

# セマンティックメタデータ

— 技術情報／製品情報交換システムへの適用 —

三木 敬 松平 正樹  
大熊 好憲

近年、情報に対してその意味を記述したメタデータを利用した情報統合への取り組みが始まっている。これらの動向を紹介するとともに、実際に製品情報の属性情報、仕様情報、知識情報、検索者の視点などをメタデータとして付与し、利用者へのナビゲーション機能を強化したLSI製品情報統合ポータルを試作したので報告する。

## 意味情報による情報統合

インターネットの発達により大量の情報が流通する中、必要な情報を抽出、選択、統合するための仕組みが注目されている。情報統合システムにおいては、従来人手に頼っていた情報の抽出、選択、統合処理を情報システムで取り扱えるようにすると同時に、情報システムの可読性を高めること、すなわち人の思考の流れに沿う形で情報の受け渡しを行うことが要求される。これらの要求を実現するためには情報の意味（セマンティクス）をふまえた処理が必要になる。

意味情報による情報統合の仕組みは、データベース、既存文書、Web、情報システムなどの情報リソースに対して、その概要やキーワード、情報の提示方法、情報と情報の関係性などをメタデータとして付与し、このメタデータに基づいて情報統合に向けた機能拡張を順次積み上げていく設計手法が主流である。メタデータで記述した意味属性や情報間の関係性を考慮した処理を行うことにより、情報系システム開発で工数が膨らみかつ再利用が困難な部分、例えばサービスプロセスの逐一的記述や異種データへの変換テーブル設定、あるいは人手作業と連携させるためのインタフェース開発部分を減らすことができる。逆にこのような処理を含む図1のようなアプリケーションでは、意味情報による情報統合の効果が大きいと考えられる。

## セマンティックメタデータ

従来、情報の意味を取り扱うことは、いわゆる人工知能技術の範疇と見られ、情報システム系の技術者にはとっつきにくい一面があったが、現在では意味情報を記載す

- 情報検索系  
Webポータル、情報ポータル、ナレッジマネジメントシステム
- 情報管理系  
- 部品・設計情報管理、情報管理、文書管理、知識管理  
- 電子政府、情報公開システム  
- eラーニングシステム
- 情報交換・結合系  
Eコマース、EAI (Enterprise Application Integration)、エージェントシステム、Webサービス連携

図1 セマンティックメタデータの適用領域

るメタデータの仕様、すなわちセマンティックメタデータがXML上で規定されているので取り組みやすくなった。セマンティックメタデータの設計（エディタ）や解釈（クエリー機能）を扱う実用的なソフトモジュールやツールも開発されており、また現実問題として重要な相互運用性（別々に記述した意味情報の統合）を確保する仕組みもあり、十分に利用障壁は下がっているといえよう。以下、代表的なセマンティックメタデータ仕様であるRDFとXTMについて紹介する。

## RDFとSemantic Web

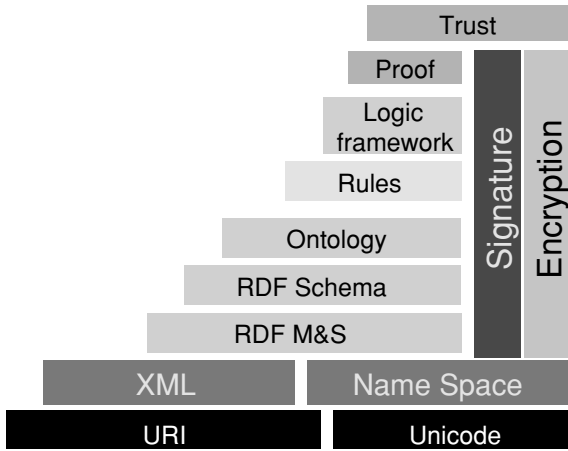
HTMLは主に人が読み理解するための文書の体裁を記述するメタデータ仕様であったのに対し、RDF (Resource Description Framework) は文書内容の意味情報を記述するメタデータ仕様である<sup>1)</sup>。RDFは主題と述語と対象（目的語）の三つ組みで構成される宣言文（言明）であり、ステートメントと呼ばれている。RDFの記述例を図2に示す。



```

<rdf:RDF>
<rdf:Description about="http://www.oki.com/">
<s:Publisher>OKI Electric Industry Co., Ltd.</s:Publisher>
</rdf:Description>
</rdf:RDF>
  
```

図2 RDF記述例



WWW2002,W3C trackより引用

図3 Semantic Webの階層構造図

RDFの設計思想は意味情報を機械で処理することを志向しており、処理そのものには曖昧性がないように設計されている。しかし意味自体には曖昧性、多様性、階層性があり、これら意味情報に特有な性質を取り扱う仕組みとして、図3に示される階層構造のフレームワークが検討されている。これらのフレームワークにより、Web自体を巨大な知識システムとして取り扱えるようにしようというのがSemantic Webのビジョンである<sup>2)</sup>。

現在、RDFで記述したメタデータがどのような意味を持つかを規定するRDF Schema、その上位の機構としてOWL (Web Ontology Language) の標準化まで進んでいる<sup>3)</sup>。OWLは一階記述論理をベースに、意味を機械で推論する処理を担う。今後、規定と規定の関係性を元に推論する機構 (論理表現層:Logic Framework)、関係性の正当性や矛盾を取り扱う証明層 (Proof)、情報の信頼関係を扱う信頼層 (Trust) が精力的に検討されるであろう。Semantic Webフレームワークのどの層まで取り込むかはアプリケーションの目的、将来の拡張性を考慮して決める必要がある。

### Topic MapsとXTM

Topic Mapsは利用者の持つ概念体系に合わせて情報を分類、整理するための手法である。元来はSGML (Standard Generalized Markup Language) の活動の中で、さまざまな文書のインデキシングとインデックスのマージをしたいという考えがあり、これを発展させ、文書のインデックスから概念の記述とそのインデキシングにまで広げた仕様がTopic Mapsである。2001年1月にISO13250規格として標準化された。図4がTopic Mapsの概念図である。人間が認識する概念/主題をコン

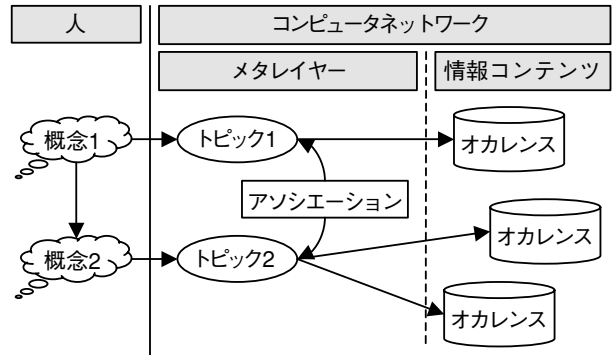


図4 Topic Mapの概念図

(内藤,「トピックマップとPublished Subject」<sup>5)</sup> 図1を元に作成)

ピュータ上で具体化したトピック、個々のトピック間の関連を示すアソシエーション、実際の情報コンテンツであるオカレンスから構成される。具体的な記述に関しては、Topic MapsをXMLのシンタックスによって記述するXML Topic Maps (XTM) の仕様がTopicMaps.Orgにより公開されている<sup>4)</sup>。

### セマンティックメタデータの社内システム適用

セマンティックメタデータをイントラネットポータルに応用した例を紹介する。弊社の場合にもさまざまな製品群と多くの関連文書を保有しており、特に製品種類が多いのはLSI製品である。LSI製品情報は社内技術者および社内営業向けイントラネット、および顧客向けインターネット上に多様な形式で存在しており、それぞれ「搭載製品一覧」や「価格一覧表」といった大きい情報のかたまりで掲載されているので、特定製品のデータシート (技術仕様) からその製品の搭載製品や価格を知りたい場合には、現状では全体の情報から辿らなければならない。この問題の背景には、製品情報が拡大するにつれ、当初設定した分類との不整合が生じ、その不整合を簡単に修正する仕組みがないために、大きな情報のかたまりのまま掲載することが多くなったと考えられる。

### LSI製品情報の分析

今回、LSI製品情報とその情報格納形式を調査するとともに、各情報の作成ワークフローについて設計部門、企画部門、営業部門にヒアリングを実施した。その結果、図5のように多様な情報を各部門が中心となって作成していることが判明した。情報格納形式に関しても、部門ごとに異なる形式、通常の文書ファイルやPDFあるいはWebページ (HTML) で蓄積されているケースが少なくない。これら部門所有の情報の項目表記を統合し、データベース等で一括管理することは膨大な手間とコストがかかり現

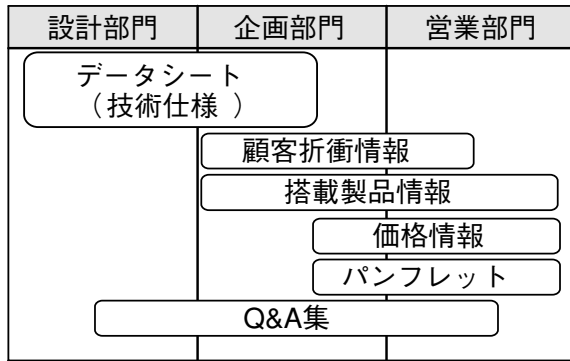


図5 LSI製品に関する情報管理部門

実的でない。そこで製品に関する知識の記述、および検索者の視点を表す知識（たとえば、営業員は製品情報で在庫・納期・価格・顧客折衝情報を重視し、設計者は製品スペックを重視するなど）を、部門所有のドキュメントの上位にセマンティックメタデータとして付与し、これらのメタデータを使って製品情報を再配置し、適切なナビゲーションを行う閲覧ページを自動生成することを目標とした。

メタデータ記述様式は、今回の活用目的が情報の閲覧であり、短期間にシステム構築とその評価が求められたこと、機械による推論や異分野との概念拡張・統合機能などが当面は求められていないことを勘案して、XML Topic Maps (XTM) を用いた。

## LSI製品情報システム

図6が開発したLSI製品情報システムの構成図である。LSI情報システムは、メタデータにより記述された製品に関するオントロジ (Ontology) 層およびナビゲーション

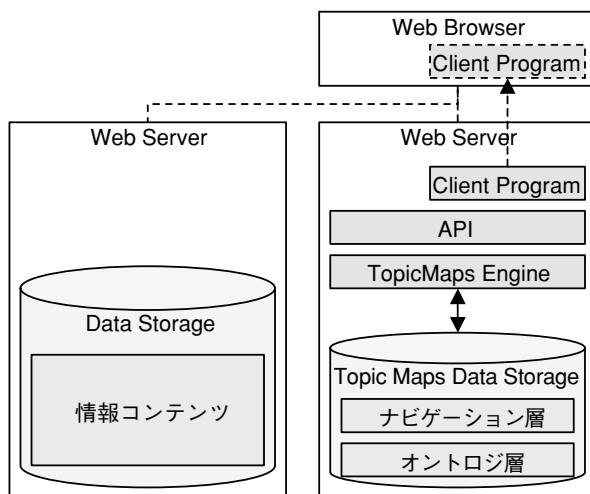


図6 LSI製品情報システム構成図

層（本稿では、両者を併せてLSI製品情報オントロジと呼ぶ）、情報コンテンツ、および、XTMを処理するエンジンとクライアント用プログラムから構成される。オントロジとは、セマンティックメタデータを階層的に記述する際によく使われる専門用語で、一般的には概念間の関係の定義とされているが、ここでは製品情報にかかわる用語の階層的定義の意味で用いる。

## LSI製品情報オントロジ

オントロジ層はLSI製品情報オントロジの中心をなす部分で、トピック、アソシエーション、オカレンスをXTMで記述したものである。構築プロセスを以下に示す。

- ①それぞれの情報コンテンツ内を体系化し、トピックおよびアソシエーションとして定義する
- ②体系間で関係のあるトピックをアソシエーションとして定義し、同等のトピックを統合する
- ③必要に応じて関連するオントロジを再利用・構築し、トピックおよびアソシエーションとして定義する  
またコンテンツ全般について、

- 情報コンテンツはXML化する以外は変更しない。

- 汎用的なオントロジを部分的に再利用する

という2つの方針によって構築した。しかし、後者については、LSIの属性、および導入事例の製品分類として利用しようとしたRosettaNetの辞書<sup>6)</sup>や総務省の日本標準商品分類は、情報コンテンツとの項目の過不足や一意に変換できないといった問題があり、独自にオントロジを構築せざるを得なかった。また、例えば録音再生LSIには「録音時間」という属性があり、音源LSIには「音源種類数」という属性があるように、LSI製品の分類によってトピックが異なる。そのため、オントロジを製品一般、LSI製品一般、音声LSI等の階層ごとに定義し、統合して利用できるようにしている。

オントロジ層の上位には、利用者の視点別に必要な情報を提供するためのナビゲーション層を設けた。ヒアリングによって得られた要求をもとに、製品別視点や用途別視点等を定義し、それぞれの視点で必要な情報とその関係性をトピックおよびアソシエーションで定義した。

## 情報コンテンツ

実際の情報コンテンツは、オントロジ層のトピックにあわせてXML化した。

- ④情報をトピック単位に分割してXML化する
- ⑤XMLデータを特定するURI (Uniform Resource Identifiers) をオントロジ層のオカレンスとして指定する

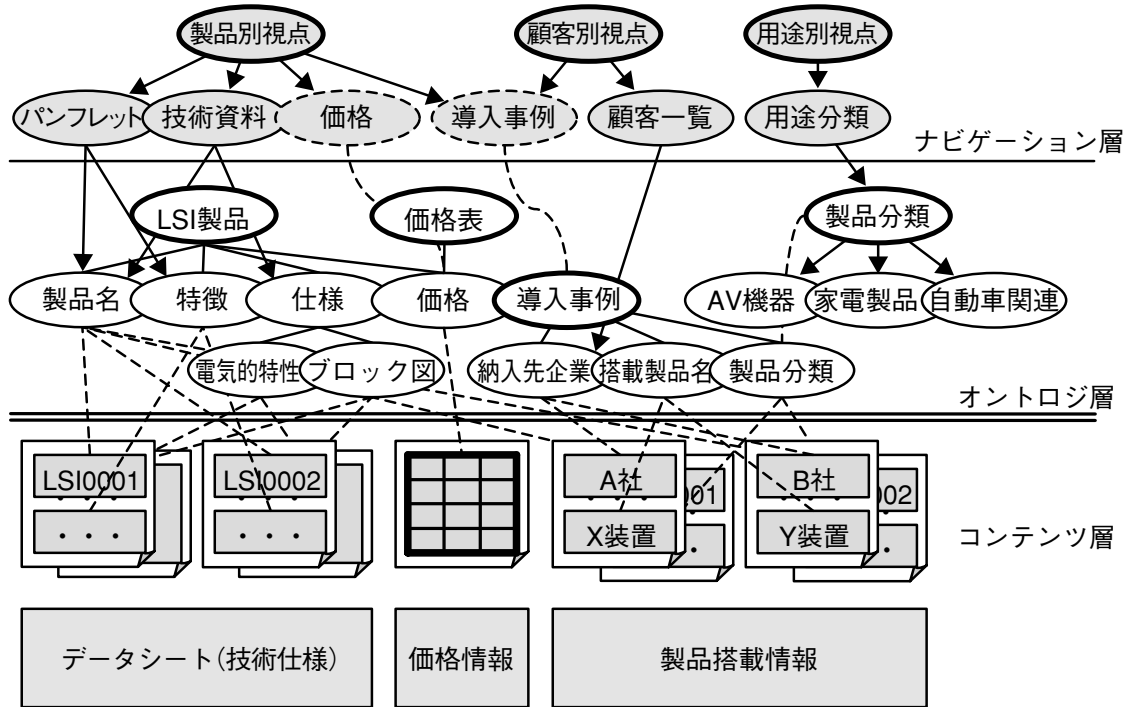


図7 LSI製品情報オントロジの概念図

図7が構築したLSI製品情報オントロジの概念図である。最上位がナビゲーション層，中間層がオントロジ層である。最下層がオカレンスである情報コンテンツである。

### XTM処理エンジンとクライアントプログラム

XTMを処理するエンジンを独自に開発し，そのAPI (Application Program Interface) を用いてクライアント側で情報を閲覧するJavaScriptプログラムを開発した。APIはトピックの追加，削除およびアソシエーションの編集機能を提供しており，利用者が容易に情報を追加したり，視点を編集したりすることができる。

### まとめ

セマンティックメタデータを用いたLSI製品情報システムを試作した。試作システムでは，必要な情報を得るためにあちこちのページを見て回る操作が少なくなり，ユーザの視点に合わせた分類で情報が提示される，製品ごとの情報がまとめて見られるなど，直感的で分かり易いと好評である。課題としては，情報コンテンツとオントロジの対応付けを各部門の従来のワークフロー内でどのように実現するかといった運用面での詰めが残されている。

今後は，情報コンテンツとオントロジの対応付けのためのツールを開発し，構築したシステムを他の製品に展

開するとともに，LSI製品情報オントロジとRosettaNet等の企業間情報交換の標準化仕様との相互運用性を確保する仕組みを整備していく予定である。◆◆

### 参考文献

- 1) W3C, Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- 2) INTAP, 平成14年度セマンティックWeb技術の調査報告書, 2003年
- 3) W3C, OWL Web Ontology Language Reference, <http://www.w3.org/TR/owl-ref/>
- 4) TopicMaps.Org, XML Topic Maps (XTM) 1.0, <http://www.topicmaps.org/xtm/1.0/>
- 5) 内藤, トピックマップとPublished Subject, 人工知能学会第1回セマンティックウェブとオントロジ研究会, 2002年
- 6) RosettaNet, RosettaNet Technical Dictionary (RSTD), <http://www.rosettanet.org/>

### 筆者紹介

三木敬 : Kei Miki. 研究開発本部 ヒューマンインタフェースラボラトリ プロデューサ  
 松平正樹 : Masaki Matsudaira. 研究開発本部 ヒューマンインタフェースラボラトリ  
 大熊好憲 : Yoshinori Ohkuma. ファイナンシャルソリューションカンパニー ITインキュベーション本部