

アプリケーションの操作と入力に基づく 副次アノテーションとその応用

伊藤 一成[†] 齋藤 博昭[†]

[†] 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,hxs}@nak.ics.keio.ac.jp

あらまし 近年、メディアデータを効率よく検索したり要約する手段として、メタデータ技術が注目を集めている。我々は、メディアデータのための汎用アノテーション記述言語 MAML (Multimedia Annotation Markup Language) を提案している。MAML は、“如何なるデータ、事象、物体、人物や行為に対して同一の仕様でアノテーションできる”ことを目標に掲げている。アノテーションはほとんどの場合、通常の映像・音声・内容転記など、付与自体を目的とする明示的作業によって生成されるが、何らかの別の作業から、副次的に生成されるアノテーションデータというのも非常に重要であると考えられる。そこで、本稿では MAML の適用事例として、ユーザのアプリケーション利用に係る一般的な操作や入力をすべてアノテーション行為と考え、アノテーションデータを副次生成するプロセスをメールとブラウザを例に解説する。
キーワード アノテーション、副次生成、MAML

A Secondary Annotation Based on Operations and Inputs of an Application.

Kazunari ITO[†] and Hiroaki SAITO[†]

[†] Department of Science and Technology, Keio University
Hiyoshi 3-14-1, Kouhoku-ku, Yokohama, Kanagawa, 223-8522 Japan

E-mail: {k_ito,hxs}@nak.ics.keio.ac.jp

Abstract A meta data technology calls our attention as means of retrieving or summarizing multimedia information efficiently. We have proposed MAML (Multimedia Annotation Markup Language), a generalized annotation description format for media data. MAML sets a goal that anyone can annotate any data, events, objects, person and acts by a unite specification. In almost all cases, an annotator does the job explicitly. But annotation which is generated secondarily by a primery work is also important. In this paper, we illustrate the process of generating data in MAML secondary, where general operations and keyboard inputs to an application are regarded as annotation acts. We have implemented a mailer and a browser to exemplify our approach.

Key words Annotation, Secondary Generation, MAML

1. はじめに

メタデータ技術が注目を集めている。これは、コンテンツの内容に関する特徴を予め記述しておき、その記述

データを直接の処理対象とすることで処理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と呼ばれる。メタデータに類似する表現にアノテーションデータがあげられる。アノテーションデータとは、メタ

データと同様にコンテンツに関する内容を別に記述したデータのことをいう。人間が生成、理解及び活用することを主目的とするものであり、計算機による意味理解を視野に入れたデータ構造にする必要はない。そういう意味で、計算機の意味理解を一義として付与するメタデータとは性質を異にする。

当初は、動画などの非テキストデータのコンテンツ記述を目的としたものが主であった。MPEG-7[1]はその代表例である。近年では、HTMLをはじめとするテキストコンテンツに対するアノテーションに関する研究も盛んになってきた。W3CのAnnotea Projectにて、開発が進められているWebアノテーションシステムAnnotea[2]は、任意のWebページに対して不特定多数のユーザがアノテーションを付与でき、またそれを、Amayaと呼ばれるブラウザで閲覧することもできる。

また、URIによって書籍のISBNコードなど、インターネット上に存在しないリソースを指し示すことが可能になり、またRFIDの登場により人やオブジェクトを一意に認識し、またさらにその行動や関係までコンピュータで容易に追跡、管理出来るようになってきた[3]。実世界の人物に関するメタデータを記述し、そのつながりを公開、共有するためのプロジェクトはその一例である[4]。このように、アノテーションを付与する対象はファイルに限らず、想定しなければならない領域は広がるばかりである。そこで我々は、“如何なるデータ、事象、物体、人物や行為に対して同一の仕様でアノテーションできる”ことを目標に掲げた汎用アノテーション記述仕様MAMLを提唱している。ファイルをアノテーションの対象とした既存の研究では、アノテータの明示的作業によりアノテートされているか、逆に計算機処理によって完全に自動付与されているという前提であったが、今後は、人の行為に基づいて副次的に生成されるアノテーションデータというものが重要となってくると考えられる。

2. 汎用アノテーション記述仕様 MAML

本章では、我々が提案する汎用アノテーション記述仕様MAMLについて概説する。

MAMLではデータ形式やファイルフォーマットによるファイルの区分は行わず、すべてのファイルは音声情報(audio)、映像情報(visual)、内容情報(contents)を有するメディアとみなす。ここでいう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やテキストデータに対して人の知識、主観、推論によって導かれる情報という。MAMLでは、メディアの分類は、この3情報の中

でどの情報をアノテーション可能であるかにより決定する。記述可能な情報と、定義されるメディアの種類の対応を表1に示す。

また、人物、書籍や出来事などのオブジェクトやイベントも仮想ファイルと考え、アノテーションの対象とする。メディアに付加されるアノテーションデータは基本単位(エレメントと呼ぶ)の集合体として構成する。

MAMLはXML[5]で表現する。最上位にmamlタグ、その下層にmediaタグを、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。図1にMAMLの記述例を示す。

図1は動画ファイルと暑中見舞いはがきを対象メディアとしている。この例では、動画ファイルのメディア時刻32.56秒から38.11秒にかけて、赤い帽子をかぶった女性が映っている。また、このシーンに対してk_ito@nak.ics.keio.ac.jpがデータを付与している。さらに、そのアノテーションデータに対して、kato@nak.ics.keio.ac.jpが書いた暑中見舞いはがきの文面を引用する形でアノテートされている。

2.1 メディア情報の記述

メディアに関する情報は、mediaタグの属性として記述する。表2に属性の種類を示す。type属性値にはメ

表2 mediaタグの属性
Table 2 Attributes of media tag.

属性名	説明
type	メディアの種類
maml-location	MAMLファイルの存在場所をURIで記述
media-location	対象ファイルの存在場所をURIで記述
media-word	対象ファイル特定する単語(文節)を記述
creator	対象ファイルの作成者を記述
annotator	アノテータを記述
contents-basetime	コンテンツ時間の基準値
annotation-basetime	アノテーション時間の基準値
duration	メディア時間長

ディアの種類別、maml-location属性値には、MAMLファイルのURIを記述する。対象の特定方法としては、URIで指定できるものについてはmedia-location属性値でそれを記述する。対象を特定づける単語あるいは文章の形式でmedia-word属性値として表わすことも可能である。

2.2 エレメントの記述

エレメントは、最上層にelementタグを記述する。表3に属性の種類を示す。各エレメントを識別するための

表3 elementタグの属性

Table 3 Attributes of element tag.

属性名	説明
id	エレメントを特定するための識別子
begin	メディア時間におけるエレメントが対象とする開始時刻
end	メディア時間におけるエレメントが対象とする終了時刻
contents	コンテンツが表現している実世界上の時刻
annotation	エレメントが生成された実世界上の時刻
target	アノテーション対象の指定

表 1 メディアが有する情報と名称の定義

Table 1 Classification of media.

メディア時間	包含する情報			メディアの種類	記述可能な情報			実ファイルの例	仮想ファイルの例
	音声	映像	テキスト		音声情報	映像情報	内容情報		
	x		/x	movie	x			動画ファイル	ビデオテープ中の動画
		x	/x	video				映像ファイル	ビデオテープ中の映像
x	x		/x	audio	x	x		音声ファイル	カセットテープ中の音声, 人物の声
x		x	/x	image				画像ファイル	写真, カセットテープの形状, 人物の外見
x				text	x	x		テキストファイル	文書, 書籍
x	x	x	x	null	x	x		圧縮ファイル	人物の性格, イベント

```

<?xml version="1.0" encoding="UTF-8"?>
<maml>
  <media type="movie"
    maml-location="http://www.xxx.jp/a.maml"
    media-location="http://www.xxx.jp/a.mpg"
    contents-basetime="2003-05-07"
    duration="01:42.56">
    <element id="1">
      :
    </element>
    :
    <element id="10" begin="32.56" end="38.11"
      contents="2003-01-01/2003-01-02">
    <visual>
      <character>
        赤い帽子をかぶった女性。
      </character>
    </visual>
    </element>
    <element id="11" target="10"
      annotation="2003-05-06T05:30"
      annotator="k_ito@nak.ics.keio.ac.jp">
    <contents>
      <supplementation>
        このシーンに映っている女優と結婚しました。
      </supplementation>
    </contents>
    </element>
  </media>
  <media type="text"
    maml-location="http://www.xxx.jp/a.maml"
    media-word="加藤君から来た暑中見舞いはがき"
    creator="kato@nak.ics.keio.ac.jp"
    annotation-basetime="2003-07-18">
    <element id="12" target="11">
    <contents>
      <copy>
        例の映画見ました。
        ご結婚されたんですね。おめでとうございます。
      </copy>
    </contents>
    </element>
  </media>
  :
</maml>

```

図 1 MAML の記述例

Fig. 1 An example of MAML.

id 属性, さらに対象メディアがメディア時間を有する場合, エレメントの対象開始時刻を begin 属性で, 対象終了時刻を end 属性で記述する。また, target 属性値によって図 2 に示す 4 通りのアノテーションを定義できる。第一に, target 属性がない場合は, メディアに対する直接的なアノテーションとみなされる (図 2 の (1))。第二に, target 属性値が id 番号のみの場合は, 同一 MAML ファイル中の特定のエレメントに対するアノテーションとみなされる (図 2 の (2))。第三に, target 属性値が他のメディアの URI の場合は, 対象メディアに関する何らかの記述が, 同時に target 属性値で指定されるメディアへのアノテーションになっていることを示す (図 2 の (3))。最後に, target 属性値が URI+"#" + id 番号の書式になっている場合は, アノテーションの対象が, 他のメディアを対象とする MAML ファイルの特定のエレメントであることを意味する (図 2 の (4))。

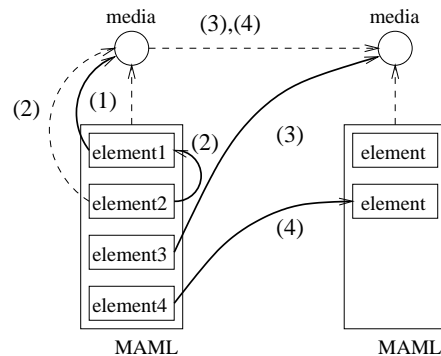


図 2 target 属性によるエレメントの意味付け

Fig. 2 Definition of an element by a target attribute.

2.3 エレメント内容の記述

エレメントタグの下層に音声, 映像, 内容情報を区別するための種別タグ, さらにその下層にそれぞれの種別毎にアノテーションの内容をさらに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 4 に示す。クラスに基づく具体的な内容は, クラスタグの下層に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)。

2.3.1 音声情報の記述

人が発生した言葉をそのまま記述する“発話”クラス,

表 4 情報の種別

Table 4 Classification of information.

種別	クラス
音声情報 (audio)	発話 (utterance)
	擬音 (imitation sound)
	音楽 (music)
映像情報 (visual)	物体 (object)
	人物 (character)
	文字 (letter)
	背景 (background)
	場所 (place)
内容情報 (contents)	標題 (subject)
	説明 (explanation)
	属性 (attribute)
	補足 (supplementation)
	要約 (summary)
	複製 (copy)

その他の音情報で言葉に直接変換できる場合、それを記述する“擬音”クラス、言葉に直接変換出来ない音情報はすべて“音楽”クラスとしてその音情報に関する文章表現を記述する。

2.3.2 映像情報の記述

移動可能なオブジェクトに関しては人とそれ以外に分類し、人に関する情報を“人物”クラスで、それ以外を“物体”クラスで記述する。移動不可能なオブジェクトまたはその集合を“背景”クラスで定義する。また映像中の文字情報をテキストで抽出する“文字”クラス、背景の地理情報を記述するための“場所”クラスを設定している。

2.3.3 内容情報の記述

コンテンツのタイトルを記述する“標題”クラス、コンテンツの内容を自由に記述する“説明”クラス、対象ファイルの属性に関する情報を記述する“属性”クラスがある。テキストメディアに含まれるテキストの一部をMAMLファイルに転記する場合、それを“複製”クラスの要素で表現する。他のクラスの要素内で記述されたテキストの要約表現を“要約”クラスで表現する。対象メディアに関する直接の要約表現は、先に解説した“説明”クラスで表現する。また、個々のユーザがデータを管理する上で必要な補足情報や、一時的なコメントなどは“補足”クラスで表現する。

3. アノテーションデータの副次生成

通常の映像・音声転記や内容記述は、アノテーション付与自体を目的とする明示的作業によって生成される。よってそれに係る時間コストがアノテーション生成及び流通の大きな障害となっている。アノテーション関連技術の研究自体は盛んであるが、実際に稼働されているシステムや製品がなかなか出現しない理由もそこにある。ところで、アノテーションの概念を用いた大規模システムの成

功事例に Google のイメージ検索 [6] が挙げられる。画像がエンベッドされている HTML テキストをアノテーションデータと考えることにより、キーワードによるイメージ検索を実現している。ユーザが自らの意志で作成したホームページのテキスト自体がアノテーションデータになっている。つまり成功の背景には、自然発生的にデータが生成されている点が挙げられる。MAML は、このように本来の作業から副次的に生成されるアノテーションデータも対象領域としている。本稿ではその一例として、“アプリケーション利用に関するすべての操作及び文章入力をアノテーション作業とみなす”ことによってアノテーションデータを生成するプロセスをメーラ、ブラウザを例に提示する。

3.1 メーラ利用からの MAML 生成

はじめに、メーラを使う際の操作、入力について列挙してみる。

- (1) 受信ボタンを押して新着メールを受信する。
- (2) アドレス帳にユーザを登録する。
- (3) メールへのヘッダ一覧から本文を参照したいメールを選択する。
- (4) メールを新規に作成する。
- (5) メールに返信する。
- (6) メールを転送する。

これら一連の操作をすべてアノテーション行為とみなし、MAML データを自動生成する機能を有するメーラ Annotation Mailer を実装した。スクリーンショットを図 3 に示す。見た目は普通のメーラであり、当然メーラとしての機能は実装されている。

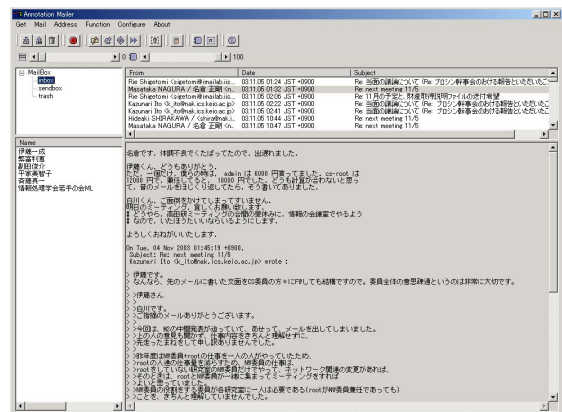


図 3 Annotation Mailer のスクリーンショット

Fig.3 A screenshot of Annotation Mailer.

Annotation Mailer では、例えば“メールのヘッダ一覧から本文を参照したいメールを選択する”という操作は事象“本文を読んだメール”に対するアノテーション

斉藤研の伊藤です。

>高田研の名倉と申します .
>
>> 斉藤（博）研の伊藤といひます。
>
>伊藤さんと私で、昔ネットワーク管理者してました .
>
>> 以下の形式でもいいと思うけど、僕らのころは、情報
>> 工学科の会議室にノート PC を置いて、各自で PPT の
>> コピー及び動作検証をしていました。
>
>という感じでした . 特に昔は、図形の横線が出なかつ
>たりとか、本人の意図してない色で出てしまったりと
>か言うトラブルがあったので、本人が確認してました .

これ以外にも PPT のバージョンの違いで意図しない表示
になっていたりました。

図 4 メール本文の例

Fig. 4 An example of an e-mail text.

と考える . また , “ 新規にメールを作成する ” という行為
はいわば , メールタイトルに記述した内容に関するアノ
テーションデータをメール本文として作成していると考え
える . さらに , “ 受信メールに返信する ” ということは ,
受信メールのメールタイトル又は , 本文 (の一部) に対
して , アノテーションしていると考ええる . 例として , 図 4
に示す返信メールから図 5 に示す MAML が Annotation
Mailer 内部で自動生成される .

3.2 ブラウザ利用による MAML 生成

次に , ブラウザを使う際の操作 , 入力について列挙し
てみる .

- (1) アドレスバーに URL を入力する .
- (2) お気に入りリストに登録する , お気に入りから
ページを参照する .
- (3) リンクを参照する .
- (4) ページを見る .
- (5) 検索バーに検索単語を入力する .

これら一連の操作も同様にすべてアノテーション行為と
みなし , MAML データを自動生成する機能を有するブラ
ウザ Annotation Browser を実装した . スクリーンショッ
トを図 6 に示す . 見た目は普通のブラウザであり , 当然
ブラウザとしての機能は実装されている . Annotation
Browser では , 有用なホームページをお気に入りリスト
に登録する操作は , 事象 “ お気に入りのホームページ ”
に対してアノテーションするとみなす . これにより図 7
のような MAML が自動生成される . また図 8 に示すよ

```
<?xml version="1.0" encoding="UTF-8"?>
<maml>
  <media type="text"
    maml-location=
      "file://c:\data\mail\5b7d729.maml"
    media-location=
      "file://c:\data\mail\5b7d531.mail"
    media-word="11/5 meeting"
    annotation-basetime="2003-11-09T14:15">
    <element id="1"
      annotator="k_ito@nak.ics.keio.ac.jp">
      <contents>
        <comment>斉藤研の伊藤です.</comment>
      </element>
    <element id="2">
      <contents>
        <copy>高田研の名倉と申します.</copy>
      </contents>
    </element>
    <element id="3">
      <contents>
        <copy>斉藤（博）研の伊藤といひます.</copy>
      </contents>
    </element>
    <element id="4" target="3">
      <contents>
        <copy>伊藤さんと私で , ...</copy>
      </contents>
    </element>
    <element id="5">
      <contents>
        <copy>以下の形式でも...</copy>
      </contents>
    </element>
    <element id="6" target="5">
      <contents>
        <copy>という感じでした . 特に...</copy>
      </contents>
    </element>
    <element id="7"
      annotator="k_ito@nak.ics.keio.ac.jp"
      target="6">
      <contents>
        <comment>これ以外にも PPT の...</comment>
      </contents>
    </element>
  </media>
</maml>
```

図 5 電子メール本文の MAML 記述例 (本文は一部省略)

Fig. 5 An example of MAML description of the e-mail text.

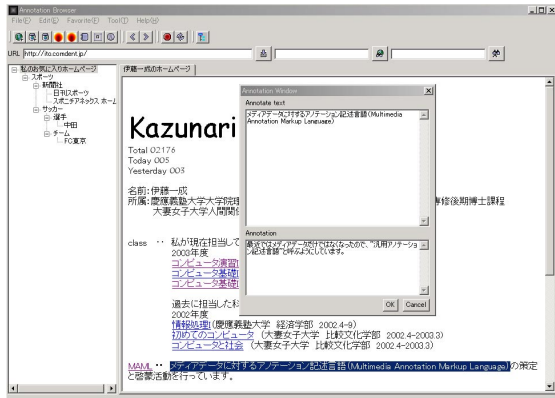


図 6 Annotation Browser のスクリーンショット
Fig. 6 A screenshot of Annotation Browser.

うな MAML が、通常のナビゲーションの過程から Annotation Browser 内部で順次自動生成される。図 8 において id 番号 1 のエレメントは、直接 URL をキーボード入力してホームページをアクセスした例である。その場合、アノテーションテキストはアクセス先の HTML ソースの TITLE タグ内を抜き出す。id 番号 2 のエレメントは、リンクを参照することで別のホームページにアクセスした例である。参照元のアンカー文を参照先のホームページのアノテーションテキストとする。つまりホームページ製作者が、参照先ホームページに対してアンカー文の形式でアノテーションしているとみなすわけである。id 番号 3 のエレメントは、お気に入りリストから選択してアクセスした場合である。アノテーションテキストは、以前お気に入りリストに登録する際にユーザが入力した(アノテートした)タイトルである。id 番号 4 のエレメントは検索エンジンに問い合わせを行った際に生成されたものである。この場合、アノテーションテキストは検索キーワードとした。

4. アノテーションデータの構造化と処理プロセス

MAML ファイルに含まれるエレメントは、メディアに対する直接のアノテーションと、他のエレメントに対するアノテーションとに大別される。あるエレメントを子とし、そのアノテーションの対象を親とする関係を定義できる。あるいは、エレメント間に時間的な包括関係が成り立つ場合が存在し、それも親子関係と見なす。すると、エレメント全体はメディアをルートとするツリー構造とみなすことができる(図 9 の矢印)。また、各エレメントのアノテーションテキスト間で自然言語処理に基づいて関連性を定義する(図 9 の点線)。MAML では、タグ構造

```
<?xml version="1.0" encoding="UTF-8"?>
<maml>
  <media type="null"
    maml-location=
      "file:///c:/data/browser/favorite.maml"
    media-word="お気に入りのホームページ">
    <element id="1"
      annotation="2003-11-12T09:35"
      target="http://www.asahi.com/">
      <contents>
        <explanation>アサヒ・コム
      </explanation>
      </contents>
    </element>
    <element id="2"
      annotation="2003-11-17T20:30"
      target="http://ito.comdent.jp/">
      <contents>
        <explanation>伊藤一成のホームページ
      </explanation>
      </contents>
    </element>
  </media>
</maml>
```

図 7 お気に入りリストの MAML 記述の例
Fig. 7 A description of a favorite list written by MAML.

による機械の意味理解は想定しない記述方式となっている。よって表層的な自然言語処理技術を駆使したデータ構造化により、検索、大域要約、条件付要約、関連情報抽出、エレメント群の自動クラスタリング等の処理プロセスを提供する。我々はすでに、この概念を基に、MAML データを用いたメディア検索システムおよび要約システムを構築している [7] [8]。システムでは、条件を基に図 9 のグラフ構造から、個々のエレメントのスコアを算出し、検索結果の提示や要約表現の生成を行っている。

5. 応用事例

前章で解説した概念を基に、MAML 汎用処理モジュールを実装し、それを用いた様々な応用事例を提案している。本章ではこれらを簡単に説明するとともに MAML の利点について述べる。

5.1 複数メール要約機能

Annotation Mailer は複数のメールを一つに融合し、かつ任意の要約率に圧縮できる機能を有する。複数メールを要約した場合のスクリーンショットを図 10 に示す。Annotation Mailer では、すべての受信メール及び送信メールに対して図 5 のような MAML 形式でデータが保

```

<?xml version="1.0" encoding="UTF-8"?>
<maml>
<media type="null"
maml-location=
"file://c:\data\browser\profile.maml"
media-word="ブラウジング履歴">
<element id="1"
annotation="2003-11-23T03:45"
target="http://www.asahi.com/">
<contents>
<subject>アサヒ・コム</subject>
</contents>
</element>
<element id="2"
annotation="2003-11-23T03:46"
target="http://www.asahi.com/international/
/update/1123/003.html">
<contents>
<explanation>次回6者協議「安全保証」の文書
化見送り 日米韓中口</explanation>
</contents>
</element>
<element id="3"
annotation="2003-11-23T03:46"
annotator="k_ito@nak.ics.keio.ac.jp"
target="http://ito.comdent.jp/">
<contents>
<explanation>伊藤一成のホームページ
</explanation>
</contents>
</element>
</media>
<media type="text"
maml-location=
"file://c:\data\browser\profile.maml"
media-word="入力した検索キーワード">
<element id="4"
annotation="2003-11-23T03:48"
target="http://www.google.co.jp/">
<contents>
<copy>京都 研究所</copy>
</contents>
</element>
</media>
</maml>

```

図 8 ブラウジング履歴からの MAML 生成の例

Fig.8 An example of MAML generated from a browsing history.

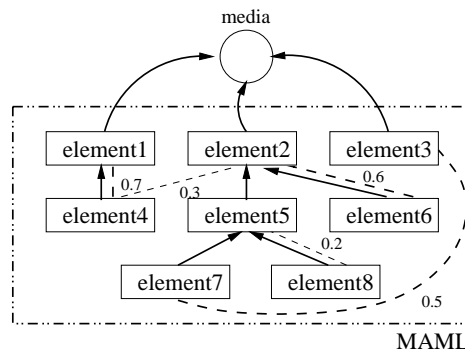


図 9 エレメント間の関連性の定義

Fig.9 Definition of a relevance between elements.

存されている。初めに、ユーザが指定したメールのサブジェクトと同一のメール群それぞれを対象にした MAML ファイル群を結合する。生成された単一 MAML ファイルに対し、同様に図 9 で示したグラフ構造化を施す。図 10 の左上のスライダーバーから自由に要約率を設定でき、重要度の低いエレメントから順に要約率を満たすまで削除し、残ったエレメント集合を基にメール本文を再合成する。メール本文を対象とする MAML の親子関係は、親が本文の一部であり、子がそれに対する返答文に相当する。よってエレメントを削除する際、その子エレメント群を再帰的に削除する。

自然言語処理分野における重要文抽出手法のみで、電子メールのような対話形式のテキストデータの要約を行うと、その対話構造が保持されない。しかしながら、MAML では文書データに対するアノテーションを考えた場合、どの文がどの文に対する返答文か、つまりアノテーションかという構造が記述されているので、どのような要約率を設定しても、対話構造を保持した要約文書を生成できる。

5.2 ユーザプロファイルに基づく Web 検索機能

Annotation Browser は、ユーザ適応型 Web ページ検索機能を有する。検索を実行した場合のスクリーンショットを図 11 に示す。Annotation Mailer において新規作成及び返信したメールの MAML 群 (図 5 参照)、Annotation Browser のお気に入りリスト (図 7 参照)、Annotation Browser のブラウジング履歴の MAML (図 8 参照) をユーザのプロファイルとみなしグラフ構造化する (以後グラフ A と呼ぶ)。一方ユーザから入力された検索キーワードから Google API [9] を利用して、検索一覧を入手する。Google API から入手した個々の Web ページのタイトルと要約表現をそれぞれエレメントとし、検索キーワードを media-word 属性値とする MAML ファイルを生成し同様にグラフ構造化する (以後グラフ B と

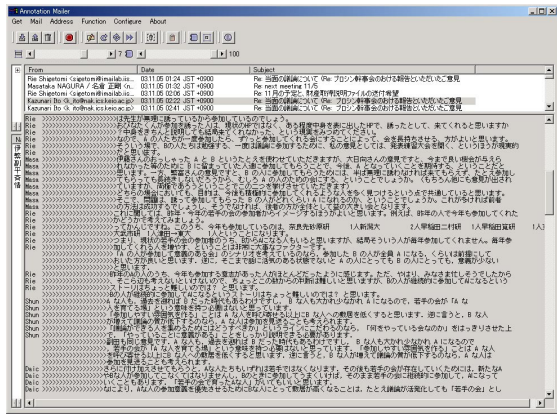


図 10 複数メールの要約提示機能使用時のスクリーンショット
Fig.10 A screenshot when summarized mails are displayed.

呼ぶ)。グラフ A をクエリーとしてグラフ B に対して検索処理を行うことで、グラフ B 中の個々のエレメントに対してスコアを算出する。これにより Google の検索一覧から自分に関連性の高い Web ページの情報を選択抽出し、Google の検索結果表示画面と同様なフォーマットで提示する。

このように、応用システムを構築する上でデータ形式に MAML を採用することで、対象メディアや生成形態が異なる様々なデータを容易に流用および併用することが可能である。さらに一元的に処理可能である点が最大の利点であり特徴であると考えられる。

本章で概説した応用事例の数々は、すべて同一の MAML 汎用処理モジュールを用いて実現されている。MAML はそのデータ形式から、既存の自然言語処理分野で培われてきた検索、要約、言い換え、翻訳等の様々な技法をほぼそのまま取り入れて活用できるという利点がある。自然言語の最たる魅力はその表現の無限なる豊かさであり、それを保持したままでのデータの処理や提示が可能である。

6. ま と め

本稿では、我々が提案している、汎用アノテーション記述仕様 MAML を用いて、日常のアプリケーション利用に係る自然な操作や入力をアノテーション行為と考え、副次的にアノテーションデータを生成するプロセスについて例示した。またそのデータを用いた応用事例について解説した。MAML は本稿で解説した以外にも様々な適用事例が想定される。

謝 辞

本成果は、平成 15 年度 IPA 未踏ソフトウェア創造事

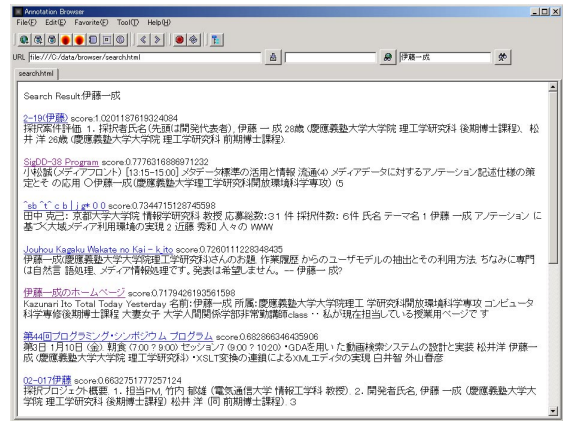


図 11 ユーザプロフィールに基づく Web 検索機能使用時のスクリーンショット
Fig.11 A screenshot when the user-oriented web search is executed.

業の一部である。IPA（情報処理推進機構）及びプロジェクトマネージャーの京都大学 田中克己教授に深く感謝いたします。

文 献

- [1] The MPEG Homepage, <http://mpeg.telecomitalia.com/>
- [2] Kahan, J., et al.: Annotea: An open RDF infrastructure for shared Web annotations. Proc. , the 10th International World Wide Web Conference (WWW-10), (2001).
- [3] 椎尾一朗, 早坂達: モノに情報を貼り付ける: RFID タグとその応用 情報処理学会誌, Vol. 40, No. 8, pp. 846-850, (1999)
- [4] The 'friend of a friend' project: FOAF Homepage, <http://rdfweb.org/foaf/>
- [5] Extensible Markup Language(XML) Homepage, <http://www.w3.org/XML/>
- [6] Google イメージ検索 ホームページ, <http://www.google.co.jp/imghp/>
- [7] 伊藤一成, 齋藤博昭: メタデータ解析に基づくメディア検索システム, 情報処理学会研究報告, DBS131-69, pp. 515-520, (2003).
- [8] 伊藤一成, 酒井康旭, 齋藤博昭: メタデータ解析と自然言語処理を併用した要約動画の生成, 情報処理学会研究報告, DBS132-6 (2004).
- [9] Google API Homepage, <http://www.google.com/apis/>