

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

湯本 高行[†] 吹野 直紀[†] 馬 強[†]
角谷 和俊[†] 田中 克己[†]

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 ページの任意の箇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 ページから意味構造を抽出し、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。

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

Takayuki Yumoto,[†] Naoki Fukino,[†] Qiang Ma,[†]
Kazutoshi Sumiya,[†] and Katsumi Tanaka[†]

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 extract 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いえない。

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を実現するアプリケーション 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・表示するものである。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。今回、提案する Video-Augmented Web では対象をサッカーのニュース記事とし、目的をニュース記事と同時に試合のハイライトシーンを見ることとする。その目的に合わせ、ニュース記事から意味構造を抽出し、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。

さらに Video-Augmented Web では、Web ページの長所であるインタラクティブ性に注目し、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。

本稿の構成を述べる。第 2 節では、関連研究について述べる。第 3 節では、Video-Augmented Web の動機や概要について述べる。第 4 節では、今回採用した Web ページからの意味構造の抽出手法について説明し、第 5 節では、この意味構造を用いたクエリの生成手法について説明する。第 6 節では、Video-Augmented Web の実行例について説明する。第 7 節では、本稿のまとめと今後の課題について述べる。

[†] 京都大学大学院 情報学研究所 社会情報学専攻
Department of Social Informatics,
Graduate School of Informatics, Kyoto University

2. 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 WebTelop¹⁾ 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。呈示させるコンテンツは、リアルタイムに検索し、動的に決定される。また、寺田らのアクティブカラオケ²⁾ は、歌詞に合わせて動的に映像を呈示するというものである。どちらも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、これに対して、Video-Augmented Web では、ユーザが文を選択するという行為によって、統合の結果に影響を与えることが可能である。

また、Video-Augmented Web はビデオから効率よく映像を探すと側面もある。これに近いアプリケーションとしては、Munisamy らの TV2Web³⁾ 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うことによって、ユーザの見たい箇所を、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。

統合を実現する枠組としては、SMIL⁴⁾、Cuyper⁵⁾、Macchiate⁶⁾ などが存在する。SMIL は表示に関して高い表現能力を持っており、複数のメディアを時間的に同期させて呈示することが可能であるので、静的な統合は可能であるが、呈示されるコンテンツを動的に決めることはできない。また、Cuyper の Multimedia Formatting Vocabulary では、マルチメディアスタイルシートを実現し、メタデータを利用したクエリや制約が記述可能なため、単純な動的な統合は可能である。Macchiate では、後述する意味構造などを利用したクエリやインタラクションに応じたクエリ、さらにコンテンツの隣接状態などの制約も記述可能なため、Cuyper よりもさらに動的統合の表現能力は高いと言える。また、Macchiate では、統合アプリケーションを簡単に記述することを目的としているため、表示に関する表現能力はあまり高くはないものの、Macchiate は XML で設計されているため、詳細なプレゼンテ-

ションを記述したい場合は、SMIL などの XML ベースの言語と併用することによってそれを補うことも可能である。本研究では、Macchiate をベースとして統合の枠組を実現する。

3. Web ページのビデオによる動的拡張

3.1 動機

近年、Web ページ、ビデオコンテンツ共に増加しているが、これらは専ら別々に利用されている。これら結びつける利用方法としては、前述の WebTelop¹⁾ などのコンテンツ統合が有効である。ここでは統合の目的について、以下の 2 点に注目する。

- メディア毎に異なる表現能力の補完
 - 探索インターフェースとしての Web ページの利用
- まず、前者であるが、文章 (Web ページ) と映像 (ビデオコンテンツ) では表現に向き/不向きがある。このため、Web ページとビデオで表現することに向く内容を表現することによってより表現能力が高いメディアを生成するという目的がある。

続いて、後者について説明する。ビデオコンテンツと Web コンテンツを比較した場合、ビデオコンテンツから受けとる情報は、順序つきでかつ決まった速度で受けとるが、Web コンテンツから受けとる情報は、文書構造による順序こそあるものの、斜め読みや戻り読みなど受けとる情報の順序や速度はユーザに任意である。このため、Web ページは情報を能動的に探索するための手段としてはビデオコンテンツよりも適していると言える。よって、Web ページをビデオの探索インターフェースとして使うことによって、ビデオコンテンツの弱点を補うという目的もある。

また、ユーザの立場から考えると、コンテンツの統合を行う際に、呈示されるコンテンツに対する要求は多様である。しかし、ユーザはその要求を複雑なクエリとして表現することは嫌う。そのため、簡単な操作でユーザの要求を汲みとる必要がある。

3.2 アプローチ

このように、Web ページのインタラクティブ性を生かしつつ、両メディアの表現能力の違いを補い合うために、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粒度によって、Web ページに含んでいるトピックの数は変わってくる。そのため、Web ページと呈示されるビデオコンテンツは 1 対 1 の対応ではなく、Web ページの部分とビデオコンテンツを 1 対 1 に対応する

べきである。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。この意味構造の抽出手法の一例を第 4 節で説明する。この意味構造を利用して、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。これに対する詳細な説明は、第 5 節で行う。

3.3 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ンショットを示す。また、Web ページ表示部分の文字列の一部

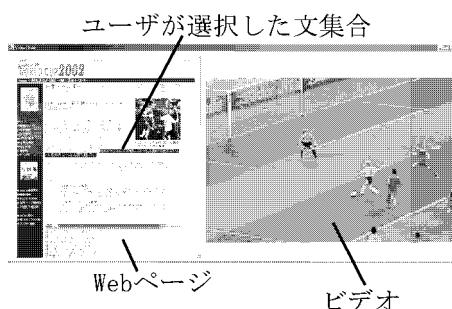


図 1 Video-Augmented Web のイメージ

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。そのため、Web 文書があるイベントの要約である場合

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。

4. Web ページからの意味構造の抽出

Web ページの各部分に対し自動的に映像クリップを対応付けるためには、Web ページの内容を基に映像クリップを特定するためのクエリを生成する必要がある。適切な映像クリップを得るための詳しいクエリを生成するためには、Web ページの意味構造を理解する事が必要になる。

4.1 キーワード表現の形式

キーワードによる表現には、いくつかの形式がある。

- keyword set
一般的な Web 検索と同様に、検索目標を関係のあるキーワードによって特定したい場合、その問い合わせはキーワード集合によって表現される。
- keyword sequence
Web ページに対応付ける対象である映像データは時系列データである。そのため、時系列上に並んだキーワードによって、前後関係を考慮しながら対応クリップを特定した方が適切な場合がある。その場合、質問はキーワードシーケンスによって表現される。
- keyword graph
求める対象の条件にキーワードの親子関係が含まれる場合がある。例えば、キーワード A に該当する部分の中で、キーワード B に該当する部分が欲しい、といった場合である。そのような場合、質問はキーワードをノード、キーワード同士の親子関係をエッジで表わしたキーワードグラフによって表現される。

Video-Augmented Web では、Web ページをこれらが複合した形で表現し、質問を構成する。

4.2 各シーンに関するキーワードの抽出

Web ページは一般に複数のシーンを記述している。それら複数のシーンを切り分けた上で、各々のシーンに関するキーワードを抜き出さなければならない。1 シーンが Web ページ上の 1 文で記述されていると仮定すると、図 2 のような手順で抜き出す事ができる。1 文は keyword sequence によって表現されるが、キーワードの種類によっては順序が関係ない場合もある。

4.3 段落構造に基づく意味構造の抽出

多くの Web ページの文章はいくつかの段落で構成されている。サッカーの試合に対するニュース記事を例に挙げると、まず最初の段落で試合全体について説

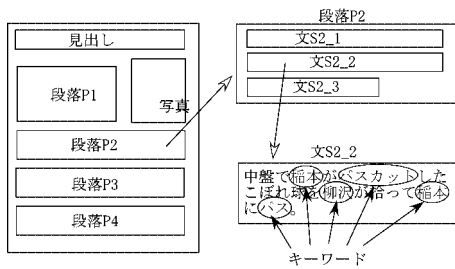


図2 記事構造・キーワードの抽出

明し、次の段落では前半の概要、次の段落は前半の特定シーンについて記述されている、といった構成が考えられる。Web ページの内容に関してこれらの段落構造から推測できる特徴に、次の2点がある。

- 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。

これらの関係をグラフによって表現する。まず、各段落を1つのノードとし、親段落・子段落の関係をエッジで結ぶ。各ノードは複数の文を含む。各文は keyword sequence により表現されている。

4.4 意味構造の修正

Web ページから判断された意味構造と映像のメタデータを対応づける際に矛盾が発生する場合がある。原因は、主に次の2点である。

- 1文と1シーンが対応していない
- 段落構造の推測が間違っている

1つ目は、1文に対応する映像クリップが長すぎる場合や隣り合う2つの文に対応する映像クリップが時間的に近すぎる場合に発生する。2つ目は、段落同士の包含関係が成り立っていない場合に発生する。例として図3(a)と映像のメタデータを比較した結果“文2.2と文2.3は1つのシーンを表現している”、“文3.2は2つのシーンを記述している”、“段落4は段落2に包含されている”という事が分かった場合、図3(b)のように修正される。修正の結果、Web ページの各部分と映像中の各シーンが1対1に対応する。

5. 意味構造を利用したクエリ生成

5.1 選択した文集合とノードとの対応づけ

選択された文集合とノードを対応づけし、対応するノードに割り当てられたクエリを実行することによって、呈示されるビデオクリップを決定する。選択された文集合は、単一のノードだけに対応するだけでなく、複数のノードの対応することもある。そのため、クエリはノードまたはノード集合に対して、参照的に

記述する。クエリを付加する対象となるノードは、列挙などによる内包的な記述、条件指定(親子関係のあるノードなど)による外延的な記述が可能である。意味構造を利用して、例えば以下のように分類して外延的な記述が可能である。

- (1) $input \subseteq sentences(v)$
 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
- (2) $(sentences(v_1) \subseteq input) \wedge (input \subseteq sentences(v_1) \cup sentences(v_2))$
 ただし、 $(i = 1, 2, j = 2 - i)$
 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
 - (a) $parent^+(v_1) = v_j$
 さらにノード間に親子関係がある場合
 - (b) $parent(v_1) = parent(v_j)$
 さらにノード間に兄弟関係がある場合
 - (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。 $parent$ 関数の1回以上の繰り返しである。つまり、以下ようになる。

$$parent^+(v) = (parent \circ \dots \circ parent)(v)$$

図4に選択文集合の分類例を示す。

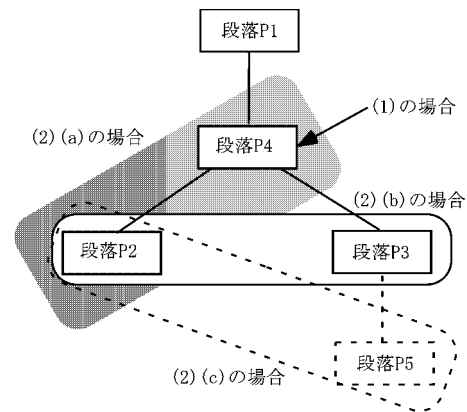


図4 選択文集合の分類例

5.2 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。

- キーワードによって、対応するビデオクリップを検索する。

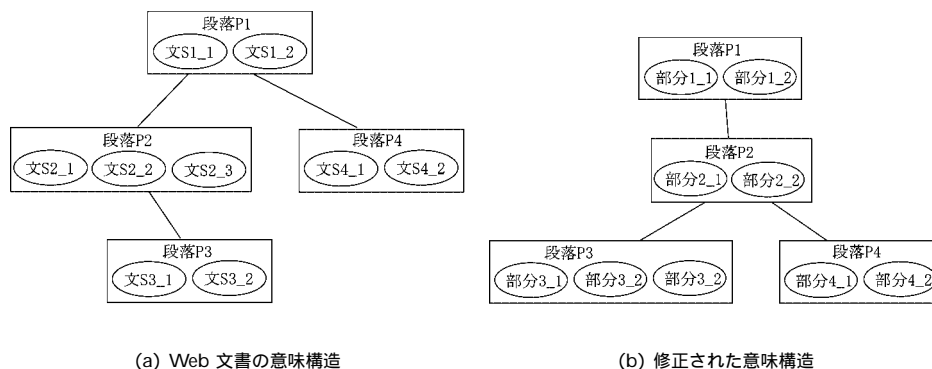


図 3 意味構造

- ビデオの URI とビデオクリップの開始時間、終了時間を指定することによって、表示するビデオクリップを特定する。
- 他のノードに割り当てられたクエリの結果に対して演算を適用する。
- クエリの結果に対して、フィルタリングを実行する。

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 $r^{(i)}$ ($i = 1, \dots, n$) のリストからなるとする。つまり、 $R = [r^{(1)}, \dots, r^{(n)}]$ と表せる。

5.3 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⁷⁾の演算や一般的なグルー結合演算とは若干、定義が異なっている。ここで、 $R_j = [r_j^{(1)}, \dots, r_j^{(n_j)}]$ ($j = 1, 2$) と定義する。演算対象には、素材と結果の 2 種類がある。前者の場合は、 $r_j^{(i)}$ 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 $R = [r^{(1)}, \dots, r^{(n)}]$ とおいたとき、必ず $i < j \Rightarrow start(r^{(i)}) < start(r^{(j)})$ を満たす。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を表 1 に示す。

演算名	演算子	対象
連結	\circ	結果
区間和	\cup	素材
区間積	\cap	素材
差	$-$	素材
ペアワイズグルー結合	$\cdot /$	素材
パワーセットグルー結合	$\cdot /^*$	素材
条件つきグルー結合	$\cdot /_c$	素材
並列化	\parallel	結果

連結 ($R_1 \circ R_2$)

結果 R_1, R_2 を素材ビデオ内での位置関係とは無関係に連結する。つまり、次式 1 のようになる。

$$R_1 \circ R_2 = [r_1^{(1)}, \dots, r_1^{(n_1)}, r_2^{(1)}, \dots, r_2^{(n_2)}] \quad (1)$$

区間和 ($R_1 \cup R_2$)

結果 R_1, R_2 に含まれる任意のビデオクリップ $r_1^{(i)}, r_2^{(j)}$ のそれぞれに対して、代数的ビデオ演算における区間和をとったものである。

区間積 ($R_1 \cap R_2$)

結果 R_1, R_2 に含まれる任意のビデオクリップ $r_1^{(i)}, r_2^{(j)}$ のそれぞれに対して、代数的ビデオ演算における区間積をとったものである。

差 ($R_1 - R_2$)

結果 R_1 から結果 R_2 に含まれるビデオクリップを除いたものである。また、一般には可換ではない。

これらの演算の例を図 5(a) に示す。

ペアワイズグルー結合⁸⁾ ($R_1 \bowtie R_2$)

結果 R_1, R_2 からひとつずつ要素を選び、区間グルー結合⁸⁾ を実行したものである。(式 2 参照)

$$R_1 \bowtie R_2 = \{r | r_1 \bowtie r_2, r_1 \in R_1, r_2 \in R_2\} \quad (2)$$

区間グルー結合の結果は、以下を満たす単一区間になる。

$$x \bowtie y = [r] \quad (3)$$

$$start(r) = \min(start(x), start(y))$$

$$finish(r) = \max(finish(x), finish(y))$$

ただし、関数 $start(x), finish(x)$ はそれぞれ区間 x の開始時間、終了時間とする。

パワーセットグルー結合⁸⁾ ($R_1 \bowtie^* R_2$)

R_1, R_2 のそれぞれからひとつ以上の区間を取り出して、区間グルー結合演算を適用したものである。つまり、式 4 のように定義できる。

$$R_1 \bowtie R_2 = \{r | R'_1 \subseteq R_1, R'_2 \subseteq R_2,$$

$$R'_1 \neq \phi, R'_2 \neq \phi, r = \bowtie(R'_1 \cup R'_2)\} \quad (4)$$

ただし、 $\bowtie(\{r^{(1)}, \dots, r^{(n)}\}) = r^{(1)} \bowtie \dots \bowtie r^{(n)}$ である。

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。

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、式 5 のように定義できる。

$$R_1 \bowtie_c R_2 = (\cup \sigma_c(R_1 \bowtie R_2)) \cup R_1 \cup R_2 \quad (5)$$

ただし、 $\cup S = s_1 \cup \dots \cup s_n$ ($S = \{s_1, \dots, s_n\}$) 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。

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

$R_1 \bowtie_{diff \leq t_2} R_2$ は、 R_1, R_2 に含まれるビデオクリップのうち、時間的な距離が 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 5(b) にグルー結合演算の例を示す。ただし、 $t_1 > t_2$ である。

並列化 ($R_1 \parallel R_2$)

R_1, R_2 の開始時間を合わせて、同時に表示する。

これらの記述は、Macchiate の仕様を拡張することで記述可能である。

6. Video-Augmented Web の具体例

6.1 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。

6.2 クエリの例

第 5 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 4 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。

(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、区間グルーを行い、ビデオクリップの間の映像も呈示する。クエリは以下ようになる。

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

ただし、 $result(sentences(v))$ は以下のように定義する。

$$result(sentences(v))$$

$$= \{R | R = result(s), s \in sentence(input)\}$$

(2)(a) 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、区間グルーを行う。 $result(v_1) \bowtie_{diff \leq 10s} result(v_2)$

ただし、以下が成り立つとする。

$$nodes(input) = \{v_1, v_2\}, v_1 = parent^+(v_2)$$

$node(input)$ は、選択された文集合に対応するノード集合である。

(2)(b) 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_j のクエリの結果のそれぞれに対して、区間グルーを行う。共通の親ノードで呈示される内容とは別の内容を求めていると考えられるので、共通の親ノードを選択したときに呈示されるビデオクリップは除く。クエリは以下ようになる。

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

$$- result(parent(v_1))$$

ただし、 $nodes(input) = \{v_1, v_2\}$ である。

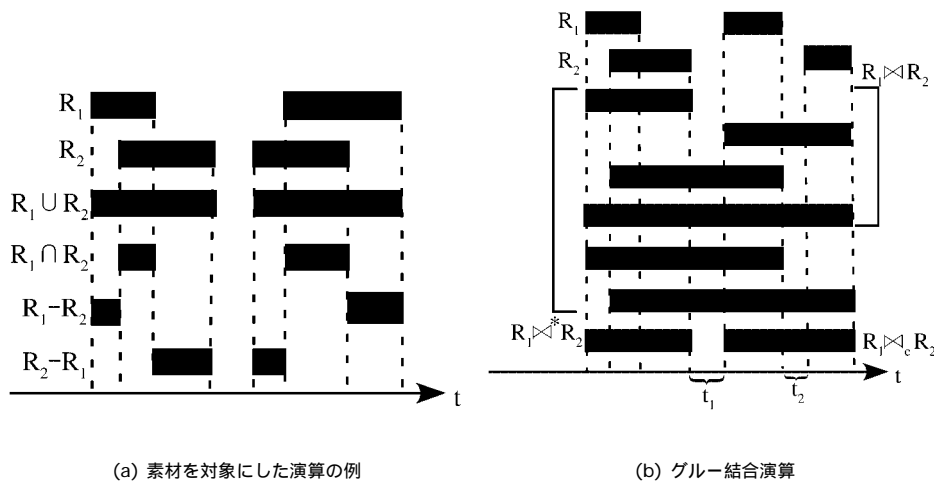


図5 演算の例

(2)(c) の場合

互いに関連性が薄いと考えられるので、結果を連結して呈示する。以下のようなクエリを付加する。

$$result(input) = result(v_1) \circ result(v_2)$$

6.3 実行例

http://www2.asahi.com/2002wcup/group_H/jpn_bel/K2002060402121.html とその記事から抽出した意味構造をそれぞれ、図 6(a)、図 6(b) に示す。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 $r_{2-1} \bowtie r_{2-1}$ となる。「部分 2.1、部分 2.2、部分 3.1、部分 3.2」の場合、 $[r_{2-1} \bowtie r_{2-1}, r_{3-1}, r_{3-2}]$ となる。

7. おわりに

本稿では、Web ページとビデオストリームの動的な統合アプリケーションである Video-Augmented Web を提案した。Video-Augmented Web は、ユーザが呈示された Web ページの一部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 Web ページから意味構造を抽出し、意味構造を利用したクエリを生成するための枠組を定義した。また、ユーザが選択した文を意味構造のノードに対応づけることによって、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。

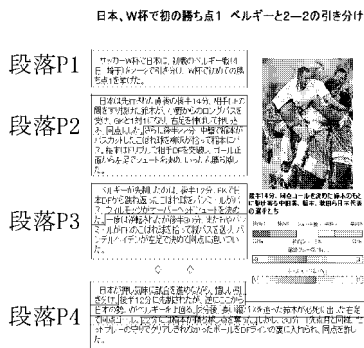
- さまざまな用途でのアルゴリズムの開発と有用性の検証

- インタラクションの種類の追加
- 統合コンテンツの呈示方法の評価、検討

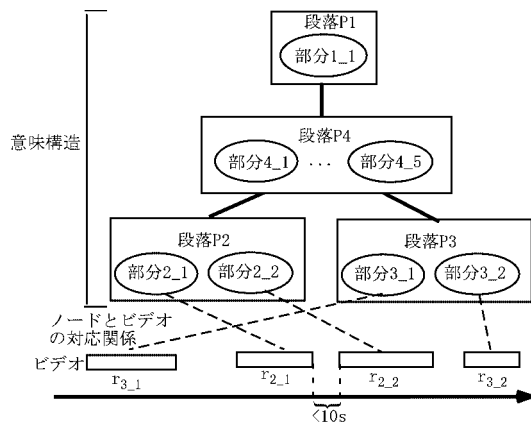
謝辞 本研究の一部は、平成 15 年度科研費基盤研究 (B)(2) 「蓄積型放送のためのパーソナル視聴の研究」(課題番号:14380177, 代表:角谷 和俊) 及び平成 14,15 年度基盤技術研究促進事業 (民間基盤技術研究支援制度) 「クロスメディアコンテンツ基盤技術の研究開発」によるものです。ここに記して謝意を表すものとします。

参考文献

- 1) Qiang Ma, Katsumi Tanaka, "WebTelop: Dynamic TV-content augmentation by using web pages", Proceedings of IEEE International Conference on Multimedia and Expo (ICME2003) (to appear) (2003).
- 2) 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).
- 3) 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).
- 4) W3C Synchronized Multimedia Home page, <http://www.w3.org/AudioVideo/> (2002).
- 5) Jacco van Ossenburg, Joost Geurts, Lynda Hardman, Lloyd Rutledge, "Towards a Multimedia Formatting Vocabulary for Time-Based



(a) 素材となる Web ページ



(b) 意味構造とビデオとの対応

図 6 実行例

Hypermedia", Proceedings of The Twelfth International World Wide Web Conference (WWW2003), <http://www2003.org/cdrom/papers/refereed/p383/p383-ossenbruggen.html> (2003).

- 6) 湯本高行, 馬強, 角谷和俊, 田中克己, "製作者の意図を反映したコンテンツ統合", 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/3-P/3-P-06.pdf> (2003).
- 7) Ron Weiss, Andrzej Duda, and David K, "Composition and Search with a Video Algebra", IEEE MultiMedia, Vol. 2, No. 1 (1995).
- 8) プラダン スジツ, 田島敬史, 田中克己, "ビデオデータ検索のための区間グレ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 (1999).