

## 意味データを解釈するエージェントを用いたデータ活用環境の開発\*

佐伯嘉康†

東京大学 大学院工学系研究科†

相田仁‡

東京大学 大学院工学系研究科‡

## 1 はじめに

本研究では、膨大な量のデータを効率的に利用者に提供する手段として、意味データが適切に付与されたデータの流通の場としてのデータ活用環境を構築する。本環境で、データに適切な意味を付与するための意味データの要素の発見、意味データの公開についてエージェントの支援を行うことで、データの利用効率が向上することを示す。

## 2 意味データ

意味を記述したデータとして盛んに用いられているデータは、メタデータである。メタデータとはその名の通り、データの上位概念として、対象となるデータについての情報を記述したデータである。

## 2.1 Web に於ける意味データの普及

メタデータは Web の普及以降広く使われるようになり、メタデータの書式や内容の記述方式等を、出来るだけ多くの場合に適応出来るような汎用性のある性質を持つように、あらかじめ定め、利用者に周知してもらう必要性が現れた。そのために、幾つかのメタデータ標準技術規格が策定された。

RDF(RDF Schema)[1] と Topic Maps[2] は最も標準的なメタデータ規格として、広く活用されている技術である。共にメタデータのモデルを定義しており、モデルを記述するための書式は複数ある。RDF は W3C によって、Topic Maps は ISO/IEC 13250 で定められている。活用例として、RSS(Ver. 1.0 / RDF Site Summary) の書式として XML で記述した RDF モデルが使われている

メタデータに関する最新の話題として、W3C に

よる Microdata[3] がある。元は HTML5 技術の 1 つとして技術策定に向けて進められていたが、今年に入り HTML5 技術からは独立することになった。前身は RDFa と呼ばれる RDF のメタデータモデルを HTML 文書内でタグの属性として記述するための技術である。

Microdata と類似の技術として microformats[4] がある。microformats は RDF モデルではない。そのメタデータのが扱う文書メタ種類は全て、microformats 団体が管理している Web 上の Wiki で決定、公開される。

## 2.2 意味データ技術に関する既存の問題点

問題点 1 現在の Web には、メタデータの要素定義の存在を周知させる手段が欠けている。

先に挙げたメタデータ技術について、メタデータそのものの存在を周知するための手段に欠けている。メタデータが Web 上に公開されても、それを広く知らせ、活用してもらうための術は無い。メタデータ作成者の能動的な行動から、つまり、検索エンジンを通じた発見に頼るといのが現状である。

また、メタデータを記述するためには、そのメタデータが何を示すものなのかを表すためのメタデータ要素定義が必要となる。例えば、RSS をメタデータとして記述するためには、title、description 等のメタデータ要素を必要とするが、これら要素の定義については、RSS-DEV Working Group によって管理されている <http://purl.org/rss/1.0/> で参照することが出来る。

しかし、この定義を発見することもまた、メタデータの発見同様、困難である。microformats の様に、権威ある団体が管理している要素定義集 (Wiki) のみを用よという姿勢は 1 つの問題解決手段ではあるが、RDF や Topic Maps などが持つ汎用性/拡張性が失われることになる。

問題点 2 現在の Web 上のメタデータは、相互利用/再利用性が十分に活かされていない。

\* Development of system for utilization of semantic data with agents

† SAEKI Yoshiyasu, Dept. of Electrical Engineering and Information Systems, School of Engineering, The University of Tokyo.

‡ AIDA Hitoshi, Dept. of Electrical Engineering and Information Systems, School of Engineering, The University of Tokyo.

現在のメタデータは、そのメタデータを記述したファイル 1 つで完結させている傾向がある。この傾向が生まれた原因として、やはり、メタデータの発見手段が不十分であることが挙げられる。メタデータというデータは、これまで利用されてきたデータよりも上の汎用性の高さを保証した技術/概念であり、複数のデータに跨って再利用されてこそ、その有用性を発揮することが出来る。

### 3 エージェントを用いた意味データ活用環境

#### 3.1 既存の問題点の解決

第 2.2 節で挙げた問題点を解決するために、本研究では、意味データの発見の役割と、意味データの定義を利用者に提供するための役割を持つプログラムとしてエージェントを用いる。複数のエージェント間で意味データの存在、コンテンツを広告を通信することで、意味データの再利用効率を上げる環境を構築する。

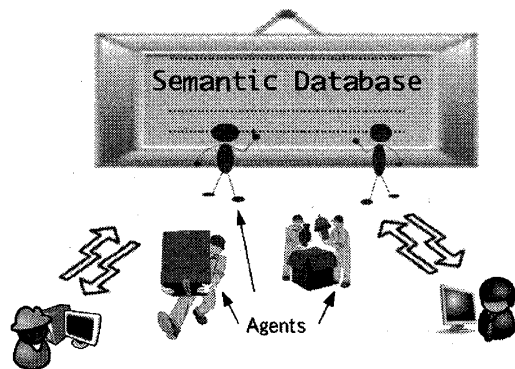


図 1 An image of this system

#### 3.2 設計

エージェントは、個々のコンピュータに常駐する。ここで言うコンピュータとは、データを活用する利用者が利用している PC 等を指す。利用者はコンピュータを通して、データと共に、意味データを生成、公開する。エージェントが個々のコンピュータから意味データの発見、理解、配布を行うことで、データの利用を利用者に促し、データの利用効率を向上させることを本環境の目的としている。

既に、Web 上には意味データとしてメタデータが多く流通している。これらの意味データとしてのリソースは今後も活用されるべきである。そこで、意味データの参照利用を高めるための、参照構造の格納のためのデータベース機構を設ける。

意味データは利用者が記述する必要がある。しかし、意味データを記述するための要素の定義などを、検索エンジン等で発見しても、意味データ要素が正確であるかという保証は無い。前に述べたように発見自体困難である。そこで、エージェントが意味データ要素定義の発見を担う。利用者が記述し公開したいデータに対し、適切な意味データ要素の発見は、データベースにクエリを投げることで参照する。ここでは、データに含まれる単語等から特徴、他のデータ/意味データとの関連を算出、適切な意味データ要素を判断、利用者に通知する。

利用者はエージェントから推薦された意味データを利用することが出来、独自の意味データ要素の定義を作成することも出来る。その場合、エージェントは、メタデータ要素の定義の再利用を促すために、その意味データ要素をデータベースに登録し、他エージェントに参照可能な状態にする。

このエージェントの支援によって、メタデータとその要素の定義と共に、メタデータの再利用可能な状態を維持しながら、メタデータを通した利用者のデータ利用効率向上に繋がる。

### 4 まとめと今後の計画

本稿では、まず意味データの現在に至る歴史について概要を述べ、意味データ活用環境として現在の Web が機能不十分であることを述べた。次に、本研究の主題として、利用者が作成するデータに付与される意味データを解釈するエージェントと呼ぶプログラムを通し、意味データの発見と再利用を促す環境を開発した。

第 3.2 節で述べたデータベースは、現在 Web アプリとして開発中である。従って、機構全体として Web の一部として実装を進めているが、今後、P2P ネットワーク上での新たな Web として実装していくことも計画している。

#### 参考文献

- [1] W3C: "Resource Description Framework (RDF) / W3C Semantic Web Activity," <http://www.w3.org/RDF/>, accessed: 2010-01.
- [2] TopicMaps.Org: "TopicMaps.Org Home Page," <http://www.topicmaps.org/>, accessed: 2010-01.
- [3] HTML5 Working Group: "5 Microdata - HTML5 (including next generation additions still in development)," <http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html>, accessed: 2010-10.
- [4] microformats: "Microformats," <http://microformats.org/>, accessed: 2010-01.