションに関する研究が数多く行われている．馬らの WebTelop[2]は，ビデオコンテンツに同期させてビデオの内容を補完する Web ページを呈示するというものである．呈示されるコンテンツは，リアルタイムに検索し，動的に決定される．これに関連して，Henzinger らは TV closed caption から自動的に質問を生成し，検索を行う Query-Free Search 方式について研究を行っている[2]．また，寺田らのアクティブカラオケ[3]は，歌詞に合わせて動的に画像を呈示するというものである．これらの研究では，アプリケーション利用中にユーザとインタラクションをすることができないが，Video-Augmented Web では文の選択というインタラクションが可能である．また，Video-Augmented Web にはビデオから効率よく映像を探すという側面もある．これに近いアプリケーションとしては，Munisamy らの TV2Web[4]がある．これは，ビデオと字幕データから Web ページを生成し，ズームイン/ズームアウトメタファーによって Web ページ-ビデオ間のシームレス変換を行い，ユーザの見たい個所を，見たい情報粒度でビデオまたはWebページとして閲覧するというものである．TV2Web はインデックスとして字幕データをそのまま表示しているため，ハイライトシーンがわかりづらいが，その点，Video-Augmented Web では Web ページをインデックスとして利用しており，Web ページにはハイライトシーンに関する情報が優先的に記述されているので，ハイライトシーンの視聴が効率的に行える．

2.2 アプローチ

メディアの統合は，互いの表現能力の違いを補い合うことができるという利点がある．また，Webページはビデオページに比べて，インタラクティブ性が高く，ユーザが自由に閲覧できるという利点がある．そこで，本研究ではWebページを中心として，Webページの内容に応じたビデオを呈示する方式を採用する．また，ユーザの要求のうち，ユーザが欲している情報の粒度に注目する．ユーザが注目している情報の内容やその粒度によって，呈示されるコンテンツは異なるべきであるが，その組み合わせは膨大である．さらに，コンテンツにもよるが，ビデオコンテンツはWebコンテンツに比べて，コンテンツの概要をつかむのに時間がかかる．そのため，ビデオを呈示する際にユーザに何らかのインタラクションを要求することによって，何に対するビデオが呈示されるの

かを意識させることが重要である．よって，ユーザの欲する情報の内容と粒度を汲みとるための手段として，注目している文章をマウスで選択する方式を採用する．

選択された文章に関係したビデオシーンの決定には，Webページの意味構造を利用したクエリを用いる．意味構造とは，本研究では，ノードに各段落を対応させ，段落間の話題の包含関係をエッジとして表したツリーである．ユーザによって選択された文に対するクエリを定義することによって，統合時に呈示されるコンテンツを決定する．

2.3 Video-Augmented Webの概要

これらを実現するアプリケーションがVideo-Augmented Webである．Video-Augmented Webは，Webページ表示部分とビデオ表示部分からなる．Webページ表示部分では通常のブラウザと同じようにWebページのナビゲーションができる．図1にVideo-Augmented Webのプロトタイプのスクリーンショットを示す．

図1 Video-Augmented Webのイメージ
Fig.1 Running example of Video-Augmented Web

また，Webページ表示部分の文字列の一部をマウスで選択することがトリガーとなって，ビデオ表示部分にビデオが表示される．ビデオ再生中は選択された文字列はハイライトされる．また，Webページごとに何らかの意味構造が与えられているものとする．

Video-Augmented Webは，ビデオ単体で視聴することが目的ではなく，Webページのユーザが注目している情報を，ビデオで補強することが目的である．また，Webページがある映像の要約になっている場合もある．このような場合には，Video-Augment WebはWeb要約文書を使った映像要約と位置づけることも可能である．

3. 意味構造を利用したクエリ割り当て

クエリは意味構造のノードに明示的に割り当てられる場合とノードのパターンに対して割り当てられる場合がある．これに対して，クエリの処理は以下の3つの手順によって行われる．

- 選択文集合と意味構造のノードとの対応づけ
- 該当するノードまたはパターンの特定
- クエリ内に表れる各ノードに対するクエリの実行

ユーザが選択する文のバリエーションはさまざまであるが，ここでは，それらに対して如何にクエリを割り当てるかについて述べる．

3.1 意味構造

意味構造とは，Webページの段落間の内容的な包含関係を木構造で表したものである．意味構造のノードには段落が，エッジには内容的な包含関係が対応し，親ノードは子ノードの内容を含むものとする．また，ノードの中には段落を構成する文が格納されている．

Web ページからの意味構造の抽出にはさまざまなアルゴリズムが考えられる．対象をサッカーの試合に関する記事とすると，次の2点の特徴がある．

- 親段落に対応する映像区間は子段落に対応する映像区間を包含している．
- 1つの段落内では，イベントが時系列に沿って書かれている．

これらと共に文ごとの特徴キーワードとそのシソーラスを用いることによって，意味構造を抽出する．

3.2 ノードに対するクエリの対応づけ

クエリはノードまたはノード集合に対して割り当てることが可能であるが，その方法は列挙などによる明示的な記述，条件指定(親子関係のあるノードなど)によるパターンによる記述の2種類がある．意味構造を利用して，例えば以下のようなパターンに分類して記述することが可能である．

(1) 選択した文集合が，単一のノードを構成する文集合の一部のとき

$$input \subset sentences(v)$$

(2) 選択した文集合が，複数のノードを構成する文集合に含まれるとき

$$input \cap sentences(v_i) \neq \phi \wedge$$
$$input \subset sentences(v_i) \cup sentences(v_j)$$

ただし，(i=1,2, j=3-i)

(a) さらにノード間に親子関係がある場合

$$parent(v_i) = v_j$$

(b) さらにノード間に兄弟関係がある場合

$$parent(v_i) = parent(v_j)$$

(c) ノード間に上記の関係がない場合

ただし，選択された文集合をinputとし，ノードvに対応する文集合をsentences(v)，ノードvの親ノードをparent(v)とする．

図2に選択文集合の分類例を示す．

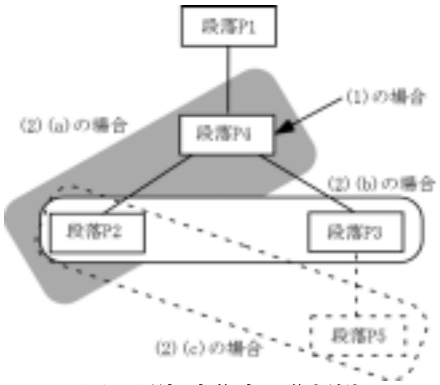

図2 選択文集合の分類例
Fig.2 Example of relationships between selected sentences and semantic nodes

このように意味構造上のパターンに応じてクエリを割り当てることによって，統合に用いるコンテンツを変更してもクエリの割り当て方法は同じものが使える．つまり，統合のテンプレートが実現できる．

3.3 クエリの種類

クエリは，アプリケーションの設計者が記述し，エンドユーザはクエリを意識せずにアプリケーションを利用できる．ノードに割り当てることができるクエリの種類としては，以

下のようなものがある.

- キーワードによる対応するビデオクリップの検索
- ビデオの URI とビデオクリップの開始時間，終了時間によるビデオクリップの明示的指定
- 他のノードに割り当てられたクエリの結果に対する演算の適用
- クエリの結果に対するフィルタリング

フィルタリング関数は，σ_c と表記する. c はフィルタリング結果が満たすべき条件である．c では，ビデオクリップのキーワードや時間的な長さだけではなく，他のクリップとの隣接関係などについての条件も表現できる.

クエリの結果 R は，オリジナルのビデオに含まれるビデオクリップ $r^{(i)}$ (i=1,⋯,n)のリストからなるとする. つまり，$R=[r^{(1)},\cdots,r^{(n)}]$と表せる.

3.4 合成演算

意味構造でのノードの位置を考慮したクエリとして，ビデオクリップの合成演算は非常に重要であり，表 1 のようなものが定義されている. 代数的ビデオ[5]の演算や一般的なグルー結合演算[6]とは若干，定義が異なっている．これらの演算は，選択範囲に応じて呈示されるビデオの区間を変化させるために必要である.

演算対象には，素材と結果の 2 種類がある．ここで，$R_j=[r^{(1)}_j,\cdots,r^{(n)}_j]$ (j=1,2) と定義すると，前者の場合は，$r_1^{(j)}$ の素材となるビデオ上での時間的位置関係に基づいた演算が行われ，後者の場合は，素材となるビデオ上での時間的位置関係に基づいた演算が行われる．これは，本研究における演算が，単にビデオからシーンの抽出だけでなく，ビデオの編集も目的としているからである．また，素材が対象となる演算の結果を構成するビデオクリップのリストの順序は素材ビデオ内の時間的な順序に従うものとする．つまり，演算の結果を $R=[r^{(1)},\cdots,r^{(n)}]$とおいたとき，必ず以下を満たす.

$$i < j \Rightarrow start\ (r^{(i)}) < start\ (r^{(j)})$$

ただし，関数 start(r)は，ビデオクリップ r の素材ビデオ上での開始時間を示す．本研究で定義される統合演算とその例を表 1，図 3 に示す.

表 1 演算の種類
Table 1 Operations

演算名	演算子	対象
連結	∘	結果
区間和	∪	素材
区間積	∩	素材
差	−	素材
ペアワイズグルー結合	⋈	素材
パワーセットグルー結合	⋈*	素材
条件つきグルー結合	$\bowtie_c$	素材
並列化	||	結果

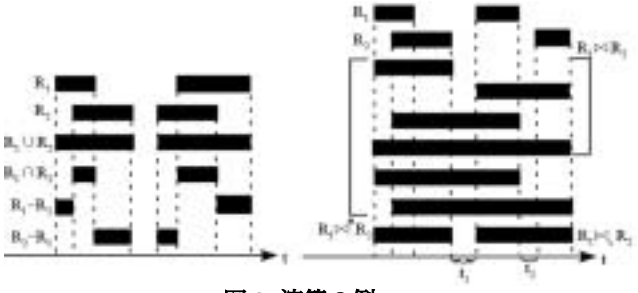

図 3 演算の例

Fig.3 Examples of operations

ペアワイズグルー結合, パワーセットグルー結合の結果はビデオクリップリストの集合である．本研究ではクエリの表現能力を高めるために条件つきグルー結合演算を定義する.

条件つきグルー結合($R_1 \bowtie_c R_2$)

R_1,R_2 のそれぞれから任意の区間から, c で表される条件が真になる組み合わせのみ，区間グルー結合を行い，結合の対象にならなかったものに関しては区間和が適用されるというものである．つまり，下の式のように定義できる.

$$R_1 \bowtie_c R_2 = (\cup \sigma_c (R_1 \bowtie R_2)) \cup R_1 \cup R_2$$

ただし，$\cup S = s_1 \cup \cdots \cup s_n \quad (S = \{s_1, \cdots, s_n\})$ である.

ここで例を示す. 関数 diff(r_1,r_2)はビデオ r_1, r_2 の時間的な距離を表す．つまり，以下の式で表せる.

$$diif\ (r_1, r_2) = \begin{cases} \min(|\ finish\ (r_1) - start\ (r_2)\ |, \\ \quad |\ finish\ (r_2) - start\ (r_1)\ |) & (r_1 \cap r_2 = \phi) \\ 0 & (r_1 \cap r_2 \neq \phi) \end{cases}$$

$R_1 \bowtie_{diff \le t_2} R_2$ は，R_1,R_2 に含まれるビデオクリップのうち, 時間的距離が t_2 以下のものは区間グルー結合演算を行うことを意味する. 図 3 にグルー結合演算の例を示す. ただし，$t_1 > t_2$ である.

4. Video-Augmented Web の具体例

4.1 前提

対象となる Web ページは，Asahi.com のスポーツニュースのうち，サッカーの試合結果についての記事である．呈示されるビデオは記事に対応している試合のビデオがメタデータつきでに蓄積されており，その一部分が呈示される．ユーザはサッカーの試合映像を見ることを目的としているが，すべてを見る時間はないので, Web ページで紹介されているようなハイライトシーンのみを見たいとする．また，意味構造はあらかじめ与えられているとし，文とビデオクリップの対応はとれているとする.

4.2 クエリの例

前節で述べた選択された文の場合分けに応じて，以下のようなクエリを割り当てる．(図 2 参照)　ただし，result(v)はノードまたは文集合 v に対して，呈示されるビデオクリップリストであるとする．ただし，各文にはビデオクリップが対応づけられているとする.

(1)の場合

10 秒以下しか離れていないビデオクリップは因果関係があるとし，10 秒以下という条件の元で条件つきグルー結合を行い，ビデオクリップの間の映像も呈示する．クエリは以下のようになる.

$$\sigma_{diff \le 10s}(\bowtie(result\ (sentences\ (input\ ))))$$

ただし，result(sentences(v))は以下のように定義する.

$$result\ (sentences\ (v)) = \{R \mid R = result\ (s), s \in sentences\ (input\ )\}$$

また，以下のように定義する.

$$\bowtie(\{r^{(1)},\ldots,r^{(n)}\}) = r^{(1)} \bowtie \cdots \bowtie r^{(n)}$$

(2)の場合

10 秒以下しか離れていないビデオクリップは因果関係があるとし, ノード v_1,v_2 のクエリの結果のそれぞれに対して，10 秒以下という条件の元で条件つきグルー結合を行う.

$$result\ (v_1) \bowtie_{diff \le 10s} result\ (v_2)$$

ただし，*input* $= v_1 \cup v_2$， $v_1 \cap v_2 = \phi$ である．

4.3 実行例

http://www2.asahi.com/2002wcup/group_H/jpn_bel/K20020602121.html とその記事から抽出した意味構造をそれぞれ，図 4，図 5 に示す．

図 4 素材となる Web ページ
Fig.4 Web page used as materials

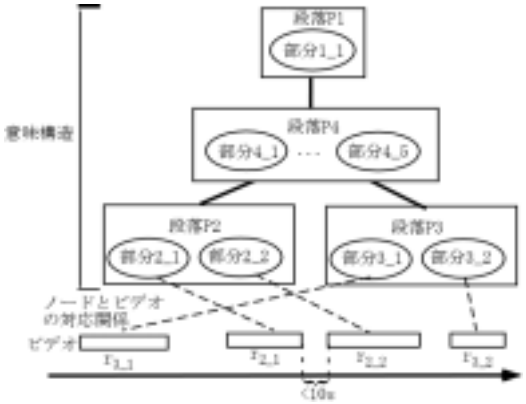

図 5 意味構造とビデオとの対応
Fig.5 Matching semantic structure and video

ただし，各部分にはビデオクリップが対応づけられているとし，それを元にしたクエリの処理が行われるものとする．

この記事からまず，「部分 2_1」をユーザが選択したとすると，結果は r_{2_1} となる．また，「部分 2_1，部分 2_2」を選択すると，結果は $r_{2\_1} \bowtie r_{2\_2}$ となる．「部分 2_1，部分 2_2，部分 3_1，部分 3_2」の場合，$[r_{3_1}, r_{2_1} \bowtie r_{2_2}, r_{3_2}]$となる．

5. まとめと今後の課題

本稿では，ビデオストリームの Web ページへの動的な統合システムである Video-Augmented Web を提案した．本システムでは，ユーザが呈示された Web ページの一部分を選択することにより，その部分についてのビデオが適切な粒度で呈示される．これを実現するために，意味構造を利用してクエリを割り当てるための枠組を定義した．これによって，ユーザからの自由度の高い入力をサポートし，ユーザが見たい部分の情報を見たい粒度で呈示することが可能になる．

今後の課題としては以下が挙げられる．

- 文字の選択履歴を利用した呈示コンテンツの情報粒度の制御
- 演算ポリシーの選択的利用についての検討
- 呈示/インタラクション方法の評価，検討

[謝辞]

本研究の一部は，平成 15 年度科研費基盤研究(B)(2)「蓄積型放送のためのパーソナル視聴の研究」（課題番号：14380177，代表：角谷 和俊）及び平成 14,15 年度基盤技術研究促進事業（民間基盤技術研究支援制度）「クロスメディアコンテンツ基盤技術の研究開発」によるものです．ここに記して謝意を表すものとします．

[文献]

[1] Qiang Ma, Katsumi Tanaka, “WebTelop: Dynamic TV-content augmentation by using web pages”, Proceedings of IEEE International Conference on Multimedia and Expo (ICME2003) v.II, pp.173-176 (2003).

[2] Monika Rauch Henzinger, Bay-Wei Chang, Brian Milch, Sergey Brin, “Query-Free News Search”, Proceedings of the Twelfth International World Wide Web Conference (WWW2003), pp.1-10 (2003).

[3] 寺田 努，塚本昌彦，西尾章治郎，”アクティブデータベースを用いたカラオケの背景作成システム”，情報処理学会論文誌，Vol. 44, No. 2, pp. 235-244 (2003).

[4] Mahendren Munisamy, Kazutoshi Sumiya, Katsumi Tanaka , ”TV2Web: Generating and Browsing Web Contents From Video With Metadata”, 第 14 回データ工学ワークショップ(DEWS2003)論文集, http://www.ieice.org/iss/de/DEWS/proc/2003/papers/8-P/8-P-09.pdf (2003).

[5] Ron Weiss, Andrzej Duda, and David K, “Composition and Search with a Video Algebra”, IEEE MultiMedia, Vol. 2, No. 1 (1995).

[6]プラダン スジット，田島敬史，田中克己，”ビデオデータ検索のための区間グルー操作と解のフィルタリング”，情報処理学会論文誌：データベース，Vol.40，No.SIG3, (TOD1) pp.80-90 (1999).

湯本 高行 Takayuki YUMOTO
京都大学大学院情報学研究科修士課程在学中．2001 京都大学工学部情報学科卒業．情報処理学会，日本データベース学会 各学生会員．

吹野 直紀 Naoki FUKINO
京都大学大学院情報学研究科修士課程在学中．2001 京都大学工学部情報学科卒業．情報処理学会，日本データベース学会 各学生会員．

馬 強 Qiang MA
京都大学大学院情報学研究科博士後期課程在学中．2000 神戸大学大学院自然科学研究科博士前期課程修了．情報処理学会，ACM，IEEE Computer Society，日本データベース学会各学生会員．

角谷 和俊 Kazutoshi SUMIYA
京都大学大学院情報学研究科助教授．1998 神戸大学大学院自然科学研究科博士後期課程修了，博士(工学)．情報処理学会，日本データベース学会，ACM，IEEE Computer Society，映像情報メディア学会各会員．

田中 克己 Katsumi TANAKA
京都大学大学院情報学研究科教授．1976 京都大学大学院修士課程修了．工学博士．主にデータベースの研究に従事．情報処理学会，日本データベース学会，人工知能学会，ACM，IEEE Computer Society 等各会員．