

[基調講演] セマンティック Web と Linked Data

武田 英明^{†‡}[†] 国立情報学研究所 〒101-8430 千代田区一ツ橋 2-1-2[‡] 東京大学 人工物工学研究センター 〒277-8568 柏市柏の葉 5-1-5

E-mail: takeda@nii.ac.jp

あらまし セマンティック Web は Linked Data の普及で新しい局面を迎えている。これまでセマンティック Web はいかに概念の体系（オントロジー）を構築して利用するかについて注力してきた。これに対して昨今、Linked Data という形で、個々の事柄に関する情報（インスタンス）の構築と利用が注目されている。本講演ではセマンティック Web の基礎概念について説明するとともに、Linked Data の役割と現状を報告する。

キーワード セマンティック Web, オントロジー, linked data

Semantic Web and Linked Data

Hideaki TAKEDA^{†‡}[†] National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan[‡] RACE, The University of Tokyo 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8568 Japan

E-mail: takeda@nii.ac.jp

Abstract Semantic Web is shifting to the next phase by diffusion of Linked Data. Research on Semantic Web has focused on ontology, i.e., the system of concepts. Now it is focusing on instances of concepts, so-called linked data. In this article, I briefly introduce the basic idea of Semantic Web and then the role and the state-of-the-art of linked data.

Keyword Semantic Web, Ontology, Linked Data

1. はじめに

Web は Tim Berners-Lee は研究者間の情報交換のために考案したものであり、その後広く人々の情報交換のために使われるようになった。その経緯についてはいまさら語る必要もないであろう [1]。

しかし、彼は現在の標準となった http と html に満足していなかった。彼はもっと知的なサービスをその当初から考えていた。Web の標準化が一旦なされたとき、それをセマンティック Web として提示した [2]。セマンティック Web は Web の標準化団体である W3C (World Wide Web Consortium) での標準化活動と情報科学の研究者の研究の両輪で進められている。

本稿ではセマンティック Web の最近の動向を概説する。

2. セマンティック Web

セマンティック Web の方向性を理解するのは "layered Cake" と呼ばれる図 (図 1) に基づくとわかりやすい。この図はかなり初期 (2002 年) のもので、その後さまざまな変化をしているが、基本構造はあまり変わっていない。いわゆる Web は XML と namespace, URI の層である。これは基本的に現在の Web を支える仕組みであると同時に、この上に Web の拡張として構築されるセマンティック Web の基盤としても重要である。とくに URI はオントロジー共有において重要な役割を果たす。

RDF M&S (Model and Syntax) 層, RDF Schema 層, オントロジー層それぞれのために RDF, RDFS, OWL というセマンティック Web のための言語が作られている。セマンティック Web として標準化されているのはこ

こまでで、これより上の部分は研究途上である。

2.1. Web における意味記述の考え方

セマンティック Web においては Web ページには明示的に意味が書かれることが期待する。そして、この意味とはオントロジー中の概念によって指し示されると考える。たとえば、あるページの中に「ある犬がある人間を噛む」という意味が含まれているならば、この中の「犬」「人間」「噛む」はオントロジー中の概念とその関係で示されると考えるわけである。

オントロジーとは本来は哲学の存在論のことであるが、90年代の人工知能においてコンピュータ科学に導入された。Tom Gruber[3]はオントロジーとは「概念化の形式的な仕様」と定義しており、具体的にはコンピュータの中で扱われる概念の体系を指している。多くの場合、オントロジーは形式論理を用いて書かれることが多く、この場合は概念自体や概念間の関係は論理の公理として書かれる。また一般に階層関係をもつことからオブジェクト指向パラダイムにおける抽象化と近いことがわかっている。

2.2. URI による意味記述

セマンティック Web の基本的な意味記述の仕組みは、URI とオントロジーの組み合わせになる。URI (Uniform Resource Identifier)とはインターネット上にある情報資源(resource)を指し示す方式のことで、普通に Web で使われる URL(Universal Resource Locator)を包含する概念である。この URI を使って個々の“意味”を指し示す。これは、ある意味に対応する URIにいくとその意味内容が書かれているということではない。単に意味の識別子を与えているに過ぎない。さらに URIは誰でも生成することができるので、一意の識別子を与えているわけではない。Web の世界は分散的であるので、セマンティック Web でも意味に一意に識別子を振ることは期待してない。多くの Web ページがある意味を指し示す特定の URI を使うことで初めて意味の共有が実現される。

URI が指し示す意味内容は URI 間の関係によって記述される。この URI 間の関係を使って意味の体系を構成する。とくに URI が個体(individual)ではなくて、概念を指している時には、概念間の関係を記述することになる。すなわち、オントロジーを URI 間の関係として与えることで、Web 上の情報資源にその意味を与えることができる。

この関係の記述の方式の違いで、以下に述べるような RDF, RDF Schema, OWL と異なる記述方式がある。

2.3. RDF (Resource Description Framework)

RDF は情報資源を指し示す URI を記述する枠組みである。基本的には主語(subject), 述語(predicate), 目的語(object)の3つ組として表現される。主語は記

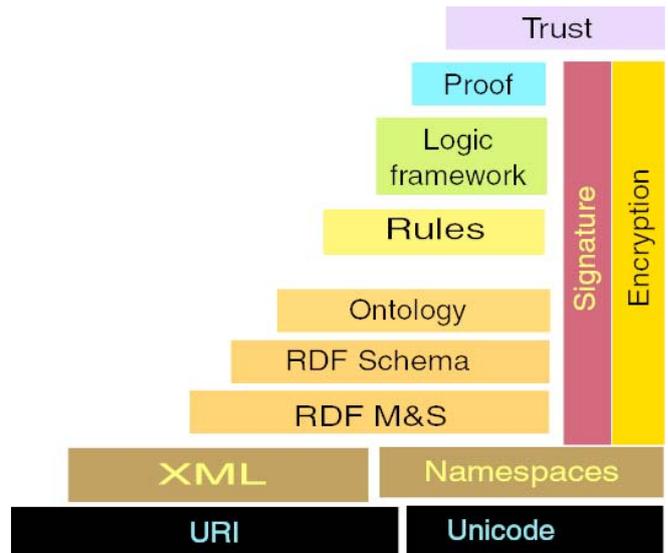


図 1: セマンティック Web の階層図

<http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

述対象の情報資源であり、もっぱら URI で示される。目的語は他の情報資源であり、URI あるいはリテラル(文字列)である。述語は二つの情報資源の関係を示すもので、これもまた情報資源 (URI またはリテラル) である。述語はプロパティ(property)とも呼ばれる。RDF はラベル付きの有効グラフを構成する。また、主語は空白ノードでも許されるので、3つ組では記述が難しい構造的な記述を空白ノードを使った入れ子構造で表現することができる。

なお、RDF はこのような情報構造のモデルを提供するだけで記述の方式は規定しない。このため、XML で書く方式(RDF Serialization)や専用言語 Notation 3 (N3)などがある。

2.4. RDF Schema (RDFS)

RDF Schema は RDF の上に構築された簡便なオントロジー記述言語である。RDF 語彙 (RDF vocabularies)とも呼ばれ、RDF で記述された情報資源を構造化するために用いられる。

その構造化とはオブジェクト指向言語あるいはフレーム言語に類しており、クラスおよびサブクラス関係を基本とするクラスの階層構造である。またプロパティにはその定義域と値域をつけることができ、サブプロパティ関係も使うことができる。

ただし、RDF Schema で書ける構造というのは限られており、制約のついたクラス定義といった複雑な記述はできない。

2.5. Web Ontology Language (OWL)

OWL もオントロジー記述言語であるが、RDF Schema より複雑な記述をすることができる。RDFS と同様にクラスやサブクラス関係を持つが、さらにク

ラス制約（あるクラスに一定の論理的制約を課すこと
によってそのサブクラスを定義する）やクラスの論理
和、論理積といった記述も可能である。またデータ
タイプとして XML Datatype を使うので、より正確な定
義が可能である。

2.6. セマンティック Web 言語の利用

RDFS や OWL による Web のためのオントロジーはだ
んだんと多く開発されるようになった。セマンティッ
ク Web 検索エンジンである [swoogle \(swoogle.umbc.edu/\)](http://swoogle.umbc.edu/)
や [Watson \(watson.kmi.open.ac.uk/\)](http://watson.kmi.open.ac.uk/)
を使えば Web 上で公開されて
いるオントロジーを検索することができ、多数あること
が確認できる。

またセマンティック Web のオントロジーは分散的
に開発されることが当然であるので、複数のオントロ
ジーをどう使っていくかが問題になる。このためオン
トロジーマッチング、オントロジーアライメントとい
った技術が現在研究開発されている [4]。

3. Linked Data

セマンティック Web の研究においては主にクラス
に相当するオントロジーに焦点が絞られてきた。しか
し、近年、インスタンスの部分により焦点があてられ
るようになった。これが Linked Data と総称される。

先に述べたようにオントロジーではクラスやプロ
パティの定義が書かれる。その定義にそって具体的
な事象が記述される。これがインスタンスに関する記
述である。たとえば、先のれいでは「ある人間」「ある犬」
に対してそれぞれ、「人間」くらす、「犬」クラスのイ
ンスタンスが定義され、その二つのインスタンスに「噛
む」というプロパティが付けられる。この文(多くの場
合 RDF)自体もインスタンスに関する記述である。

Web ページは本来このインスタンス記述が含まれて
いることが期待される。

近年、大量のデータを Web に提供するサイトがこの
流儀にしたがって RDF データを大量に公開するよう
になった。この RDF データ公開において重要なのは単
に自サイトで閉じたデータではないという点である。
オントロジー定義においてほかのサイトと共有する概
念や関係をもつことができれば、生成されるインスタ
ンス記述も当然、他サイトのインスタンス記述の関
係をもつデータになる。すなわち、データレベルに関
係することになる。このためこのインスタンス記述を
総称して Linked Data と呼ばれる [5]。

現在、Linked Data がどんな状況であるかを図 2 に示
す。この中で DBpedia (<http://dbpedia.org/>)
というのは Wikipedia の情報のうち、infobox の情報
を中心に機械的に抜き出し、RDF のデータとして書き
だしたもので

ある。現在、約 1.1 億 RDF 文が公開されている。ま
たこの中の地名に関しては GeoNames というデータの
相互にリンクがある。GeoNames 自体は約 7 千万文あ
る。この関係は図中で相互リンクとして書かれてい
る。

この例でもわかるように Linked Data は名前の通り
相互にリンクしあうからこそ価値がある。そのため
データはオープンであることが望ましい（必須ではな
いが）。そこでオープンな Linked Data を普及しよう
というプロジェクト Linking Open Data Project
([http://esw.w3.org/topic/SweoIG/TaskForces/CommunityP
rojects/LinkingOpenData](http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData))が W3C のメンバーらで行わ
れている。

このように RDF データが相互にリンクしあうこと
で巨大なデータ空間を作っている。そうするとこの
データを使うアプリケーションが可能になる。

例えば、SemaPlorer (<http://btc.isweb.uni-koblenz.de/>)
は Google map を中心に DBpedia, Geonames, flickr デ
ータをマッシュアップして作られている。ここでは単
にデータを結合して検索する以上のサービスを提供し
ている。

またすでに“Linked Planet Data Conference”と呼ば
れるビジネスカンファレンスシリーズが開かれるよ
うになっている(<http://www.linkeddataplanet.com/>) [6]。

4. Linked data の生成の仕方

Linked Data の作り方には二通りある。一つは RDF
のデータを Web ページとは別に用意しておく方法
である。前述の大きなデータセットはこの方法による。
あるいは、個別の Web ページにメタデータを埋め込
んでおくことも可能である。RDFa では RDF を直接
XHTML に埋め込むことができる
(<http://www.w3.org/TR/xhtml-rdfa-primer/>)
。Microformats (<http://microformats.org/>)は RDF ではない
が、メタデータを同様に直接 XHTML に埋め込む。

リンク可能な RDF を作るには一種のルールが必要
である。[http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTu
torial/](http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial/)). とくに URI の使い分けには注意が必要である。

5. 我々の取り組み

筆者の研究グループではどのように柔軟なメタデ
ータを入力できるかについて研究している。socioblog
ではブログに書誌データを埋め込むことで、メタデ
ータの公開や利用の新しい使い方を提案した [7]。
STYLID では多様な概念定義を共存できる仕組み
を提案している [8]。このシステムは freebase
(www.freebase.com/)に似ているが、だれでも自由に概
念を定義してそのインスタンス記述を追加しておく
ことができる。しかも類似の概念をまとめて半自動で関

