

GDA を用いた動画検索システムの設計と実装

Design and Mounting of a Multimedia Retrieval System using GDA

伊藤 一成 松井 洋
Kazunari ITO Hiroshi MATSUI

慶應義塾大学 大学院理工学研究科
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
E-mail: k_ito @ nak.ics.keio.ac.jp

ABSTRACT. This paper proposes a media retrieval system based on a meta (annotation) data technology. The system uses MAML(Multimedia Annotaion Markup Language) for the annotation data. MAML is XML based description specification which is independent of the kind of media and formats. Hence, such media as images, sounds and videos can be handled simultaneously. Moreover, the text retrieval techniques proposed in the natural language processing field are applicable, because the annotation data is based on document structures. Our system is capable of ambiguous retrieval by natural language sentence input in addition to the simple matching retrieval by keyword input. In a text retrieval system, the number of hits becomes huge according to the given keyword sets. A time cost of confirmation work which checks whether each result satisfies a demand or not is not questionable in the case of texts, but become a serious problem when it comes to time series data. Therefore, it is important to improve the precision, and is necessary to restrict the hypotheses as much as possible. We confirmed that accuracy is improved by the proposed retrieval method including the case and syntactic analysis between the input text and the annotation data.

1 背景

アノテーション技術を利用したメディア検索の代表例に、Google のイメージ検索が挙げられる [1]. IMG タグによってエンベッドされる画像ファイルに対し、その参照元 HTML テキストをアノテーションデータと考え、キーワード入力による画像検索を実現した。しかしながら、この例のように、多くのユーザが、HTML という統一の記述仕様で、且つ非意識的なアノテーション作業によって膨大な量の検索対象データが作成されている事例というのは、ほとんど無いに等しい。アノテーションに基づく応用システムの実現を考える際に、定められた記述仕様に即したアノテーション生成のプロセスをどう克服するかという困難な問題に直面するため、近年まで実用に結びつく関連研究は皆無に近い状態であった。しかしながら、メディアに対する検索や要約の需要は増すばかりであり、動画等の非テキストデータのコンテンツ記述を目的とした MPEG-7[2] や、Web 上に点在するファイルを対象に機械的処理を前提とする付与データの記述方式である Semantic Web RDF[3] 等、メタデータを意識した統一仕様が発表されはじめてきた。さらには、TREC に Video Retrieval Track が追加され [4]、メディア検索に関する議論が今後ますます活発になるのはいうまでも無い。

2 目的

テレビドラマ、ニュース、語学教材などの動画・音声データはエンターテインメントや教育等にまつわる用途をそれぞれとして持っているが、これに映像情報や意味内容に基づくインタラクティブな検索や提示が可能となれば、実用的な価値を更に高められる。そのような対話的なアクセスが可能な知的コンテンツは機械翻訳、情報検索、自動要約、検索応答、知識発見システムなどを実用化する上で、今後

益々必要になってくると考えられる。テキストを、音声や動画と有機的に結び付けることは、これらの技術の基礎研究と応用開発の推進に寄与するであろう。また情報化社会の今、膨大な量のコンテンツの中からユーザが自分の目的とするデータを自分で探しだし、能動的にアクセスする作業が必要である。これらマルチメディア情報を効率よく検索したり要約する手段として、近年メタデータ技術が注目を集めている。これは対象とするマルチメディア・コンテンツからその検索対象となる特徴を予め記述しておき、記述データを直接の検索対象とすることでマルチメディア・コンテンツの検索を代替しようというものである。ここで、記述データはメタデータなどと呼ばれる。特に注釈の意味でつけられるメタデータをアノテーションデータと呼ぶ。現在インターネットに代表される情報空間上には、テキスト、画像、音声、動画など様々な形態のメディアデータが混在しており、そしてメディア毎に見ても、非常に多くのファイルフォーマットが存在する。これらメディア群を一括に利用処理するには、メディアデータの種類の依存しない、統一的なアノテーションが必須である。本プロジェクトでは、アノテーション記述言語 MAML を利用したマルチメディア検索システムを提案した。ところで、従来のキーワード入力による単純マッチングをベースとした検索システムでは、利用者が指定するキーワードが単純なもの、一般的なものであるほどヒット数が膨大となり、要求を満足するデータの獲得を困難にしている。テキストを対象にした検索システムの場合、個々の結果の対象ファイル全体を閲覧し、実際に内容が合致するものが見極めるのに時間的コストはほとんど要しないが、時間軸を有するデータ（動画、映像、音声）の場合、再生して確認しなければならないため、情報獲得までの総時間は爆発的な増加を招く。そのため、ファイル単位の粒度で結果を表示するのではなく、適合する時間区間のみを返す仕組みが必要である。また、ユーザの要求を満たす結果を可能な限り適確

表 1: メディアが有する情報

対象メディア	時間軸の有無	包含する情報		
		音声情報	映像情報	内容情報
動画 (video)	○	○	○	○
映像 (projection)	○	×	○	○
音声 (sound)	○	○	×	○
画像 (image)	×	×	○	○
テキスト (text)	×	×	×	○
仮想データ (null)	×	○	○	○

に、つまり適合率を最重視して結果を返さなければならない。本研究では、自然文章による検索入力を行い、格解析、構文解析を含めた独自のマッチング処理を行うことで、検索精度の向上を試みた。

3 アノテーションスキーマ

3.1 メディア情報の記述

本プロジェクトにおいて、メディア情報のアノテーション記述仕様、MAML (Multimedia Annotation Markup Language) [5] を策定した。MAML は、人間が理解及び、記述しやすい表現構造を念頭においた、メディアの種類やフォーマットに非依存な統一記述仕様である。タグによるデータの構造化は最小限にし、自然文章中心の構造となっている。MAML では、動画 (movie)、映像 (projection)、音声 (sound)、画像 (image)、テキスト (text) 等のメディアに付加されるアノテーションデータは、エレメントと呼ばれる基本単位の集合体により構成され、すべてのエレメントは時間軸と音声情報 (speech)、映像情報 (projection)、内容情報 (contents) の 3 層から構成される空間上にマッピングされる。音声情報とは人の聴覚から得られる情報、映像情報とは人の視覚から得られる情報、内容情報とはコンテンツに対して人の知識、主観、思考によって導かれる情報をいう。例えば音声データは、時間軸を有する音声情報と内容情報の 2 層上にエレメントがマッピングされ、画像データは時間軸を有しない映像情報と内容情報の 2 層上に配置される。それぞれのメディアが有する情報をまとめた表を表 1 に示す。文書、HTML、XML 等のテキストデータに対しても同様にアノテーション付けできる。また既存のテキストデータ自体も仮想メディアファイル (null ファイルと呼ぶ) に対するアノテーションデータと考える。これにより、複数のメディアを一元的に記述・処理可能な環境の構築が可能となる。MAML は XML (eXtensible Markup Language) [6] の規格に基づいている。図 1 に記述例を示す。最上位に <maml> タグ、その下層に <media> タグを記述する。<media> タグの属性には対象メディアについての情報が記述される。さらにその下層にアノテーションの基本単位となる <element> タグを列挙していく。<element> タグの属性には各エレメントを識別するための id 属性、及び対象メディアが時間軸を有する場合、エレメントの開始時刻を begin 属性で、終了時刻を end 属性で記述する。その下層に種別タグ、さらにその下層にそれぞれの種別毎にそのアノテーションの内容を分類するためのクラスタグを記述する。クラスの種類を表 2 に示す。MAML は、一般の人々でも容易にタグ付けできることを目的としており、クラスの各項は名前から直感的に把握でき、且つ分類数は最小限にとどめている。実際のコンテンツは自然言語で記述する (以下アノテーションテキストと呼ぶ)。普段使っている自然言語で記述するので、当然専門的な知識が無くともアノテーション作業を行うことができる。また表現の自由度が格段に高まる。本システムでは、MAML をリソースに用いて、タグ構造に基づく記号処理とアノテーションテキストを対象にした自然言語解析両面からの検索処理を試みる。エレメン

```
<?xml version="1.0" encoding="UTF-8"?>
<maml>
  <media type="video"
    maml-location="http://www.xxx.jp/abc.maml"
    media-location="http://www.xxx.jp/abc.mpg"
    duration="01:42.562">
    <element id="s1">
      :
    </element>
    <element id="s2">
      :
    </element>
    :
  </media>
  :
</maml>
```

図 1: MAML の記述例

表 2: 情報の種別

種別	クラス
音声情報 (speech)	発話 (utterance)
	擬音 (imitation sound)
	音楽 (music)
映像情報 (projection)	物体 (object)
	人物 (character)
	文字 (letter)
	背景 (background)
	場所 (place)
内容情報 (contents)	説明 (explanation)
	要約 (summary)
	補足 (supplementation)
	タイトル (title)
	その他 (other)
	アンカー (anchor)

トの記述例を図 2 に示す。この例は、6.789 秒から 9.225 秒にかけて「伊東は全域で晴れていた。」という発話が存在することを意味する。

3.2 言語情報の記述

検索処理の過程でアノテーションテキストに対して自然言語解析を行う。その際必要な言語情報を予め GDA (Global Document Annotation) [7][8] によって記述することにした。これにより検索処理の高速化が図られる。GDA は、産業技術総合研究所の橋田らが提唱する、多言語間に共通の統語・意味等に関する XML タグ集合である。GDA タグセットは、品詞、係り受け、代名詞の指示対象、多義語の意味など、かなり細かい情報まで記述できる。本件では、形態素情報、構文情報、単語の読み、動詞の語幹、助詞の種別を GDA タグにより付与する。図 3 は、図 2 のエレメント記述に対して GDA タグを付与した例である。名前空間の記述は、紙面のスペースの都合上割愛してある。<su> は文を示す。<n>、<np> は名詞および名詞を主辞とする語句。<n> は他の語句の係り受け対象となることができ、<np> はならない。他にも同様である。<v>、<vp> は動詞、助動詞、終助詞、またはそれらを主辞とする語句。<ad>、<adp> は、終助詞以外の助詞、副詞、連体詞、接続詞、およびこれらの投射である。prn 属性値は単語の読み仮名、bfm 属性値は動詞の基本形、sem は助詞の種別を表している。GDA タグに関する詳細な情報は文献 [7]、[8] を参照されたい。

3.3 参照構造の記述

参照構造の記述とは、エレメント間の関係を定義するために付与するデータのことである。GDA には、照応、代入、省略を明記する関係属性と呼ばれるタグが存在する。図 4 に省略の記述例を示す。agt 属性と id 属性から、「活躍し」の行為者が「イチロー」であることを示している。GDA の関係属性は、本来、文間の意味的、構造的なつながりを記述しているが、MAML 上で見れば、それはエレメント間の関連付けを施す役割も果たしている。一方、

```

<element id="s10" begin="6.789" end="9.225">
<speech>
<utterance who="p1">
伊東は全域で晴れていた。
</utterance>
</speech>
</element>

```

図 2: エレメントの記述例

```

<element id="s10" begin="6.789" end="9.225">
<speech>
<utterance who="p1">
<su>
<adp>
<placementp prn="いとう">伊東</placementp>
<ad prn="は" sem="係助詞">は</ad>
</adp>
<adp>
<np prn="ぜんいき">全域</np>
<ad prn="で" sem="格助詞">で</ad>
</adp>
<v>
<vp prn="はれ" bfm="晴れる">晴れ</vp>
<ad prn="て" sem="接続助詞">て</ad>
<v prn="い" bfm="いる">い</v>
<v prn="た" bfm="た">た</v>。
</v>
</su>
</utterance>
</speech>
</element>

```

図 3: エレメント中の GDA タギング例

MAML が定義する参照構造とは、単にファイルやエレメント間の参照を目的とするもので、XHTML1.0[9] で定義されている anchor タグと機能的に等価である。参照元はアノテーションテキストの一部である。参照構造は `<a>` タグによって、対象とするテキスト部分を囲む形式で表記する。`<a>` タグは `href` 属性を持ち、属性値は参照先ファイルの URI である。特定のエレメントに対象を限定する場合、URI の後に `#` とフラグメント識別子と呼ばれる ID 名を追加する。記述例を図 4 に示す。この例では、“イチローのマリナーズ入りが決まった。”という文に対して、`http://www.xxx.jp/ichiro.maml` 中の `id` 属性値が“p6”であるエレメントへの参照を定義したことになる。図 4 の下部に示した例のように、参照先は MAML ファイル以外でもよい。参照先が同一 MAML ファイルの場合、`href` 属性の URI 記述部は省略可能である。

4 メディア統合処理システム

我々は今年度のプロジェクトにおいて、MAML と GDA をアノテーションデータに用いて、メディアデータの検索、要約、対話処理等の実現を目的とするシステム MDAS (Multimedia Data Acquisition System) を開発した。システムの概要図を図 5 に示す。本システムは現在の所、アノテータが利用する MDAS Annotation Tool、検索者が利用する MDAS Retrieval Tool、複数のモジュールと XML-DB なる MDAS Server から構成されている。個々の構成要素について解説する。

4.1 MDAS Annotation Tool

MDAS Annotation Tool は、アノテータの MAML ファイル生成を視覚的操作により効率化する統合的アプリケーションである。MDAS Annotation Tool の画面構成を図 6 に示す。図 6 の下部に表示されているアノテーションボードは、MAML ファイルの情報を視覚的に表示したイ

```

<element id="s23" begin="13.521" end="16.385">
<speech>
<utterance who="p1">
<a href="http://www.xxx.jp/ichiro.maml#p6">
<persname id="g1">イチロー</persname>
のマリナーズ入りが決まった。
</a>
</utterance>
</speech>
</element>
:
<element id="s28" begin="21.013" end="23.385">
<speech>
<utterance who="p1">
<a href="http://mlb.mlb.com/index.html"
メジャーリーグ
</a>
でも<vp agt="g1">活躍し</vp>てくれるだろう。
<persname id="p1">村田</persname>
がお伝えしました。
</utterance>
</speech>
</element>

```

図 4: 参照構造の記述例

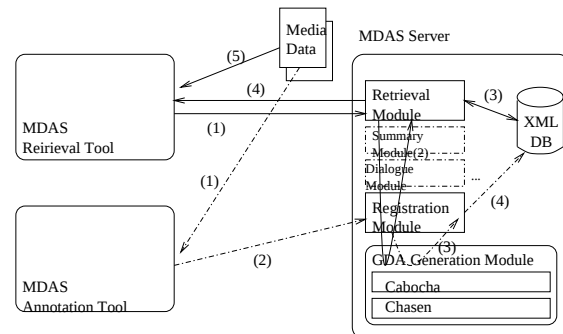


図 5: システム構成及び登録検索処理の手順

ンタフェースである。横軸はメディア上の時間軸を表し、各行(トラック)はそれぞれ「音声情報」、「映像情報」、「内容情報」に関するエレメント群が時系列に沿って表示される。アノテーションボードに構築されたエレメント情報は、MAML ファイルとして保存できる。既存の MAML ファイルを読み込んで、アノテーションボード上に情報を配置することも可能である。このボード上で、マウス操作によるエレメントの追加/削除、およびタイムスタンプの修正を行うことができる。生成されたエレメントはアノテーションボード上でクラス毎に異なった色の長方形で表示され、その長さや位置はメディア上の時間情報を表している。アノテーションテキストはその内部に表示される。

4.2 MDAS Retrieval Tool

MDAS Retrieval Tool は、検索要求及び検索結果表示を行うための標準ツールである。画面構成を図 7 に示す。図 7 の左下に表示されているウィンドウを検索結果表示ウィンドウと呼ぶ。このウィンドウの下部には、検索条件を入力するための各種フィールドが配置されている。またウィンドウ上部に、検索の結果返されたエレメント群の情報を検索スコア、時間長、アノテーションテキストをカラムとする表形式で表示する。表の各行をクリックすることで、その対象メディアの特定区間を再生することができる。

ここで本システムにおける検索要求表現について述べておく。検索要求表現とは、検索条件を記述したデータをいう。検索要求表現は検索文章、種別制約、時間長制約から構成される。検索要求表現は、図 8 に示される XML 形

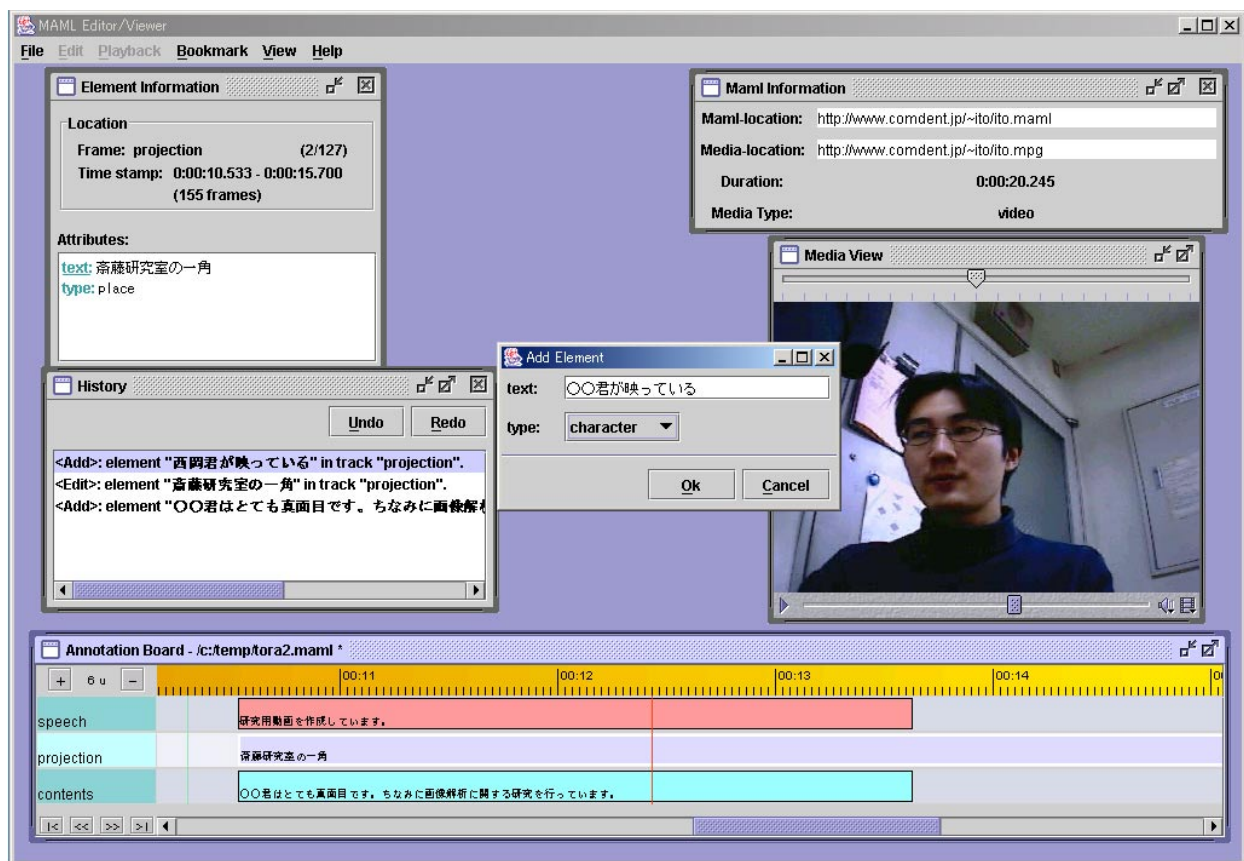


図 6: MDAS Annotation Tool のスクリーンショット

```
<request type="retrieval">
  <sentence>
    イチローが佐々木や野茂といった
    日本人投手と直接対決している場面。
    ヒット、ホームラン。メジャーリーグ。
  </sentence>
  <mediatype>
    <video/>
    <projection/>
  </mediatype>
  <infotype>
    <projection/>
    <contents/>
  </infotype>
  <length>
    <longer-about>120</longer-about>
    <shorter>300</shorter>
  </length>
</request>
```

図 8: 検索要求表現の例

式でサーバ側に送出される。検索文章とは、ユーザの要求を自然文章で表現したものである。複文で構成されていてもよい。検索文章中のある文が句点によって区切られる名詞（複合名詞を含む）または未知語だけで構成されている場合、その検索文がキーワードの組で与えられたとみなす（以下キーワード入力文と呼ぶ）。例えば図 8 に例示されている“イチローが佐々木や野茂といった日本人投手と直接対決している場面。ヒット、ホームラン。メジャーリーグ。”という検索文章は“イチローが佐々木や野茂といった日本人投手と直接対決している場面 and （ヒット or ホー

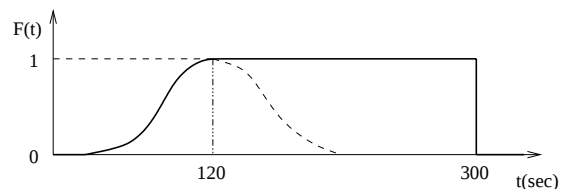


図 9: 時間長制約関数の例

ムラン) and メジャーリーグ”を意味することになる。ユーザは、文章入力によるあいまい検索とキーワード入力による厳密なマッチング処理を必要に応じて選択・融合することができる。GDA Generation モジュールにより検索文章に対して GDA タグ付与された XML データが、実際の検索要求表現としてサーバ側で処理される。また種別制約により、検索対象とするメディアの種別及び、情報の種別を指定する。図 8 では、メディアの種別として動画または映像、情報の種別として、“映像情報”または“内容情報”に対象を限定している。さらに時間長制約により、返されるメディアの全長に制限を課す。時間長制約の記述は、ファジー関数 $F(t)$ に変換される。ここで t は時間長（秒）である。 t 秒以上、 t 秒以下の表現には、 t 秒を閾値とする単位ステップ関数で定義する。また約 t 秒の表現は、中心 t 、中心値 1、標準偏差 $t/3$ のガウス関数で定義する。図 8 は、“約 2 分以上 5 分以下”の制約を課した例である。このファジー関数表現を図 9 に示す。

4.3 MDAS Server

MDAS Server は、Annotation Database 及び複数のモジュールから構成される。

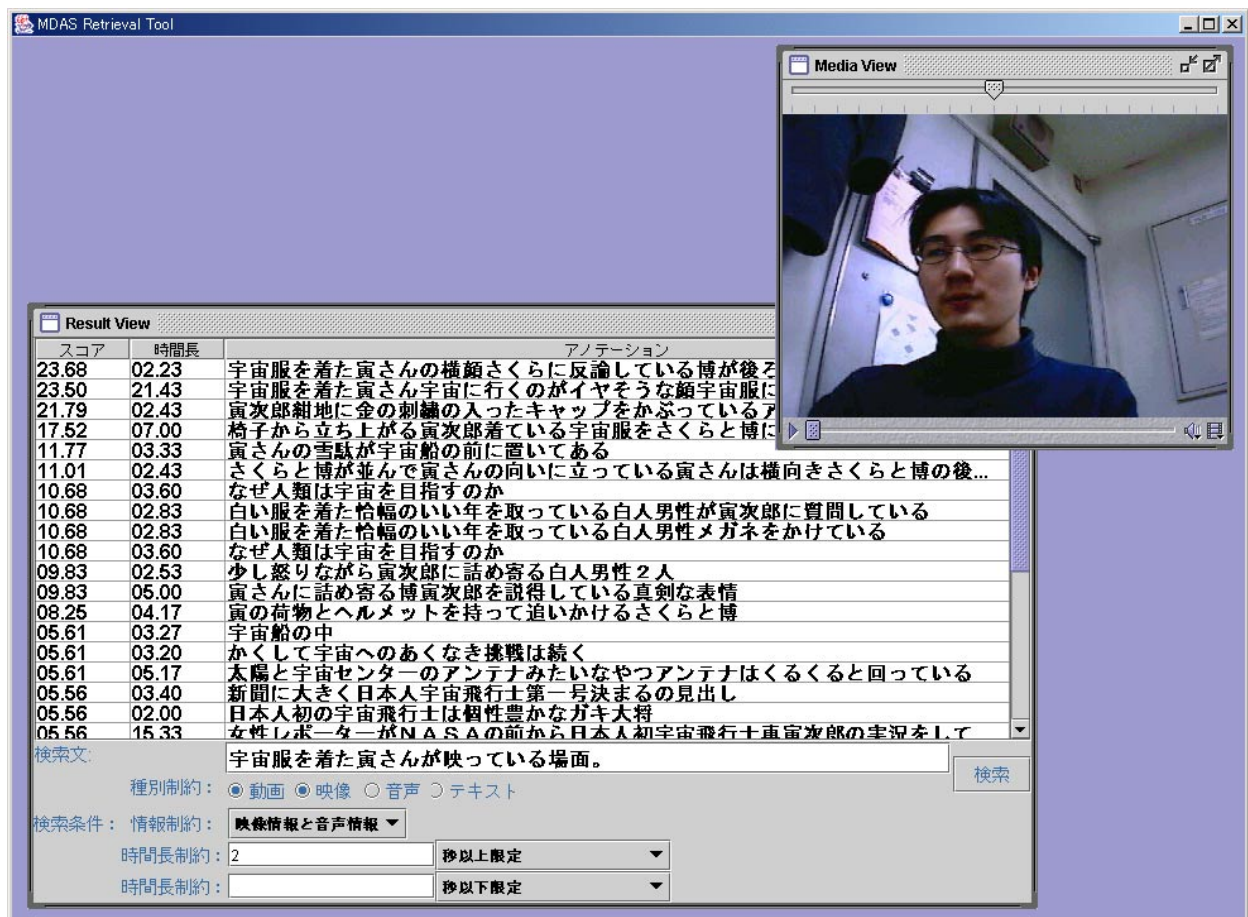


図 7: MDAS Retrieval Tool のスクリーンショット

4.3.1 Annotation Database

生成されたアノテーションデータは、一括してデータベースに格納される。データベースは PostgreSQL と XMLPGSQL (PostgreSQL を XML-DB として利用するためのプログラム) [10] による XML-DB を使用した。またタグ構造によるデータの抽出には、XML 文書に含まれるタグとその値を指定して、検索できるメソッドを用意することが必要である。そこで SQL の知識が無くとも XMLPGSQL にアクセス可能な Java ライブラリを実装した。

4.3.2 GDA Generation モジュール

GDA Generation モジュールは、MAML ファイルのアノテーションテキスト及び検索文章に GDA タグを付与する機能を有する。機械処理によるタグ付けが比較的容易な、文章の形態素情報、構文情報、単語の読み仮名、動詞の基本形及び助詞の種別に関してタグ付けする。形態素解析ツール Chasen[11]・構文解析ツール Cabocha[12] を用いて機械的処理を施し、単語の品詞や、文節間の大雑把な係り受けについて出力された情報から GDA タグ構造への自動変換を行っている。ちなみに、先に図 3 で示したデータの GDA タグは、このモジュールによって付与されたものである。

4.3.3 各種処理モジュール

処理目的に応じて、検索 (Retrieval) モジュール、要約 (Summary) モジュール、対話 (Dialogue) モジュールなどをサーバに追加していくだけで、多様なユーザの要求を実現できる仕組みになっている。本プロジェクトでは検索モジュールを実装した。また MAML ファイルのデータベー

ス登録処理を行う登録 (Registration) モジュールが実装されている。

4.4 データ登録・検索処理の手順

データの登録処理は、図 5 の点線の矢印で示す手順で行う。

- 1) MDAS Annotation Tool で対象メディアファイルを読み込み、アノテーションを付与する。
- 2) 生成した MAML ファイルをサーバに送信する。
- 3) MAML ファイル内のすべてのアノテーションテキストに GDA タグを付与する。
- 4) Annotation Database に登録する。

一方、データ検索・閲覧処理は図 5 の実線の矢印で示す手順で行う。

- 1) MDAS Retrieval Tool から検索要求を送信する。
- 2) 検索文章に GDA タグ付けする。
- 3) GDA 化された検索文章と、Annotation Database に格納されているアノテーションテキスト間で、XML 構造のマッチングをベースにした検索処理を施す。
- 4) 検索要求を満たすエレメント集合の情報を MDAS Retrieval Tool に送信する。
- 5) 閲覧要求に対して対象メディアファイルを取得し、エレメントに対応した時間範囲部分を再生する。

5 検索処理

本章では検索候補の選定手順について解説する。

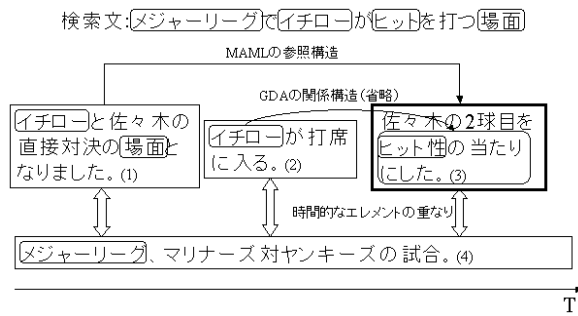


図 10: 検索候補の選定

5.1 処理の流れ

はじめに個々のエレメントに対して、GDA タグ付けされた検索文とアノテーションテキスト間の XML 構造を比較し、格解析、構文解析によるテキストの類似性を判定し、検索スコアを算出する。表層的なテキスト解析のみによって算出されたスコアを、検索文との意味的適合の度合いの指標とするのはかなり難しい。そこで本手法では、エレメント間の時間的な重なり、MAML の参照構造、GDA の関係構造を解析し、検索スコアの補正を行い検索精度の向上を試みる。一連の流れを例を挙げて説明する。図 10 は、検索対象として(1)～(4)のエレメントが存在し、“メジャーリーグでイチローがヒットを打つ場面”という検索文を入力した例である。すべてのアノテーションテキストが部分的に検索文と一致しており、順位付けは困難である。しかしながら、エレメント(1),(2),(3)は記述が無くても、エレメント(4)との重なりから“メジャーリーグ”に関連性があると考えられ、またエレメント(3)は GDA の参照構造から“イチロー”という単語が省略されていることも分かる。さらにエレメント(3)は、リンクによって参照されているので重要性が高いと予想される。つまりこの例では、エレメント(3)が最も適すると判定する。このように言語要素以外の関連性もスコアに反映させる。検索スコアが一定値以上のエレメントを検索結果として、MDAS Retrieval Tool に送信する。

6 考察

自然文章の入力を許容し、GDA のタグ情報に基づいた格解析、構文解析を行うだけでも、キーワードマッチング処理に比べ平均精度が大幅に向上することは実験により確認している。現状の課題として、検索文章とアノテーション間の単語マッチングによって初期の検索候補エレメントの選定をしているので、アノテーションテキストに構文的に近い記述を検索文書として入力しなければならない点が挙げられる。また、キーワード、助詞、動詞との完全一致を基本とする簡易的な処理によるスコア算出しが行っておらず、より厳密な言語解析が今後必要であると考えている。例えば“日本一高い山の景色”から“富士山の風景”という結果は抽出されない。“景色”と“風景”が意味的に近いこと、及び“日本一高い山”と“富士山”が同一の記述に相当することをシソーラス辞書を利用し、単純な単語、文節の置換処理を施す程度での質問文章の拡張は組み込んでいく必要があると考えられる。ただし、適合率の低下や検索処理時間の増加が予想され、今後の検討課題となっている。検索文章の意味解析による検索結果の出力は、自然言語処理技術の中でも未だ未開の分野であり、計算機処理による手法が確立されるまでは難しいであろう。

7 まとめ

本プロジェクトでは、アノテーションに基づく動画検索システムの設計及び実装手法を提案した。MAML, GDA は検索に限らず様々な利用方法が想定される。MAML, GDA による高度マルチメディア利用環境の実現を目指して、システムの拡張と言語解析手法の検討を中心に進めていきたいと考えている。

参考文献

- [1] Google イメージ検索 ホームページ, <http://images.google.co.jp/>
- [2] The MPEG Homepage, <http://mpeg.telecomitalialab.com/>
- [3] Berners-Lee, James Hendler and Ora Lassila: The Semantic Web, Scientific American May 2001, (2001).
- [4] TREC Video Retrieval Evaluation Homepage <http://www-nlpir.nist.gov/projects/trecvid/>
- [5] 伊藤一成, 斎藤博昭: メディアデータに対するアノテーション記述言語 (MAML) の策定とその応用, 情報処理学会研究報告, FI70-4, pp. 19-26, (2003).
- [6] Extensible Markup Language(XML) Homepage, <http://www.w3.org/XML/>
- [7] 橋田浩一: GDA 意味的修飾に基づく多用途の知的コンテンツ, 人工知能学会論文誌, Vol. 13, No. 4, pp. 528-535, (1998).
- [8] The GDA Tag Set Homepage, <http://www.i-content.org/gda/>
- [9] The Extensible HyperText Markup Language(XHTML) Homepage, <http://www.w3.org/TR/xhtml1/>
- [10] 小松 誠: RDB と ODB を融合する XML-DB フレームワーク, 第 43 回情報処理学会プログラミングシンポジウム論文集, (2002).
- [11] 日本語構文解析ソフト 茶筌 ホームページ, <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>
- [12] 工藤拓, 松本祐治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842, (2002).