

汎用アノテーション記述言語 MAML の提案と その生成・処理プロセス

伊藤 一成[†] 齋藤 博昭[†]

近年、メディアデータを効率良く検索したり要約したりする手段として、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) を策定した。MAML では、アノテーションの対象はファイルだけではなく、人物、オブジェクト、イベント等までも含めて、統一的な仕様で記述可能な点が大きな特徴である。いうまでもなく、ファイルの種類やフォーマットも限定しない。さらに、アノテーションデータの生成者および利用者たる人間が、理解・記述しやすい文章中心の表現構造を採用している。ユーザの明示的な作業によるデータ付与は、その時間的コストが問題視されるが、アノテーション支援ツールの実装によって大幅に削減可能となった。また、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のアノテーション行為と考え、MAML データを生成するプロセスについてもメールとブラウザを例に解説する。さらに、MAML データを用いた応用例について例示する。MAML を適用することにより、メディアの種別に依存しない様々な応用アプリケーションの実現が期待できる。

MAML: A Generalized Annotation Description Language for Media Data

KAZUNARI ITO[†] and HIROAKI SAITO[†]

A meta data technology calls our attention as means of retrieving or summarizing multimedia information efficiently. This paper proposes MAML (Multimedia Annotation Markup Language), a generalized annotation description format for media data. The annotated subjects in MAML include not only a file but a person, an object, an event, etc., all of which can be annotated by a united specification which is independent of media type or format. MAML allows annotation in semi natural language style so that the user of the annotated data, as well as an annotator, can easily understand the data. Against the high cost of human's explicit annotation work, we implement a supporting tool to cut it down significantly. We also 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. A variety of applications will be realized by MAML.

1. はじめに

情報化社会の今、膨大な量のコンテンツの中から目的とするデータを探し出し、能動的にアクセスできる仕組みが必要となってきた。近年、これらマルチメディア情報を効率良く検索したり要約したりする手段として、メタデータ技術が注目を集めている。これは、コンテンツの意味・内容に関する特徴を別に記述しておき、その記述データを直接の計算機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、メタデータの中で計算機による意味理解を一義とせず、人間が生成・理解および活用することを目的としたデータをいう。

現在、テキスト、画像、音声、動画をはじめとした様々な形態のメディアデータが混在しており、またメディアごとに見ても、非常に多くのファイルフォーマットが存在する。これらを一括に利用・処理するには、メディアデータの種別に依存しない、統一的なアノテーション仕様が必須である。また多量のアノテーションデータが生成されていくためには、データ構造の理解が一般の人々でも容易な記述仕様でなければならない。この理念に基づいたアノテーション記述仕様を策定することには大きな意義があると考えられる。以下 2 章

[†] 慶應義塾大学大学院理工学研究科

Department of Science and Technology, Keio University

で研究の背景および目的, 3章で MAML の概要, 4章で MAML の記述仕様, 5章で MAML の特徴, 6章および 7章でその生成プロセス, 8章でその処理プロセスについて順次解説した後, 9章で応用事例, 10章で関連研究, 11章でまとめについて述べる。

2. 背景, 目的

本章では, 汎用アノテーション記述言語 MAML (Multimedia Annotation Markup Language) の提案に至る背景と目的について概説する。

近年新しく規格化される仕様の中には, メタデータを意識したものがいくつか見受けられる。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-7¹⁾ はその代表例である。MPEG-7 は, 度重なる仕様の拡張にともなって, テキストファイルに対するメタデータ付与や数種にわたるコンテンツを统一的に記述可能となっているが, 仕様が膨大でかつ複雑化していることが問題となっている。

他には, Web 上に点在する様々なデータに統一の意味情報を付与するのは不可能であるという前提から, 個々のユーザが独自に定義し記述したメタデータを付与し, 人間と機械とのコミュニケーションの実現を目指す Semantic Web²⁾ があげられる。Semantic Web では, RDF³⁾ によってリソースのプロパティやリソース間の関係を記述する。ホームページ自体の概要表現に関する諸項目を RDF の語彙定義によりメタデータとして簡潔に記述する RSS (RDF Site Summary)⁴⁾ も策定されている。しかしながら, 計算機による意味理解を想定する場合, 結局はセマンティクスが決められている必要があり, 多様なデータすべてに対して定義されているとは現状ではいい難い。

URI によって書籍の ISBN コード等, 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, 管理できるようになってきた⁴⁾。よって, メタデータ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ぬ領域は広がるばかりである。実世界の人物に関するメタデータを RDF によって定義し, 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⁵⁾。このような情勢の中, 様々な記述能力を有するメタデータ規格が次々と発表されているが, それらの中から目的や用途に応じて選択し, その仕様を理解して初めて生成のための作業が可能であり, その初期コストが作業者にとって大変な負担となっている。

一方, アノテーションデータは, その意味合いから人間が記述および理解しやすい表現構造を念頭におくのが望ましい。一般的に, 機械が解析しやすいデータは可読性に乏しく, 概して記述も難しい。その逆も然りで, 相反するものである。よって, タグ等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に主眼を置くのが妥当であろう。この理念に基づいて設計された記述フォーマットに Dublin Core⁶⁾ があげられる。Dublin Core では任意のファイルフォーマットに付与可能な情報として, 基本 15 項目を設定している。しかしながら, 項目は“作成者”, “日付”, “資源タイプ”をはじめとしたファイルの属性に関する記述と内容記述に限定されている。

MAML は, 上記の点を考慮した汎用アノテーション記述言語である。“いかなるデータ, 事象, 物体, 人物や行為に対しても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。

3. MAML の概要

ファイルは, 通常データ形式によって分類される。たとえば, 画像, 映像, 音声, 映像と音声を合わせたメディアである動画, テキスト, HTML, XML 等があげられる。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) を有しうるメディアと見なす。人物, 書籍や出来事等のオブジェクトやイベントもアノテーションの対象とし, これらも含めアノテーションの対象は, 総称してメディアと呼ぶ。また, 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をいう。MAML では, 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メディアを分類する。記述可能な情報と, 定義されるメディアの種類に対応を表 1 に示す。たとえば, 音声 (audio) メディアは, 音声情報と内容情報を, 画像 (image) メディアであればメディア時間情報を有しない映像情報と内容情報をアノテーション可能なメディアと定義する。ここでメディア時間とは, メディアの再生経過時間のことをいう。また表 1 中の null メディアとは, そのメディアからは音声, 映像, テキストいずれの情報も抽出できず, 結果として内容情報に関する記述のみが可能なメディアと定義する。この規則によりすべてのメディアは, 表 1 に示した 6 種類のメディア種類のいずれかに一意に分類される。それぞれのメディ

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	抽出可能なデータ			メディアの種類	記述可能な情報			ファイルの例	ファイル以外の例
	音声	映像	テキスト		音声情報	映像情報	内容情報		
	x		/x	movie				動画ファイル	ビデオテープ中の動画像
		x	/x	video	x			映像ファイル	ビデオテープ中の映像
			/x	audio		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	x			画像ファイル	写真、カセットテープの形状、人物の外見
x	x	x		text	x	x		テキストファイル	文庫、書籍
x	x	x	x	null	x	x		なし	人物の性格、イベント

アが付与可能な情報も表 1 に併記してある。どの情報をアノテーション付与対象と見なすかで、同一のメディアでもその種類が異なることもありうる。たとえば、書籍を考えた場合、書かれている文章だけに注目すればテキスト (text) メディアとなり、形状や外観にまで対象に含めれば画像 (image) メディアとなる。movie, video, audio, image メディアに関して、ファイル埋め込み型のテキストデータ (たとえば ID3 タグや MPEG-7) も存在しうるので表 1 の項「テキスト」は「 /x 」となっている。

また時間に関する情報は、アノテーションデータを処理するうえで重要な役割を果たす。MAML ではメディア時間に加え、コンテンツ時間、アノテーション時間の計 3 種類の時間情報を定義し、これを記述可能とする。コンテンツ時間情報とは、アノテーションデータが実世界上の時刻で何時のことを表現しているかを記述するものである。たとえば、複数ファイルから要約生成を行う場合、その時間的順序を加味する必要性があり、コンテンツ時間を基に整合性を保持する。一方、アノテーション時間情報とは、アノテーションの付与が実世界上の時刻で何時になされたかを記述したものである。あるメディアに対して、多人数が長期にわたって感想や意見等のアノテーションデータを追記していくようなケースも MAML は想定しており、個々の情報が生成された時間順序がデータを解析するうえで非常に重要になるからである。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 (以後エレメントと呼ぶ) の集合体として構成する。以上の概念に基づいて、複数のメディア情報を一元的に記述・処理可能な環境の構築を目的とする。

4. MAML の記述仕様

MAML は XML⁷⁾ で表現する。XML は、人間が読むことができるシンプルなタグを使って、データをマークアップするために使われる汎用的なシンタクスを定義した規格である。近年 RSS¹¹⁾ や FOAF⁵⁾ 等、RDF の語彙定義によるメタデータ仕様も散見されるが、そもそも RDF は、機械が意味理解可能とする手段を提供するためのフレームワークである。MAML

は、機械が意味を理解することはまったく念頭に入れておらず RDF の目的とは相反すると考えられる。そこで、前章で述べた MAML の概念に基づいて独自のタグセットを新たに策定する。

MAML では、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、動画メディアのメディア時刻 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おり、この女性が“あけましておめでとうございます”と発話している。また、このシーンに対して k_jito@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ートされている。

4.1 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。メディア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。メディアを特定づける単語あるいは文章の形式で media-word 属性値として表すことも可能である。コンテンツ時間の基準値を contents-basetime 属性値に、アノテーション時刻の基準値を annotation-basetime 属性値として記述できる。個別のエレメントがここで指定した時刻と異なる場合は、次節で解説するエレメント記述方式により、さらに時間指定をする。また、対象メディアがメディア時間を有する場合は duration 属性にその全長も併記しておく。

また media タグを列挙することで、複数のメディアのアノテーション情報を同一 MAML ファイル内に記述できる。

4.2 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

```

<?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">
      <audio>
        <utterance id="p1">
          あけましておめでとうございます .
        </utterance>
      </audio>
      <visual>
        <character id="p1">
          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>
          先日、MAML Editor/Viewer で例の映画見ました .
          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.

めの 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 1 に, target 属性がない場合は, media-location または media-word で指定されるメ

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	メディアの存在場所を URI で記述
media-word	メディアを特定する単語(文節)を記述
creator	メディアの作成者を記述
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

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

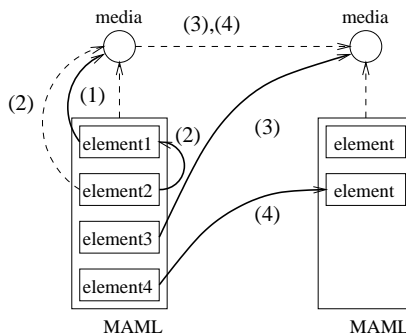


図 2 target 属性によるエレメントの意味付け
Fig. 2 Definition of an element by a target attribute.

ディアに対する直接的なアノテーションと見なされる(図 2(1)). 第 2 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見なされる. つまり, 対象メディアへの間接的なアノテーションを定義していることになる(図 2(2)). 第 3 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. つまり, メディアとメディア間の直接的な関係をアノテーションデータで定義することになる(図 2(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

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)

アを対象とする MAML ファイルの特定の要素であることを意味する。対象メディアに関する何らかの記述が、target 属性値で指定される MAML ファイルの対象メディアに関する記述へのアノテーションとして定義される。つまり、メディアとメディア間の間接的な関係をアノテーションによって定義することに相当する(図 2(4))。

4.3 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ごと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)。自然言語で主観に基づいて記述するので、当然専門的知識はまったく必要としない。また表現の自由度が飛躍的に高まる。単に時間情報やクラス分類の情報だけを設定したいのであれば、アノテーションテキストは空白でもよい。また、1つのエレメント内に複数のクラスの情報をまとめて記述することも可能である。

4.3.1 音声情報の記述

音声情報は、人が発生した言葉をそのまま記述する“発話”クラス、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、言葉に直接変換でき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。

4.3.2 映像情報の記述

映像情報は、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。

4.3.3 内容情報の記述

内容情報は Dublin Core の基本 15 項目にほぼ準拠する形となっている。コンテンツのタイトルを記述する“標題”クラス(Dublin Core の“Subject”に相当)、コンテンツの内容を自由に記述する“説明”クラス(Dublin Core の“Description”に相当)、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”クラスのエレメントで表現する。他のクラスのエレメント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。たとえば、クロードキャプションは発話文を要約して表示しているが、そのような用途を想定している。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、アノテーションデータは、コンテンツの製作段階で付与される映像・音声転記や、流通の段階で必要な属性の情報に限らず、ユーザがコンテンツを活用する段階で付与されるデータも含まれる。この段階のデータは、個々のユーザがデータを管理するうえで必要な補足情報や、一時的なコメント等があげられる。それらを“補足”クラスで表現する。

ちなみに Dublin Core の一部は内容情報のクラス以外でも定義できる。“Date”は element タグの annotation 属性で、“Coverage”は place クラスあるいは element タグの contents 属性で、“Identifier”は media タグの media-location か media-word 属性で、“Creator”は media タグの creator 属性で、記述可能である。このように、MAML では異なる情報の分類形態を用いているが、Dublin Core で定義される 15 項目すべてを記述可能である。

5. MAML の特徴

MAML は、一般の人々でも容易にアノテーションできることを主目的としているので、クラスの名称は、名前からその意味を直感的に把握でき、またクラスの分類数は最小限にとどめている。前章の説明や図 1 の例から、MAML のクラス体系がいかに単純かが理解可能であろう。また、各クラスタグが属性を持つことも極力避けた。現段階では、対象を特定したい場合の記述に id 属性を持つことだけを許容した。また、種別およびクラスタグに対してセマンティクスを定めることは想定していない。

分類規則を多層化・複雑化すれば、その決定に専門的知識と時間を要し、また選定ゆれも発生する。タグセットを限定することで情報の詳細が設定できない場合でも、自然文章中に何に関する情報であるかを含めることで、情報の種別を細分化できると考える。しかしながら、アノテーションとはメディアの内容に関する情報だけではなく、それに含まれる映像や音声の転記情報についても記述対象とするものであり、MAMLもそれを踏襲している。ここで注意しなければならないのは、自然文章で種別情報まですべて記述しようとする、逆に何についてのアノテーションが限定できなくなる場合が生じる。たとえば、“おはようございますという会話がある？”とアノテートした場合、音声発話を転記したもの(<utterance> おはようございます。</utterance>)なのか、映像中のキャプション(<letter> おはようございますという会話がある。</letter>)なのか、単なる内容説明(<explanation> おはようございますという会話がある。</explanation>)なのか分類できない。これを解決するためにはアノテーションテキストの表現構造に制約を与えなければならない。そのルールの定義は非常に困難であり、さらにアノテータへのルールを強いることによる負担は非常に大きい。MAMLで定義されているクラスは、あくまでこれを人間が区別できるようにタグとして明記するものである。

また、MAMLでは画像の一部領域やテキストの一部をタグや属性値によって指定することは想定していない。明示したければ、アノテーションテキスト中に自然言語で対象領域を表記するか、“複製”クラスを用いることとなる。MAMLで時間領域の情報について定義し、空間領域の情報を定義しないのは、アノテーションデータと対象メディアを同時にユーザに対して提示することを考えた場合、ユーザによってその情報の対象範囲を検出するための時間的コストに大きな差異があるからである。ある女性が映像に映っていたとして、メディア時間の指定がされていなければ、ユーザは全ビデオ区間を順再生して確認しなければならない。一方、女性が映っている画像が提示されたとして、その女性がどの画像領域のどの範囲に映っているかは、アノテーションテキストにその女性の特徴が書かれていれば、瞬時に把握できる。文書に関しても同様である。

MAMLの最大利点は、アノテータは対象とするメディアの種類や目的に応じて異なる仕様で記述しなければならないという煩わしさから解放されることにある。アノテーション作業において、専門的知識をまっ

たく必要としない記述体系も大きな特徴である。

次にアノテーションデータの生成プロセスについて言及する。6章では、ユーザの明示的作業によるMAMLデータ生成のための方策について解説する。7章では、ユーザの暗示的作業からアノテーションと見なせるデータを抽出し、MAMLを生成するための方策について例示する。

6. アノテーションデータの明示的生成

通常アノテーションデータは、その付与自体を目的とする明示的操作によって生成される。その作成にかかるコスト削減は、非常に重要な課題である。MAMLは、タグエディタでファイルすべてを記述可能なほど、シンプルで簡易な仕様である。とはいえ、時間的コストを考えるとやはり現実的とはいえない。MAMLのような多用途の知的コンテンツは、それが専門的知識を有しない一般人でも簡単に生成できなければならない。そのためには、オーサリングツールの整備が不可欠である。そこで、ユーザのアノテーション作業を簡略化するオーサリングツール「MAML Editor/Viewer」を実装した。MAML Editor/Viewerは、MAMLファイル生成や閲覧を視覚的操作によって効率化する統合アプリケーションである。MAML Editor/Viewerの画面構成を図3に示す。MAMLファイルを生成するためには、メディアのメディア属性値を指定し、さらにエレメント情報を複数記述しなければならない。メディア属性は、図3の右上に表示されている「メディア情報表示ウィンド」上で指定する。次にエレメントの生成手順を説明する。以下の3項目を設定する必要がある。

- (1) 情報の種別とメディア時間の対象範囲の指定(メディア時間を有する場合)
- (2) エレメントのクラス
- (3) アノテーションテキスト

これらを図3中の「アノテーションボード」から設定する。「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ンタフェースである。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表示される。横軸は、対象メディアがメディア時間を持つ場合、それを表し、時系列に沿ってエレメントが配置される。(1)の操作はいずれかのトラックの一部をクリックすることにより、まず情報の種別が決定される。メディア時間を有する場合、さらにマウストラック操作により時間範囲を指定する。(2),(3)は(1)の設定後、表示される「エレメント追加ウィンド」上で行う。プルダ

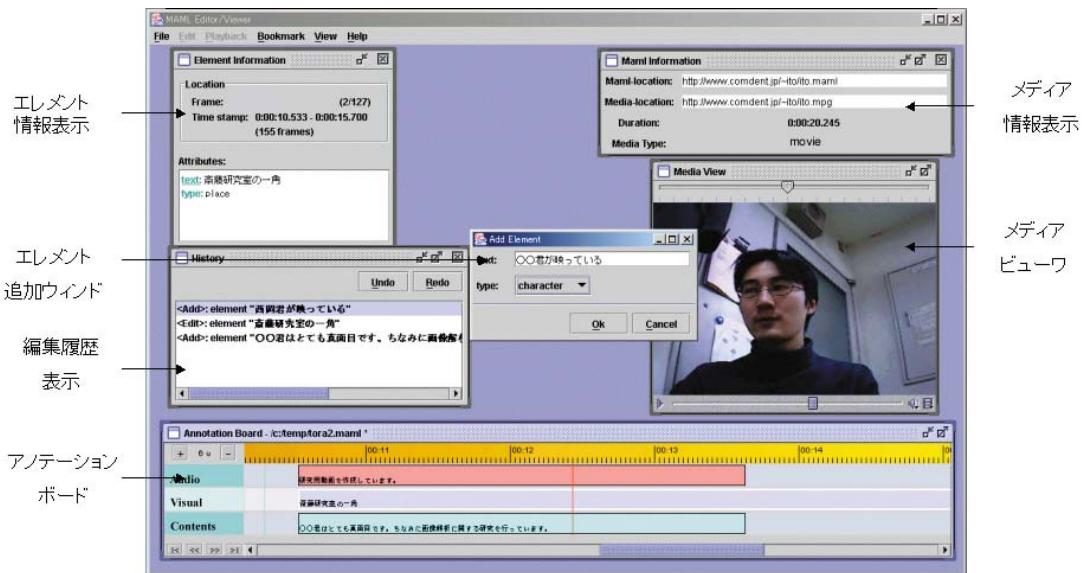


図 3 MAML Editor/Viewer のスクリーンショット

Fig. 3 A screenshot of MAML Editor/Viewer.

ウンメニューによりクラスを選定し、さらにテキストフィールド上にアノテーションテキストを記述する。エレメントはアノテーションボード上でクラスごとに異なった色の長方形で表示され、長さや位置は、メディア時間を持つ場合、メディア時間長と対象開始(終了)時間を表している。アノテーションテキストはその内部に表示される。以後はこの簡易な操作を繰り返すだけで、エレメント情報がアノテーションボード上に構築されていく。そして、メディア情報表示ウィンドの情報とあわせて MAML ファイルとして保存できる。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し再編集や閲覧することも可能である。現時点では、テキストメディアの一部に対するアノテーションや、エレメントに対するアノテーション機能は未実装である。

MAML Editor/Viewer の試用実験で、メディア時間範囲の指定が必要な場合、それがアノテーションに要する時間の大部分を占めることが判明した。空テキストのエレメントを、画像解析や音声認識により自動生成すれば、ユーザはアノテーションテキストの入力作業だけ行えばよいことになる。これはアノテーション作業の大きな省力化につながる。画像解析処理および音声認識処理を利用したアノテーション自動生成に関する研究もさかんに行われるようになってきている⁸⁾。我々も、動画・映像メディアにおける映像情報層に注目し、画像解析処理により個々の登場人物の出現範囲およびショットを抽出し、“人物”、“背景”クラス

のエレメント情報に関する MAML ファイルを自動生成するシステムを開発している⁹⁾。音声情報に関しても、音声波形の解析による“発話”クラスのエレメントの自動時間範囲抽出に加えて、音声認識によってアノテーションテキストの自動設定が可能であると考えられる。まだ精度の面からいえば技術的な問題も多く存在しているが、計算機処理による情報付加をベースに、人手によってそれを補完する半自動アノテーションの仕組みが非常に重要になってくるであろう。

7. アノテーションデータの副次生成

前章で解説した、明示的作業によって生成されるデータの量は、絶対的に少ないのが現状である。アノテーション関連技術の研究自体はさかんであるが、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。

アノテーションの概念を取り入れた大規模システムの成功事例に Google のイメージ検索¹⁰⁾ があげられる。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。このように本来の意図は別に存在するが、結果としてアノテーションと見なすことが可能なデータを副次生成されたアノテーションデータと呼ぶことにする。MAML は、このように副次的に生成されるアノテーションデータも記述

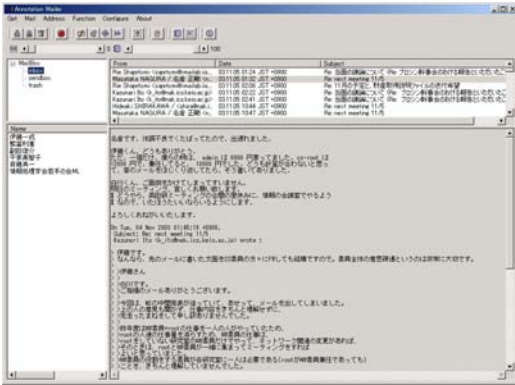


図 4 Annotation Mailer のスクリーンショット
Fig. 4 A screenshot of Annotation Mailer.

対象領域としている．本章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および文章入力をアノテーション行為と見なす”ことにより副次生成されるアノテーションデータが，MAML によって記述されるプロセスを，メーラとブラウザを例に提示する．

7.1 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) メールヘッダー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見なし，MAML データを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 4 に示す．

Annotation Mailer では，たとえば“メールのヘッダー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と考える．この操作により図 5 に示すような MAML を生成する．“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られる．“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または本文(の一部)に対してアノテーションしていると考えられる．例として，図 6 に示す返信メールから図 7 に示す MAML が Annotation Mailer 内部で生成される．

```
<?xml version="1.0" encoding="UTF-8"?>
<maml>
  <media type="null"
  maml-location=
  "file://c:\data\mail\checked.maml"
  media-word="本文を読んだメール">
  <element id="1" annotation="2003-11-09T07:12"
  target="file://c:\data\mail\5b7d531.mail">
  <contents>
    <subject>11/5 meeting</subject>
  </contents>
  </element>
  <element id="2" annotation="2003-11-17T11:19"
  target="file://c:\data\mail\5b9c8a2.mail">
  <contents>
    <subject>未踏ソフトウェア報告会</subject>
  </contents>
  </element>
  ...
</media>
</maml>
```

図 5 受信メールの閲覧によって生成される MAML の例
Fig. 5 An example of MAML generated by inspection of receive mail.

斉藤研の伊藤です．

```
>高田研の名倉と申します．
>
>> 斉藤(博)研の伊藤といいます．
>
>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，本人の意図していない色で出てしまったりと
>が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
になっていたりしてました．
```

図 6 メール本文の例
Fig. 6 An example of an e-mail text.

7.2 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


```

<?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>伊藤くんと私で , ...</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 7 電子メール本文の MAML 記述例 (本文は一部省略)

Fig. 7 An example of MAML description of the e-mail text.

為と見なし , MAML データを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 8 に示す .

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見なす . これにより図 9 のような MAML が生成され

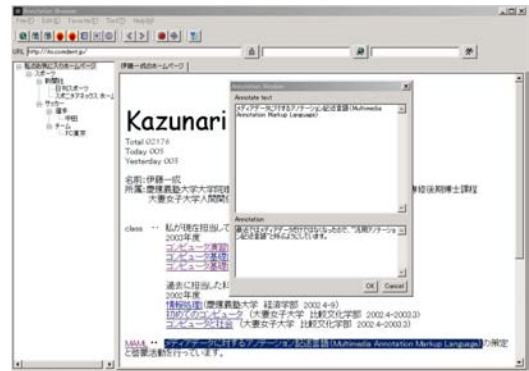


図 8 Annotation Browser のスクリーンショット

Fig. 8 A screenshot of Annotation Browser.

```

<?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 9 お気に入りリストの MAML 記述の例

Fig. 9 A description of a favorite list written by MAML.

る . また図 10 に示すような MAML が , 通常のナビゲーションの過程から Annotation Browser 内部で順次生成される . 図 10 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 タグ内を抜き出し , subject クラスを適用した . ホームページ作成者がそのページに付与したタイトルをアノテーションテキストと見なし , 引用するわけである . 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 . 参照元のアンカ文を参照先のホームページのアノテーションテキス

```

<?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</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 10 ブラウジング履歴からの MAML 生成の例

Fig. 10 An example of MAML generated from a browsing history.

トとした。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文の形式で説明（アノテーション）していると見なし、クラスには explanation クラスを適用した。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）説明と見なした。id 番号 4 のエレメントは検索エンジンに問合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。キーワードを入力した時点で Google のホームページのテキストフィールドには“京都 研究所”というテキストが存在しており、それ引用しているということで copy クラスを適用した。

ここに列挙したユーザの操作をそれぞれどのように MAML で定義するかは、人であるアプリケーション開発者が設計する。逆に MAML データからアノテーションを生成する基となったユーザの操作を推定できる必要はまったくない。どうしても明示したければ supplementation クラスのエレメントを生成し、その定義を自然文章で書いておけばよいのである。

RSS¹¹⁾ と同様な要約表現も MAML で記述できる。図 11 に HTML テキストから TITLE, Hn, A タグを自動抽出して生成した要約表現の例を示す。さらに、Annotation Browser では、HTML ファイル全体または一部に対してユーザの明示的操作により、アノテーションデータを付与することも可能である。図 11 における id 番号 9 のエレメントがその例である。

前章で解説した MAML Editor/Viewer を用いたアノテーションの生成過程も、結局はアプリケーション利用に係る一連の操作なわけである。Annotation Mailer や Browser も MAML Editor/Viewer と同じく MAML アノテーションツールと考えることができよう。

8. 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見なすことができる（図 12 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 12 の点線）。MAML では、タグ構造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。

```

<?xml version="1.0" encoding="UTF-8"?>
<maml>
<media type="text"
maml-location=
"file:///c:/data/browser/ito.comdent.jp.maml"
media-location=
"http://ito.comdent.jp/"
annotation-basetime="2003-11-23T03:46">
<element id="1">
<contents>
<subject>Kazunari Ito's homepage</subject>
</contents>
</element>
<element id="2">
<contents>
<copy>コンテンツ一覧</copy>
</contents>
</element>
<element id="3"
target="http://ito.comdent.jp/diary/">
<contents>
<copy>日記</copy>
</contents>
</element>
...
<element id="8"
target="mailto:ito@comdent.jp">
<contents>
<copy>メール</copy>
</contents>
</element>
<element id="9" target="3"
annotator="hxs@nak.ics.keio.ac.jp">
<contents>
<supplementation>全然更新してないな .
</supplementation>
</contents>
</element>
</media>
</maml>

```

図 11 MAML によるホームページコンテンツの要約表現の例
 Fig. 11 An example of summary expression of a homepage contents by MAML.

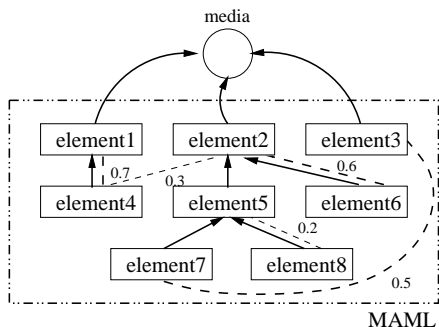


図 12 エレメント間の関連性の定義
 Fig. 12 Definition of a relevance between elements.

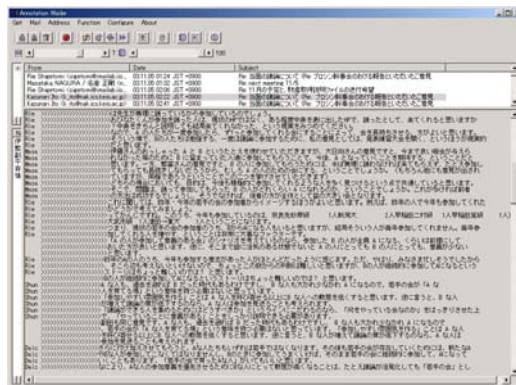


図 13 複数メールの要約提示機能使用時のスクリーンショット
 Fig. 13 A screenshot when summarized mails are displayed.

9. 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．

9.1 動画検索システム¹²⁾ および動画要約システム¹³⁾

このシステムでは，図 12 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．動画検索システムではアノテーションテキストと検索文間の単純な単語マッチングだけではなく文の係り受け関係等も考慮して検索処理を行っている．また要約システムでは，utterance クラスのアノテーションテキストに対して文内要約した結果を音声合成エンジンで再合成して，要約映像と同期して発話させることが可能である．このように，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．

9.2 複数メール要約機能

Annotation Mailer は複数のメールを 1 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 13 に示す．Annotation Mailer では，すべての受信メールおよび送信メールに対して図 7 のような MAML 形式でデータが保存されている．はじめに，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 12 で示したグラフ構造化を施す．図 13 の左上のスライダ



図 14 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. 14 A screenshot when the user-oriented web search is executed.

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、子がそれに対しての返答文に相当する。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。

9.3 ユーザプロフィールに基づく 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 14 に示す。Annotation Mailer において新規作成および返信したメールの MAML 群 (図 7 参照)、Annotation Browser のお気に入りリスト (図 9 参照)、Annotation Browser のブラウジング履歴の MAML (図 10 参照) をユーザのプロファイルと見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 API¹⁴⁾ 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 と呼ぶ)。グラフ A をクエリとしてグラフ 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うえでデータ形式に 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。また自然言語の最たる魅力はその表現の無限な豊かさにあり、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。

10. 関連研究

近年マルチメディアアノテーション技術に対する研究が活発に行われている。多くは、本論文と同様にアノテーションスキーマを策定し、そのうえでアノテーション支援ツールの実装や、スキーマの特徴を生かした応用システムを構築した研究事例が多い。

アノテーション作業をビジュアル環境で行うための支援ツールは、言語解析、音声認識、ゼスチャ認識、表情認識等の方面で用途やタグ仕様に準じた開発がなされている。MATE Workbench¹⁵⁾ は MATE (Multilevel Annotation, Tools Engineering) プロジェクトの一環として開発され提供されているアノテーションツールである。特に、音声コーパスに関する音韻や談話構造にわたるレベルのアノテーションを対象としている。アノテーションには独自の XML 形式を採用している。

Anvil はジェスチャや表情に関するアノテーションが可能なツールである¹⁶⁾。時間軸を持つ多層構造により情報を表現する。層名と属性をユーザが自由に指定できるので、任意のアノテーション体系に対応できるという特徴がある。しかし、独自にタグ仕様を設定できるということは、アノテーションを共有するうえで大きな障害ともなる。

以上のように、前もってメディアの種類や記述対象を限定して、それに特化したタグ仕様を策定あるいは利用する試みがほとんどである。

他には、Web データマイニング分野を中心に、Web コンテンツにリンク情報をはじめとするメタデータを付与する手法や、その実装システムに関しても提案がなされている。廣津らは、様々なメディアデータを統一的に取り扱うことができる Web ブラウジングシステム Cmew (Continuous MEdia with the Web) の構築を行っている¹⁷⁾。このシステムはプロキシとして

機能し、動画像のある時点での画像中の物体や、音の特定の区間に対してリンク情報を付与したブラウジング環境を提供している。しかしながら、扱っている情報はリンク情報に限っている。

W3C の Annotea Project¹⁸⁾ で開発が進められている Web アノテーションシステム Annotea 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。しかし、対象は Web コンテンツに明示的に付与するアノテーションに限定され、またコンテンツ間の関連定義については考慮されていない。

関連研究は、1) メディア特有の記述に関する研究、2) 言語的アノテーションに関する研究、3) メディア間の関係記述に関する研究に大別される。MAML は任意のメディアを対象として、これらすべてを包含できるのが大きな特徴である。

11. ま と め

本論文では、メディアの種類やフォーマットに依存しない統一的なアノテーション記述仕様 MAML を提案した。また明示的なアノテーション生成が必要な場合、そのコストを大幅に削減するツールを実装した。さらに、ユーザの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、データを副次的に生成するプロセスについて例示した。またいくつかの応用事例について簡単に言及した。アノテーション仕様に関しては、以下に述べる 3 点が重要である。

- (1) 一般の人々がデータを生成可能であること。つまり生成過程が容易であること。
- (2) データが広く流通し、応用アプリケーションが数多く生成されること。
- (3) その結果、ユーザがアノテーションしたことによる恩恵が多分に得られる環境が実現されること。

この 3 要素はスパイラルの関係にあり、すべてを満たさなければ、その仕様は受け入れられないであろう。

MAML は本論文で解説した以外にも様々な適用方式が想定される。今後は MAML による高度メディア利用環境の実現を目指して、仕様の再検討、アノテーションツールの機能強化、各種応用アプリケーションおよびシステムの構築を継続していく予定である。

謝辞 本成果は、平成 14 年度 IPA 未踏ソフトウェア創造事業「未踏ユース」および平成 15 年度 IPA 未踏ソフトウェア創造事業の一部である。IPA (情報処

理推進機構)、プロジェクトマネージャーの電気通信大学竹内郁雄教授および京都大学田中克己教授に深く感謝いたします。

参 考 文 献

- 1) The MPEG Homepage.
<http://mpeg.telecomitalia.com/>
- 2) Hendler, B.-L.J. and Lassila, O.: The Semantic Web, *Scientific American*, May 2001 (2001).
- 3) W3C RDF Homepage.
<http://www.w3.org/RDF/>
- 4) 椎尾 一郎, 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).
- 5) The 'friend of a friend' project: FOAF Homepage.
<http://rdfweb.org/foaf/>
- 6) Dublin Core Metadata Initiative (DCMI) Homepage.
<http://dublincore.org/>
- 7) Extensible Markup Language (XML) Homepage.
<http://www.w3.org/XML/>
- 8) 長尾 確, 白井良成, 橋田浩一: アノテーションに基づく知的マルチメディア処理, 情報処理学会研究報告, ICS120-6, pp.41-48 (2000).
- 9) 西岡伸紘, 伊藤一成, 斎藤博昭: 動画像からの人物情報記述データ自動生成システムの提案, 電子情報通信学会信学技報, PRMU2002-201, pp.1-6 (2003).
- 10) Google イメージ検索ホームページ.
<http://www.google.co.jp/img/hp/>
- 11) RDF Site Summary (RSS) 1.0 homepage.
<http://web.resource.org/rss/1.0/spec>
- 12) 伊藤一成, 斎藤博昭: 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).
- 13) 伊藤一成, 酒井康旭, 斎藤博昭: 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, pp.41-48 (2004).
- 14) Google API Homepage.
<http://www.google.com/apis/>
- 15) McKelvie, D. and Isard, A.: The MATE Annotation Workbench—an annotation tool for XML coded speech corpora, *Speech Communication*, Vol.33, No.1-2, pp.97-112 (2000).
- 16) Kipp, M.: Anvil—A Generic Annotation Tool for Multimodal Dialogue, *Proc. Eurospeech 2001*, pp.1367-1370 (2001).
- 17) 廣津登志夫, 高田敏弘, 青柳滋己, 佐藤考治, 菅原俊治: Web アノテーション共有システム Cmew/U の設計と実装, 情報処理学会論文誌, Vol.42, No.10, pp.2466-2475 (2001).

- 18) Kahan, J., et al.: Annotea: An open RDF infrastructure for shared Web annotations, *The 10th International World Wide Web Conference (WWW-10)* (2001).

(平成 15 年 12 月 20 日受付)

(平成 16 年 4 月 6 日採録)

(担当編集委員 春本 要)



伊藤 一成 (学生会員)

1974 年 7 月 18 日生。2001 年 3 月慶應義塾大学大学院理工学研究科 計算機科学専攻前期博士課程修了。現在同後期博士課程在学中。マルチメディア情報処理に関する研究に従事。電子情報通信学会，電気学会，日本データベース学会各学生会員。



斎藤 博昭 (正会員)

慶應義塾大学工学部数理工学科卒業。現在同大学理工学部情報工学科専任講師。工学博士。自然言語処理，音声言語理解等に興味を持つ。言語処理学会，日本音響学会，電子情報通信学会，ACL 各会員。