

ていない<sup>(10)</sup>。そのため、履行されるか確信がもてないアーカイバルアクセス条項を、契約時に購読者はさほど重視しない(したくても、できない)。

こうした事情から言えるのは、電子ジャーナルの十全なアーカイビング体制は二重の意味で重要である、ということだ。コンテンツの保存とアクセスを保障することは、出版社の不測の事態に備えるためだけでなく、アーカイバルアクセスが購読契約の際に真剣に協議される前提としても必要なのである。各事業の進展に合わせ、「電子ジャーナルのアーカイビング」をレビューすることは今後も欠かせないが、その際は、そうした二つの重要性を踏まえた上で書かれる必要があるように思われる。

(青森中央短期大学：<sup>こととしゆき</sup>後藤 敏行)

- (1) Portico. "Why Archive?". (online), available from <<http://www.portico.org/about/why.html>>, (accessed 2006-03-12).
  - (2) Steenbakkers, Johan F. Digital Archiving: A Necessary Evil or New Opportunity? *Serials Review*. 30(1), 2004, 29-32.
  - (3) Koninklijke Bibliotheek. "The current state of production". (online), available from <<http://www.kb.nl/dnp/e-depot/dm/standvanzaken-en.html>>, (accessed 2006-03-12).
  - (4) Koninklijke Bibliotheek. "National libraries of the Netherlands and Japan sign joint operating agreement". (online), available from <<http://www.kb.nl/nieuws/2005/japan-en.html>>, (accessed 2006-03-12).
  - (5) Elsevier. "Elsevier Acts to Safeguard E-Journals". (online), available from <[http://www.elsevier.com/wps/find/authored\\_newsitem.cws\\_home/companynews05\\_00370](http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_00370)>, (accessed 2006-03-12).
  - (6) LOCKSS. "About LOCKSS". (online), available from <<http://lockss.stanford.edu/about/about.htm>>, (accessed 2006-03-12).
  - (7) Case, Mary M. A Snapshot in Time: ARL Libraries and Electronic Journal Resources. *ARL Bimonthly Report*. (235), 2004. (online), available from <<http://www.arl.org/newsltr/235/snapshot.html>>, (accessed 2006-03-12).
  - (8) Watson, Jennifer. You Get What You Pay for? Archival Access to Electronic Journals. *Serials Review*. 31(3), 2005, 200-205.
  - (9) *Ibid.*
  - (10) Lewis, Nicholas. 'Are we burning our boats?' Survey on moving to electronic-only. (online), available from <[http://www.sconul.ac.uk/pubs\\_stats/newsletter/31/19.rtf](http://www.sconul.ac.uk/pubs_stats/newsletter/31/19.rtf)>, (accessed 2006-03-12).
- Ref: Koninklijke Bibliotheek. (online), available from <<http://www.kb.nl/>>, (accessed 2006-03-12).  
 Portico. (online), available from <<http://www.portico.org/>>, (accessed 2006-03-12).  
 LOCKSS. (online), available from <<http://lockss.stanford.edu/>>, (accessed 2006-03-12).  
 後藤敏行. 電子ジャーナルのアーカイブ. アクセス

の観点からみた集中・分散の2方面戦略 . 情報管理. 48(8), 2005, 509-520.

CA1598

## 動向レビュー

### ウェブ・オントロジーの可能性と図書館

#### 1. はじめに

ウェブ上の情報を組織化して活用するというWWWが設計された当初からの夢<sup>(1)</sup>は、現在「セマンティック・ウェブ」(CA1534 参照)と呼ばれて要素技術の標準化が進められています。その核となるウェブのオントロジー(ontology)は、ばらばらに記述、蓄積された知識の連動を重視しており、電子図書館どうし、あるいは図書館と関連コミュニティ<sup>(\*)</sup>との連携、さらに書誌情報にとどまらない豊富なウェブ資源との統合といった面での応用が期待できます。

(\*)ここでいうコミュニティは、地域社会に限らず、同じ関心領域や活動を共有する集合全般を指します。

#### 2. セマンティック・ウェブとオントロジー

##### 2.1. オントロジーとその働き

オントロジーとは、知識を共通の認識に基づいて体系化、形式化し、計算機で扱うことができるように記述したものを指します。本来オントロジーは、ものが「存在する」ことを問う哲学を意味するわけですが、人工知能や知識表現分野でのオントロジーも、対象とする領域に存在するものごととその関係を、概念(クラス)の階層を中心に体系立てて記述し、共通理解の基盤とするものです。たとえば、遺伝子研究に不可欠のオントロジーである Gene Ontology では、“Molecular Function”(分子の作用)“Biological Process”(生物学的プロセス)“Cellular Component”(細胞の構成要素)という3つのクラスから出発して、2万に及ぶ概念がクラスツリーとして体系化されています。これらの概念とその記述法をオントロジーの利用者が共有することで、対象世界の知識について語ったり調べたりするときに食い違いが無くなり、コンピュータによる推論などの知的な処理が可能になるわけです。

オントロジーには、哲学的な視点で世界を捉えようとするものや高度な論理操作の基盤となる本格的なものから、対象領域の語彙を整理したというレベルの「軽量オントロジー」まで、様々なものがあります。「ページからページを駆け巡るエージェントがユーザに代わって高度なタスクを実行する」<sup>(2)</sup>といった世界の実現

には、精緻なオントロジーと論理規則が求められるため、まだ今後の研究開発が必要です。しかし、データベースの横断検索や情報の自動交換といったコミュニティ間での知識の流通には、語彙体系の共有だけでも力を発揮します。

2.2. セマンティック・ウェブの技術スタック

セマンティック・ウェブとは、ウェブ上の文書だけでなく様々なデータをコンピュータの力でうまく利用できるようにしよう (Web of data) という試みで、オントロジーはその核となる要素です。ただしウェブでの知識や情報は、断片的で不完全なものしか得られず、さまざまなコミュニティで独自に体系化され、また素朴に信頼できるとは限りません。これらを踏まえて、モジュール化した技術を積み重ね、柔軟で現実的なアプローチで「データのウェブ」に取り組みようとしています<sup>(3)</sup>。

図1 セマンティック・ウェブの技術スタック：出典(4)

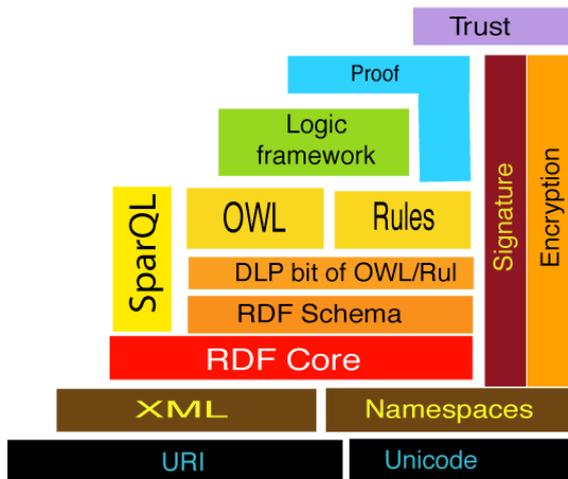


図1は、セマンティック・ウェブの技術アプローチを説明するためにしばしば引用されるものです。一番下にはウェブ全体を支える技術であるURI (Uniform Resource Identifier :ウェブ上のリソースを特定する識別子)とUnicodeが置かれ、その基盤の上に「データのウェブ」実現に必要な技術が、汎用的なデータフォーマットから応用的な記述言語、論理表現という順で積み重ねられています。

この中で、オントロジーを記述し知識を形式化するための技術は、RDF (Resource Description Framework) および RDF スキーマ (RDFS) から OWL (Web Ontology Language) までの層です。オントロジーの基本であるクラス階層は RDFS で記述され、クラスを定義する制約条件や他のオントロジーとの関

連を記述する語彙は OWL が提供します。

この横に位置づけられている SPARQL は、RDF のデータを検索するためのクエリ (問い合わせ) 言語とプロトコル (やり取りの方法) です。Rule, Logic といった層と合わせ、オントロジーに従って記述、蓄積したデータから有益な情報を取り出す方法を標準化しようとしています。

また、コンピュータに推論を任せようというセマンティック・ウェブにとっては、ネットワークで流通する情報の改竄やなりすまし、あるいは情報そのものの信頼性は非常に大きな課題です。この部分を、図1の右側で全体を貫く Signature, Encryption 層と、Logic の上に置かれる Proof 層でカバーし、最上位に位置する Trust を得ようというのが、ここで描かれているモデルです。

これらの技術のうち、OWL 以下の層が 2004 年 2 月までに標準化されました。SPARQL は現在討議が行われている最中で、順調に進めば 2006 年中には W3C の勧告となると思われます。続く推論規則 (Rules 層) については、2005 年末に W3C に RIF (Rule Interchange Format) 作業部会が設置され、2006 年 3 月にその要件とユースケース (応用の具体例を挙げて機能を説明したもの) を提示する初の草案が公開されました。また、言語仕様の標準化と並行して、オントロジーの相互運用性を高めるためのガイドラインなどを検討する Semantic Web Best Practices and Deployment 作業部会が活動しています。

2.3. オントロジーの共有とマネジメント

中央集中型のコントロールが行われないウェブにおいては、類似の概念が各地でばらばらに記述されることが避けられません。セマンティック・ウェブでは、資源名 (識別子) や語彙 (クラス, プロパティ) を全て URI として表現することで、分散記述を可能にしつつ名前の衝突や意味の混乱を避けるようにしています。

たとえば「Opera」という単語が、コミュニティ A では歌劇、コミュニティ B ではブラウザの種類として用いられている場合、それぞれを a:Opera, b:Opera と表すことで両者を曖昧さなく区別できます (本稿では語彙の名前空間 URI を a:, b:などの接頭辞で略記します)。同様に、title というプロパティも dc:title, foaf:title と表現すれば、前者は「表題」、後者は「敬称」と使い分けが可能になります。

語彙の連動のためには、違いを区別するだけでなく、類似の概念を束ねる機能も必要です。一方が他方の下位概念であるときは、RDFS で定義されている

subClassOf, subPropertyOf で、同一であるならば OWL で定義されている equivalentClass, equivalentProperty を用いて関係を示します。コミュニティ C が作曲家を表すためにダブリン・コアの creator よりも詳細な c:composer という語彙を使っているとすれば、

```
c:composer rdfs:subPropertyOf dc:creator .
```

と表現できるでしょう。コミュニティ D で歌劇を表すのに「Oper」を用いているときは、

```
a:Opera owl:equivalentClass d:Oper .
```

で両者が同じ実体集合を持つクラスであると示すことができます。概念レベルではなく個別の資源実体が同じである場合には、owl:sameAs によって同一性を記述します。

また、対象領域の知識の変化に対応した改訂など、オントロジーにもマネジメントは欠かせません。OWL には、priorVersion, DeprecatedClass などのバージョン管理語彙が用意されており、既存のデータを作り直すことなく、改訂版への移行や統合を進められるようになっています。

RDF/OWL を処理するプログラムは、通常こうした基本的な包含、同一関係を利用して推論を行うためのライブラリ(プログラム部品のパッケージ)を備えています。異なるコミュニティが独自に開発した語彙であっても、オントロジーによるマッピングを行えば、標準的なツールを使って連携させることが可能になるのです。

### 3. セマンティック・ウェブと図書館

#### 3.1. オントロジーによる電子図書館の協働

図書館のメタデータ交換は、たとえば OAI-PMH (CA1513 参照)によってすでに実践されていますが、これは書誌情報の最大公約数をシンプルなダブリン・コアに置き換えて提供するというもので、キャンベル (Grant Campbell) らが「図書館は“本の番人”、インターネットは“情報ハイウェイ”という古い図式」と述べるように<sup>(5)</sup>、蓄積した豊かな情報を十分生かすまでには至っていません。それぞれの持つデータをより有効に活用するため、RDF/OWL を応用して相互運用性を高める試みが行われています。

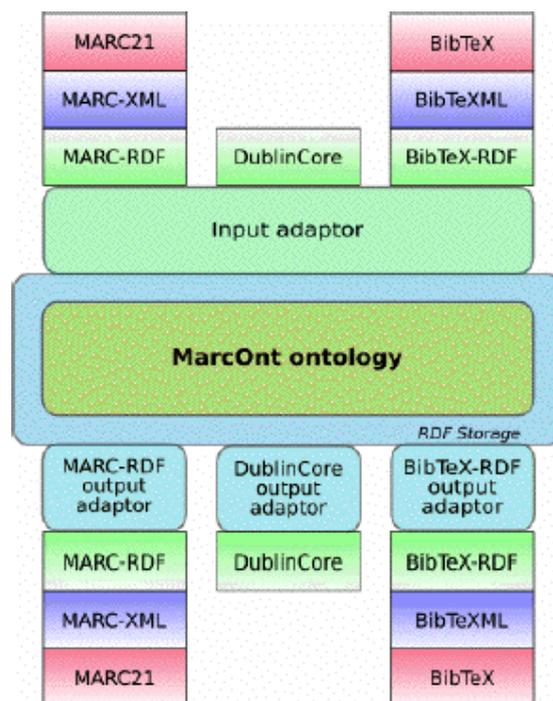
DSpace (CA1527 参照)の発展形である SIMILE (Semantic Interoperability of Metadata and Information in unlike Environment) プロジェクトは、RDF を用いてメタデータの記述方法を拡張可能にし、

異なるコミュニティで構築されたレポジトリを取り込んだり連動させることを目指しています。同プロジェクトのバトラー (Mark Butler) らは、XML で記述されたスキーマを RDF に変換してデータを相互に結びつける、人物名などの固有名は OCLC の典拠レコードを用いて同定する、統制語彙は SKOS<sup>(6)</sup>を用いてオントロジーに組み込む、関連情報を Wikipedia から取得するなどの方法で、2つのメタデータ群を統合し横断検索を可能にするデモを行いました。その結果、文字列データから URI への変換などでの難しさが浮き彫りになりましたが、連携コンセプトの実証としては一定の成果が上がったことが報告されています<sup>(7)</sup>。

同様の試みは、EU の文化遺産に関する電子図書館の協働プロジェクトである BRICKS (Building Resources for Integrated Cultural Knowledge Services)でも行われています。ここでは、異なる電子図書館のデータを収集して OWL のオントロジーにマッピングし、横断検索を提供する実験が行われ、柔軟で低コストなメタデータのマネジメントが可能になることが確認されました<sup>(8)</sup>。

北アイルランド大学の DERI (Digital Enterprise Research Institute)では、ダブリン・コア、MARC21、BibTex といったフォーマットを連動させるオントロジー MarcOnt の構築が進められています<sup>(9)</sup>。これはまだ開発途上ということですが、うまく実現すれば、電子図書館だけでなく通常の図書館も含めた幅広い書誌

図2 MarcOnt の変換モデル：出典(9)



データの共通言語としての働きも期待できるでしょう。

これらはいずれも、個別のメタデータ体系を XML 経由で RDF/OWL に変換し、推論エンジンが共通オントロジーのデータを生成、追加するという手法を採っています (MarcOnt の場合は、固有の推論規則も加えた形でオントロジーへの変換を行っています)。このアプローチは、XSLT (CA1552 参照) などの汎用ツールが利用できて効率がよい上に、オリジナルの持つ豊富な情報が失われません。セマンティック・ウェブの技術をメタデータの連動に適用する出発点として、有効なモデルのひとつと言ってよいでしょう。これに SPARQL, RIF などの技術が標準化されて加われば、いっそう柔軟で効果的なメタデータの共有が期待できます。

### 3.2. 書誌メタデータを越えて

セマンティック・ウェブでは、文書や画像などの資源だけでなく、人間、自動車から思想、芸術やサービス、さらに興味や知人といった関係性までを「リソース」として記述します。キャンベルらは、この表現力とウェブの膨大な情報資源を取り込むことで、図書館のサービスが大きく変貌するというシナリオを描いています<sup>(10)</sup>。

たとえば、よく知られた「軽量オントロジー」である FOAF (Friend of a Friend) が記述するデータは、人の様々な関心事や知人関係などです。クルーク (Sebastian R. Kruk) らは、この FOAF と電子図書館のメタデータを組み合わせ、協調フィルタリングやパーソナル図書館を通じてより適切な資源情報をユーザに提供するという試みを行ってきました<sup>(11)</sup>。FOAF の応用は、図書館ごとに個人プロフィールを登録する必要がない、知人の関心事項までを踏まえた資源をリストアップできる、プライバシーや信頼度のバランスを取る技術を利用できる、などの利点を生むことが報告されています。これを発展させれば、FOAF だけでなくイベントカレンダーや位置情報などの多様な RDF メタデータを組み込んだ、現実世界と連動したサービスも考えることができるでしょう。

セマンティック・ウェブの技術は、独立したシステムの間でも、また書誌情報と個人プロフィールといった質や粒度の異なるデータ間でも、事前の取り決めなしに連携を可能にするところに特徴があり、大きな潜在力が秘められています。そして、これらを結びつける上でのキーとなるのが、オントロジーという共通項なのです。

( www.kanzaki.com : かんざきまさひで 神崎 正英 )

- (1) Berners-Lee, Tim. Information Management: A Proposal. 1989. (online), available from <http://www.w3.org/History/1989/proposal.html>, (accessed 2006-03-31).
- (2) Berners-Lee, Tim et al. The Semantic Web. Scientific American. 284(5), 2001, 34-44. (online), available from <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809-EC588EF21>, (accessed 2006-04-25).
- (3) Miller, Eric. Digital Libraries and the Semantic Web. 2001. (online), available from <http://www.w3.org/2001/09/06-ecdl/>, (accessed 2006-03-31).
- (4) Berners-Lee, Tim. WWW at 15 years: looking forward. 2005. (online), available from <http://www.w3.org/2005/Talks/0511-keynote-tbl/>, (accessed 2006-03-31).
- (5) Campbell, Grant et al. Academic Libraries and the Semantic Web: What the Future May Hold for Research-Supporting Library Catalogues. The Journal of Academic Librarianship. 30(5), 2004, 382-390.
- (6) Simple Knowledge Organisation System (SKOS). (online), available from <http://www.w3.org/2004/02/skos/>, (accessed 2006-03-31).
- (7) Butler, Mark H. et al. Data conversion, extraction and record linkage using XML and RDF tools in Project SIMILE. 2004-08-31. (online), available from <http://www.hpl.hp.com/techreports/2004/HPL-2004-147.html>, (accessed 2006-03-31).
- (8) Haslhofer, Bernhard et al. Metadata Management in a Heterogeneous Digital Library. 2005. (online), available from <http://www.brickcommunity.org/discussion\_area/papers/paper.2006-01-05.5885506496/file/>, (accessed 2006-03-31).
- (9) Kruk, Sebastian Ryszard et al. MarcOnt - Integration Ontology for Bibliographic Description Formats. 2005. (online), available from <http://www.marcont.org/marcont/pdf/DC2005skmskz.pdf>, (accessed 2006-03-31).
- (10) Campbell, Grant et al. *op. cit.*, (6).
- (11) Kruk, Sebastian Ryszard et al. FOAFRealm: Making Social Collaborative Filtering Real. 2005. (online), available from <http://www.marcont.org/marcont/pdf/eswc2005\_foafrealm.pdf>, (accessed 2006-03-31).