

セマンティックWebとオントロジ

三菱電機(株) 情報技術総合研究所 音声・言語処理技術部
(財団法人 情報処理相互運用技術協会セマンティックWebタスクフォース委員)

今村 誠

imamura@isl.melcaco.jp

・セマンティックWebにおけるオントロジ

・オントロジとは

ECにおけるXML文書情報交換 -商品カタログの例-

・オントロジへの期待(応用メリット) -電子商取引の例-

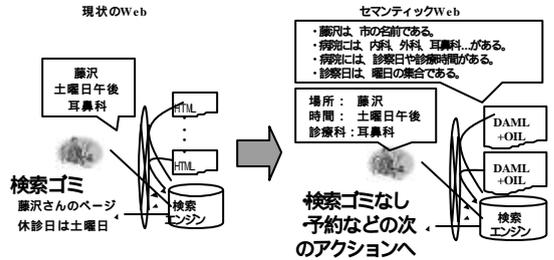
・Webコンテンツ記述言語の階層

従来のAIの知識表現システムとの差異

今後の課題

セマンティックWebにおけるオントロジ -現状のWebとの差異

- ・ Web コンテンツに意味情報を付与することにより、利用者が必要とする情報を手に入れやすくする

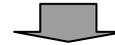


オントロジとは?

- ・ 意味情報を表現するための方法の一つ。
- ・ 元来、「ある」ということの意味を問う哲学分野だったが、Web やAIの分野で、概念間の関係の明確な定義の集まり」という意味で使われるようになった。
- ・ セマンティックWebでは、概念どうしの関係や、それらを解釈するための論理的なルールを定義する文章の集合」という意味で使う。典型的なオントロジは、「分類体系」と「推論規則集」である。

ECにおけるXML文書情報交換 - 商品カタログの例 -

- ・ メーカー横断のカタログ検索を実現するために、商品の分類体系や、分類カテゴリ毎の商品属性定義をXMLで表現



- ・ 複数の企業や組織からなる業界がオントロジを標準化するのはたいへん
- ・ 異なるオントロジをもった業界間での情報交換が難しい

XMLにおける情報表現の多様性

```

<自動車>
  <車名> カローラ </車名>
  <価格> 150万円 </価格>
</自動車>

<製品>
  <タイプ> 車 </タイプ>
  <名前> カローラ </名前>
  <値段 単位="万円"> 150 </値段>
</製品>

```

セマンティックWebによるオントロジ記述

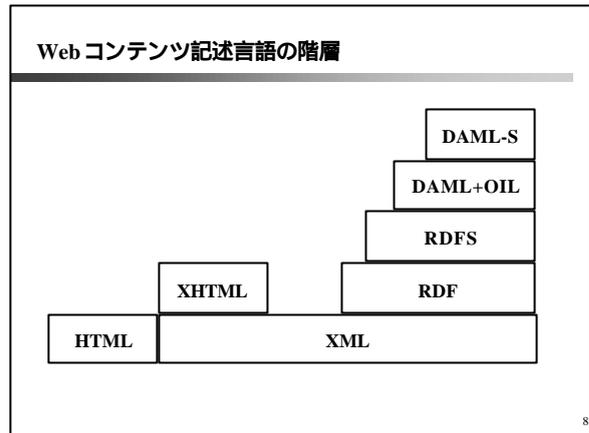
- ・ 複数のオントロジ標準が並存する環境下での情報交換を実現
- ・ 複数Webサイトのコンテンツを連携させるような自動サービスの実現



- ・ オントロジ間の相互変換が容易になるように、概念間の階層関係や概念定義間の整合性などを自動計算可能なオントロジ記述言語を提供

オントロジへの期待 (応用メリット) - 電子商取引 -

- 商品分類や商品属性のオントロジ記述により、自由な用語による商品検索
- 利用者嗜好のオントロジ記述により、利用者の要求にあった商品検索や情報配信
- オントロジ変換により、異なるオントロジ標準の商品カタログ比較検索
- オントロジを用いた推論により、複雑な質問への回答



オントロジ記述言語を用いた実験システム (On-To-Knowledge プロジェクト)

- OILをベースとして、非構造データや半構造データに対してアノテーション(補足情報)を自動付与するツールや、異種形式のWebコンテンツに対する統合検索ツールを開発
- それらのツールを用いて、保険、通信、およびエネルギーの3つの産業分野を対象として実験システムを開発

DAMLプロジェクトの実験システム ITTALKS

Maryland大は、DAML-Sを用いて、IT分野の講演に関する情報(場所、発表者、主催組織、トピックなど)を提供するWebポータルであるITTALKSを開発。

- 講演案内のWebページに付与されたメタ情報を参照することにより、利用者の興味、スケジュール、および住所などに応じて、参加が推奨される講演を利用者のカレンダーに反映する機能を提供。
- ACM(Association for Computing Machinery)のトピックとUMBC(University of Maryland Baltimore Country)のトピック間のオントロジ - 変換を実現。

従来のAIの知識表現システムとの差異

	80年代 AIの知識表現システム	セマンティック Webの知識表現システム
管理方式	集中方式	分散方式
矛盾に対する考え方	矛盾がないように、固有のルール体系を推論に利用	矛盾があっても実用上困らないように、ルールの運用に柔軟性をもたせる
オントロジの多様性	一つの普遍的なオントロジを想定	多様なオントロジの存在を容認
知識の作成と利用	少数の人が作成	多数の人が作成。他のサイトのオントロジを参照
推論の実行	単一計算機で実行	Webの分散環境で実行

今後の課題

- オントロジの作成**
単純なオントロジ作成エディタだけでなく、オントロジの相互変換、統合、および改版などを容易にするようなオントロジ構築の方法論の確立
- オントロジの質**
オープン、分散、そして動的に変化するWebの世界において、どのオントロジが信頼できるかを知る仕組み、また、互いに矛盾がある場合にそれを柔軟に処理する仕組みの確立

セマンティック Web とオントロジ

今村 誠

E-mail: imamura@isl.melco.co.jp

三菱電機（株） 情報技術総合研究所 音声・言語処理技術部
(財団法人 情報処理相互運用技術協会 セマンティック Web タスクフォース委員)

最近、次世代の Web ビジョンとして、Web コンテンツに意味情報を付与することにより、Web の有用性を飛躍的に高めようとするセマンティック Web が注目されている。セマンティック Web では、この意味情報を表現するための方法としてオントロジに注目し、オントロジを記述する新しい言語を提供する。本稿では、電子商取引(EC: Electronic Commerce)における XML(eXtensible Markup Language)を用いた文書情報交換を例にあげて、オントロジの必要性を説明した後、セマンティック Web におけるオントロジ記述言語の現状について述べる。

1 はじめに

90 年代の初めに提案された HTML(HyperText Markup Language)は、インターネットを介して世界中で公開されている様々な情報にアクセスできる WWW(World Wide Web)を生み出した。その一方で、WWW の爆発的な普及は、情報の氾濫を引き起こし、利用者が自分の必要とする情報を手に入れる手間も次第に大きくなってきた。セマンティック Web は、この情報の氾濫を解決するために、現状の Web コンテンツに機械が理解可能な意味情報を付与することにより、Web の有用性を飛躍的に高めようとする次世代の Web ビジョンである。

そして、この意味情報を表現するための方法の一つが「オントロジ」である。オントロジとは、元来は哲学の用語で、「ある」ということの意味を問う学問分野のことであったが、人工知能や Web の分野では、「概念間の関係の明確な定義の集まり」という意味で使われている用語である。オントロジを記述する言語(以下、オントロジ記述言語と呼ぶ)は、セマンティック Web を実現するための中核技術であり、米国の DAML プロジェクト (<http://www.daml.org>)、欧州の On-To-Knowledge プロジェクト(<http://www.ontoknowledge.org>)を中心に、研究開発が進められてきた。

本稿では、オントロジの必要性とセマンティック Web におけるオントロジ記述言語の現状について述べる。以下、2 章では、EC における XML 文書情報交換を例にあげて、従来の辞書を用いた XML コンテンツの標準化とその問題点について述べる。3 章では、2 章で述べた問題点を解決する技術としてオントロジが

期待されていることについて述べる。4 章では、セマンティック Web におけるオントロジ記述言語の現状について述べる。最後の 5 章では、まとめを述べる。

2 EC における XML を用いた文書情報交換

XML を用いた EC では、企業間で交換される XML 文書中のタグ名やタグの内容のデータ型などを標準化する必要がある。商品カタログの例でいえば、メーカー横断の商品検索を実現するために、商品の分類体系や、各々の商品分類カテゴリごとに商品属性を規定する辞書を標準化する必要がある。

例えば、電子機器部品の業界では、まだ XML が W3C の勧告になる以前から、SGML(Standard Generalized Markup Language)形式の商品カタログを作るための辞書の整備が進められており、最近のパソコンなどの情報機器と電子部品(半導体)のサプライチェーンの効率化を目的とする RosettaNet (<http://www.rosettanet.org>) に受け継がれている。電子部品用の辞書では、例えば、「周波数シンセサイザやミキサーは、チューナーの一種である」といった商品の分類体系や、「チューナーは、入力オフセット電圧や電源電圧変動除去比という属性をもつ」といった知識が記述される。

しかし、複数の企業や組織からなる業界が辞書を標準化するには大きな労力と時間が必要となるので、同じ業界内でも複数の辞書標準ができる可能性がある。また、ある業界での標準化が達成されても、異なる業界では、同じ概念を別の辞書体系で表現されることは避けられないといえる。結果として、現実の世界では、異なる辞書体系をもつ企業間での情報交換を実現する必要が生じてくる。

ところが、現状の XML の枠組みだけでは、複数の辞書体系が並存する世界では問題が生じる。なぜなら、XML を用いた商品カタログの記述には、入れ子構造の深さや、タグの名称、属性の使用法などにより、情報の表現方法に任意性があるため(図 1 に同じ自動車仕様の二つの表現例を示す)、本質的には同じ情報が異なった形式で表現された場合に、機械処理ではそれを認識することができないからである。

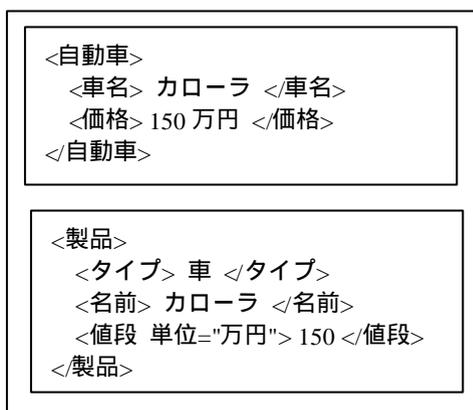


図 1 自動車の仕様の XML による二つの表現例

そこで、異なる辞書体系をもつ企業間の情報交換では、XML/XML 変換プログラムが必要になるが、現状の XML の枠組みでは、この変換プログラムは、DTD(Document Type Definition)毎に個別に開発するしかないという状況である。この XML/XML 変換プログラムの作成をできるだけ容易にするしくみが望まれるが、オントロジはこの要望に応える技術として期待されている。

3 オントロジ

セマンティック Web では、複数の辞書体系が並存する環境下での情報交換を実現するために、異なる辞書体系をもつ XML 文書間の変換が容易になるように、概念間の階層関係や概念定義間の整合性などを自動計算できるようなオントロジ記述言語を提供するというアプローチを採用している。さらに、オントロジ記述言語は、オントロジの機械処理を容易にするという特徴もあるので、複数 Web サイトのコンテンツを連携させるような自動サービスの実現を容易にすることも同時にねらっている。

本章では、EC における従来の辞書とオントロジとの差異を明確化するために、オントロジとは何か、また、オントロジに期待する効果について述べる。

3.1 オントロジとは

オントロジにはいくつかの定義があるが、本節では、代表的なオントロジの定義として Gruber の定義を紹介する。また、計算論的な意味付けの観点からオントロジのレベルについて述べる。

(1) Gruber のオントロジの定義

On-To-Knowledge プロジェクトにおけるオントロジ記述の枠組みである OIL(Ontology Inference Layer: <http://www.ontoknowledge.org/oil/>)の開発を主導している Fensel らは、彼らの主張に合致したオントロジの定義として、Gruber の定義「共有される概念化 (conceptualization) の形式的(formal)かつ明示的(explicit) な仕様」をあげている[1]。

ここで、概念化とは、対象とする現象の抽象的なモデルのことである。このモデルでは、現象中で興味をもつ概念と、それらの概念間の関係が表現される。形式的とは、オントロジが機械に理解されるものでなければならないことを意味する。明示的とは、概念のタイプと、概念間の制約関係が明示的に定義されていないことを意味する。

現状の XML を用いた EC における辞書も、オントロジと呼ぶこともできるが、異なる辞書体系をもつ企業間の情報交換を促進するには、XML/XML 変換をできるだけ機械的に支援しやすくするように、Gruber の定義でいう「形式性」や「明示性」が重要になる。

(2) オントロジのレベル

[2]では、知識ベースの立場から、「人工システムを構築する際のビルディングブロックとして用いられる基本概念/語彙の体系(理論)」と定義し、オントロジを、問題解決固有のタスクに関わるタスクオントロジと、タスクが実行される領域(ドメイン)に関わるドメインオントロジの2種類に分類している。そして、計算論的な意味付けの観点から、以下に示す3つのレベルを設定している。

(i) レベル 1

レベル 1 は、概念階層の記述である。オントロジの基本的な機能は、対象世界に存在する概念の切り出しと、それらの関係の記述なので、概念階層の記述は、最もプリミティブなオントロジにあたる。

(ii) レベル 2

レベル2は、概念の意味定義(制約)や関係の記述(公理的記述)である。

(iii) レベル3

レベル3は、オントロジを用いて記述されたものが実行されるときの振る舞いに関する記述である。

XMLを用いたECにおける辞書は、主にレベル1を扱っているが、分類体系毎の属性や属性値の型を記述しているという意味で、レベル2に一部ふみこんでいるといえる。セマンティック Web が目標とする複数 Web サイト連携の自動サービスを実現するには、概念や属性間の制約などのより高度なレベル2の記述や、サービスの手順や前提条件などのレベル3の記述が必要になってくる。

3.2 ECにおけるオントロジへの期待

ECからみたオントロジに期待する効果としては、例えば以下のものがある。

- ・商品分類や商品属性のオントロジ記述により、カタログに含まれない自由な用語による商品検索が実現できる。
- ・利用者の嗜好のオントロジ記述により、利用者の要求にあった商品検索や情報配信が実現できる。
- ・オントロジ相互変換の自動化ないし半自動化により、異なるオントロジ標準で記述された商品カタログの比較検索システムが実現できる。
- ・複数 Web サイトのコンテンツを連携させるような自動サービスが実現できる。

4 オントロジ記述言語

本章では、セマンティック Web におけるオントロジ記述言語の現状と、今後の課題について述べる。

4.1 セマンティック Web におけるオントロジ記述言語の現状

図2に、Webコンテンツを記述するための言語の階層図を示す。セマンティック Web では、機械による自動処理をさらに進めるために、Webコンテンツに意味情報を付与するための言語として、XMLの上に、RDF(Resource Description Framework)、RDF Schema(図中ではRDFS)、DAML+OIL、そしてDAML-Sを提案している。以下、順に簡単に説明する(より詳細な解説は、[3]に譲る)。

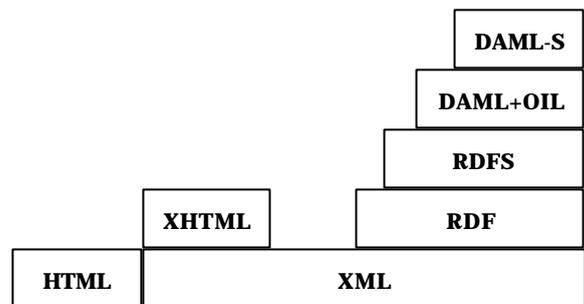


図2 Webコンテンツ記述言語の階層

4.1.1 RDF

RDF(<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>)は、データに対するデータ、すなわちメタデータを記述するためのW3Cの標準である。RDFでは、有向グラフによるデータモデルを用いて、メタデータをリソース、プロパティ(属性に相当する)、および値(属性値に相当する)と呼ぶ3つのデータの組で表現する。例えば、「Aさんは、ホームページBの作者(Creator)である」という言明の場合は、「ホームページB」がリソースに対応し、「Creator」の部分がプロパティに対応し、そして「Aさん」の部分が値に対応する。

4.1.2 RDF Schema

RDF Schema(Recommendation, <http://www.w3.org/TR/2000/CR-rdf-schema-20000327>)は、特定のプロパティをもつリソースの集まりを表現する概念(クラスと呼ぶ)や、プロパティの値のとりうる値に対する制約を記述する言語である。3.1節で述べたレベル1のオントロジを記述することができ、現時点(2001/10/13)では、W3Cの勧告候補になっている。

RDF Schemaでは、例えば、「乗用車やトラックは、自動車的一种である」といった概念間の包含関係や、「自動車は、乗員や重量というプロパティをもつ」といった概念がもつ属性を表現することができる。

4.1.3 DAML+OIL

DAML+OIL(<http://www.daml.org/2001/03/daml+oil-index.html>)は、DAMLプロジェクトで開発されていたオントロジ記述言語DAML-Oと、On-to-Knowledgeプロジェクトで開発されていたOILを統合した言語である。現時点(2001/10/13)での最新版は、2001年3月の版(Ver4.1)である。開発途上の言語であるが、レベル2のオントロジ記述を目的としているといえる。

DAML+OIL は、RDFS をベースとしてクラスやプロパティに対する制約の記述能力をさらに向上させた言語であり、言語の数学的基盤である意味論を明確に定義していることを特徴としている。Ver4.1 では、RDF Schema に対して、「クラスのプール結合演算や排他的関係」や「プロパティに対する制約の拡張」などの記述能力の拡張がなされている。今後は、推論ルールを表現できるように、プロパティの値間の関係を規定する条件文や、変数や限量子(「すべて」や「ある」などの変数の束縛に関する制約)を表現できるように拡張される予定である。

DAML+OIL の Ver4.1 では、例えば、「中古車とは、自動車かつ中古商品という意味である」とか、「自動車にはタイヤがあり、数は4つである」といった概念や属性に対する制約を記述できる。今後は、「自動車の重量に応じて、税金が異なる」といった推論ルールを記述できるように拡張されることが期待される。

4.1.4 DAML-S

DAML-S(<http://www.daml.org/services/daml-s>) は、Web 上のサービスを実現するために必要なオントロジを記述するための言語である。現時点(2001/10/13)での最新版は、2001年5月の Ver.0.5 である。開発途上の言語であるが、レベル3のオントロジ記述を目的としているといえる。

DAML-S では、WEB 上のサービスの発見、起動、構成、結合、実行、およびモニタリングなどのタスクを自動化するために必要な情報を、サービス・プロファイル(Service Profile)、サービス・モデル(Service Model)、およびサービス・グラウンディング(Service Grounding)の3つに分けて記述する。

DAML-S では、例えば、「車の購入は、車種の選定と、支払からなる」といったサービスを構成する手続きや、「点検の無料サービスを受けるためには、会員のメンバでなければならない」とか、「車を購入すると、自動的に会員メンバに登録される」といった手続きの前提条件や結果などを記述できる。

4.2 今後の課題

オントロジ記述言語は、セマンティック Web を支える中核技術の一つであるが、3.2 節で述べたようにその期待が大きいだけに、解決しなければならない課題も必ずしも簡単なものではない。以下、今後の課題として、「オントロジの作成」と「オントロジの質」について述べる。

(1) オントロジの作成

セマンティック Web の応用メリットを享受するには、膨大なオントロジ記述が必要と思われるが、それをどのようにして作っていくかという課題がある。また、セマンティック Web が目標とする複数オントロジ並存の情報交換を実現するためには、単純なオントロジ作成エディタだけでなく、オントロジの相互変換、統合、および改版などを容易にするようなオントロジ構築の方法論が必要である。

(2) オントロジの質

セマンティック Web では、インターネット上に分散したオントロジを利用して、人間の複雑な要求に応えられる自動サービスを実現しようとしているが、オントロジの質はまちまちであり、お互いに論理的に矛盾している可能性もある。そこで、オープン、分散、そして動的に変化する Web の世界において、どのオントロジが信頼できるかを知る仕組み、また、互いに矛盾がある場合にそれを柔軟に処理する仕組みを確立する必要がある。

5 おわりに

本稿では、オントロジの必要性とセマンティック Web におけるオントロジ記述言語の現状について述べた。

セマンティック Web は、本格的な研究が始ってからまだ1年足らずということもあり、技術的に本当に実現可能か、さらに、実用技術として広く普及するようになるかの評価はまだ定まっていないが、情報氾濫を解決する次世代の Web ビジョンとして魅力的である。今後、セマンティック Web が普及するためには、HTML がそうであったように、RDF、RDF Schema、DAML+OIL、および DAML-S などの言語によるコンテンツを簡単に作成できるツールと、その作成手間を上回るメリットが利用者に直観できるようなアプリケーションが必要になるだろう。今後の動向が注目される。

【参考文献】

- [1] Dieter Fensel 他, OIL: An Ontology Infrastructure for the Semantic Web, IEEE Intelligent Systems MARCH/APRIL 2001 pp38-45 (2001)
- [2] 溝口 理一郎, オントロジ工学 -内容指向研究の基礎-, 1999年情報学シンポジウム pp 1-8(1999)
- [3] 今村 誠, セマンティック Web におけるオントロジ記述言語 -DAML+OIL と DAML-S-, セマンティック Web コンファレンス予稿集, pp63-72 (2001)