# セマンティック Web 技術と次世代電子政府での活用方法

に関する調査研究

調査報告書

平成14年3月

財団法人 ニューメディア開発協会

# 目 次

1		調査	の背景	景と目的	. 3
	1	. 1	背景.		. 3
	1	. 2	目的.		. 3
2		セマ	ンテ	ィック Web の研究開発・標準化の動向と社会的影響に関する調査	. 4
	2	. 1	セマ	ンティック Web の現状	. 4
	2	. 2	米欧(	か研究動向	. 7
		2.2	2.1	米国の研究動向	. 7
		2.2	2.2	欧州のセマンティック Web 動向	15
		2.2	2.3	欧州の電子政府におけるメタデータ応用の動向	19
		2.2	2.4	欧州におけるオントロジー技術の動向	27
		2.2	2.5	欧州におけるセマンティック Web 関連システムの動向	31
	2	. 3	標準値	と機関の動向	33
		2.3	3 . 1	W3C O Semantic Web Activity	33
		2.3	3.2	WOWG のオントロジー記述言語の動向	33
		2.3	3.3	OWL の利用事例	37
	2	. 4	活用	事例と課題	43
		2 . 4	1.1	ナレッジマネジメント	43
		2.4	1.2	e-learning	45
		2.4	1.3	e-ビジネス	49
		2.4	1.4	コミュニティー支援	50
		2.4	1.5	課題	52
	2	. 5	応用:	システム開発時の考慮点	54
	2	. 6	社会I	こおよぼす影響	56
3				子政府での活用方法に関する調査研究	
	3			ンティック情報(メタデータ)の付与に関する調査	
				国内の行政公開情報の調査	
		3 . 1	1.2	海外政府のメタデータ付与に関する調査	71
		3 . 1	1.3	メタデータ階層モデル調査	73
				メタデータ定義調査	
				利用者プロファイルのメタデータ定義に関する調査	
	3	. 2	情報(	のセマンティック検索に関する調査	94
		3.2	2.1	メタデータ管理システム調査	94
				半自動メタデータ生成システム調査	
				対話型メタデータ生成システム調査10	
				利用者プロファイル定義システム調査1	
		3.2	2.5	セマンティック検索システム調査1	14

# 1.調査の背景と目的

#### 1.1 背景

来るべき高度情報処理社会のあり方を抜本的に変える可能性を秘めた新たな Web 技術の開発が、欧米において急速に進行中である。その技術は、セマンティック Web 技術である。

現在の Web 技術が、Web 情報を人間が読んで、人間が理解し、人間が操作することを前提にしているのに対し、セマンティック Web は、Web 情報をエージェントにより自動処理させることを目指している。

セマンティック Web の基本原理は、Web を含むあらゆるデータと情報の意味をマシンリーダブルなメタデータで記述し、人間の代わりをするソフトウェア(エージェント)で自動処理させることである。

セマンティック Web は、物理的な事物をも含めた膨大な情報、ハードウェア、ソフトウェア及び機能と言ったあらゆるものを記述可能とし、人間が簡単な指示を行なうだけで、エージェントにより、自律的且つ自動的に処理をすることを目指している。

セマンティック Web はインターネット上での単調で機械的な作業を自動処理してくれるので、経済活動の効率化と社会や家庭の利便性の向上とを飛躍的にもたらす。

米国や EU では、政府の財政的補助の下にセマンティック Web の研究開発プロジェクトが推進されており、その中でセマンティック Web の国際標準規格の検討も進められている。

このセマンティック Web の国際標準規格が、米国と欧州によって、我が国の事情を考慮することなく決められてしまう事のないように、我が国も早急にこうした研究活動に取り組み、その国際標準規格作りへのコミットメントを行なう必要がある。

また、セマンティック Web の研究開発プロジェクトの成果を先取りした新たなビジネスが、欧米では生まれつつある。、我が国の情報技術が世界をリードするためにも、セマンティック Web で拓ける新しい Web の世界への研究に注力することが肝要である。

# 1.2 目的

上記の背景を踏まえて、当調査研究では、平成 13 年度にセマンティック Web 技術の現状と動向とを調査すると共に、次世代の電子政府にセマンティック Web 技術を活用する視点から、セマンティック Web 技術の本質的な利点の究明、利点を生かす上での基盤技術、ツール技術、アプリケーション技術等について調査研究する。セマンティック Web 技術の実用化に向けた問題点や必要な技術要素の洗い出しを行ない、将来の利用に向けた課題を明らかにする。

# 2. セマンティック Web の研究開発・標準化の動向と社会的影響に

# 関する調査

# 2.1 セマンティック Web の現状

Web の発明により、インターネットを中心とする IT 革命が引き起こされた。この IT 革命により空間的・時間的制約が取り払われ、経済活動の効率化や新規産業の創出など経済社会構造の転換が生じてきている。

この IT 革命のキーとなった現在の Web 技術は、Web 上の情報を人間が目で見て、人間が理解し、人間が指示することを基本原理としている。このため人間が細かな指示を与えなければならないし、人間が理解できない場合は有効に活用することができない。

このような人間だけが理解できる情報をヒューマンリーダブルな情報というが、さらに Web の情報を、コンピュータなどの機械がその意味を理解できる情報(これをマシンリーダブルな情報という)にし、コンピュータで自動処理させる研究とその応用が、次世代 Web 技術として、W3C (World Wide Web Consortium)や DARPA(米国国防高等研究計画局)を中心に米欧で強力に推進されている。これがすなわちセマンティック Web である。

セマンティック Web の基本原理は、あらゆるデータと情報について、マシンリーダブルかつ構造化されたメタデータ<sup>1</sup>を付加することによってその意味を記述し、エージェント<sup>2</sup>で自動処理させることである。このように、コンピュータがエージェントなどを使って、自動的な意思決定を行なえるような情報空間のことをセマンティック Web という。

Web 技術の標準化団体である W3C の設立者である Tim Berners-Lee は、WWW 立ち上げ時からセマンティック Web を構想しており、2 年ほど前からはかなり具体的にセマンティック Web を提唱している。W3C では 2001 年 2 月に Semantic Web Activity を立ち上げており、現在の W3C の中心的な活動は、このセマンティク Web である。Tim Berners-Lee、James Hendler、及び Ora Lassila の言葉によれば、「セマンティック Web は現在の Web を拡張したものだ。Web に記述される情報に明確な意味の定義を与え、コンピュータと人間とがうまく協力して作業できるようにする。現在の Web 構造に意味関係を織り込んでいく初歩的な作業がすでに進みつつある。近い将来、今は画面に表示されているだけの情報をコンピュータがきちんと"理解"してもっと有効に利用し、非常に強力な機能を実現できるようになるだろう。」ということである。

すでに米国及び EU は調査開発に国家予算を当てており、例えば、米国は DAML プロジェクトに国家予算として 2000 年から 3 年間で 7000 万ドル (約 80 億円) の拠出をして いる。 EU 委員会は、On-To-Knowledge プロジェクトに 2 年半で 134 万ユーロ (約 1.5 億円)を、IBROW プロジェクトに 3 年間で 110 万ユーロ (約 1.2 億円) の拠出をしてい

<sup>1</sup> メタデータとは、データを記述するデータのことをいう。メタデータを記述するデータのことも、メタデータという。

<sup>2</sup> エージェントとは、人間の代わりをするソフトウェアのことをいう。

<sup>&</sup>lt;sup>3</sup> Tim Berners-Lee, James Hendler, Ora Lassila. "The Semantic Web." *Scientific American*, May 2001 より。日本語訳は「自分で推論する未来型ウェブ」日経サイエンス 2001 年 8 月号。

る。また、セマンティック Web に関する技術開発において米国と EU はジョイントプログラムをスタートさせ、オントロジー言語の DAML+OIL を作成した。米国では DARPA、MIT、スタンフォード大学、メリーランド大学が、EU ではフライエ大学(オランダ)、カールスルー工大学(ドイツ)、ブリストル大学(英国)、INRIA(フランス)などが中心となってこれらの研究開発を進めており、多くの成果を出している。

セマンティック Web は、W3C の策定した XML (eXtensible Markup Language) と RDF(Resource Description Framework)を基盤技術とする。Tim Berners-Lee は、2000 年 12 月の講演資料4において、セマンティック Web のアーキテクチャとして、図 2-1のようなレイヤー構造を提示した。

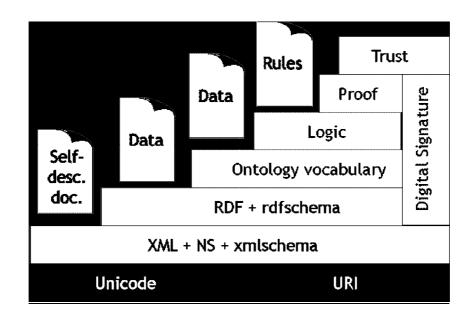


図 2-1 セマンティック Web のアーキテクチャ

以下に、主要なレイヤーについて概要説明する。

XML・・・Web 作成者等が<zip code>等のタグを自ら定義し、Web ページやテキストにタグを付加することにより、それらに注釈を付けることができる。XML によって Web 作成者は文書に任意の構造を付加できるが、その構造が何を意味するかについては XML は何も規定しない。

RDF・・・RDF は XML によって文書に付加された構造に対して、その意味を付加するものである。すなわち、メタデータを記述するための仕様である。RDF では、メタデータは図 2-2のような主語(リソース)、述語(プロパティ)、目的語(値)の 3 つ組(トリプル)で記述される。すなわち、主語: 「http://www.w3.org/Home/Lassila」、述語: 「Creator」、

<sup>4</sup> http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

目的語:「Ora Lassila」である。これを日本語の文章で述べると、「http://www.w3.org/Home/LassilaのCreator(作者)はOra Lassilaである」となる。

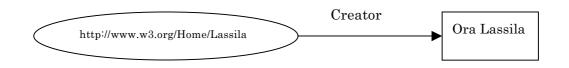


図 2-2 RDF の3つ組

オントロジー(Ontology)・・・メタデータの表現方法が異なる場合(すなわち、二つの語彙体系が同じ要素・概念に対して異なった名前を付けている場合)、意味が通じなくなる。複数の語彙体系間で意味の共有化を図るための、「用語間の関係を正式に定義している文書またはファイル」をオントロジーという。

論理(Logic)・・・エージェントはメタデータを処理することで Web ページの内容を理解し、人が発した複雑な質問に的確な回答を導き出すことができる。このような、メタデータに関する処理を「推論」と呼ぶ5。論理レイヤーは、推論を行なうための仕組みを提供する。セマンティック Web では、これまでの検索エンジンにおける and と or のみならず、三段論法や否定、述語論理で用いるような量化記号も取扱えるようになる。

6

 $<sup>^5</sup>$  萩野達也 「セマンティック Web の現状と課題」 データベースと Web 情報システムに関するシンポジウム論文集、 $p74 \ (2001)$ 。

# 2.2 米欧の研究動向

- 2.2.1 米国の研究動向
- 2.2.1.1 DAML (DARPA Agent Markup Language)

#### (1) DAML 概要

DAML はオントロジーを記述するための言語である。RDF スキーマをベースに拡張が 施されている。

DARPA から 7000 万ドル (5 年間) 6の資金提供を受けている。MIT (W3C) を含む複数のプロジェクトからなる。開発は 2000 年から開始され、正式なキックオフは 2000 年 8 月であった。

プロジェクトの目的は、セマンティック Web の実現を容易化するための言語(オントロジー記述言語)の設計と、ツール(オーサリングツール等)及びソフトウェア(知的インテリジェント・エージェント)の開発である。

MIT、スタンフォード大学、BBN、Nokia Research Center 等 18 の技術開発チームが参加し、それぞれのプロジェクトを担当している7。DAML のプロジェクトマネージャーは米メリーランド大学の James Hendler 教授である。各プロジェクトの年間予算は、平均して 100 万ドル程度であると思われる8。

#### DAML 仕様書

2000 年 10 月 10 日に"DAML-ONT"仕様書の初版9がリリースされた。

2000 年 10 月に Joint US/EU アドホック・エージェント・マークアップ言語コミッティー (Joint Committee) 10が DARPA / メリーランド大学の James Hendler と EU IST プログラムの Hans-Georg Stork によって結成された。

2001 年 1 月 27 日に Joint Committee は DAML+OIL (December 2000) <sup>11</sup>という仕様書をリリースし、DAMLとOILとの統一化を行なった。2001 年 3 月 27 日には"DAML+OIL (March 2001) <sup>12</sup>"として改版がなされた。DAML+OIL は RDF スキーマをベースとしたオントロジー記述言語であり、RDF にはない機能が追加されている。

# DAML-S version 0.5

2001年5月30日に DAML-S version0.5 がリリースされた<sup>13</sup>。DAML-S は DAML ベースの Web サービス向けオントロジー記述言語であり、Web サービス提供者に対して、曖昧性がなくコンピュータが解釈可能な形式で Web サービスのプロパティとケーパビリティとを記述するためのマークアップ言語を提供するものである。DAML-S の目的は、Web サービスの発見(discovery)、起動(invocation)、構成(Composition)とサービス間

<sup>&</sup>lt;sup>6</sup> INTAP「WWW10調査報告書」より。

<sup>&</sup>lt;sup>7</sup> http://www.daml.org/researchers

<sup>&</sup>lt;sup>8</sup> INTAP「BBN Technologies 出張報告」より。

<sup>9</sup> http://www.daml.org/2000/10/daml-ont.html

<sup>10</sup> http://www.daml.org/committee/

<sup>11</sup> http://www.daml.org/2000/12/daml+oil-index.html

<sup>12</sup> http://www.daml.org/2001/03/daml+oil-index

<sup>13</sup> http://www.daml.org/services/daml-s/2001/05/

操作(Interoperation)、及び実行モニタリング(Execution Monitoring)などのタスクを自動化するための記述を提供することである<sup>14</sup>。DAML-S version0.5 は DAML+OIL (March 2001)の上に構築されている。

#### (2) DARPA による DAML プロジェクトの位置付け

現在 DARPA において国家予算によって運営される 30 件のリサーチ・プログラムの中に、「PE0602301E Computing Systems and Communications Technology」というプログラムがある(大分類)。本プログラムは 6 件の研究プログラムから成り立つが、このうちの 1 件が「ST-11 Intelligent Systems and Software」である(中分類)。本プログラムは、2001 年計画で 6 件、2002 年計画で 7 件のプログラムを含むが、DAML はそのうちの 1 プログラムに当たる(小分類)。なお、DAML については 2001 年度がプログラム初年度になる(2000 年度の成果報告に DAML の記述はない)。(図 2-3参照のこと。)

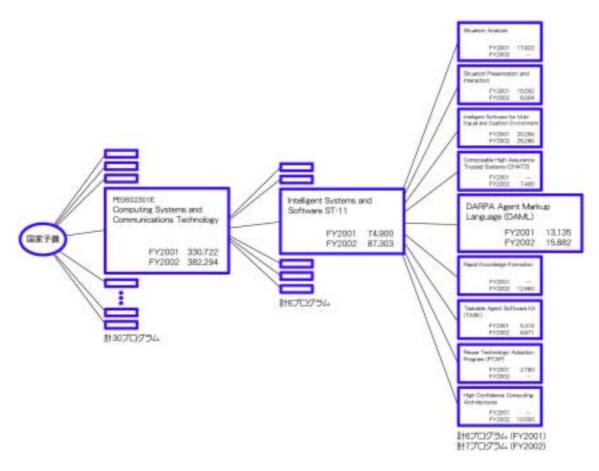


図 2-3 DARPA による DAML プロジェクトの位置付け

「DARPA Agent Markup Language (DAML)」 (小分類)の実施計画及び予算(2001年度)

-

<sup>14</sup> INTAP「SWWSの聴講メモ(暫定版)」より。

- DAML 言語仕様の完成
- Intelink<sup>15</sup>向けブリーフィング・ツールのワーキングバージョン・リリース
- Intelink 上の DAML 検索ツールのワーキングバージョン・リリース
- Intelink 上の DAML オントロジー作成ツールのワーキングバージョン・リリース
- non-pre-planned エージェント間の相互運用をサポートするための DAML 向け要求 仕様定義
- 陸軍戦訓センター (Center for Army Lessons Learned) のアーカイブ情報の保存、アクセスや編集能力を向上させる DAML オントロジー作成ツールの効用に関するデモンストレーション
- データ抽出及びリモートでキャッシュされる Web 情報の更新率に関する一階述語論 理 (first-order rule)の開発
- Robust Open Source の開発モデルに準拠し、組み立て可能な高アシュアランス・高 信頼性システム実現に向けた新しいアプローチの研究
- 高アシュアランス・高信頼性インプリ言語やツール開発に関するフィージビリティと 新しいアプローチの研究
- 高アシュアランス・高信頼性を備えたシステム・プロテクション・プロファイル及び 高信頼性言語・ツール開発に関する新しいアプローチの研究

計 10 項目 \$M13.135 (約 16 億円弱)

「DARPA Agent Markup Language (DAML)」 (小分類)の実施計画及び予算(2002年度)

- DAML 技術を C2 に応用したツールセットの定義
- Intelink DAML ブリーフィング・ツールの実験的な分析の実施
- 実務的な Intelink ノード上での DAML 検索ツールの展開
- 軍や国の情報機関向け Web アプリケーションのための DAML オントロジー作成ツールのプロトタイプのデモ実施
- 陸軍戦訓センター (Center for Army Lessons Learned) における検索 (search and retrieval)ツールの能力を向上させた選りすぐりの DAML ツールのプロトタイプ作成
- Millennium Challenge への参加を含む、海軍及びジョイント C2 のインターオペラビ リティーへの DAML の応用に関する実験的な分析の実施
- World Wide Web の実験的評価や設計に関する 150 万件以上の DAML ステートメントのリポジトリ作成
- dynamic prioritization management 向け技術の開発
- 帯域状況が変化する中でのインテリジェントな情報配信技術の開発

計 9 項目 \$M15.882(約19億円)

<sup>15</sup> FBI、CIA、DEA(米麻薬取締局)、NSA(米国家安全保障局)等で収集した諜報情報の相互利用ネットワークのこと。Intelink のテストベッド運用は 1994 年に開始された。

#### (3) DAML 各プロジェクト概要

#### **BBN** Technologies

BBN は DAML プロジェクトにおいて、プロジェクト全体を統合する「オーバーオール・コントラクター」の地位を与えられており、商用化の役割も担っている。BBN は DARPA から 2000 万ドルの年間予算を得ている<sup>16</sup>。

#### スタンフォード大学及びドイツ・カールスルー工大学

Web 上のインテリジェント・エージェントを可能にする OnTo Agents プロジェクト。インフラストラクチャ・コンポーネントの開発(OnTo Agent セマンティック Web ページ・アノテーションツール、オントロジー表現ツールキット、推論システム)や、アノテーションされた Web ページを用いた、貨物輸送スケジューリング用のアプリケーションである OnToCargo-Agent のデモを行なう。

#### CMU 及び Nokia Research Center

ATLAS (Agent Transaction Language for Advertising Services)プロジェクト。 ATLAS とは、広告サービス向けの DAML ベースのエージェント処理言語である。広告したり、要求したり、発見したりするエージェントサービスのための DAML 言語とツールの開発を行なう。また、広告用オーサリングツールや、ミドルエージェント向け仲介アルゴリズムの開発も行なう。

DAML では、各プロジェクトでそれぞれの分野の ontology を構築している $^{17}$ 。上部集合としては、 $^{\circ}$ Cyc が無料で公開しているものを使うとのことである $^{18}$ 。

#### (4) DAML 関連ツール

#### DAML Viewer

2000 年 11 月 17 日には DAML Viewer がリリースされた<sup>19</sup>。同ツールでは、DAML で書かれた Web 上のステートメントを表示することができる。BBN Technologies の Mike Dean 氏と Kelly Barber 氏が開発した。

#### OntoEdit

2001 年 3 月 7 日にはドイツの Ontoprise 社が OntoEdit v1.03 をリリースした $^{20}$ 。同ツールでは、DAML+OIL のフォーマットでオントロジーを作成することができる。OntoEdit については、3 . 2 . 3 節で紹介する。

<sup>16</sup> INTAP「BBN Technologies 出張報告」より。

<sup>17</sup> http://www.daml.org/ontologies/を参照のこと。

<sup>18</sup> 浦本「Semantic Web Working Symposium 参加報告メモ」より。

<sup>19</sup> http://www.daml.org/viewer/

<sup>20</sup> http://www.ontoprise.de/

#### DAML Crawler

2001 年 5 月 4 日には DAML Crawler がリリースされた<sup>21</sup>。同ツールでは、DAML で書かれた Web 上のステートメントを収集することができる。BBN Technologies の Mike Dean 氏と Kelly Barber 氏が開発した。

#### **DAML Validator**

2001年5月11日には DAML Validator がリリースされた<sup>22</sup>。同ツールでは、DAML を使用して作成されたオントロジーファイルの文法をチェックすることができる。 BBN Technologies の Dave Rager 氏が開発した。

#### その他の DAML ツール23

その他の DAML ツールとして、以下のものがある。

- ・DAML オントロジー・ライブラリ:クリアリングハウス(150以上のオントロジー)
- ・RDF API: Java パーサ及びシリアライザー
- ·OilEd:オントロジー・エディター
- · Chimaera: オントロジー・アナライザー
- ・DAML Viewer、PalmDAML<sup>24</sup>、HyperDAML:ナビゲーターGUI
- ・HAIRCUT: text/DAML 検索エンジン

# DAML アプリケーション<sup>25</sup>

DAML アプリケーションとして、以下のものがある。

- · UMBC IT Talks
- ISI todo
- · Horus

#### IT Talks

IT 関係の研究者の講演案内の Web ページに対し、メタデータを付与し、個人のプロファイル、カレンダー、地図情報(物理的にその講演を聴きにいけるか否か)と組み合せて、興味のある講演を検索し、カレンダーに反映させるシステムである<sup>26</sup>。

# (5) DAML の課題と今後の見通し

DAML の課題としては、DAML オントロジーの作成やマークアップ、アクセスのためのツールがまだ十分でない<sup>27</sup>。

BBN の意見では、DAML のツールはリサーチレベルのものに留まっている。来年には

<sup>21</sup> http://www.daml.org/crawler/

<sup>22</sup> http://www.daml.org/validator/

<sup>23</sup> http://www.daml.org/2001/06/swday-daml/Overview.html

<sup>&</sup>lt;sup>24</sup> PalmDAML とは、PalmOS 向けに DAML Viewer を実装したもの。

<sup>25</sup> http://www.daml.org/2001/06/swday-daml/Overview.html

<sup>26</sup> 浦本「Semantic Web Working Symposium 参加報告メモ」より。

<sup>&</sup>lt;sup>27</sup> INTAP「BBN Technologies 出張報告」より。

W3C で標準が出てくるので、民間のデベロッパーが商用のツールを出してくるだろう。また、オントロジー言語も普及するであろう。商用プロダクトは、2、3 年で出てくるのではないか。DAML は XML と置き換わるものではなく、より豊かな理解のための道具として、エージェントとともに普及するであろう。重要な点は、DAML を使うことでソフトウェアエージェントを軽量かつ汎用にできることである。FIPA(Federation for Intelligent Physical Agents エージェント技術の国際標準化団体)は、とりわけ W3C で標準化された後には、DAML の標準を受け入れるであろうということである<sup>28</sup>。

#### 2.2.1.2 DCMI (Dublin Core Metadata Initiative)

#### (1)概要

DCMI は 1994 年にシカゴで開催された W3C の年次コンファレンスである WWW5 から本格的に活動を開始した。

DCMI は 700~800 の会員組織及び個人の会員を持つ。国家レベルではオーストラリア、デンマーク、フィンランド、英国、アイルランドが加入済みである。ニュージーランド、カナダが地方政府レベルで加入しており近々国家レベルで加入予定である。米国は地方政府ならびにエージェンシーレベルでの加入は行なっているが、連邦レベルということになると最後の加入者になるかもしれない29。

DCMI の目的は、相互運用可能なメタデータ標準の普及促進と、情報資源を記述するための特定のメタデータ・ボキャブラリの開発である。

Web 上の各種の情報資源を統一的に記述するためのメタデータ(Title、Subject といった基本エレメントの集合)を定義することを目指している。メタデータの記述には RDF を用いる。

DCMI の活動において中心的な役割を担う OCLC (Online Computer Library Center) の年間予算は、116万ドルであり、その内の 90%は米国の図書館コンソーシアムのメンバーから得ている。OCLC には 25 名のフルタイムの研究者がいる。

DCMI が策定したダブリンコアとは、情報資源を記述するための基本的な 15 エレメントのことである。多くの人たちが共同に使うためのポリティカルな合意の結果 15 個のメタデータエレメントとなった。これはメタデータ記述の共通になる部分として 30 ヶ国(25 言語)間でコンセンサスが取られている。ダブリンコアの「コア(Core)」とは、このようにメタデータ記述の共通部分、基本部分という意味である。この基本部分に加えて、用途ごとに拡張を行なうこと(拡張エレメントを作成すること)が可能である。

Creator	Title	Subject
Contributor	Date	Description
Publisher	Type	Format
Coverage	Rights	Relation
Source	Language	Identifier

<sup>&</sup>lt;sup>28</sup> INTAP「OCLC (DCMI) 出張報告」より。

\_

<sup>&</sup>lt;sup>29</sup> INTAP「OCLC (DCMI) 出張報告」より。

各エレメントの詳細については、3.1.4節を参照のこと。

ダブリンコアの拡張性としては、アプリケーションごとに拡張部分を作成している。例えば、「インターナショナル・デジタル・ライブラリ・プロジェクト(バージニア工科大学)」「Dublin Core・ガバメント・グループ」「Dublin Core 教育グループ」では 15 エレメントの上にそれぞれが拡張を行なっている。お互いの互換性は、拡張部分においては必ずしもなくてよい。また、完璧な互換性も必要ない30。

2000 年 10 月、DCMI は活動領域の拡大を行なった。DCMI のミッションは、以下の活動を通じてインターネットを用いた情報資源検索を容易化することとなった。

- ・境界を超越した検索のためのメタデータ標準の開発
- ・メタデータ集合の互換性のための枠組みの定義
- ・コミュニティーの発展を促進する、または規範となる特別なメタデータ集合の開発 を促進する

DCMI の 2001 年の Workshop は 10 月 22 日~26 日に東京で開催された。

#### (2)今後の展望

ダブリンコアに基づいた技術適用の例としては、RDF を使用した Adobe 社の Acrobat 製品がある。また、ダブリンコアの技術を使ったビジネスデータの交換・共有は有望である(文書管理システム)。例えば、サンマイクロシステムズ、ノキア、自動車業界はその実現について模索中である<sup>31</sup>。

ダブリンコアの標準化は欧州や米国では進みつつある。例えば、政府文書にダブリンコアに準拠したメタデータを付与しようという取り組みが各国でなされている。各国政府のメタデータ標準については3.1.2節を参照のこと。

<sup>&</sup>lt;sup>30</sup> INTAP「OCLC (DCMI) 出張報告」より。

<sup>&</sup>lt;sup>31</sup> INTAP「OCLC (DCMI) 出張報告」より。

#### 2.2.1.3 RSS (RDF Site Summary)

#### (1)概要

コンテンツ・シンディケーションのためのメタデータ記述フォーマットである。メディ ア・コンテンツ・ポータルを開発するために使用する。標準化されたボキャブラリ(ダブ リンコア)の上に構築される。RDFシンタクスを使用している。

RSS は、Netscape が 1999 年に自社ポータル( My Netscape 32 )のために開発した RSS0.9 が基となっており、この時点では RDF とは無関係であった。その後 RSS は、Netscape と My Userland の Dave Winer33によって拡張され、他のポータルサイトでも利用される ようになるが、ボキャブラリの問題などから 2000 年中ごろには使われなくなっていた。 しかし、RSS-DEV ワーキンググループが結成され、RDF を中心としてモジュールによる 拡張機能を備えた RSS1.0 仕様書34が 2000 年 12 月に発表された。また、RSS1.0 とともに、 ダブリンコア・モジュールと、いくつかのサポート用ツールが発表された35。

#### (2) RSS の応用システム: O'Reilly Network

O'Reilly Network

オライリー・ネットワーク36とは、新しいプラットフォーム、プログラム言語、OSを含 むオープンな次世代技術に関心を持つ開発者向けのポータルサイトである。同ネットワー クは専門技術者に技術情報を明確かつ整合的に提供する。また、専門技術者のコミュニテ ィーのためのフォーラムを作成し、彼らが互いにコミュニケーションを行なったり、提携 サイトとコミュニケーションしたりする機会を提供する。

#### O'Reilly Network O Meerkat<sup>37</sup>

Web ベースのシンディケート化された RSS ベースのコンテンツ・リーダーである。 RSS は、ニュースや製品紹介、討論のスレッド、その他の多種多様なコンテンツを、いくつか のチャンネルに分類して配信するために用いられる。Meerkat は、これらのストーリー (個々のニュース等)に対して単一のインターフェイスを提供する。Meerkat では、最新 のストーリーはそのソースに関わりなく、常にトップの位置に置かれる。

Meerkat はデータベースにチャンネルのリストを保持している。1 時間に 1 回、Meerkat はそれぞれのチャンネルに関連付けられた RSS ファイルを訪問し、データベースに新たな ストーリーを追加する。

Meerkat を利用することにより、利用者は自分の関心のあるストーリーを専ら表示する ことができ、サイトをあちこち飛びまわる必要がなくなる。

<sup>32</sup> Netscape の My Netscape はチャンネルラインに沿って組織された大きな「スタート」ページを提供 している。チャンネルを好きなように追加したり、削除したりすることで、利用者は自分のページをカス タマイズすることが出来る。

<sup>33</sup> Dave Winer の My.Userland.Com はより広範な焦点を持っており、Userland と EditThisPage.com (Manila の Web サイトの成長するコミュニティー)でホストされた WebLogs からコンテンツを取り入 れている。

<sup>34</sup> http://groups.yahoo.com/group/rss-dev/files/namespace.html

<sup>35</sup> INTAP「WWW10調査報告書」より。

<sup>36</sup> http://www.oreillynet.com/

<sup>37</sup> http://www.oreillynet.com/meerkat/

#### 2.2.2 欧州のセマンティック Web 動向

欧州におけるセマンティック Web 技術の研究・開発にあたっては、予算年度が 1998 年から 2002 年の 5 年間にわたる EU の第 5 次 RTD (research and technological development) フレームワーク・プログラム (FP5) によって予算措置が講じられている。 FP5 は 4 つの個別プログラムと 3 つの横断的プログラムからなり、全体で 23 の、いわゆる主要なアクションや一般的な研究活動をカバーしている。これらすべての活動のために、5 年間で 149 億 6000 万ユーロが投じられることになっている。 FP5 は、エネルギー政策にからむものまで含め、非常に広汎な領域の研究活動を支援しているが、セマンティック Web 技術に関わる研究・開発は、上記 4 つの個別プログラムのうち、最大の予算を持つ「ユーザーフレンドリーな情報社会の創造 (IST)」プログラム (5 年間で 36 億ユーロ) に含まれる。

FP5 はプログラム予算の策定が 1997 年以前に遡るため、当初セマンティック Web に直接的に言及するプログラムはなかったが、2001 年になって、セマンティック Web 技術に関するアクションラインを発表した。しかし、具体的かつ直接的なセマンティック Web 関連技術の研究・開発は後述する第 6 次フレームワーク・プログラム (FP6) において行われることになろう。FP5 の IST プログラムにおいては、別途メタデータを用いたプログラムが 20 件以上活動している。

EU 委員会は 2001 年 2 月 21 日、FP6 に向けたプロポーザルを発表した。このプロポーザルにおいても、IST (Information Society Technologies)、いわゆる IT が主要テーマ領域の一つに位置付けられている。FP6 の予算年度は 2002 年から 2006 年までで、5 年間に 175 億ユーロが投じられる予定である(対 FP5 比 16% 増)。

IST 領域において実施される予定の様々な活動に備え、EU 委員会の情報社会総合理事会(DG Information Society)では、一連のミーティングを持った。ミーティングの目的は、FP6 において IT のコンテンツや実装手段に関し産業界や学術的研究機関に対して意見を求めることである。このうち、2001年4月27日にブラッセルで開催された知識関連技術に関するミーティングには、学会ならびに産業界をリードする知識技術の専門家22名が一堂に会した。

本ミーティングにおいては、知識関連技術の研究・開発を進めるに当たり、主要な取り 組みテーマとして以下の4つを挙げた。

- ・ 知識システムの可用性 (Usability of knowledge systems)
- ・ 知識としてのコンテンツ (Content as knowledge)
- ・ 標準とインターオペラビリティー (Standards and InterOperability)
- ・ ナレッジ・コミュニティーとポータル (Knowledge Communities and Portals)

また、ミーティングの最終報告書において、セマンティック Web が概念的には広範な支持を獲得しつつあり、RDF を採用することが知識システムのためのコンテンツ構築に向けて重要なステップであるといった、現在のマーケット状況や研究状況についての総括がされている。アプリケーションやマルチメディア・モバイル開発環境へのリンクについては、2000 年 11 月にルクセンブルクで開催されたセマンティック Web ワークショップに

#### おいて報告された。

欧州でのセマンティック Web 技術に関する理解は総じて以下のようなものである。

- ・ セマンティック Web はマシンが意味情報を含んだドキュメントやデータを理解 できるようにする。
- セマンティック Web の主要コンポーネントは、知識表現システム・技術、オントロジー、そしてインテリジェント・エージェントである。
- 学会においては、個別領域向けのオントロジーを開発する動きが高まりつつあるが、強固な競争力を確保するには至っていない。
- ・ 「知識システムの可用性」に関しては、「意思決定支援システム」を、セマンティック Web 技術を用いて構築すれば、ビジネスにおいてユーザーのニーズに直接的に応えることができるのではないかという議論が行われている。
- ・ 意思決定において、マシンに委ねられる部分が多くなれば、人間は本当に重要な部分のみに関与すればいいという、10年前のSIS(戦略情報システム)で想定されたことが現実味を帯びることになる。
- ・ 人間が重要部分への関与のみで意思決定が遂行されるようになれば、今度はユーザーが「楽しんで」、このプロセスを実行できるようにしなければならない。そのためには、単純にコンテンツやメタデータを作ればいいというわけではない。知識発見プロセスをモデル化し、知識についてのライフサイクル全体の中にユーザーの行動パターンを適用していく必要がある。
- ・ マシン・トゥ・マシン (デバイス・トゥ・デバイス) の環境が重要なのは、最終的なゴールが「ネットワーク化された状態でのヒューマン・トゥ・ヒューマンなインタラクション」だからである。
- ・ 「知識としてのコンテンツ」に関しては、テキスト・ドキュメントに限定されないマルチメディア・コンテンツや、マルチメディア・コンテンツ作成・流通の際に不可欠な認証・公証といったことが議論されている。
- ・ 企業内の知識システム活用については、組織知ないしは集合知の保存ならびに利用といったことが議論されている。
- ・ 「標準とインターオペラビリティー」に関しては、セマンティック Web 研究者 にとっての重要領域とされる、別々のセクターもしくはサブセクターで構築された オントロジー同士をリンクさせるための「上位レベルのオントロジー」といった概 念を設定する必要性について論じられている。しかし、こうした議論を展開すると、 オントロジー構築に当たってはボトムアップ型にすべきかトップダウン型にすべきかという議論を呼んでしまう。本件については双方のアプローチをハーモナイズ すべきであると認識する。
- ・ オントロジーとメタデータは徐々に自動生成されていくべきであるという点は 一般的な合意事項となっている。新たなオントロジーを生成し、校正し、相互運用 性を持たせるに当たってはデータマイニングの技術が有用である。
- ・ 「ナレッジ・コミュニティーとポータル」に関しては、ユーザーインターフェイスやアクセス環境といった議論が交わされているが、これらの議論はナレッジ・コ

ミュニティーや知識ベースのポータル環境といったテーマに帰着する。こうした、ユーザーが情報を得ようとすれば必ず見に行く環境においては、必要としないコンテンツをフィルタリングしてくれる機能や、楽しんで見に行けるようなビジュアライゼーション機能も求められる。こうした機能は e ビジネスにおいても活用が可能である。

【参考 】カールスルー工大学のステューダー教授は開発すべき四つのキー・テクノロジー領域として以下のものを挙げている。

- ・ 知識の抽出 (Knowledge Extraction )
- ・ 知識の維持・管理 (Knowledge Maintenance)
- ・ ナレッジマネジメント (Knowledge Management)
- ・ 知識の表現 (Knowledge Presentation)

上記のプロセスを行なうに当たり、オントロジー構築やメタデータ整備を手作業で行なうとなると、すぐにもボトルネック化することになるだろう。したがって、オントロジーやメタデータを半自動的に生成するツールが求められる。個々のオントロジーは不均一に成立していき、異なったビジネス・シナリオやアプリケーション領域に適合する必要があるという前提に立つなら、オントロジー間の連携も喫緊の課題である。オントロジーやメタデータを更新し続け、その進化や拡張をサポートするためにもツールが必要である。

ステューダー教授はまた、たとえばエージェント技術を用いたナレッジマネジメント、セマンティック (メタ)ポータルやインテリジェント・サービスといったアプリケーション候補についても強調している。モバイル・ユーザー向けには、新しい情報機器用のコンテンツに対応させる必要があるが、メタデータとオントロジーによって、伝送データ容量を減らすことができる。もはやドキュメントベースのアプローチは捨てて、相互接続された「知識の断片」を提供していくやり方に集中していく必要がある。また、ビジネスプロセスをナレッジマネジメントに統合していく必要もある。そして、古くなった情報を見つけ最新の知識と関係付けられるようにしなければならない。

セマンティック・ポータルに期待されるのはセマンティック・クエリイングを提供することである。同様に、メタデータ・ポータルでは一連のメタデータ・リソースへのアクセスを提供することが求められる。また、インテリジェント・サービスはセマンティック・メタデータを開発していく役割を担う。複数サービスを組み合わせる新しい方法を早急に生み出せるようなソリューションが求められる。

【参考 】サレルノ大学のカプアーノ博士 ( Dr. Capuano ) は次の三つのアクション

を求めている。すなわち、第一に、分散サービスやセマンティック Web のアプリケーション向けにミドルウェアを提供するセマンティック・グリッドを作ることである。まったく新しい世代の知識ベースの技術やアプリケーションによって、たとえば、ユーザープロファイルに則ったリソース・セレクションやフィルタリングなどを行なうためのインテリジェント・モバイルエージェントやインテリジェント・インターフェイスが可能になるかもしれない。第二に、オントロジー間連携実現の一つの方法としては、あるドメインに特化したオントロジーを構築する場合にベースとするようなスタンダード・コア・オントロジーを通じて連携させるというものである。第三に、インテリジェント Web インターフェイスやそのアプリケーションに関しては更なる研究が必要である。特に、遠隔学習、情報探索、インテリジェント・コンシューマー・サポート・コールセンターのようなアプリケーションに関連する領域においてである。カプアーノ博士はまた、ナレッジ・モデリング、ナレッジ・レプレゼンテーションのための、複雑系システム向けのシミュレーションモデル、最適化モデル、数学的モデルに準拠した手法やツールをサポートするための集中的な努力が必要であると訴えている。

# 2.2.3 欧州の電子政府におけるメタデータ応用の動向

#### 2.2.3.1 英国の動向

#### (1)英国概要

IT 分野で出遅れていた英国では、トニー・ブレア首相のリーダーシップの下、1997 年から本格的に行政情報の電子化に取り組んでいる。1999 年からは、電子政府化を推進する「e-government 担当大臣」を新設し、イアン・マッカートニー大臣を任命した。

英国では 2000 年 7 月の段階で行政サービスの 33%がオンライン化されているが、2005 年までに行政サービスを 100%電子化することを目標にしている。

2000 年 3 月の段階で、英国で自宅を保有する人のうちインターネットに接続している人の割合は 25% (1 年前は 13%) に過ぎない。そこで政府は 38 億ポンド(約 6000 億円)を投じてインターネット端末を普及させ、それを使いこなせるように国民を教育するプログラムを開始した。図書館や郵便局に端末の導入を始めており、2002 年までに英国内に6000 ヶ所の公的なアクセス拠点を開設することになっている。

マッカートニーe-government 担当大臣は、「電子政府を導入するには、新しい技術と公的サービスの文化の改革が必要だ。我々のゴールは、国民の指先で 24 時間オープンしている便利な政府を作り出すことだ」と述べている。

## (2) e-GMF (e-Government Metadata Framework)の概要

e-GMF は英国内閣府の Office of the e-Envoy が推進している。

2001 年 5 月に、Office of the e-Envoy は e-GMF を発表した<sup>38</sup>。e-GMF は、公的部門にあまねく適用されるメタデータ標準を開発し、実施するための政府の政策を策定するものである。この標準は公的部門の情報システムのすべてに適用される予定である。同フレームワークでは次の 2 つの主要な決議事項が発表された。

- ・英国政府は、ダブリンコアを英国政府メタデータ標準 (UK Government Metadata Standard (e-GMS)) として採用する。ただし、ダブリンコアだけではニーズに応え 切れないので、同フレームワーク内で規定された規則に基づいて、更なる要素の追加と改善を行なう。
- ・汎政府シソーラス(同意語辞典)を開発すること。シソーラスは構造化された用語とキーワードのリストから成り、我々が情報を正確に定義することと、情報をより 迅速に発見することを手助けする。汎政府シソーラスは、政府、産業、および市民が利用できる予定である。

e-GMF に従うことは義務であり、段階的に施行される予定である。ダブリンコアの 15 エレメントについては直ちに義務付けられ、他の拡張エレメントと汎政府シソーラスについては、それらが開発され承認された際に義務付けられる。

e-GMF のインプリメンテーションは現在進行中の仕事であり、それには、開発のための中央集権的な(central)サポートとメカニズムを必要とする。そのポータルとなるのが後述の UK GovTalk の Web サイト<sup>39</sup>である。

\_

<sup>38</sup> http://www.govtalk.gov.uk/documents/UK%20Metadata%20Framework%20v1%202001-05.rtf

<sup>39</sup> http://www.govtalk.gov.uk

英国には Citizen Portal<sup>40</sup>という、政府(国と地方)の情報とサービスへのクイックアクセスを提供するポータルサイトがあり、 汎政府シソーラスはとりわけこの Citizen Portalのニーズを念頭において開発される予定であった<sup>41</sup>。

Office of the e-Envoy は、Metadata Working Group (MWG)を持ち、3ヶ月に1度の 頻度でミーティングを行なっている。

e-GMFのミッションは、一般市民がデジタルテレビやキオスクなどあらゆる媒体を通じて 24 時間 365 日すべての国家機関の情報にアクセスできる環境を実現することであり、また、そのためにすべての国家機関のみならず地方自治体の参加を仰ぎつつ、国家情報へのメタデータ付与を進め、アクセシビリティーと相互運用性の向上を図るためにシステムの標準化と近代化に努めることである。

これまでの成果としては、年間 50 億トランザクション、20 の政府機関と 340 の地方自 治体が参加し、高い評価を得るにいたっている。現状としては、各国の状況を様子見して いるが、英国はセマンティック Web といったような新規技術を追うよりは既存技術の活 用を通じた展開を志向している。

e-GMFの展開方法としては、各国共通の問題であるが、英国でもネットへの信頼性すなわちセキュリティーへの不安が叫ばれていた。そこで、UKOnline.gov という国民による政府情報アクセスならびに省庁間のトランザクションのためのポータルサイトを立ち上げ、約1年間実施してきた。セキュリティーを高め認証も行なえるようにするためにゲートウェイサーバも立ち上げた。

UKOnline.gov は二つのサイトを持っており、一つは内閣府内の情報交換を目的とした e-Envoy、もう一つが省庁間の情報交換を目的とした UK GovTalk である。UK GovTalk は国民から広くコメントを求めたり、新しい企画を持ち込んでもらうための Web サイトである。

英国においては通常、省庁間の情報流通は活発とはいえない。その背景には個人情報保護法の規定によって、省庁間の個人情報流通が規制されているという法的側面も影響している。ただし、Knowledge Network という省庁データの閲覧システムは運用されており、簡単な政策概要程度は閲覧することができる。現在は役人しか見ることができないが、年内にはネットで公開される予定である。しかし、現在でも政策立案者が大臣に上げた情報を他省庁の人が見ることができるという状況に抵抗もあり、そうした文化の改革が求められる。具体的には、当面は強制的に情報公開ならびにメタデータ付与を積極的に推進するよう担当者にプレッシャーをかけたり、広報担当官にクレームを上げたりする方法が考えられる。

Knowledge Network では初期バージョンのメタデータを使っている。データを大中小のカテゴリーに分類するにあたっては、ダブリンコアの Subject エレメントを採用している。メタデータを付与することにより、検索の容易化を意図している。

\_

<sup>40</sup> http://www.ukonline.gov.uk/

 $<sup>^{41}</sup>$  しかし、汎政府シソーラスの開発については、実現可能性について検討した結果、中止されることとなった。将来的にアプリケーションがたくさん出てくるとシソーラスが必要になるであろうが、その際には多言語対応している Eurovoc が採用される見込みである。 Eurovoc とは、15 の言語に対応した多言語シソーラスであり、20 年の歴史を持つ。将来は EU 域外の言語も含め言語の標準化が行われていく予定

メタデータの拡張に関するドキュメントには 2 バージョン(英国及び EU)があり、2002年 1月現在で、UK GovTalk 上でコメントを集めている。

e-GMF において 2001 年にメタデータ利用のメリットについて検証したところ、非常に 良好な結果が得られた。このため、政府機関においてはメタデータ付与を義務付けている が、すべての機関が使っているわけではない。

上記のメタデータ利用メリットの検証については、具体的には以下の 5 つの標準検索フォーマットを作って国税庁、地方自治体、e-GMFで検証が行われた。

全自治体の名前

生後 6 ケ月の子供がいる場合に親のパスポートに併記することが可能か(もしくは 独立して必要か)

PITO (Police Information Technology Center) とは何か(警察関連の団体)

ペットを EU 圏内の旅行に帯同した場合に動物用パスポートがあれば検疫免除になるかどうか(昨年、当該法令が施行)

インターネット経由でのオンライン投票実施時期

という 5 つの項目について、役所で働く 50 人を選び、メタデータ未使用時と使用の双方について、「検索エンジンは何を用いたか」「検索に成功したか」といった質問をした。 また、検索に要した時間、答えの正しさ、結果としてどこまで進行したかなどを調べた。

メタデータの自動生成システムについては、2002 年 1 月現在で国税庁の Web サイトとロンドンのルイシャムという自治区のサイトがメタデータを付与して運用しているが、これらは EDS が開発したメタデータの自動生成システムを使用している。

省庁がメタデータを付与することについては、法律による強制力はないが、ある省庁がこれを無視した場合には IT 関連の予算配分を却下する権限を、内閣府は Guidance (行政指導)という形で持っている。

メタデータ付与が強制されるのは政策、書簡といったすべての政府発行情報についての みであり、その目的は以下の通りである。

国民による政府情報へのアクセスの容易化

電子データ管理の容易化

データの自動処理の容易化(一部)

メタデータを付与した省内の公文書に関する情報のレポジトリについては、現在は Public Document Office が主導する形で各省庁が個別に対応している。

紙の通信文を電子化するような部署はなく、メタデータ付与の対象はすでに電子化された情報のみである。

英国には公文書電子化についての法律がある。2000年末に法案が通過している。この法律により電子ドキュメントにも効力が認められるようになった。これに加え、女王による署名も電子的に行なうことが可能になった。

現状では Westminster (= 英中央政府、England、Wales) 政府のみが e-GMF を採用しており、Scotland 政府は採用を検討中である。Northern Irland は特に検討していない。

である。

<sup>42</sup> 本報告書の付録 2 と付録 3 を参照のこと。

#### 2.2.3.2 スウェーデンの動向

#### (1) スウェーデン概要

スウェーデンでは、北極圏に入っている北部は人口密度が著しく低く、南部の主要都市に人口が集中する傾向があり、全人口の半数以上が、日本と同規模の国土面積中 3%以下の地域に住んでいる。このような自然環境ゆえ、道路や鉄道などの交通網整備には時間とコストがかかるため、より安価に情報だけでも伝達できる電信、電話網は、発明から時をおかずしてスウェーデン社会に広く普及していた。

現在では世界の IT 立国として高い評価を得ているスウェーデンであるが、80 年代後半から 90 年代初めにかけては、「バブル経済とその崩壊」を日本とほぼ同様に経験していた。金融自由化と高い輸出競争力から、80 年代後半にかけて世界的に資本が流入することで景気が過熱、ついに 90 年代初めにバブルが弾け、深刻な景気後退に見舞われた。

しかし、90 年代半ば以降、IT 産業を基幹産業として育成することに成功し、「デジタル経済」へとアメリカ以上に見事に対応することで、バブル経済から立ち直った。

政策面では、バブル経済崩壊後、社会全体が情報・知識の集積・集約に基づく産業活動、 社会的活動中心に移行したとの認識を明確にし、情報ネットワークを基軸に社会を再構成 することを政策の中心的課題とした。このことは、1994年に当時のビルト首相が表明した、 「万人が、素早く、容易に、安全に、安価に、時と場所を選ばずに、情報を電子的に引き 出し、互いにコミュニケートできるスウェーデンを作る」とのメッセージに具現されてい る。

1996年にはIT立国のビジョンを示した「IT法案」が制定され、Confidence(信頼)、Competence(能力)、Accessibility(アクセス)という情報ネットワークに関する三条件を市民に確保することが重要であるとの見解が示された。こうした基本方針が示され、施策の立案、実行が進んだが、最も影響力の大きな施策が実施されたのは 1998年であった。

まず、第一に「パソコン法案」(企業が PC および周辺機器を購入し、従業員が家庭で使えるようにリースするプログラム)が成立し、家庭のパソコン保有率が 48%から 67% へと急激に上昇した。特にブルーカラー世帯でのパソコン保有が進んだという。

次いで、初等教育、中等教育の教員 10万人のうち 4万人に無償でパソコンを配布する 3年間のプロジェクトが開始されたことである。これは、一般家庭へのインターネット普及が進むにつれ、父母と子供たちに教員と電子メールで連絡をとることへの要求が強まったことが大きな影響を与えている。

スウェーデンは上記のようにデジタル経済を成長させ、失業率では、98年以降毎年1%ずつ低下させることに成功し、2000年にはついに日本を逆転した。財政面では、93年以降累積債務も増やさず、景気回復による税収の増大で98年から財政収支は黒字に転じ、累積債務も相対的に減らしている。

# (2)スウェーデン政府における情報化の概要

情報化の進展を 4 段階で表現すると、Information Interaction Transaction Integration となる。スウェーデンも他国と同様、情報化は緒に着いたばかりであり、Information の段階にある。

スウェーデンの e-government ポリシーとしては、以下の2つがある。

「24/7 e-government」(24 時間 365 日いつでもアクセスが可能という意味)

「ライト・タッチ」(中央省庁が地方やエージェンシーに対して強制的な影響力を振るわない、振るえないという意味)

中央政府(Central Government)はゴールだけを決め、課税自主権を持った県 (Regional Government)や市(Local Government)レベルが独立して行政を執行する。

# (3)行政情報へのメタデータ付与の具体的事例

労働市場局 (National Labour Market Board)

企業や政府機関などの求人サイドと求職中の個人との間で行われる求人情報の掲示やそれに応募するにあたっての履歴書作成・提出といった 1 件あたり 2、30 分の処理時間を要し年間 50 万件に上る処理件数をメタデータ付与によって効率化し活発化させることを目的とする。

構築にあたっては HR-XML<sup>43</sup>のフレームワークを用いた。RDF を使わなかった理由は、たまたま HR-XML があったから使ったというきわめてプラグマティックな理由による。

#### 高等教育庁 (National Agency for Higher Education)

スウェーデン国内に 40 の大学があり合計で 14000 のコースが存在する。これをいちいち個々の大学のホームページに見に行く代わりに、ワンストップで検索・閲覧可能なコースカタログを国が作ろうとする試み。

システムの仕組みは国が立ち上げたセンターデータベースに置かれたソフトウェア・ロボットが、各大学が作るデータベースを巡回し、各校のデータベース内のメタデータを収集して回り、センターデータベースを更新していく。各校のメタデータはダブリンコアに加え5つの拡張エレメントを採用している。

ライト・タッチの方針では各校にメタデータを付与させるのが難しいとも考えられるが、スウェーデンの教育機関はすべて国立といってよく、国からの予算配分は学生数に応じて行われているため、各校に共生するまでもなくセンターDBに自校の情報が反映されることが死活問題となる。したがって、導入に先駆けて各校の担当者を集めて説明を行なったところ、各校がスムーズに理解してくれたという。

メタデータの付与を行なうためには膨大な工数を要する。各校は具体的にどのように予算や工数を捻出したのかいうと、大手 25 校は元々データベースを持っていたため、フィルターを作り一括して移した。フィルターの開発とメタデータ付与は各校の IT 部門が行ない、そのための開発予算は各校の IT 予算から支出された。中小15 校はゼロから作った。

これらの具体的事例から得られる結論として、スウェーデン ICT 委員会は以下の2つ

-

<sup>43</sup> http://www.hr-xml.org

を挙げている。

セマンティクスはシンタクスよりも重要である。

- 国が何でもやらなければならない
- 「自分が持っているものから始めよ」
- HR-XML、ダブリンコア、iCal

また、失敗から学ぶための事例もある。商用車用の運転免許証発行システムの事例である。スウェーデンでは県単位で発行認可手続きを行なっており統一フォーマットがない。そこで政府が資金を拠出して"SHS"(スウェーデン語で Spreading & Picking up System)というシステムを作った(取り扱い窓口は郵便局)が、国際標準の SOAP や UDDI といった Web サービスに準拠していないのが問題となった。今度開発する e-ID では同じ轍を踏まないようにすることが必要であるという。

#### 2.2.3.3 アイルランドの動向

#### (1)アイルランド概要

北海道とほぼ同じ面積に人口わずか 360 万人あまりのアイルランドは、農業従事者が 90 年代後半でも労働人口の 1 割を占めており、主要資源が鉛、泥炭、ジャガイモなど、主要産業が食品加工、織物、自動車組立てといったように、高付加価値を創出できる産業構造が歴史的に見てもなかった。

そこで、73 年の EEC (欧州経済共同体)加盟以降、徐々に海外との競争力の低さがアイルランド経済を悪化させることになる。80 年代前半にはインフレが年率 20%程度に跳ね上がり、それに反比例して社会の雇用力は急激に低下、80 年代半ばには失業率が 18%にまで達するほど深刻な不況に見舞われた。

この事態を克服するため、90年代に入りアイルランド政府は、外資導入、情報通信分野への重点的投資と人材育成などの施策を実行に移し、アイルランド経済は見事に立ち直った。

1992 年から 1998 年までに 20 万人以上の雇用が創出され、失業率は 1993 年の 15.9% から 1998 年の 7.8%にまで減少した。1994 年から GDP は年率 7%で上昇し、国家財政赤字は 1986 年に GDP の 122%に達していたものが、1998 年には 55%へと改善した。

1998 年にアイルランドは米国を抜いて世界第一位のソフトウェア輸出国へと成長した (米国:30 億ドル弱、アイルランド:33 億ドル弱。アイルランドは欧州各国にとっての オフショアソーシング先として成長してきた)。これは 90 年代の IT 政策の賜物であり、1100 社以上の外資を導入することに成功した結果といえる。

ここで重要なのは、90年代のIT政策を中核とするアイルランドの政策は、1987年から、 労働者、雇用者組織、労働組合、企業、市民などの社会的主体間で結ばれている「社会協 約(Social Partnership Agreement)」の精神に基づく点である。それは 1987年の『国 家再生計画(Programme for National Recovery)』に始まり、アイルランド国民が戦略 的目標を定め、それぞれがどのような役割を担うかを認識し、貧困と不平等に一致して立 ち向かうことを目標にしたものである。そこでは、競争を無視することなしに、適切な賃 金の上昇と雇用の増大を社会で分かち合うことが謳われている。

#### (2)アイルランド政府における情報社会実現に向けた政策概要

90 年代に国の競争優位確保のために、首相主導でいくつかのイニシアティブができた。 内閣官房では様々なエージェンシー横断の e-Strategy を策定するためにハイレベルの組 織編制が行われ、歳入、土地登記、教育などに関するイニシアティブが結成された。

2000 年 7 月 10 日に「電子商取引法(eCommerce Act)」がマッカリース大統領署名によって制定され、ビジネス向けの電子商取引に適した環境整備と電子的な政府サービスの提供を可能にするフレームワークが整備されることになった。

具体的なイニシアティブとしては、以下のものがある。

「OASIS (On-line Access to Services Information and Support)」 2000 年 11 月 16 日にプロトタイプ版の Web サイト44の運用が開始された。国民 への情報提供を目的とする。ボキャブラリ管理の実現可能性を検証する。

「BASIS (Business Access to State Information and Services)」 2001年5月4日に本プロジェクトの第1モジュールが完成した。企業向けの情報提供を目的とする。

「Metadata」

中央政府および地方政府の公的サービスへのメタデータ適用を目的とする。

電子政府を推進する EU 各国とのベンチマークの結果、アイルランドが No.1 という結果が出ている。

電子政府の全体政策については、首相直属の機関(首相府)が策定しており、電子政府 政策の立案・普及促進と法律問題について所管している。電子政府政策のほか、情報社会 のあり方といった問題も研究している。

#### (3) メタデータ・イニシアティブ

2000 年 10 月に開始された、公共サービスへのメタデータ適用を目的としたプロジェクトである。ボキャブラリ管理、タクソノミー構築などを行なう。

ダブリンコアのエレメントセットを採用し、オーストラリア政府の策定した AGLS (Australian Government Locator Service) の追加エレメントも採用している。

プロジェクトの最大目的はリソース・ディスカバリーである。

2001 年 3 月に公開された「メタデータ --- 共通の標準を目指して45」は、アイルランド公共サービス・メタデータフレームワークのドラフト文書であり、オンラインの情報資源にメタデータを付けるための標準案を提示している。メタデータの普及により、利用者は公共サービスの Web サイト上で入手できる情報をより容易に、より迅速に見つけることができるとされる。

同文書の主眼は、情報とサービスをオンラインで提供する公共サービス機関に対し、アイルランド公共サービスメタデータ標準の基盤としてダブリンコアのメタデータエレメントを適用することを勧告することである。

<sup>44</sup> http://www.oasis.gov.ie.

<sup>45</sup> http://www.gov.ie/metaconsult/ipsms.rtf

同文書は公開され、文書中で提示される複数のプロポーザルに関して、関連する組織からのパブリックコメントが求められた。アイルランド政府のメタデータ・コンサルテーション・グループは最終標準を承認する前に、すべてのコメントを吟味することになっている。パブリックコメントの提出期限は2001年5月4日であった。パブリックコメント集はhttp://www.irlgov.ie/metaconsult/consultation responses.htm で公開されている。

メタデータを適用している分野については、プロジェクトの主要目的がリソース・ディスカバリーであるため、Web ベースで記録されたデータへのメタデータ付与がメインである。使用しているツールについては、非常に初歩的なものしか作っていないという。

英国では公文書へのメタデータ適用の効果を検証しているが、アイルランドではメタデータの適用自体が非常に初期的な段階にあるので、具体的な検証は行なっておらず、むしる国民に対しレガシーデータをデータ化することの意義について教育することに重点を置いているという。

メタデータを適用して政府関係のデータベースを横断的に検索できる検索サービス (Infoseek ベース)を構築する予定もある。

- 2.2.4 欧州におけるオントロジー技術の動向
- 2.2.4.1 オランダのフライエ大学

#### (1)フライエ大学の研究状況

フライエ大学でのセマンティック Web 関連研究は、Artificial Intelligence 学科及び Business Infomatics 学科の 2 ヶ所で行われている。AI 学科では基盤技術の研究を、BI 学科では応用の研究を、強く連携を取りながら行なっている。

フライエ大は本分野では比較的先進的な方であり、他には、カールスルーエ大学、英マ国マンチェスター大などが先進的である。その中でも最も早く手がけている。その他、ブリストル大学は、他大学だけでなく英国 HP 社とも連携して良い研究を行なっているという。

欧州の第 6 次フレームワークプログラムのみならず、トランスアトランティックな研究としては、米国 DARPA と DAMIL + OIL の研究を行なっている。OIL はアムステルダムで開発され、W3C に提言を行なっている。

フライエ大学と共同研究している企業については、初期は AIdministrator のような大学からスタートアップした小さな会社と共同研究開発を行なってきたが、最近は、BT、ドイツテレコム、ダイムラークライスラー、スイスライフ、オラクル、フィリップスのような大きな企業も参加するようになっている。

# (2)オントロジー研究開発状況

フライエ大学ではオントロジーの応用領域として以下の3つを検討している。

- (1) 知識管理:電子政府も検討されたが、民間への適用を中心に検討中である。BT、スイスライフなど大企業と共同研究をしており、大規模なイントラネットでの知識流通のための研究を行なっている。
- (2) 文書管理(検索エンジン):現状の検索エンジンには様々な限界がある。Web のグローバル規模での文書共有の世界で、より知的に管理できるようにしたい。 但し、より知的に管理するためには、企業内での利用も重要である。
- (3) 電子商取引(EC)と Web サービス: OBELIX のプロジェクトがある。電子商取引をWeb サービス関連の研究では、ORACLE とも提携している。

# 多くのプロジェクト(およそ10)があるが、主なものは以下

OnToKnoledge プロジェクト: 初期のプロジェクトで OIL と Open Sesame などのツールを作成した。

IBROW: サービス統合 / データ統合のために、Web 上のコンポーネントの知的 Brokering サービスを行なう。

SWAP: セマンティック Web と PtoP 技術を統合する(中心サーバをおかずに直接交換する)。

EU のプロジェクトとして提案を予定しているおり、2002 年 1 月現在検討中である。

WonderWeb:オントロジーマッピングを研究する。複数のオントロジーの対応をとるもので、オントロジーの翻訳に近い。

OntoWeb:

研究プロジェクトではなく、普及のための Web サイトである。大学、企業などとの連携や、ワークショップや会議の開催などがある。OntoWeb については2.4.4.2節を参照のこと。

#### (3) Web サービスや人工知能(AI)との関係

セマンティック Web と Web サービスの違いについては、フライエ大学の研究者の見解では、Web サービスはセマンティック Web の階層図(図 2-1参照のこと)の中で見るとプロトコルや情報バインディングなど低レベルの技術的な統合はできるが、意味的統合はできない。Web サービスは、内容に基づく事はやっておらず、現在の活動は重要であるが、拡張が必要である。オントロジーを導入して、コンテンツに関する意味的な統合を行なうことが重要であるとのことである。

また、電子商取引ではオープン性が特徴の PtoP アプローチがある。しかし、スケーラブルサポートが必須であり、ロゼッタネットのようなメディエーション機能が必要となる。このために、セマンティック Web が必要となる。ビジネスロジックのような多様な動的な構造に対応するためには、データ発見や変換、コンフィグレーションなどを用いて支援する仕組みが必要であるとのことである。

セマンティック Web と人工知能 (AI) との違いについては、フライエ大学の研究者によれば、以下の 2 つの違いがあるという。

(1) オントロジーの交換ができる点。

昔(AI研究が盛んであった頃)はゼロから作る必要があった。現在は他のオントロジーを再利用し改良することが現実的な状況になっている。

(2) セマンティック Web のオントロジーは、AI よりももっと単純で、構造は簡単で (場合によってはタクソノミーレベル)あり、そのかわり量は多い。

オントロジーは歴史的には AI から出たが、現在はコミュニケーションや自然言語システム、電子商取引など様々な場面で用いられ多様化しているので、オントロジーでは厳密な(厳密な定義による/厳密な理論的背景を持つ)オントロジーに限らない。

# 2.2.4.2 ドイツのカールスルーエ大学

#### (1)カールスルーエ大学の概要

カールスルーエ大学には AIFB (Applied Informatics and Formal Description Methods)と FZI (Research Center for Information Technologies)という2つの研究拠点がある。

AIFB は、セマンティック Web と人工知能関連の研究を行なっている。オントロジーに関しては、欧州最先端の研究所であるとの自負がある。

カールスルー工大学では、商用化を意識して研究活動を行なっており、開発しているシステムは、商用に十分耐えるものである。AIFB で研究開発した先進技術の産業界への移転は、FZI を通じて行われている。

セマンティック Web 研究開発のキーパーソンであるオランダ・フライエ大学の Dieter Fensel やスタンフォード大学の Stefan Decker もカールスルーエ大学の出身である。

欧州におけるオントロジー及びセマンティック Web 研究のグループには、カールスルー工大学のグループ(オランダのフライ工大学はこのグループに属する)、英国のマンチェスター大学のグループ、及びイタリアのピサ大学のグループがある。

研究プロジェクトには、EU のプロジェクトとドイツ独自のプロジェクトとがある。以下に AIFB が参画する研究プロジェクトを挙げる。

- 1) セマンティック Web 関連
  - Wonderweb(EU)
  - ・ PHT などの教育プログラム(EU)
- 2) 知識管理関連
  - OntoKnowledge(EU)
  - OntoLogging(EU)
  - Consenser(BMBF)
  - · Ontowise(DFG)
  - · SWAP(EU)
- 3) ビジネス・インテリジェンス関連
  - ・ ドイツテレコム(DT)
  - ・クライスラー
- 4) e-learning 関連
  - · PADLR
  - Saxony(BMBF/L3S)

# (2)開発ツールの紹介

カールスルーエ大学で開発しているオントロジー関連のツールに OntoEdit と KAON (Karlsruhe Ontology) がある。

#### OntoEdit

3.2.3節を参照のこと。

#### **KAON**

KAON は、オントロジーとセマンティック Web の利用基盤であり、RDF をベースとしたオントロジーとメタデータのツール群である。Java,J2EE 及び python を用いて作られている。次の三階層のツールから構成されている。

(1) アプリケーション&サービス層 (オントロジー作成及びウェブポータル作成ツール等):

アノテーションツールやオントロジーエディタ、オントロジー/メタデータクローラ、既存データベースのデータを RDF に変換するツール、及びオントロジー/メタデータから HTML データを生成するツールなどがある。

(2) ミドルウェア層(API 及びサーバ等): RDFS データやオントロジーデータを操作するための API、RDF データサーバ などがある。 (3) データ&リモートサービス層 ( P2P、RDB 等 ) : データ&リモートサービス層は、KAON を用いて実現されるサービスである。

- 2.2.5 欧州におけるセマンティック Web 関連システムの動向
- 2.2.5.1 イタリアの EU 委員会共同開発センター (JRC)

#### (1) JRC の概要

JRC は、EU 委員会 の一部であり、20 年前に設立して、いろいろなテーマに取り組んでいる。Web 発祥の研究所である CERN(欧州原子核共同研究機関)とは直接関係ないが、CERN 出身の人が多い。

JRC 内部の組織としては、原子力、エネルギー(核、プルトニウム)、安全保障、環境、健康・食品衛生、シンクタンク、情報社会などの分野がある。この中で EU の第 5 次フレームワーク・プログラムは、情報関係など、色々な研究に出資を行なっており、この中で、セマンティック Web 関係のテーマも扱われている。

W3C が技術規格を策定する所であるのに対し、JRC はソフトウェアを作って実践的に 行なう所という位置付けを行なっている。

#### (2) セマンティック Web 関連のシステム

ETB (EuropeanTreasuryBrowser) 46

2.4.2.2節を参照のこと。

RDFStore (PerlAPI for RDFStorage) 47

RDFStore のプロジェクトは、2001 年 10 月から開始された。RDFstore は、RDFの構造を解析、格納、管理する Perl のライブラリーである。これは、RDFの主語・述語・目的語という 3 つ組(図 2-2参照のこと)の検索を効率的に行なうことを目的としている。検索にはコンテキストが重要であり、コンテキストも考慮した検索が可能である。3 .2 . 1 節を参照のこと。

#### 2.2.5.2 フランスの国立情報処理自動化研究所(INRIA)

# (1) INRIA の概要

2002 年 1 月現在、85 人の研究者、300 科学者(ポストドクターや海外からの客員研究員)のメンバーで研究を行なっている。W3C の欧州におけるホスト機関である<sup>48</sup>。30 個のプロジェクトが進行しており、以下の 4 つの大きなテーマに大別できる。

ネットワークとシステム

ソフトウェア工学シンボリックコンピューティング

ヒューマンインターフェイス、画像処理、知識システム

複雑系システムのシミュレーションと最適化

これらのプロジェクトからスピンオフしたものはいくつか(ISTAR, RealViz など)あり、研究内容が実用化されている。セマンティック Web 関連のプロジェクトとしては、ACACIA があり、これは上記 に含まれる。この ACACIA プロジェクトでは、企業にお

\_

<sup>46</sup> http://etbrowse.jrc.it

<sup>47</sup> http://rdfstore.sourceforge.net

 $<sup>^{48}</sup>$  W $^3$ C の米国におけるホスト機関が MIT、アジアにおけるホスト機関が慶應義塾大学 SFC 研究所である。

ける知識管理の方法論とツール作成を研究している。

次世代の電子政府のイメージとしては、市民が必要なフォーマットに書きこめば、必要な情報を得られるというものであり、中央政府と地方政府の情報共有が主な目的だとしている。

# (2) セマンティック Web 関連のシステム

#### ACACIA

2 . 4 . 1 . 1節を参照のこと。

#### RDF-PIC

RDF-PIC は、画像データにメタデータを RDF 形式で付加することにより、画像データの検索の際に、実際の画像を検索することなしにメタデータのみで必要とする画像データを獲得する試みである。これによりネットワークの負荷を軽減することが可能となる。

# 2.3 標準化機関の動向

Web 技術の国際標準化機関である W3C (World Wide Web Consortium) の活動状況について調査を行なった。

#### 2.3.1 W3C O Semantic Web Activity

2001 年 2 月 9 日に Semantic Web Activity グループが設立した。従来の Metadata Activity グループ<sup>49</sup>からの名称変更である。2003 年 2 月に終了予定である。

Semantic Web Activity は、セマンティック Web の実現のために、アプリケーション間のデータの自動化、統合、及び再利用を支援する技術の共同開発と標準仕様化でリーダーシップを担うために設立された。

W3C の Semantic Web Activity は DARPA の DAML プロジェクトから活動支援を受けている。また、EU 委員会からの資金提供をうけている。<sup>50</sup>

Technology & Society Domain<sup>51</sup>に属する。Web の技術が社会で実際に利用されるためには、単に技術的な問題を解決しただけでは不十分であり、社会的な様々な問題を解決しなくてはいけないが、Technology & Society Domain ではこのような内容に関する活動が行われている<sup>52</sup>。

Web オントロジー・ワーキンググループ (Web Ontology Working Group, WOWG) 2001年8月に立ち上げられた新しいワーキンググループである。WOWGの活動目的は、各研究開発コミュニティー間のデータの統合と相互運用性を提供するような、構造化された Web ベースのオントロジーを定義する単一の言語を開発することである。

# 2.3.2 WOWG のオントロジー記述言語の動向

W3CのWOWGでは、オントロジー記述言語の標準化検討を実施している。W3Cにおいて検討されているオントロジー記述言語としては、2001年12月にノートとして提出されたDAML+OILと、2002年1月のFaceToFace Meetingで提案されたDAML+OILの後継言語であるOWL(The Web Ontology Language)がある。OWLには、まだ、詳細な仕様の議論がなされていないので、本資料では、(1)でDAML+OILの仕様概要について53説明した後、(2)では、現時点でのOWLの概要について54述べる。

<sup>49</sup> http://www.w3.org/Metadata/

<sup>50</sup> http://www.w3.org/Consortium/

<sup>51</sup> http://www.w3.org/TandS/

<sup>53</sup> Fikes,R and McGuinness,D,: An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL (March 2001) 18 December 2001, Richard Fikes, Deborah L. McGuinness (DAML+OIL Web Ontology Language Submission) http://www.w3c.org/TR/daml+oil-axioms、Horrocks,I, Harmelen,F, and Patel-Schneider,P 編集, A Model-Theoretic Semantics for DAML+OIL (March 2001) http://www.w3c.org/TR/daml+oil-model、及び Horrocks,I, Harmelen,F, and Patel-Schneider,P 編集, DAML+OIL (March 2001), http://www.daml.org/2001/03/daml+oil-index.html (2001)を参考にした。
54 Peter F. Patel-Schneider: The Web Ontology Language (OWL)

http://lists.w3.org/Archives/Public/www-Webont-wg/2002Jan/att-0061/01-swol.textを参考にした。

# (1) DAML+OIL

DAML+OIL は、DARPA からの資金援助を受けた DAML プロジェクトで開発されていたオントロジー記述言語 DAML-O と、EU 委員会 IST プログラムからの資金援助を受けた On-to-Knowledge プロジェクトで開発されていた OIL (Ontology Inference Layer) を統合した言語である。2001年10月13日時点での最新版は、2001年3月の Ver4.1 である。

DAML+OIL は、RDF スキーマ(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、RDFS に対して、「クラスのブ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。以下、順に説明する。

#### クラスのブール結合演算や排他的関係

クラスに対する以下の演算や関係を表現することができる。

- ・二つのクラスのいずれかに属するというクラスを定義する演算(unionOf)
- ・二つのクラスのいずれにも属するというクラスを定義する演算(intersectionOf)
- ・あるクラスには属さないクラスを定義する演算(complementOf)
- ・二つのクラスに共に属する要素がないという関係(disjointWith)
- ・二つのクラスが等しいという関係(sameClassAs)

図 2-4に、DAML+OIL の記述例を示す。以下、各々の記述例を順に説明する。

- ・クラス Women の定義では、「女性(Women)は、男性(Man)でない人(Person)である」ことを意味する。
- ・クラス TallMan の定義では、「身長の高い男性(TallMan)は、高いもの(TallThing)かつ男性(Man)である」ことを意味する。
- ・クラス HumanBeing の定義では、「クラス HumanBeing とクラス Person は等しい」ことを意味する。

<daml:sameClassAs rdf:resource="#Person"/>

</daml:Class>

#### 図 2-4 DAML+OIL の記述例 1

# プロパティに対する演算や制約の拡張

RDFS では、プロパティのドメインとレンジしか規定できなかった。すなわち、プロパティをとるリソースや値が属するクラスしか規定できなかったが、プロパティに関する制約記述の向上を目的として、以下に示すような拡張がなされている。

- ・プロパティの値そのものを規定する制約(hasValue, hasClass)
- ・プロパティの値をいくつとることができるかを規定する制約(cardinality, minCardinality, maxCardinality)
- ・プロパティが推移律を満たすという制約(TransitiveProperty)
- ・プロパティのドメインの各々の要素に対して、その値が一つしかないという制約 (UniqueProperty)
- ・二つのプロパティが等しいという関係(sameProperAs)
- ・プロパティのドメインとレイジを交換したプロパティを作る演算(inverseOf)

以下、図 2-5に、DAML+OIL の記述例を示す。以下、各々の記述例を順に説明する。

- ・クラス Animal の定義では、「動物(Animal)は親をもち(hasParent)、かつ、親の数は2つである」ことを意味する。
- ・プロパティ hasChild の定義では、「プロパティ hasChild(子供をもつ)と、プロパティ hasParent(親をもつ)は逆の関係である」ことを意味する。
- ・プロパティ hasMom の定義では、「プロパティ hasMom とプロパティ hasMother は等しい」ことを意味する。

<daml:ObjectProperty rdf:ID="hasMom"> <daml:samePropertyAs rdf:resource="#hasMother"/>

</daml:ObjectProperty>

#### 図 2-5 DAML+OIL の記述例 2([3]より)

# (2) OWL(The Web Ontology Language)

OWL は、W3C の WOWG において、Peter F. Patel-Schneider を中心として検討され ている言語である。また、WOWG の検討フェーズとしては、要求仕様を明確にしている 段階と思われるが、以下では OWL の概要について述べる。

OWL の DAML+OIL からの基本的な変更点は、以下の3つである。

- (1) OWL の意味は、model theory for RDF55と矛盾しないようにする。
- (2) OWL の構文は、the XQuery 1.0 and XPath 2.0 Data Model<sup>56</sup>と矛盾しないように する。
- (3) データ型の取り扱いを単純にする。

意味の変更の背景にある基本的なアイデアは、サブクラスのような意味論の観点から扱 うべきコンストラクトを構文としてもつ RDF アプローチを採用するだけでなく、解釈に おける関係を明確にすることにある。しかしながら、意味的な矛盾を引き起こすような言 語全体の拡張は避けなければならないので、意味的な矛盾が存在しないことを確信できる ようにするという観点から、解釈における関係を明確にできるコンストラクトの数を制限 する方針とのことである。

また、構文は XQuery Data Models なので、XQuery の標準的な実装を用いることによ り、XML Schema のデータ型のほとんどすべてを提供できるようになるとしている。

<sup>55</sup> Patrick Hayes , RDF Model Theory 14 February 2002, http://www.w3c.org/TR/rdf-mt/ 56 Mary Fernández, Jonathan Marsh, Marton Nagy: XQuery 1.0 and XPath 2.0 Data Model W3C Working Draft 20 December 2001. http://www.w3c.org/TR/query-datamodel/

## 2.3.3 OWL の利用事例

## (1) コレクション・マネージメント

コレクション・マネージメントは大規模なデータ集合体(テキスト、イメージ、マルチメディア、Webページ等)の管理のことで、ドメインへの依存性やスケーラビリティの問題を含んでいるため、OWLの有望な応用分野と考えられている。WOWGの2002年1月のFace-to-Face Meetingでは、次の5つの具体的な事例が紹介された。

Use Case 1: "ARKive: catalog of endangered species descriptions"

ARKive<sup>57</sup>は英国 Wildscreen Trust が構築している絶滅危機生物のマルチメディアデータベースで、学校の生徒からこの分野の専門家までが、Web サイトを通じてアクセスできるようになっている。実際のデータベースシステム開発には HP 社が参画している<sup>58</sup>。

Use Case 2: "EDS Web page landfill"

企業の社会とのコミュニケーションやコーポレートメモリの管理に関するもの。具体的な事例は、EDS 社が製品化している PLM(Product Lifecycle Management)ソリューションで、新聞発表、製品提供とケーススタディ、企業内手続き、製品説明と比較、ホワイトペーパー、プロセス記述等を対象としている。EDS 社の代表的な製品には Teamcenter 59 がある。

Use Case 3: "Aerospace Engineering Data Modelling"

航空機産業における企業の社会とのコミュニケーションやコーポレートメモリの管理に関するもの。特に、製造工程管理が中心で、航空機設計ドキュメント、製造工程ドキュメント、テスト工程ドキュメント、メンテナンスドキュメント、イラスト等を対象とし、設計エンジニアが利用する。具体的な活用製品としては、Stilo Technology 社及び子会社のOmniMark Technologies Corporation 社が提供する Content Engineering ソリューション OmniMark®がある。

Use Case 4: "Art-image collections"

絵画コレクションのデータベース構築において、オントロジーを利用して、データのインデクシングと検索をサポートすることを目的としたもの。オランダのUniversity of AmsterdamのIBROWプロジェクト<sup>61</sup>(Web上での情報ブローカリング技術の開発が目的)で取り扱っている応用事例。この分野では規約化に関する多くの既存のリソースがあり、ここでもVRA(Visual Resources Association<sup>62</sup>)で定めているVRA Core Categories, Version 3.0や、ゲッティ財団によって構築されたAAT(Art and Architecture Thesaurus、芸術作品(芸術カテゴリ、材料、スタイル、カラー等)を記述するための120,000用語の

<sup>57</sup> http://www.arkive.org.uk/default.asp

<sup>58</sup> http://www.hpl.hp.com/news/arkive.html

<sup>59</sup> http://www.eds.com/products/plm/teamcenter/

<sup>60</sup> http://www.stilo.com/casestudies/cs005.html

<sup>61</sup> http://Web.swi.psv.uva.nl/projects/ibrow/home.html

<sup>62</sup> http://www.vraWeb.org/index.html

階層から成る)などを利用している。

Use Case 5: "Conceptual Open Hypermedia"

学術機関が構築する研究成果データベースのような巨大なハイパーメディア / Webページをブラウジングする際のナビゲーションを対象としたもの。英国University of SouthamptonのAKTプロジェクト<sup>©</sup>で取り扱っている事例を参考にしている。

これら 5 つのコレクション・マネージメント事例の分析から、OWL に対する要求事項 として、次のような項目がリストアップされている。各項目の右側の数字は、上記のどの 事例から導かれた項目であるかを示している。

1) 表現力に関する要求	八事リ	텻
--------------	-----	---

- ・Class hierarchy with properties/attributes and relations ( attribute と一般的な relation との明確な区別
- Default knowledge ( )
  - ・生き物は飛ばない
  - ・後期ジョージ王朝風の整理だんすは通常マホガニー製である
- Part-whole relation ( )
  - ・翼の桁は翼の一部分である
  - ・整理だんすは独自のスタイルの脚をもつ
- · Constraints ( )
  - · wing-spar.length < wing.length
  - urniture.style = Late-Georgian <=> furniture.culture = British AND furniture.date-created 1760-1811
  - ・利用事例 4 における後の例に対する DAML+OIL の解決策についてのコメントにも注意。
- Instance specification / instances as classes (
  - ・Mammal は Species のインスタンスであるが、またそれ自身クラスでも ある。
  - ・A380 は Aircraft のインスタンスであるが、また A380 のインスタンス

<sup>63</sup> http://www.iam.ecs.soton.ac.uk/projects/akt/

集合も表している。

- ・参照: Protege-2000 の metclass 概念。注: これは RDF の特徴である。
- ・属性値による匿名インスタンスの仕様。
- ・Mechanism for relation typing ( ) さらに、RDF feature (property のために rdf:type を定義することができる。)
- ・Abstract classes() インスタンスのないクラス: 応用事例4の color の階層の例を参照。
- ・Synonyms / lexical term -> concept ( )
  用語とそれが表すコンセプトを区別する必要がる。 [ これは OWL の問題ではないかもしれない。 ]

## 2) 他の要求事項

- Provenance (1, 5)
   典型的な例: 専門家と非専門家による表記を区別すること、ハイパーリンクによる出典参照。
- ・Version management (1)オントロジーを拡張・変更する能力。
- ・Query support (1) 偽の仮定による推論を行なう能力、例えば「魚としての鯨」。
- ・Support for content standards (1, 3)

  一般的なシソーラス(WordNet、TGN)が、ドメイン固有のシソーラス(AAT、ICONCLASS)と同様、しばしば表記を標準化するために使われる。

## (2)一般的な利用事例

2002 年 1 月に W3C にて行われた WOWG Face-to-Face Meeting において議題となった一般的な利用事例を以下に説明する。なお、この利用事例はドラフトであり、今後はこれらの場面において OWL の利用が期待できるとするものである。

## 1) デバイス相互運用性

計算機能のあるデバイスでは、デバイスに意味が分かる(セマンティック)レベルで相互運用ができなければならない。こういったデバイスには、PC やラップトップ、PDA、プリンタやその他ネットワークデバイスなどが考えられる。これらのデバイスは旅行の計画などのような同じ目的で使用できる。例えば、ドメインへ適

切にアクセスする場合、オントロジーの同じ集合を使用できるが、デバイス特有の限定された低域幅や制限されたメモリなどを含む特定の要件にマップされたり、カスタマイズされたりするべきではない。

また、デバイスは動的に変化する環境に対してもセマンティック的に採用するべきである。例えば、新しいプリンタを規定のデバイスのネットワークに追加する場合や、PDAが異なる計算や、アクセス、そしてセキュリティー要件がある新しい場所に導入された場合などである。

更に、デバイスは Web サービスによって制御可能である。この Web サービスはオントロジーを使用してタスクの分解とコントロールフローを表示することできるようになる。

## 2) インテリジェントなエージェント

さまざまなエージェント基準にはオントロジー言語の利用が必要である。この言語は(複数のオントロジーからの)オントロジー表現交換と(あるオントロジーか他のオントロジーへの)オントロジー翻訳をサポートしなければならない。さらに、交換や翻訳されるオントロジーは異なるエージェントプラットフォーム、インフラストラクチャー、あるいはドメインで使用できる。

また、交換/翻訳されるオントロジーは異なる言語で表現可能であるし、異なるレベルであることもできる。例えば、上位オントロジーや下位のドメインオントロジーなどである。複数の上位オントロジーや上位オントロジーの一部分が交換/翻訳、そして、判断されなければならない場合もある。同じ個人/インスタンス/データが異なるオントロジーにマップされなければならない場合もある。また、SQL、RQL などの場合、照会言語間でマッピングされる必要がある場合もある。

## 3) インテリジェントな管理

ビジネス組織、財閥のWebコミュニティーといった場面では、複数の組織の単位でWeb上でお互いに通信を行わなければならない。これらの単位には異なるOM(Organizational Memories)や、知識リポジトリが含まれることがあり、複数のデータベースや文書のセットをサポートしなければならないこともある。これらのセットはすべてこの異なるオントロジーで表示することができる。組織に対してさまざまなレベルがあることがあるため、データ/情報固有のドメインから大企業のレベルまで、さまざまな情報のモデルとそれに相当する語彙がある。

さらに、組織と組織がサポートしているオントロジーは動的に変化することがある。組織の中には、例えば財閥のコミュニティーなど、仮想的であり他の既存の組織の合成や分離であったりすることもある。しかし、こういった組織や組織の単位は、オントロジーマッピングを使用して、異種の語彙やモデルとセマンティック的に相互運用ができなければならない。必要であれば、これらのマッピングを見つけることができ、保持・更新も可能となり、破棄もできることがある。

#### 4) インテリジェントな検索、電子商取引

Web 上の B2B、及び B2C の電子商取引では、買い手はアプリケーションやカタログ、新旧の B2B または B2C エンジンやインフラストラクチャーのさまざな範囲から将来の製品やサービスを探し、見つけ出すことができる。 こういったリソースはすべて、分類や分類学、あるいはオントロジーによってモデル化されることがある。結局、それらはオントロジー言語を使用してモデル化されるのである。

従って、問題となる点は、Web 上に異質にモデル化されたり、明示的にまたは 黙示的に分類されたりする製品やサービス(そして、他のドメインのオブジェクト) をユーザーが検索できること、及びこれらのモデルや分類化システム、またはオン トロジー間を概念的にマッピングすることである。

将来の買い手は、個々の会社のカタログや文書にさまざまに表示されている語彙、及び意味を使用して製品やサービスを探したいと考える。一方で、将来の売り手は、将来の買い手が探すことができるように、製品やサービスを分類し、適切にモデル化するのである。このような概念的な検索/ナビゲーション(分類)には、セマンティック的にマップされるオントロジーが必要である。

## 5) インテリジェントな統合(異種のオントロジー情報の統合)

Web や Web にアクセス可能なアプリケーションに表示されている情報は、ドメイン毎や、アプリケーションが解決しようと試みている問題毎、あるいはモデルの複雑さの度合いに応じてさまざまな方法でセマンティック的に表示されることがある。オントロジーの概念とその概念が特徴付けている個人、インスタンス、データにはセマンティック的な合成・統合・分離の方法、及び、対応する構文上の形式の変換が数多く必要である。

この使用事例はインテリジェントなエージェントの使用事例と関連している。

## 6) 相互評価

例えば、国防総省の武器、ネットワーク、コマンドと制御システムといったドメインの例においては、多くの場合、特定基準のセマンティックモデル(オントロジー)の基準や適合検査の実現に関してその特定の基準のセマンティックモデル(オントロジー)において難しい点がある。これらの重要な三つのコンポーネントもすべて、お互いに相互関係なしに動的に変化するものであり、原則では、モデルで強制されべきである。基準と一致していると称するアプリケーションは、作成可能な一連の適合検査によって、その適合に関して検査すべきである。これには、モデルと基準と一連の基準の検査間でのセマンティックマッピングが必要である。

さらに、複雑な点としては、基準には異なるバージョンがあるので、モデルとその一連の適合検査には異なるバージョンがあることを想定すべきということである。

#### 7) 文脈と表示マッピング

センサー検知、 CAD 技術、マルチメディア作成といった場面においては、アプ

リケーションは、文脈や表示を元にオントロジーの情報をフィルタリングしたり、表示したりすることによって、ユーザーへのコンテンツ表示を合わせる必要がある。センサー検知と CAD 技術は、ユーザーが関心を持つ特定の表示を表すために文脈上の情報に適合する場面である。これらの表示は、例えば個人化やローカル化といった場合、アプリケーション表示の要件、または、ただ単に与えられた時間、あるいは照会への回答において、オントロジーの中の特定面に重点を置くという要求をもとに抑制やフィルタリングをする必要がある。こういった表示や文脈は合成されたり、操作されたりする必要があるため、オントロジーのアプリケーション理論として持続し、主要なものとなる。

#### 8) 相互運用性

コマンドとコントロールを目的とした防衛関連のアプリケーション、軍事派遣計画、ロジスティックスなどは、サービスと組織ユニット間でマルチドメイン情報を使用し、普及することが必要である。多くの場合、オントロジーは明確で、セマンティック的にほとんど同じ概念に対しても異なる用語や意味を使っている。さまざまなサービスや組織間で命令や情報が行き交う場合や、異なる目標を達成するために同じ情報をしようする場合、セマンティックをできる限り保持するために異なるオントロジーをマップする必要がある。

# 2.4 活用事例と課題

セマンティック Web 技術、特にオントロジー技術の活用事例と課題について述べる。オントロジーの活用事例に関しては、着手が始まった段階で、事例といっても研究の方向性を示すものや、プロジェクト提案レベルのものに留まっている。現時点では、セマンティック Web の有用性の実験とそのために必要な要素技術の蓄積を行なっている状況で、今後の調査・検証が必要である。調査は主にプロジェクトを中心に行なったが、EU 委員会の PF5 の IST にセマンティック Web 関連プロジェクトがあるため、欧州のものが多くなっている。

セマンティック Web 技術の活用の方向性としては、以下のものが見られた。

- 1) ナレッジマネジメント
- 2) e-learning
- 3) e-ビジネス
- 4) コミュニティー支援

一つは、知識管理の直接的な応用として、ナレッジマネジメントがあげられる。以下の事例では明確に示せていないが、統一されたオントロジー/システムを用いることができる企業内での応用と、その後に、個々のオントロジーを持った企業間同士でのマッピングの問題などがある。また、e-learningでは、多種多様な教材の蓄積を活用するためにセマンティクWebやメタデータの技術が用いられている。

e-ビジネスへの適用は、オントロジー技術を用いて、取引関係の価値連鎖 (バリューチェーン)を記述したり、部品のメタデータの組み合わせから顧客に適した製品を作る際の、カタログの自動生成や製品の自動分類などに活用しようとしている。

また、知識交換の場としてのコミュニティーに関しては、セマンティック Web 技術を用いる

## 2.4.1 ナレッジマネジメント

ここでは、フランス INRIA のナレッジマネジメントシステムである ACACIA とその関連研究、及びナレッジマネジメントの一環としてドイツのダイムラークライスラー社の人材支援システム (HRMore) におけるセマンティック Web の活用事例について述べる。

#### 2.4.1.1 ACACIA

ナレッジマネジメント関連の研究では、フランスの 国立情報処理自動化研究所 (INRIA)における ACACIA (Knowledge Acquisition for Explainable, Multi-Expert Systems) <sup>64</sup>プロジェクトがある。ここでは、ACACIA プロジェクトと関連技術研究の概要について述べる。

ACACIA は、企業情報構築・管理・普及のために、暗黙知を形式知として具現化し、情報の共有と再利用を目指す、ナレッジマネジメントシステムである。このためのセマンティック Web のアプローチとしては、まず、企業情報の蓄積のために XML 文書を作成し、オントロジーに基づいた意味的な夕グ付けを行ない、Web 上で配布する。タグ付けは RDF

43

<sup>64</sup> http://www.inria.fr/recherche/equipes/acacia.en.html

を用いて行なうが、その蓄積と検索のために、CORESE (Conceptual Resource Search Engine)という検索エンジンを開発している。これは、RDF の検索エンジンのプロトタイプであり、グラフ表現への変換や RDF 表現の検索を可能にする。もちろん、グラフ表現をそのまま検索の質問として入力することも可能である。推論エンジンはグラフ推論に基づくものを採用している。(図 2-6参照のこと。)

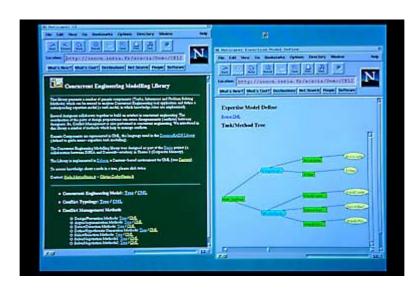


図 2-6 ACACIA における知識記述の例

次に、マルチエージェントの役割を担う、 $CoMMA(Corporate\ Memory\ Management\ through\ Agent)$ がある。 $CoMMA\ はオントロジーの作成・管理と\ CORESE\ のユーザーインターフェイスとなるマルチエージェントシステムから成る。これらのシステムを用いて新入社員用の企業情報の提供の応用システムを作成している。$ 

さらに、知識表現やオントロジーはさまざまな目的で多種につけることが可能である。これらを再利用して新しい目的に活用することが考えられる。知識表現やオントロジー、タグ付けした情報の表現法として、いろいろな表現方法が考えられるため、それらの言語間の変換が必要となる。これを行なうのが EXMO(Computer mediated exchange of structured knowledge)<sup>65</sup>である。

## 2.4.1.2 HRMore

ビジネス・インテリジェンス関連の適用事例としてダイムラークライスラー社の HRMore<sup>66</sup>がある。HRMore プロジェクトでは、人材支援システム(Human Resource System)への適用を試みている。HRMore では、スキル活用にあたって、手書きの Job Application を OCR で読み取った記述内容やその他のドキュメントを HR システム用のオントロジーに基づき分析し、スキル情報を抜き出している。そのスキル情報を元にイント

44

<sup>65</sup> http://www.inria.fr/recherche/equipes/exmo.en.html

<sup>66</sup> http://wim.fzi.de/wim/projects/hrmore

ラネット上で、スキルの需要と供給のマッチングを行なう。実験システムは、OntoPrise 社67の OntoBroker を用いており、2000 年 11 月時点で 27 個の概念、56 個の関係、4 個の規則(公理)からなる ProPer オントロジーが公開されている<sup>68</sup>。ここでの対称となる「概念」は、「ドキュメント」やそのサブクラスである「プロジェクト報告書」や「HomePage」、その他、「人」や「分類」などの概念を定義している。例えば、「ドキュメント」は、ストリングタイプの値を持つ「URL」というプロパティや、「記述言語」、「作成日」、「著者」、「関連するスキル」、「関連する分類」などのプロパティを持つ。関係は、「ドキュメント」とその「URL」、「ドキュメント」とその「記述言語」といった関係を定義している。公理は、例えば[「HomePage」とその「著者(has\_autor)」プロパティの値との関係と「作業者」とその「所有 HomoPage(has\_Homepage)」プロパティ値との関係は、逆関係(109頁の脚注<sup>117</sup>「反対の関係」参照)になっている(「ドキュメント」Dの「著者」プロパティが Pであれば、「作業者」Pの「所有 Homepage」の値は Dである)] などの規則を定義している。

## 2.4.2 e-learning

e-learning は IT 技術を用いて教育を行なう事である。教育用メディアを電子化することによりネットワークを通した配布や遠隔授業などが行なえる。教育用メディアは多種複雑であり、現状包括的で有効な内容検索手段®は存在しない。このため、メタデータの付与が期待される分野である。

#### 2.4.2.1 PADLR(Personalized Access to Distributed Learning Repositories)

e-learning への適用も試みられている。PADLR(Personalized Access to Distributed Learning Repositories)<sup>70</sup>と呼ばれるプロジェクトでは、セマンティック Web の技術を用いて、複数のモジュールから成る教育用メディアの交換基盤の構築を試みる。年度の関係でドイツとスウェーデンでは 2001 年 4 月から、米国スタンフォードでは 2001 年 9 月から始まった 2 年間のプロジェクトである。

現代では、科学技術の革新により、生活のさまざまな面においても学習することが要求される。このような環境の中で、教員や教材の不足が起こっている。特に学習者に合わせた教材を的確に効率的に提供することが求められる。このような背景を踏まえ、各機関にある教材などのリソースを相互運用することが重要となる。

実際、大学などの多くの高等教育機関では、教材など教育上のリソースを蓄積して効率 化が図られている場合が多い。しかしながら、これらの教育上リソースの管理は、個別に 行われることが多く、各大学などの機関がそのコントロールを手放すことはなく、単一の 中央サーバに蓄積する形態は実現しないという。

このために、各組織が管理している教育上のリソースを相互運用 (活用) する必要があ

<sup>67</sup> http://www.ontoprise.com

<sup>68</sup> http://ontobroker.semanticWeb.org/ontos/proper.html

<sup>&</sup>lt;sup>69</sup> 例えば、テキストメディアの内容検索としては全文検索があるが、画像を含むと適切な内容検索手段が存在しない。教育用メディアは基本的にマルチメディアであるので、有効な内容検索手段が存在しない。

<sup>70</sup> http://www.learninglab.de/pdf/L3S\_IWF-cori-4b.pdf

る。教材の共有に関連する話としては、「オープンアーカイブイニシアチブ」がある。そこでは最近、ライブラリー・サーバーからディジタル・ドキュメントに関するメタデータ情報を検索するために HTTP ベースのプロトコルを定義した。しかしながら、それはサーバー検索プロトコル及びその基本的なスキーマとしてはダブリンコアに基づいたもの(教育用に拡張した意義は大きいが)に留まっており、十分とはいえない。

PADLR では、ピア・ツー・ビア (P2P) ネットワークに基づいたアプローチで、教育メディア交換用のネットワークを構築しようとしている。しかしながら、ナップスター (Napstar)や Gnutella のような純粋な P2P アプリケーションのアプローチでは不十分である。なぜなら、例えば、Gnutella の中のメタデータはファイル名及びパスに制限されており、「マドンナの Like a Vergin」のようにタイトルが付いて(その文字列が世間一般に知られて)いるファイルにはうまく動くかもしれないが、「代数入門、 講義 23」のようにタイトル文字列だけでは中身がわからないファイルではうまく動かないからである。このため、PADLR では、P2P ネットワーク基盤として、Edutella と呼ぶオープンソースのシステムを開発 / 採用し、その基盤上にオントロジー技術の活用を試みる。

# 2.4.2.2 ETB (European Treasury Browser)

イタリアにある EU 委員会共同開発センター(JRC)では、セマンティック Web 技術 の応用システムとして ETB ( European Treasury Browser ) 71という名のプロジェクトに おいて、先生が学校で教えるための教育プログラムを蓄積したデータベースをつなげて相 互利用をはかっている。ETB の目的は、欧州域内の学校のための教育用情報資源のための メタデータネットワーク基盤を構築することである。5~6年前に XML ベースの教育用オ ーサリングツールを開発しており、これを利用して、学校にコンテンツの配信を行なって いた。初期のものは、NetNews システムを用いて配信されていた。教育用情報資源の蓄 積は、これ以前のものも含め古くからあり、欧州域内に 5~6 個の主要なリポジトリデー タベースを構築されていたが、歴史的経緯より管理方法は統一されていなかったし、メタ データに抜けも存在する。スキーマの異なるデータベースのためには、RDF を使って項目 の相互関係を定義している。教育関係のシソーラス (オントロジー)については、12 言語 (EU以外の言語も含む)から成る約1,000語のシソーラスをRDFを用いて作成している。 ETB 向けのメタデータ標準はダブリンコアを拡張して利用している。これにより、例えば、 英語とイタリア語の同義語をリンク付け、各々の言語から他の言語の同義語を引き検索に 利用できる。ETB 自体の最新版は、2002年1月現在、英語、フランス語、ドイツ語、イ タリア語、スペイン語、スウェーデン語、デンマーク語、ギリシャ語の8つの言語に対応 している。

-

<sup>71</sup> http://etbrowse.jrc.it

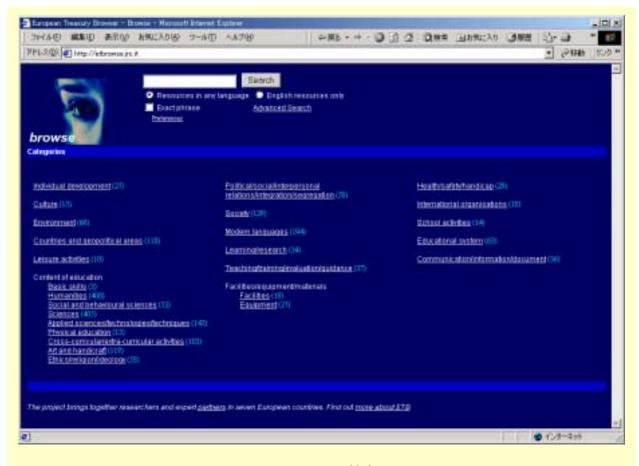


図 2-7 ETB の検索画面

図 2-7は、ETBの検索画面である。検索語として「Music」を入力し search ボタンを押すと、図 2-8のように検索結果が表示される。上部には、music を含むシソーラスの用語として classical music, contemporary classical music, modern classical music, music, music education, music listening などがあることがわかる。

JRC では RDF Site Summary (RSS)  $^{72}$ とその Viewer、さらに RDFStore の開発も行なっている。RSS は、手で書いたクエリー (RDQL/Squish $^{73}$ という独自の形式:現時点では自動的には生成できない) から Web サイトの検索を行なうものである。Web サイトのRDF のデータはそれぞれに人間がつけて (自動化はこれから) RDFStore に格納し、検索ができる。現在はデモレベルで実用レベルではないが、セマンティック Web の応用システムを作って動かしている点は注目に値する。(図  $^{2}$ - $^{9}$ 参照のこと。)

47

<sup>72</sup> http://java.jrc.it/rssfeeds/doc/scraper.html

<sup>73</sup> http://rdfstoredemo.jrc.it/rdql/

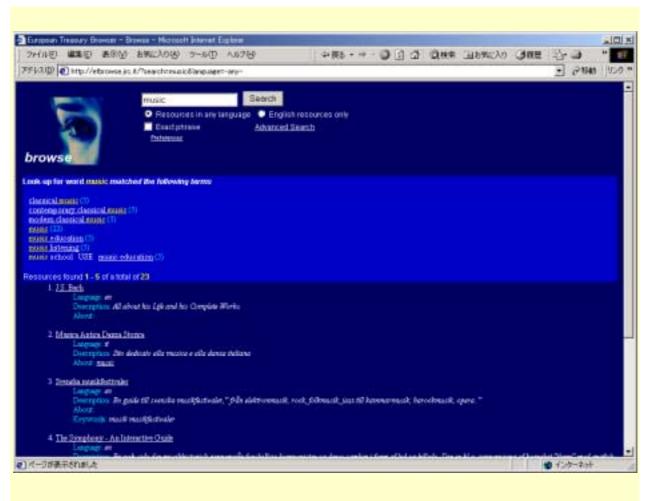


図 2-8 ETB の検索結果

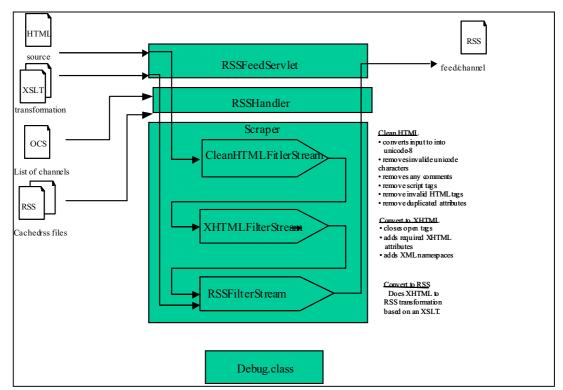


図 2-9 RSS の構成

#### 2.4.3 e-ビジネス

セマンティック Web の目的の一つが、Web 上の情報のマシンによる理解と、それによる自動化とすれば、それをビジネスに展開することも考えられる。特に電子化された製品とその流通過程で、メタデータを活用することにより、どのような事が行なえるのかは興味深い。ここではそのような観点から、e-ビジネスへの適用の試みを行なっているOBELIX<sup>74</sup>プロジェクトについて述べる。

#### 2.4.3.1 OBELIX: e-ビジネスへの適用の試み

OBELIX<sup>74</sup>は、EU 委員会の PF6 の IST プログラムに含まれるプロジェクトの一つである。スペインの LABEIN<sup>75</sup>をコーディネーターとした 2002 年 3 月から 2004 年 8 月末までのプロジェクトであり、従って、2002 年 1 月の調査時には未だ提案段階のものであった。

OBELIX では、XML などの標準化により整備されつつある基盤の上に、よりスマートな e-ビジネスを構築しようとしている。特に、インターネット時代の電子商取引や協働作業には、動的なバリューチェーン(DVC: Dynamic Value Chain)が構成される。オントロジーの技術を活用して、この DVC をモデル化し、解析することを試みようとしている。特に、電子化された複合的な製品(multi-product)に着目し、個々の部品に付与されたメタデータを、それらを組み合わせた製品に対する分類、設計、流通や解析などに利用しようとしているようである。

OBELIXでは、以下のものを開発する。

\_

<sup>74</sup> http://www.ontoWeb.org/workshop/ucecOntoWeb/obelix.ppt

<sup>&</sup>lt;sup>75</sup> http://www.labein.es/eng/index.htm

- (1)汎用のオントロジーサーバー及びオントロジーアプリケーションツール
- (2)複数コンポーネント製品の分類ツール
- (3)複数コンポーネント製品のコンフィギュレーションツール
- (4)e-ビジネスのモデリングとシナリオ解析ツール

その主な応用としては、以下の4つのものがあげられている。

- 1)デジタル音楽のバリューチェーン
- 2)エネルギー取引
- 3)イベントのオンラインでの企画支援
- 4)エネルギーサービス

例えば、デジタル化された音楽は、インターネット時代の e-プロダクトの典型的な例であり、マーケティングから販売、流通に至るまで電子化することが可能である。電子化されたことによりこれらの機能の担い手は再編成される。一方、著作権などの権利も付随する。ここでは、権利のうち二次利用を中心に検討する。その場合、音楽も通常は(権利者の異なる)複数の音楽を利用するので、複合的な製品と考えることができる。即ち、動的なバリューチェーンと複合的な製品の例と言える。

OBELIX では以下の3つの課題に対する革新的な解決策を開発することにより、最先端の技術の開発を行なうとしている。

- 1)オントロジー及びセマンティック Web 技術で Web 上の情報資源をどのようにシステムが理解できるようにするか?
  - 例えば、XML 文書電子商取引で製品が取り扱い可能になる中で、製品内容情報の自動分類やカタログ化など。
- 2 )e-ビジネスにおける Web コラボレーションの新しいアプローチをどのように促進するか?
  - 例えば、イベントや他のサービスコンフィギュレーションの設計をオンラインで 協働的に行なうためのシナリオなど。
- 3)新しい価値の組合せ(value constellation)をいかにモデル化し、解析し、シュミレーションするか?

例えば、カスタマイズされた音楽製品とその権利の提供や、電力の需要側の e-マーケットプレースでの入札など。

#### 2.4.4 コミュニティー支援

Web 技術を活用したコミュニティーの支援は、コミュニケーションから距離と時間を取り除くものとして多くのコミュニティーサイトが Web 上に存在することからも明らかなように、魅力ある Web 技術の適用例である。これらを、セマンティック Web 技術を活用して、さらに使いやすくするのは自然な流れである。ここでは、Web サイトへのオントロジー技術の試みとしてカールスルー工大学 AIFB の Web サイトとオントロジー研究コミュニティーの Web サイトである OntoWeb について述べる。

#### 2.4.4.1 AIFBのWebサイト

カールスルー工大学の AIFB の Web サイトでは、SEAL(SEmantic potAL) オントロジーと呼ばれるオントロジーが用いられている。このオントロジーは、54 個の概念、178 個の関係、4 つの規則(公理)からなっている。概念としては、研究領域や学生、プロジェクト、組織、出版物などがあり、それら概念と属性の関係が 178 個定義されている。公理系は、例えば、「あるプロジェクトのメンバーであることを記載することは、そのメンバーの所属プロジェクトにそのプロジェクトが含まれていることである」といった規則が定義されている。これらはスキーマを表しており、実際の人物やプロジェクトはこれらのインスタンスとして表現される。

これらのオントロジー定義に従い、AIFBのサイトでは、人(プロジェクト)に対して、研究領域や、プロジェクト(人)、研究グループのメニューを持った検索ページが構築されている。特に研究領域は、概念階層(クラス階層)に従い、階層的に表示されている(図2-10参照のこと)。表示文字は、属性値になっているので、ドイツ語と英語の二ヶ国語がオントロジーで定義されており、言語を切り替えることによりメニューの表示も変更される。

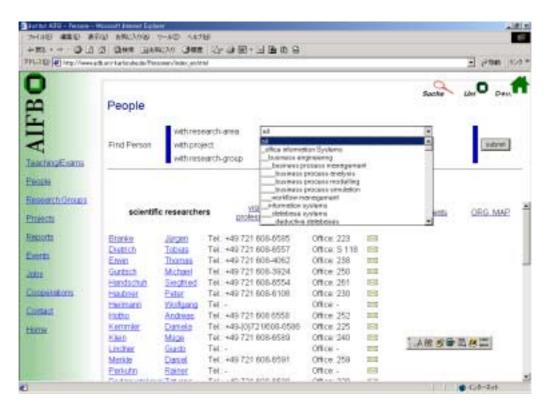


図 2-10 AIFB の研究員検索のページ<sup>76</sup>

\_

<sup>76</sup> http://www.aifb.uni-karlsruhe.de/Personen/index\_en.html

## 2.4.4.2 OntoWeb: セマンティク Web 技術のコミュニティーポータル

OntoWeb<sup>77</sup>は、セマンティック Web 技術のコミュニティーのためのポータルである。成果の公開や情報交換により、セマンティック Web 技術の研究開発の促進と、普及を目的にしている。2002年 2 月現在、まだセマンティック Web 技術をそれ自身に適用してはいないようであり、カールスルー工大学の AIFB Web サイトにあるような検索システムはないが、SWRC (Semantic Web Research Community) ontology と呼ばれるオントロジーが作成されている<sup>78</sup>。(図 2-11参照のこと。)

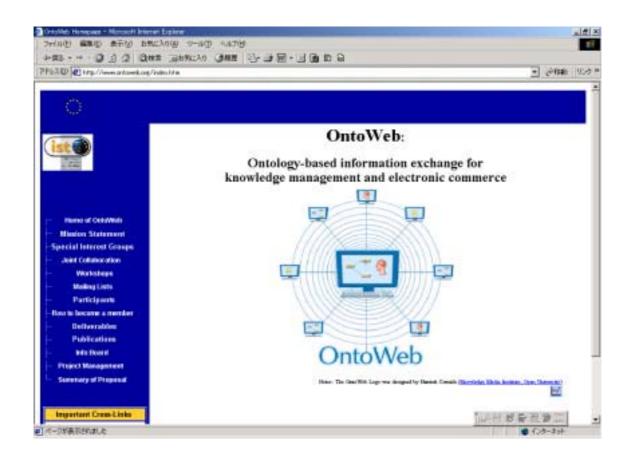


図 2-11 OntoWeb の Home Page

## 2.4.5 課題

これらの調査を通じて、得られた課題を以下に挙げる。

#### (1)メタデータの作成・管理方法

メタデータの付与には、多くの場合かなりの労力が必要となる。このために、3.2で 述べられるようなツール類のさらなる整備や、メタデータ付与者への動機付けの工夫など

<sup>77</sup> OntoWeb: http://www.ontoWeb.org/

<sup>&</sup>lt;sup>78</sup> SWRC: http://ontobroker.semanticWeb.org/ontos/swrc.html

が必要となる。具体的でわかりやすいメタデータの適用事例が望まれる。

## (2)大規模データ・知識への適用

いくつかの事例でオントロジーも作成されているが、まだサイズも小さく、実験レベルに留まっている感がある。それに基づく検索システムも同様であり、より大量の情報を扱う実用レベルの問題への適用が必要である。今後は、実用的なメタデータを使うことによって、コンテキストを考慮した検索が本当に可能かどうかを質・量の双方の観点から検証する必要があるだろう。セマンティック Web の潜在能力は高いが、応用システムが実用レベルに達するのは、ソフトウェアツールが揃うことが大切であり、研究レベルを産業レベルに引き上げることが必要であろう。

# 2.5 応用システム開発時の考慮点

セマンティック Web を活用した応用システムには、どのようなことを考慮する必要があるだろうか。セマンティック Web を活用した応用システムの多くがメタデータを使ったシステムである。ここでは、メタデータを使った電子政府サービスに限定して応用システムの考慮点を考えていきたい。

第一に考えなければならないことは、電子政府サービス用のメタデータ標準を作ることであるが、全く独自のメタデータ標準を作るべきではない。現時点で、電子政府サービスにおける世界標準メタデータ仕様なるものは、存在しない。しかし、海外各国の政府サービスの電子化におけるメタデータ標準を見渡したとき、3.1.2節で見るように、ダブリンコアを元にしたメタデータ標準を使用している国が多い。また、ダブリンコアを元にしたメタデータ標準を使用していない国でも、ダブリンコアのエレメントを個々のメタデータ標準のエレメントにマッピングすることが可能である。新たに作るメタデータ標準は、こういった現状を踏まえて、デファクト・スタンダード(例えばダブリンコアや RDF、XML など)に準拠した形で作るべきである。グローバル化が進んだ今日において、情報資源は、一つの国の中だけで限定して使用されるものではなくなっている。こうしたデファクト・スタンダード技術を使用することで、各国の情報資源を相互運用することを可能にする。また、応用システムもデファクト・スタンダードな技術を使用することで、既存のコンポーネントやライブラリーを使うことが可能になり、開発効率及び信頼性を上げることが期待できる。

一方で、できあがったメタデータ標準が未来永劫同じ形であるとは考えられない。適切な時期に見直され随時更新されるべきである。また、国内あるいは特定の分野の情報では既存のメタデータ標準だけでは要求を満たさない場合も考えられる。その場合、追加のエレメントを加える必要がある。メタデータ標準では、こういった拡張に関する問題(メタデータ標準のバージョンアップや拡張エレメントなど)に関しても十分に仕様化されている必要がある。メタデータ標準の変更は、そのメタデータ標準を使っている応用システムにとって、多くの時間と資源を消費させる。メタデータ標準の拡張に関しては、安定性と拡張性のバランスを十分考慮して行われるべきである。一方、応用システムは、メタデータ標準の変更に対して最小限の時間と資源で対応できるように考慮されるべきである。

メタデータ標準を用いることで、情報資源が、ウェブページ、電子文書、紙の文書、人(へのポインター)、データベースなどどういったものでもシームレスに扱うことが可能になる。応用システムは、この特徴を最大限に利用して開発されるべきである。例えば、オンラインで検索を行なう応用システムを考える。この場合、ネットワークを介してアクセス可能な情報のみを対象とするのではなく、紙の文書や本、CD-ROM、人など物理的な情報資源も対象になる。検索の結果は、ネットワークを介してアクセス可能な情報は、そのままオンラインで表示され、物理的な情報資源は、その情報資源の保管場所、入手方法、連絡先などが表示される。

メタデータ標準を用いることで、組織を超えて情報資源を相互運用することが可能になる。効率的に運用するためには、メタデータを組織横断的に構造化して、全ての応用システムからアクセスできるようにする仕組みが必要である。また一方では、情報資源のセキュリティーやアクセス制御も考慮されなければならない。こういった点に関しては、個々

の応用システムが管理するのではなく、全ての応用システムが共通で使う仕組みを作るべきである。

# 2.6 社会におよぼす影響

セマンティック Web はインターネットの世界を大きく変える技術であると期待されており、それだけに、実用化された暁には社会に及ぼす影響も大きいと考えられる。

あらゆる技術について言えることであるが、技術が単なる技術である限り、または個別技術の集合である限りは、社会への影響は軽微なものにとどまる。市場ニーズが技術を呼ぶのか、技術が市場ニーズを創造するのかは別として、現状ではニッチかつ限定的な状態でしか享受し得ない消費者の満足感が、複数技術の融合によって画期的なサービスに変容し、そのサービスを広く実現するための社会的な制約すらもブレークスルーしていくムーブメントになったとき、技術はメディアに脱皮し、社会に浸透し、社会に対して影響を与えるようになる。P2P技術を代表する技術といわれる Napster や Gnutella は、音楽著作権協会から提訴されるなど物議を醸すものの、音楽流通のあり様を一変させてしまうような社会的影響を及ぼすには至っていない。

セマンティック Web という技術も P2P 技術と同じように、それ単独で存在する技術ではなく、複数技術や、技術と技術を組み合わせて画期的サービスを実現するためのビジネスモデルなどの集合体である。北米ならびに欧州を調査した結果、現在「セマンティック Web」という言葉で括られる技術には大きく分けて「メタデータ技術」、「オントロジー技術」と「インテリジェント・エージェント技術」がある。

メタデータ技術とはマシンや後述するインテリジェント・エージェントが理解できるデータの意味を記述するための技術である。メタデータ技術といっても単一的に括れるわけではない。大きくはダブリンコアのように、メタデータを記述する際の標準策定とボキャブラリを作成する(技術というよりは)フレームワークに始まり、ユーザーが各自でメタタグを定義できるようにする技術の XML、用語や概念の意味をマシンリーダブル、マシン・プロセッサブルな形式で表現する技術の RDF といった言語や、RDF に準拠しながら、適用領域毎の標準を記述するための規則である CC/PP (装置やソフト等の能力を記述する)、P3P (個人情報の保護方針を記述する)、PICS (インターネット上のコンテンツ選択に際しての条件を設定する)などがある。また、恐らくは極めて労働集約的な作業になることが想像されるメタデータ付けを行なう作業を(半)自動化するための技術に至るまで多岐にわたる。

次にオントロジー技術であるが、これもメタデータ技術の一角を占める技術であるということもできる。オントロジーとは、概念どうしの関係や、相互関係についての解釈を行なうためのルールを定義した文章の集合のことである。オントロジーを記述するための言語は当初、DARPA の主管する国防領域での適用を企図した米国の標準である DAML と、欧州の学会、産業界が中心となってとりまとめ EU 委員会が予算的に援助して推進された OIL とに二分されていたが、昨年 DAML+OIL に一本化された。ただ、欧州においては、IBROW や On-To-Knowledge のように、産学を横断するオントロジー構築のためのプロジェクトが立ち上げられ、推進されているのに対し、米国独自の動きとしては DARPA での活動しか見当たらない。したがって、オントロジー構築の動きは軍事、諜報領域における戦訓や被疑者の検索精度を向上させるといった目的に閉じるきらいがある。

最後にインテリジェント・エージェント技術であるが、この技術は、究極的には適用領域に依存しない共通エージェント技術であるべきである。しかし、段階論的に言って、本

技術がセマンティック Web 関連技術の中では最も実現までの道のりが長いであろうと考えられる。したがって、初めからフルスペックのエージェント技術の実現を目指すのではなく、当初は市販の検索エンジンに、適用領域に合った機能追加を行なって使う程度のものであってよい(このことは、エージェントに次いで実現の困難なオントロジーについても言える)。

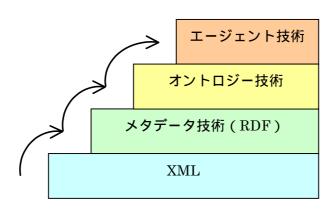


図 2-12 セマンティック Web を構成する技術

重要なことは、当初はこれら三領域にわたる個別技術、個別アプリの寄せ集めに過ぎなかったセマンティック Web が、やがて一部同士が統合され、限られた領域ではあっても、メタデータ オントロジー インテリジェント・エージェントという階段を駆け上がったとき、社会への浸透は加速することになる(図 2-12参照のこと)。また、ここでセマンティック Web 関連技術としては括らなかった P2P や IPv6 の技術は、セマンティック Web が想定するマシンリーダブルかつマシン・プロセッサブルな環境下では不可欠であろう。なぜなら、そうした環境においては、エージェントが問い合わせに行く先は現在想定されるデスクトップ PC やノート PC ばかりではなく、PDA や家電、各種センサーなど、身の回りのありとあらゆるものと通信を行なう可能性があるからである。

そうなれば、あらゆる素材に組み込まれることが見込まれるプロセッサー単位での論議が不可欠となるだろう。そうなったときに初めて、社会におよぼす影響は甚大となる。真の意味においての情報革命と言ってもよいのではないか。ただ、繰り返すが、個別技術の進化が現状では個別に進展し、適用領域も個別に設定されている今日、当初からフルスペックの夢を提示し、そこに直線的に到達するためのマイルストーンを設定するのは、批判や懐疑論ばかりを喚起するだけで現実的ではないであろう。

したがって、次に示す図 2-13のように、実現の容易度に応じ、フェーズ フェーズ フェーズ フェーズ という順序で段階的に開発を進めていくことが肝要である。

# 現Webの課題と新しいWebの世界で目指すもの

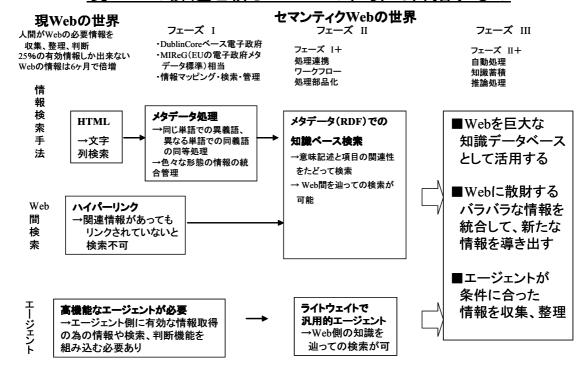


図 2-13 現 Web の課題と新しい Web の世界で目指すもの

# 3. 次世代電子政府での活用方法に関する調査研究

セマンティック Web 技術を次世代の電子政府に活用する視点から、電子政府関連の情報に対するセマンティック情報(メタデータ)の付与方法とそれを活用したセマンティック情報検索の方法について調査検討を行なった。

# 3.1 セマンティック情報(メタデータ)の付与に関する調査

各国の電子政府構築において、電子政府関連の情報に対してセマンティック情報を付与する事例としては、現状ではメタデータ付与の事例しかない。そこで、電子政府関連の情報に対するメタデータ付与に関して、下記の調査を行なった。

- ・国内の行政公開情報の調査
- ・海外政府のメタデータ付与に関する調査
- ・メタデータ階層モデル調査
- ・メタデータ定義調査
- ・利用者プロファイルのメタデータ定義に関する調査

## 3.1.1 国内の行政公開情報の調査

国内の行政公開情報の調査を行ない、国内の行政公開情報について、どのようなメタデータの付与方式とメタデータの活用方法が相応しいかその方向性を検討した。

#### 3.1.1.1 経済産業省の Web サイトの調査

経済産業省を対象として、Web サイト上で掲載されている公開情報の調査を行なった。 対象の URL は経済産業省の情報公開の説明ページ<sup>79</sup>であるが、その配下には、次の 4 つの情報へのリンクがある。

- (A) 行政文書
  - 1) 行政文書ファイル
  - 2) 特許庁行政文書ファイル
- (B) 政策情報
  - 3) 政策 HP における掲載情報の紹介
  - 4) 報道発表資料

## (A) 行政文書

2) の特許庁行政文書ファイルについては、調査期間中アクセスできなかったため検討を行なっていない。また、1)の行政文書ファイル(管理簿)は、紙などで保存されている行政文書に対する情報(これ自体がメタデータ)であり、それに対する検索システムが稼動している。図 3-1に行政文書ファイル検索システムの条件入力画面を示す。図 3-2に示すように、行政文書ファイルの(メタデータの)内容としては、分類(大分類、中分類、小分類)、日付などの項目がある。現状の検索でも、詳細検索を用いれば、日付による検

<sup>79</sup> http://www.meti.go.jp/intro/consult/disclosure/a\_main.html

索や、管理担当課室・係による分類などをメニュー的に選択しながらを検索を行なうことができる(図 3-3参照のこと)。行政文書ファイルの項目に関する考察については、3.1.4節を参照されたい。

#### (B) 政策情報

3),4)は、ホームページ上に載っており、セマンティック Web 技術を活用しやすいのはこれらの情報である。3) はどちらかと言うとフリーフォーマットに近く、PDF 文書も多く存在する。4)は報道発表なので、概要、担当、発行日、資料内容へのリンクなどからなっている(図 3-4参照のこと)。

#### (C) その他

経済産業省では、上記の(A)(B)以外にも、経済産業省 Web サイトにおいて情報の提供を行なっている。上記以外には、

- · What 's New
- ・入札
- ・公募
- ・イベント情報
- ・電子申請

などのページがある。

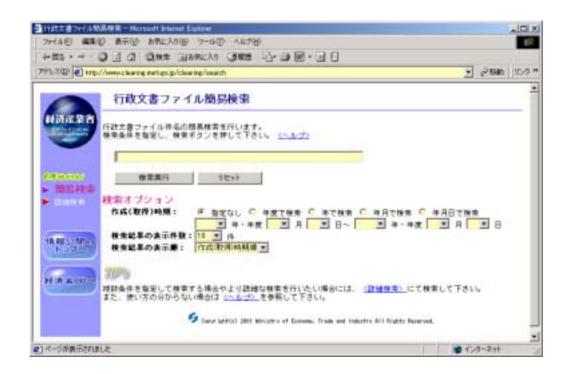


図 3-1 経済産業省の行政文書ファイルの検索画面



図 3-2 行政文書ファイルの検索結果



図 3-3 経済産業省の行政文書ファイルの詳細検索画面 分類項目をメニュー形式で選択することもできる



図 3-4 経済産業省の報道発表資料の例

## (1)簡単な検索実験

経済産業省の Web サイトには、Web ページ上の情報を検索するホームページ検索(以下 HP 検索)機能もある。紙ベースを主とした行政文書を検索する行政文書管理簿の検索とは別であり、検索結果も異なる。表 3-1は幾つかの検索語を用いて両者の検索結果(件数)の比較を行なった結果である。

検索語	行政文書 検索	HP 検索	
JEITA	1	21	
電子情報技術産業協会	0	95	
JEIDA	0	36	
日本電子工業振興協会	5	66	
電子協	2	14	

注)行政文書の2件は水産電子協会がヒットHP 検索は少なくともJEIDAの意味のものがあ る。

表 3-1 METI 公開情報の検索実験

行政情報管理簿自体が紙情報のメタデータのようなものなので、キーワード検索では、 上記のように検索結果の件数も少なくなっている。HP 検索では、文書内に電子協(JEIDA) などのように書かれるので検索できている。逆にあまり関係がない文書であっても、どこ かに記述があれば検索結果に出てくる。

このような場合、JEIDA (日本電子工業振興協会)が JEITA (電子情報技術産業協会) に組織変更されたことや、日本電子工業振興協会が JEIDA と呼ばれているという知識(オントロジー)を活用し、検索を高度化する余地がある。

## 3.1.1.2 国内の他の行政公開情報の調査

利用者側からみると、どの情報がどの省庁あるいはその他の施設等にあるかは判りにく く、省庁間で横断検索できるような機能は重要である。そのような横断検索を行なえる Web サイトとして、電子政府のポータルサイトである電子政府総合窓口がある。ここでは、 電子政府総合窓口を中心に調査を行なった結果について述べる。

## (1)電子政府総合窓口

電子政府総合窓口®では、各省庁のホームページの検索を行なうことができる。地方部局やその他施設なども検索できる81。(図 3-5参照のこと。)

全省庁ホームページ検索に対し、前述と同様の検索実験を行なったところ、表 3-2のようになった。

検索語	経済産業省		電子政府総合窓口				
	行政文書管理	HP 検索	産業省	関連省庁	関連団体	産業省	似外
JEITA	1	21	4	4	8	5	1
電子情報技術産業協会	0	95	28	39	48	56	17
JEIDA	0	36	22	22	35	35	13
日本電子工業振興協会	5	66	40	49	68	113	64
電子協	2	14	2	2	5	196	194

表 3-2 電子政府総合窓口での検索実験

表 3-2で電子政府総合窓口の「産業省」の欄は、検索対象に経済産業省を選択した場合、「関連省庁」の欄は、経済産業省に加え、資源エネルギー庁、特許庁、中小企業庁の関連庁を追加した場合、「関連団体」の欄は、さらに、地方部局や産業技術総合研究所などの関連機関を加えた場合、「産業省以外」の欄は、全本省庁で検索した結果及びそこから経済産業省関連機関を除いた結果の件数である。(図 3-6参照のこと。)

経済産業省と電子政府総合窓口の検索結果が合わないことは、対象としている Web サイトの範囲が異なることや索引の作り方による検索漏れや同一情報の重複度合などの違いが原因と推定される。

意外に「産業省以外」の検索でも、経済産業省の関連団体である日本電子工業振興協会は引用されることが多い。経済産業省以外では、総務省が多かった82。

電子政府総合窓口では、行政文書ファイルに対する検索も省庁横断的に行なう事ができる。申請様式の横断検索も行なえるが、キーワード及び分類から選択していくようになっている。

.

<sup>80</sup> http://www.e-gov.go.jp/

<sup>&</sup>lt;sup>81</sup> これに対して経済産業省のホームページ検索は http://www.meti.go.jp/ 以下のみを対象としていると思われる。

<sup>82</sup> 但し、経済産業省以外での「電子協」のヒットには、米国電子協会などでのヒットが多い。



図 3-5 電子政府のポータル「電子政府の総合窓口」

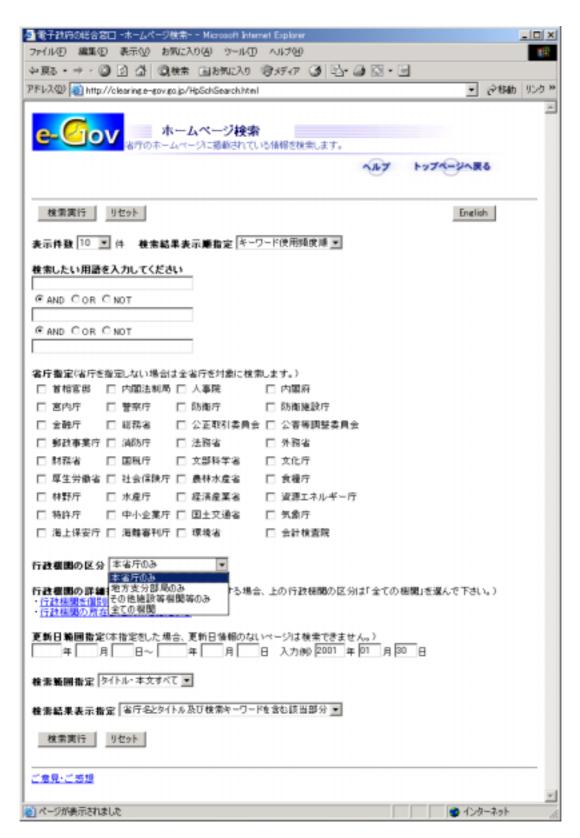


図 3-6 電子政府の総合窓口の詳細検索画面 本省庁のほか地方支分部局、その他施設等機関等も検索対象にできる

## 3.1.1.3 公開情報に対するセマンティック Web 技術活用の方向性

ここでは、上記公開情報を対象として、セマンティック Web 技術を活用し改良する一般的な方向性について述べる。

# (1)一般的な機能拡張の方向性

前節で述べた公開情報のうち、(A) については、それ自体が紙媒体による情報に対する書誌情報であり、メタデータと呼べるが、これらのメタデータを追加・改良を行なうことを考えると、人手によりデータの追加・修正を行なうことになり、大きなコストがかかる。ここでは、主に(B) 及び(C)、すなわち一般の Web を対象として、セマンティック Web の技術を用いてどのような改良が行なえるかの案を述べる。このうち幾つかは(A) の検索性能の向上にも寄与する。

セマンティック Web を用いた改良の方向性としては、以下が考えられる。

## (1)リソース発見(検索機能)の高度化

- ・検索条件の緩和:メタデータの情報を利用して、対象情報固有の類義語辞書を作成 する。特に、メタデータを対象として検索を行なう場合、使用されている語が少な いため、条件を緩和しないとヒットしにくい。
- ・知的検索機能の強化:メタデータやオントロジー情報を活用し、質問の意味内容に 沿った検索を行なうことにより、検索の精度をあげることができる。
- ・検索結果のグループ化:省庁を横断する検索システムによって検索対象となるドキュメントが豊富になる一方、検索結果として得られる項目も増大する。特に、類義語などを用いて検索条件を緩和する場合には、ヒットする項目が爆発的に増えることが予想され、利用者にとって必ずしも良い結果にはならない。したがって、検索結果のグループ分けや、不要な情報の隠蔽あるいは排除が重要になる。
- ・他人の評価情報を活用する:他人がつけたメタデータやアノテーション、投票結果 などを参考にする
- (2)蓄積情報の構造化(リソース作成、管理):メタデータやオントロジーで構造を記述することにより、一つのリソースを多角的に見せたり、構造から検索を行なったりすることが可能となる。
- (3)提示情報の構造化(受け取り手の再利用):情報を構造化することにより、利用者の望む方法での利用や再利用することができるようにする。
- (4)上記を用いた自動化エージェント

#### (2)具体的改善案

経済産業省のホームページの内容(情報公開のページに限定せず)に則してより具体的には以下が考えられる。但し、これらもまだ、方向性を示すものであり、実現性の検証がなされていないものも含まれる。

## (1) 検索の高度化

#### 1) 検索条件の緩和:

i) 電子政府総合窓口の検索では検索条件として「電子商取引」を指定した場合と「EC」を指定した場合では異なった結果が得られた。現状では、可能な全

ての同義語を利用者が順次指定して検索する必要があるが、意味情報を使うことでこの操作を自動化できる可能性がある。また、同義語間の変換を既に実現している検索エンジンにおいても、語彙の保守や拡張が容易にするために汎用のオントロジーを使うことは考えられる。

- ii) 意味情報を使って包含関係にある概念にまで検索条件を緩和することも考えられる。例えば、検索条件に「自動車」を与えた時に、「自動車」に含まれる「乗用車」や「トラック」などについての検索も自動的に行なう場合で、このことで有効な情報が収集できる場合もあり得る。
- iii)行政文書管理簿の項で述べたが、各組織の正式名称に対する俗称や、MITIから METI、JEIDAから JEITAといった組織変更など、経済産業省関連の用語の関係性が、オントロジーとして記述されているとすると、それを用いて類語辞書を作れば、「JEIDA」で検索できなかったような漏れを防ぐことができる。

## 2) 知的検索機能の強化:

- iv) 上記、ii)は、知的検索の点でも有効である。
- v) 特定の語彙が、組織名であることが推定できれば、組織名である事を利用して検索及び絞り込みを行なうことができる。
- vi) 時系列を考慮した資源内容変更の追跡 時系列情報のメタデータを活用した検索。具体的には組織の改編、消滅に伴 う名称変更や申請プロセス自体の変更により、資源に対する問い合わせ先や 資源の存在が不明になる場合がこれに相当する。
- vii) オントロジーの公理の活用によりある程度の推論を行なうこともできるようになる。例えば、親の情報が明示されていなくても、子供の情報が記載されているサイトから、親の情報を推論するなど。

## 3) 検索結果のグループ化:

viii) HP 検索などのように検索結果が多いときに、概念階層の分類を用いて、大量に出現する解を分類表示するのも有効と考えられる。

例:「JEIDA」で検索した結果、「2000年問題」\*\*件、「リサイクル関連」\*\*件・・・また、検索の絞込み条件の候補として上記分類を提示することもできる。

- ix) 意味的な関連を、検索後のドキュメント間の順位付けに利用することも可能 かもしれない。例えば、検索条件に「乗用車」を与えた場合に「自動車」や 「トラック」についても検索するが、表示順序は「乗用車」、「自動車」、「トラック」の順とする、などが考えられる。もし検索エンジンに手が入れ られるならば、検索時のスコアリングを「乗用車」>「自動車」>>「トラック」とすることで、よりエレガントな結果が得られるかもしれない。
- x) 利用者プロファイルの活用:個人に関するメタデータである利用者プロファイルを活用して、個人の興味ある情報を優先して表示する。また、セキュリティーのために、相手に応じて表示する項目を変更することも考えられる。

#### 4) 他人の評価情報の活用:

・ 他の利用者の意見やアノテーションを付加することにより、電子政府の

使い勝手を向上させることができる。

- ・ 電子投票やパブリックコメントにも関連すると思われる
- (2) 蓄積情報の構造化(リソース作成、管理)
  - 1) 概念階層を提示した検索・情報提示
    - ・ 定義された概念階層を活用して、概念階層を利用した情報提示ページの 自動作成(自動分類)が可能となる。
  - 2) 部署や団体の変更(移動、名称変更)への対応:組織名称の一括変更など
    - ・ 組織名称の一括変更や、組織と人との対応の更新など
  - 3) 同一のリソースを、多様な観点で提供することが可能:(政策、部署、日付等々):
    - ニュースリリースの検索
    - ・ 各組織住所一覧などの管理の簡略化
    - ・ 申請プロセスの生成、提示の枠組み 明文化されていない事実に関して、資源間の関連をユーザーの意向に沿って抽出し直すことにより、既存資源を異なった視点で閲覧することが 可能となる。ある実施例のプロセスが明記されていない場合に、特定の 要素に着目した処理手順の追跡やあるリポジトリに着目した場合に、存 在する情報の抽象化・俯瞰的視点を提供する枠組みがこれに相当する。
  - 4) リソース間の(半)自動リンクづけ:
    - ・ 組織変更に伴うリソース移動の自動化
    - ・ 関連組織からのリンクの更新の自動化
- (3) 提示情報の構造化(受け取り手の再利用を助ける)

例えば、以下のようなことが考えられる。

- 1) ユニバーサルアクセス対応:
  - ・ 住民からのアクセスのためには必要であり、障害者対応などを含め、省 庁関連では重要である。情報に適切なメタデータを付加することにより、 携帯電話や今後出現するモバイル端末への対応や、視覚障害者に対して 音声で応答するシステムとの連携を、文字により良い情報が提供できる ようにすることができる。
- 2) 電子申請の例:
  - ・ 例えば、財団法人ニューメディア開発協会の提案している電子申請書類 一式に対する文書のメタデータであるパッケージメタを活用すると、住 民側で申請書類作成支援ソフトウェアを用いて、文書管理機能、添付書 類の一覧表示機能、必須添付文書のチェック機能からその情報を利用で きる。
- 3) 住所にメタデータをつける。それが住所だと分かれば、それを位置情報として変換して色々なサービスを行なうことが考えられる。
  - ・ その位置を地図上に表示する。同じ地図上に複数の位置を表示すること で相対位置関係がわかる。
  - ・ その位置から最寄駅を調べる。「乗り換え案内」などと連動して、自分 の乗車駅だけを入力すればそこに行くまでのルート、時間などが分かる。

- ・ その位置から自動車でのルートを調べる。アクセスするルートが分かれば、現在のそのルート交通情報などが分かる。
- 4) 日時にメタデータをつける。
  - ・ カレンダー、年表ビューでの表示を行なう。
  - 時間軸線上にプロットし、他のイベントとの関係を見やすくする。
- 5) 数値にメタデータをつける。
  - ・ グラフ化できる。自分の好きなグラフタイプで表示する。
  - 分析のデータとして利用する。
- 6) 会議などの時間場所(住所、会議室)などの情報にメタデータをつける。
  - ・ PIM やスケジューラーとの連携を取る。
  - ・ 3)のサービスなどとの連携を取る。
- (4) 上記を用いた自動化エージェント

例えば以下のものが考えられる。

- 1) 新着情報のメールによる通知サービス<sup>83</sup>において、現在手動で行なっていると推定される作業を、ニュースリリースなどコンテンツの作成を行なうだけで、自動的に新着情報として配布できるようにする。この目的のために RSS のシンジケーションを用いることもできる。
- 2)「What's New」「入札」「公募」「イベント」などのページを定期的に見て利用者の興味ありそうな情報をメタデータにより判定し、利用者に通知するエージェント。
- 3) 例えば、各地方部局、関連施設等のスケジュール情報を定期的にチェックし、必要なスケジュールの一覧を作成するエージェント。

70

<sup>83</sup> 新着情報配信サービス (http://www.meti.go.jp/mailservices/index.html)

## 3.1.2 海外政府のメタデータ付与に関する調査

海外政府におけるメタデータ分類方法について調査を行なった。

各国政府のメタデータエレメント標準は、表 3-3の通りである<sup>84</sup>。表の中の は必須、 は任意のエレメントを意味する。

英国の e-GMF、EU の MIReG、及びオーストラリアの AGLS については、ダブリンコアの基本となる 15 エレメントに、いくつかの拡張エレメントを追加したものである。米国の GILS については、ダブリンコアを拡張したものではないが、e-GMF の文書「英国電子政府メタデータ標準」の中で、e-GMF との対応関係が記載されており、結局、ダブリンコアの 15 エレメントを GILS のエレメントにマッピングすることが可能である。このように、欧米諸国の政府機関においては、ダブリンコアがメタデータエレメントの事実上の国際標準(デファクト・スタンダード)として採用されている。

ダブリンコアの各エレメントの詳細については、3.1.4節で説明する。

<sup>84</sup>各メタデータ標準のリソースは、以下の通りである。ダブリンコアは

http://dublincore.org/documents/dces/、e-GMF は「英国電子政府メタデータ標準」、MIReG は「MIReG メタデータ要素セット」、AGLS は

http://www.naa.gov.au/recordkeeping/gov\_online/agls/user\_manual/agls\_metadata\_elements.html、GILS は http://www.gils.net/elements.html、Government of Canada は http://www.cio-dpi.gc.ca/clf-upe/6/6a\_e.asp#6.3 である。

<sup>85</sup> 付録2を参照のこと。

	ダブリンコア基本 15 エレメント	e-GMF (英国)	MIReG (EU)	AGLS(オーストラリア:Australian	GILS (米国:Government	Government of Canada (カナダ)
エレメント名	(Simple DC)	( , , , ,	, ,	Government	Information Locator Service)	
	( ,			Locator Service )		
Coverage					(複数のエレメントから成る)	
Description					(Abstract)	
Relation					( Cross reference )	
Source					(Source of data)	
Subject				1	(2つのエレメントから成る)	( Controlled Subject )
Title					(Folder title & Document title)	, ,
Туре						
Creator					(Originator)	(Originator)
Contributor					, ,	, ,
Publisher					( Distributor )	
Rights				( Availability と	( Availability & Access	
				Rights)	constraints)	
Date					(Date of Publication 等 3 つのエ	
					レメントから成る)	
Format					(Medium の詳細項目)	
Identifier				2	( Schedule number )	
Language					( Language of resource )	( Language of resource )
Audience						
Disposal						
Location						
Reservation						
Function				1	不明	
Availavility				2	不明	
Mandate					不明	
Keywords					不明	
備考		ドラフト標準。各エレメ		1「Subject」または 「Function」のいずれ	The UK e-Government Metadata Standard V2(英国電子政府メタデー	カナダ政府のサイトは上記の 5 つの のメタタグを採用しなければな
		ントの下に	ントの下に	かが必須	タ標準)において、e-GMF と GILS	らない。
		詳細項目あ	詳細項目あ	2「Identifier」または	の対応関係が記載されている。	   メタタグジェネレータ有り。
		り。各エレメ		「Avilavility」のいずれ		http://198.103.99.147/publications
		ントには、必	ントには、必	かが必須		/metagen_e.html
		須。任意、特				
			定の状況で			
		推奨、の区分	推奨、の区分			
		がある	がある			

表 3-3 海外政府のメタデータエレメント標準

# 3.1.3 メタデータ階層モデル調査

国内の行政公開情報にメタデータを付与するにあたって、どのようなメタデータの階層 モデルが相応しいかを調査した。ここでは、メタデータ付与の方向性としてオントロジー の応用から、メタデータを目的に応じて階層的に構成する方式を提案する。この方式は、 オントロジーのマッピングを用いないレベルでも有効である。

### (1)メタデータとオントロジー

セマンティック Web 技術で議論されているメタデータは、2.1節で見たように、RDF の3つ組(リソース、プロパティ、値)で表現される。さらに、その上で規則の形での制約(公理と呼ばれることもある)が付加される。

3.1.2節で見たように、いくつかの海外政府においてはメタデータ標準が策定されている。このことは、RDFの3つ組の表現で言えば、政府が所有する情報(リソース)に対するプロパティが規定され、標準化されていることにほかならない。このような標準化とメタデータ付与によって、政府所有の大量で複数のリソースから、そのリソースの「日付」や「著者」といった何らかの意味のあるデータ(プロパティと値)を取り出すことが可能になる。

セマンティック Web 技術では、これに加え、上記 3 つ組の表現を用いて、オントロジーと呼ばれる語彙体系を定義することができる。

Uschold<sup>86</sup>によれば、さまざまな分野で用いられているオントロジーに対して、オントロジーの応用を分類すると大きくは以下の3つのカテゴリになるとしている。

- 1)ニュートラル・オーサリング
- 2)情報への共通アクセス
- 3)索引付け
- 1)は、共通に抽象的にオントロジーで記述したものを複数の環境で動作させるよう変換するものである。
- 2)は、(異種性を意識していない複数の)アプリケーションから異種のリソースに共通にアクセスを可能にするものである。
- 3)は、情報検索のための応用で、いわゆる知的検索とか概念検索とか呼ばれている応用であり、オントロジーの語彙体系を用いて検索の高機能化を図るものである。

Uschold はさらに2)を、

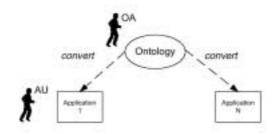
- A)共有オントロジーを用いたデータアクセス、
- B) オントロジーのマッピングによるデータアクセス、
- C)共有サービス

のバリエーションに分類している。 A ) は共有されたオントロジーに基づきデータをアクセスするもの、 B ) はそれぞれオントロジーを持つデータに対してオントロジーをマッピングすることにより相互利用を行なえるようにするもの、 C ) は共有のオントロジーの仕

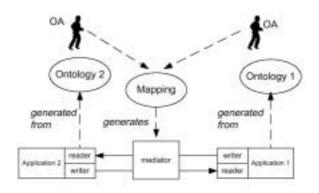
<sup>&</sup>lt;sup>86</sup> Uschold, M., Jasper, R., A Framework for Understanding and Classifying Ontology Applications, in Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5), Stockholm, Sweden, August 2, 1999.

様に従い異なった環境でのサービスを実装するものである。

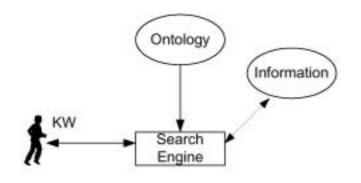
図 3-7は、Uschold<sup>86</sup>の分類の概略図である。オントロジーの一般的な適用方法の説明が目的でありさまざまなバリエーションも存在するので、セマンティック Web 技術の適用システム全体を表現するものではないが、この分類に従えば、セマンティック Web でのメタデータの応用では以下のように解釈されると思われる。1)は、メタデータ標準など典型的な応用のメタデータセットを作成して、それを基に多くの場所で活用する例と考えられる。あるいは、RDFスキーマの定義に基づき、個々のスキーマを記述するモデルと考えることができる。3)のリソース発見のための索引付けの利用でありる。2)は、この中では最も高度な活用法で、情報への共有アクセスのためのマッピングとしてオントロジーが活用される。前節まで議論されているメタデータ標準は、この2)の段階のA)のアプローチのひとつと言えよう。



1)ニュートラル・オーサリング



2)・オントロジーマッピングによるデータアクセス(情報への共通アクセスのB))



3)概念検索(索引付け)

図 3-7 オントロジーの適用パターン例(Uschold 86より)

# (2) メタデータの階層モデル

メタデータ付けの対象となるリソースにはさまざまなものが考えられる。一口に電子政府、あるいはその公開情報と言っても、定期刊行物やニュース、統計情報などその様式は多様であると考えられる。RDFの3つ組では、クラスを用いて、クラス毎に付加する情報(プロパティ)を定義することができる。

3.1.1.1節で述べたように、行政文書管理簿には、「管理担当課室・係」という属性がある。多くの場合、情報の管理責任部署が存在すると仮定するのは極めて自然である。管理部署による特有の表記方法なども存在する。その部署の管理にあった方法でメタデータを付与するのが自然である。メタデータ付与の課題となっていた、メタデータ付与者のメリットの問題も、管理責任組織が存在し、メタデータが、その管理目的に適合する場合は、問題なく、整合のとれたメタデータが保守管理されることが期待できる。逆に標準化として、使わないデータを大量に入力し・保守しなければならないとなると、普及は困難になる。

メタデータは、対象となるリソースに埋め込むことも、別のファイルにすることも、あるいは別のサイトに設置することもできる。このことは、管理部門ではないサイトのリソースや、複数のリソースサイトをまたがったリソースのメタデータを作成することもできることを意味している。

一方、利用者側からはしばしば様式や管理部署を問わず検索を行ないたいという要求がある。横断検索などを行なうためには、異種の管理体系のリソースをアクセスしてメタデータを作成・取得しなければならない。このためには、各データベースのアクセス方法や、プロパティの体系などを個別に知る必要があり、非常に煩雑な作業となる。この時、各リソースが既にあるメタデータを持っていたとするとそれを活用する事により、前述の B)のオントロジーのマッピングにより、その作業が軽減できる。この段階では、プロパティの規定に従い、プロパティどうしのマッピングを行なうレベルでもありうるが、プロパティやオブジェクト間の関係を記述した RDF によるメタデータが存在すれば、より柔軟な処理を行なうことが可能になる。

この結果、メタデータの階層が構成される。上位のメタデータは、下位層のメタデータ を利用し必要な情報をそれに付加することで構成できる。目的に応じて、柔軟にメタデー タを構成することができる。

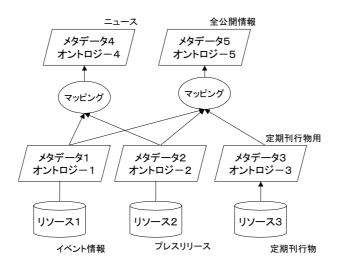


図 3-8 メタデータの階層モデル

このような構成を採ることで、もし各リソース管理部署の都合で、ある部署のオントロジーが変更になった場合や、上位のオントロジーを変更する場合などは、マッピングを変更するだけで対処することができる。

図 3-8では、マッピング機能を表現するために中間に置いている。マッピングは、上位と下位のメタデータ/オントロジーから自動生成されるのが理想的だが、最も簡易には、マッピングは上位のメタデータ/オントロジー(RDF記述)に記述されていても良い。

例えば、メタデータにより、標準用語や標準語彙を定義し、領域や組織により異なる同義語や類義語を同等に取り扱う事ができるが、メタデータで定める標準用語や標準語彙は、永久的なものではありえず、時間と共に、その意味が変化することも起こり得る。このような時にメタデータの標準用語や標準語彙が変わった時、既存のメタデータを作りなおすことは、非効率である。この場合、より上位のメタデータ(RDF記述)により、その新たな意味を定義することができれば、既存のメタデータに手を入れる必要がなくなる。

# 3.1.4 メタデータ定義調査

国内における行政機関公開情報の現状、及び行政文書に対するメタデータ付与の各国政府の現状を踏まえて、国内の行政機関公開情報に対するメタデータ定義について述べる。 意味情報検索の観点から情報資源をより有効に活用することが目的である。

# (1)国内向け行政機関公開情報に対するメタデータ定義に関する一提案

既存の各国政府メタデータエレメント標準は、前述の通りダブリンコア(DC)に準拠している(3.1.2節参照)。ダブリンコアの基本エレメントは定義内容が明確で一般に普及し易いという利点がある一方で、精度の高い情報を記述する場合やコミュニティー独自の情報を記述するためにはメタデータエレメントの拡張や精密化が必要となる。

こうした背景と各国政府の現状を踏まえ、以下にダブリンコアのエレメントを基礎として、国内の公開情報文書へメタデータを付与する場合に考慮すべき点を述べる。

既存のメタデータに対する国内行政文書の親和性([]内はエレメント名)

[Title]タイトル、情報資源に与えられた名前

一般には作者もしくは公開者によって与えられる。 国内の行政文書における「タイトル」記述に対し直接的に写像できると思われるが、検索における有効性ならびに構造を定義したタイトル付けが必要である。

例えば、ある既存の行政文書情報ページでは表題項目が[大/中/小]の3分類に加え、タイトル項目が[1/2/3]行目といった詳細な分類が施されているが(環境省の行政文書情報トップ $^{87}$ >書誌情報目次 $^{88}$ を参照)タイトル列における意味構造は見られない。この行政文書情報ページの中の分類表を一部抜粋する(表 3-4参照のこと)。

大分類名	中分類名	小分類名	タイトル 1 行目	タイトル 2 行目	タイトル 3 行目
廃棄物・リサ イクル	循環企画	リサイクル	第5、6回ごみ減量 化国民大会	-	-
廃棄物・リサ イクル	循環企画	リサイクル	分別収集促進計画、 市町村分別収集計画	愛知県	-
廃棄物・リサ イクル	循環企画	リサイクル	平成12年度	保管施設一覧	岩手県、宮城県、秋田県、 山形県、福島県
廃棄物・リサ イクル	循環企画		フェニックス法 大 綱・要綱等作成過程		昭和五五年八月 ~ 五六年 一月

表 3-4 行政文書の分類例

またこの形式では、正式名称とは異なる慣習的な名称が広く知られているプロジェクト等の文書を参照するのは困難と思われる。

既存の海外政府メタデータ標準の枠組みを直に利用するだけでも、こうした現状を大幅に改善することが可能である。[Title]エレメントには"別名"の記述が可能な詳細エレメントが用意されているため、これらを選択的に利用することで文書を効果的に同

<sup>87</sup> http://www.env.go.jp/guide/bunsho/index.html

<sup>88</sup> http://www.env.go.jp/guide/bunsho/syoshi\_mokuzi.html

定できる。さらに年度や地域情報等の記述は後述の [Description][Date][Coverage] のエレメントによる別定義、また分類は[Subject]エレメントを用いて再定義することにより、可読性が高く、かつ機械的な検索にも有効なメタデータ定義が付与できる。

例: 付与部分例(例示目的のため、ダブリンコア以外のエレメントも混在して用いている。また記載のURL、e-mailアドレス等は全て架空のものである。)

```
<rdf:RDF
   xmIns:rdf="http://www.w3.org/RDF/RDF/"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmIns:dcq="http://purl.org/dc/terms/"
     <rdf:Description about="http://www.sample.go.jp/doc/sample.html">
  <dc:title>平成n年度 ~の実態に関する調査報告書</dc:title>
 <dc:creator>制作者;
                      課
                            室 foo@sample.go.jp</dc:creator>
 <dc:creator>連絡先:
                       課
                            室 foo@sample.go.jp</dc:creator>
 <dc:creator>責任者;
                      長 director@sample.go.jp</dc:creator>
 <dc:description>この文書は、
                             計画における
                                           に関する動向調査の記録であり、××と
いう現状から、~への対処が急務であるとの結論を得た。</dc:description>
  <dc:coverage.spatial.region>
                              県,
                                    市</dc:coverage.spatial.region>
 <dc:coverage.temporal>2002-3-31/2003-3-31</dc:coverage.temporal>
 <dc:data>2002-02-21</dc:data>
 <dc:format>text/html</dc:format>
 <dc: language>ja</dc: language>
</rdf:Description>
</rdf:RDF>
```

本エレメント記述を実際に行なう上で、考慮すべき点を以下に示す89。

- ・ 情報資源の名前として一般的に知られているものを使用する。
- ・ 本エレメントは必須とされているため、情報資源のタイトルが存在しない場合に は新たに作成する必要がある。人目を引くためではなく、簡潔で意味のあるもの にするべきである。
- 情報資源が e-mail の場合には、[Title]エレメントとして電子メールの件名である"Subject"を使用する。
- ・ Alternative title (= 別称を定義するエレメント名)には、省略名や翻訳したタイトルを記載する場合もある。
- ・ 情報資源の正式なタイトルが一般に不明確な場合、意味のある[Title]エレメント を Alternative title として付けるべきである。

-

<sup>89</sup> 以下、各エレメントにおける箇条書きの項目は、そのエレメントにおいて考慮すべき点を意味する。

- ・ 何がタイトルか分からない資源に対しては、可能性のあるタイトルの[Title]エレ メントを列挙しておく用法も認められる。
- ・ 同じタイトルを持つ項目が複数ある場合、言語やバージョン番号、状態(例:草案、ドラフト)や日付などの値を[Title]エレメントに付加することで検索に役立てることができる。
- ・ 情報資源が電子文書を含む電子フォルダの場合、フォルダ名を[Title]エレメント として付けるべきである。
- ・ [Title]エレメントは、センテンスとして書いた方がその内容が明確になる。

# [Subject]主題及びキーワード

情報資源の主題あるいは内容を説明するキーワードや句を意味する。行政文書の参照方法として、大分類・中分類・小分類とカテゴリをドリルダウンする枠組みを動的に提供する場合に有効である。また、この[Subject]エレメントの定義として、全省庁を範囲とする統一的な分類カテゴリ、キーワード、体系コードに類するものを適切に策定できれば検索精度の向上が期待できる。このエレメントを用いた体系コード例(分類スキーマ)としては LCSH、IDC、MeSH、DDC、LCC、UDC 等が存在する。

- ・ 本エレメントには、情報資源の最も重要なもの及びユニークな言葉を選ぶ。利用 者が必要なものを正確に検索可能とするために必要十分な言葉を選択すべきであ る。
- ・ 一般的な用語よりも特化された用語を使用することで、情報過多になるのを防ぐ ことができる。
- ・ 本エレメントとして、同意語や言い換えた言葉、大文字と小文字を変えただけの ものなどを列挙すべきではない。

# [Creator]著者あるいは作者を表すラベル

情報資源の内容作成に関し、責任を持つ者や組織に関する記述。行政情報文書では作成者(課/室)がこれに相当する。責任者への連絡手段(e-mail)があることが望ましいが、作成元の組織改編・消滅に伴い文書追跡が困難になることも想定されるため、組織の完全な階層構造を記述しておくことが推奨される。

- ・ 本エレメントに記載された人・組織の代表者は、法律上の責任・義務を負う。また、監査を行なう場合には人名を明記しておくことが必須となる。
- ・ 役職や部署の省略名を用いることは、利用者にとって利用を困難にする恐れがあるため、正式な名前を使用するか、または(略称が記載されている)用語集へリンクを施すなどの配慮が必要である。

### [Date] 目付

情報資源が作成されたあるいは有効になった日付。この日付とは Coverage エレメント に記載されるものとは異なり、文書(情報資源)そのものが作成された日付である。文書 内容が表す時間情報とは異なる。記述形式は YYYY(年の 4 桁表記)及び YYYY-MM-DD

(年-月-日)(例 2003-04-28) といった ISO8601<sup>90</sup>準拠が強く推奨されている。元号表記はこの規格では対象外であるが、元号利用の手段としては(1)元号を正しく時間記述形式に写像できるような日時情報に関する(オントロジー変換としての)変換機構を備える (2)[Description]若しくは[Subject]エレメントに明記し文書情報の一部として扱うことが考えられる。

- ・ [Date]エレメントは、世界中の人々が認識でき、且つ、ソフトウェアが解釈できる形式でなければならない。
- 時間が必要である場合、hh:mm を付加する。hh は時間を示す(24 時間を使用した場合)mm は分を示す。(例 13:23)

# [Description]内容記述

情報資源の内容に関する説明記述。文書の場合の抄録、視覚的資料の場合の内容記述など。メタデータの付与対象(リソース)は文書に限らず動画・画像等もその範疇に入るため、今後公開情報として閲覧、検索が物理的に容易でないものも含む場合には(動画を含むメディア等)本エレメントの情報が検索に際し重要である。

・ [Description]の記述内容として推奨される例を以下に示す。

主題の補足説明

情報資源の作成理由

参照されるグループと組織

適用されるイベント

キーとなるフィールド (データベース) または章のリスト

その他有用な情報

- · [Description]エレメントの記述内容については、簡潔かつ正確であるべきである。
- ・ 複数の[Description]エレメントを記述した場合、先頭のエレメントが利用者の目 に入る可能性が高いため、一番大切な部分を同エレメント中で最初に記述する。
- ・ 他のエレメント([Title]エレメント, [Coverage]エレメント、[Subject]エレメントなど)内にある情報を[Description]エレメントと重複して記載すべきではない。

### [Publisher] 情報資源を公開する人・組織

情報資源を現在の形態で利用可能にしたことに責任を持つ人や組織。情報資源の再発行や複製の許可を得る目的で参照される。行政文書の公開に関しては管理担当課室・係がこれに相当する。

- ・ エレメントには、情報資源が存在する Web サイトを管理する組織を示す場合もある。
- ・ 人名よりも職名を記述するのが適切な場合もある。
- ・ 本エレメントは情報に関する法的な権利と責任を明記するという意味で、必須の エレメントである。

# [Contributor] 貢献者

-

 $<sup>^{90}\ \,</sup> http://www.w3.org/TR/NOTE-datetime.html$ 

[Creator]エレメントには記載のない人あるいは組織で、情報資源を作り出すに当たって知的に重要な貢献をしたもの。編集者、翻訳者等にあたる。

- ・ 本エレメントの記述内容は、Creator エレメントの記述との重複を避けるべきである。
- ・ 人名よりも役職名・組織名を記述することが適切な場合もある。

# [Type]情報資源の種類

例えば、ホームページ、ワーキングペーパー、テクニカルレポート、辞書・事典などの分類を記述する。行政機関文書では、カテゴリ別情報として提供されている分類ラベル(プレスリリース、FAQ、白書、報告書、調査研究結果、統計等)がこれに相当すると思われるが、相互利用性を保つためには、関係各省にてこのような用語の統一を図ることが必要である。

- ・ 本エレメントを[Format]エレメントと混同して用いるべきではない。[Format]エレメントは例えばワード、エクセル、一太郎といったファイル形式を記述するために用いられるが、本エレメントは情報資源の内容そのもの指す。
- ・ 公式の記録や文書は、ある大きなカテゴリの一部として存在することがある。本 エレメントはそのような場合に、どの階層に属する情報資源であるかを明記する ために用いることもできる(=より詳細なエレメントである"Type.Aggregation Level"の利用による)。
- ・ 複数の[Type]エレメントを記述することで、内容をより詳細に特定(記述)する ことができる。

# [Format]情報資源のデータフォーマット

情報資源の表示・動作に必要なソフトウェアや、場合によってはハードウェアを識別するために利用できる情報を記述する。相互利用性を保つために関係省庁で統一見解が得られた用語リストの中から選ぶことが強く推奨される。

- ・ 情報資源が別のフォーマットでも利用可能である場合には、[Relation]エレメントを併用してその旨を明記しておく。
- ・ 本エレメントは情報資源が格納されている物理的な形式を記述する。例えば「本」、 文書ファイルであれば「プレーンテキスト」、「一太郎」、「ワード」、データ ベースであれば「Access97 形式」等の記述となる。

### [Identifier]識別子

情報資源を一意に識別するための文字列もしくは番号。国際標準図書番号(ISBN)や他の標準化された名前のように全世界的に一意に定まる識別子の定義を行なう。本エレメントとして URL を使用する場合には、(URL 変更に伴う) 不整合が生じないよう注意すべき必要がある。

# [Source]元情報資源

情報資源を作り出す元になった別の情報資源に関する情報。一般にエレメントには、 当該情報資源に関する情報のみを記述することが推奨されているが、本エレメントに は当該情報資源を見つけ出すために有用である別の情報資源に関する日付、作者、形 式、識別子あるいは他のメタデータを書くことができる。実際の利用に際しては他文 書からの図、写真といった引用に近い参照の場合がこれに相当する。より抽象的な資源間の関連づけは[Relation]によって定義される。

# [Language]言語

情報資源の内容を記述するために用いられている言語(コード)。実際のエレメント記述は RFC1766 $^{91}$ に準拠した en、de、es、ja、th、zh 等の表記を用いる。日本語による文書は"ja"である。

# [Relation]参照関係

別の情報資源との関係。例として、関連するプレスリリースの定義、同一のフォルダに属する文書の定義(Relation.IsPartOf)、参照先文書のバージョン情報(Relation.IsVersionOf)、参照先文書の識別子(Relation.References)等の定義を記述しておく。多様なカテゴリ分類が想定される行政文書においては、相互利用性を得るために情報資源間の関係を表す値は省庁間における統一が必要であり、少なくとも現在定義が進められている値のリスト中から選択して与えることが推奨される。なお関連先の指定は、参照先文書を一意に特定できるような識別子による指定が望ましい。

# [Coverage] 適用範囲

情報資源の内容に関する空間的(地理的)あるいは時間的特性。空間的範囲は物理的な範囲として座標(例えば経度と緯度)や、統一見解の得られた語彙リストの中から選ばれた地名(都道府県名・市町村名)を用いる。時間的範囲は当該情報資源が表している内容に関する時間的情報であり、情報資源の作成や公開に関する日付は[Date]エレメントによる。日付・時間の表記形式は[Date]と同様 ISO8601 に基づく92

# [Rights]権利情報

権利管理に関する声明文、権利管理に関する声明文へのリンクを表す識別子、あるいは当該情報資源の権利管理に関する情報を提供するサービスへのリンクを表す識別子を記述しておく。

# [Audience]利用者情報

情報資源の利用者として意図されているカテゴリ分類の記述。対象者カテゴリとしては例えば教職員、学生、技術者、 市民、といったものが想定されている。本エレメントは情報資源の階層内における位置付けや、検索に際しある特定の人々を対象とした情報を抽出する際のフィルタリング目的で利用される。

# [Disposal]保管・廃棄情報

情報資源の保管期間、廃棄時期に関する記述。情報資源を管理する上で必須の項目である。廃棄時期を迎えた場合の資源の処理方法も本エレメントの詳細項目である"Action"に記述できる。また、情報資源がWebページである場合には、本エレメン

92 http://www.w3.org/TR/NOTE-datetime.html

<sup>91</sup> http://ds.internic.net/rfc/rfc1766.txt

トの詳細項目の一つである"Review"に更新時期を記述しておくことで、更新されずに古いまま存在しているページを管理者が容易に把握するといった利用方法も可能である。

### [Location]保存場所

情報資源の保存場所に関する記述。例えば、"執務室・棚番号 15"といった記述であり、電子的な情報であれば、DVD-ROM や DAT といった保存メディアの種別を付記しておくのも有効である。(例:地下倉庫・1-A 室・ $\mathrm{disc}$  番号  $1832\mathrm{B}[\mathrm{CD}\text{-ROM}]$ )国内の行政公開文書には必須の項目と思われる。

# [Preservation]永久保存データに関するコメント

公的文書を長期的に保存管理する上で必要な項目であり、管理者向けのエレメントである。技術の進歩に伴い新しい電子メディアが普及すると、旧来の電子媒体に蓄えられていた情報の読み出しが困難になることへの対処である。今後数 10 年、数 100 年といったスパンでの電子ファイルの読み出し・解釈に必要な情報を全て記載することを目標としているが、詳細は UK Public Records Office のメタデータ標準においても現在のところは検討中となっている。

# [Function]業務内容

情報資源が属する業務内容に関する記述。オーストラリア政府ではこのエレメントと共に Australian Governments' Interactive Functions Thesaurus (AGIFT)を用いることで、業務内容(と対応する機関)を明確に分類している。エレメントの記述内容そのものは、比較的抽象度の高い分類であり、ある機関のWebページを見た場合に、閲覧者にとって識別に役立つレベルの分類である。例えば「学校教育」「コミュニティーサービス」「チャイルドケアサービス」といった項目である。

# [Availability]情報の入手・アクセスに関する情報

主として非電子化情報の入手方法やアクセス手段に関する記述。入手先情報の詳細項目として人名(組織名)、e-mail アドレス等を記載する点では Creator、Contributor、Publisher のエレメントと類似している。本エレメントではさらに(機関の)住所、郵便番号、電話、FAX、開庁時間・曜日、情報の入手に必要となる費用(有料/無料)等の記述がある。

### [Mandate]関連条項・法律

当該情報資源の作成・公開の要求元となっている既存文書や条項、法律等についての記述。詳細エレメントとしては、法律・条令・事例(判例)等の種別がある。 (例: "http://www.austlii.edu.au/au/legis/cth/num\_act/laa1989192/"

"England v Van Donk Matter No CA 40433/97")

### [Keywords]重要語の重み付け

情報資源中でキーワードとなる語彙に付与する。このメタデータを付与した語彙は、検索システムにとって特別の重み付けをもつ語彙として扱われることを想定する。直近の検索精度向上には有効と思われるが、"意味"構造を考慮していない点において、他のエレメントとは定義の方向性がやや異なる。

# (2)海外政府メタデータ標準のエレメント名と国内行政文書との対応関係

以下ダブリンコアを軸に、海外政府で用いられているエレメント記述と、国内の行政文書の対応を表 3-5に示す。括弧書きの項目は現在の公開情報としては明確な記載が無いが、暗にその存在が仮定されるもの、記号「\*」は資源情報から定義可能なものを示す。重要度は意味情報検索の観点と実現可能性の観点から、既存の国内公開情報への付与が必須と思われる項目に「」記号を付けた。「参照」欄のメタデータ標準は次のものを表す。DC=ダブリンコア、MIReG=MIReG Metadata Element Set、AGLS=Australian Government Locator Service、Canada=Government On-Line(NR Can)である。

			•	
国内行政文書ファイ ルの情報項目	重要 度	エレメント名	参照	エレメントの概要
(編集者・翻訳者)		Contributor	DC	(Creator に次いで)貢献した人・組 織
*		Coverage	DC	情報資源内容の適用を受ける期間、 場所
作成者		Creator	DC	情報資源の作成にあたり最も責任を 有するもの
作成(取得)時期		Date	DC	情報資源そのものに関する日時情報
タイトル/(備考)		Description	DC	情報資源に含まれる情報の説明
*		Format	DC	情報資源の記述形式
*		Identifier	DC	内容に対する一意の識別番号
(ja)		Language	DC	記述されている言語
管理担当課室・係		Publisher	DC	情報資源を発行した責任者
*		Relation	DC	関連する情報資源への参照
*		Rights	DC	情報資源全般の権利情報
*		Source	DC	提示された情報資源への参照
大・中・小分類		Subject	DC	主題及びキーワード
タイトル		Title	DC	タイトル
(分類カテゴリ)		Type	DC	情報資源の種類
*		Audience	MIReG	利用者のカテゴリ
保存期間・廃棄時期 保存期間満了時の措 置結果		Disposal	MIReG	情報資源の保持と処理
保存場所		Location	MIReG	情報資源の物理的な場所
(保存期間・保存場所)		Preservation	MIReG	永久保存資源の為の備考データ
*		Function	AGLS	資源内容が関連する業務種別
*		Availability	AGLS	(主としてオフライン資源に接触す るための)資源作成者/管理者情報
*		Mandate	AGLS	資源作成の為の法的根拠、条項の記述
*		Keywords	Canada	資源のキーワード( 検索システムに 対する単語の重み付けに影響)

表 3-5 海外政府メタデータ標準のエレメント名と国内行政文書との対応関係

# 3.1.5 利用者プロファイルのメタデータ定義に関する調査

利用者によって行政公開情報の利用目的は異なる。例えば、ある利用者は生活動態調査における消費指数を知りたいと思ったり、ある利用者は環境保護に関する法令情報を知りたいと思ったりする。これら利用者ごとの用途に柔軟に且つ効果的に対応できるようにするには、利用者個人に関するメタデータである利用者プロファイルの設定が必要である。本節では、利用者プロファイル用のメタデータ定義について調査を行なった。

# 3.1.5.1 P3P調査

P3P1.0 の勧告案(The Platform for Privacy Preferences 1.0 Specification: W3C Proposed Recommendation 28 January 2002)93の調査を行なった。P3P とは、Platform for Privacy Preferences Project(プライバシー情報取扱いに対する個人の選好を支持する技術基盤)の略であり、W3C が開発中の、インターネット上のプライバシー保護を目的とした技術標準である。現状では、Web 上のプライバシーポリシー(個人情報の取扱い方針)は各社で記述形式がまちまちであり、また、その都度各社のプライバシーポリシーを読んで確認することは消費者には大きな負担となっている。P3P 標準は、プライバシーポリシーの掲載項目等を標準化し、かつマシンリーダブルな形式(XML 形式)で記述することによって、プライバシーポリシーを自動処理しようとするものである。

# (1)カテゴリ

P3P1.0 勧告案では、利用者の個人情報に関するカテゴリとして、以下の 17 要素が挙げられている。

個々の個人データ(氏名、住所、性別、生年月日、年収、趣味等)をこれらのカテゴリ種別にしたがって分類することにより、Web サイトがそれらの個人データをどのように取り扱うかという規則を、個々の個人データごとにではなく、カテゴリごとに設定することができ、設定を容易化することができる。

<physical/> | ; 実社会における連絡先情報

<online/> | ; オンライン連絡先情報

<uniqueid/> |; ユニークな識別子

<purchase/> | ; 購入情報
<financial/> | ; 金融情報

<computer/> |; コンピュータ情報

<navigation/> |; ナビゲーションとクリックストリームのデータ

<interactive/> |; インタラクティブデータ

<demographic/> | ;人口統計学的・社会経済学的データ

<content/> |; 文章の内容

<state/> | ; 状態管理メカニズム

\_

<sup>93</sup> http://www.w3.org/TR/P3P/

<health/> |;健康情報

location/> |;位置データ

<government/> | ; 政府発行の識別子

<other-category> |; その他

各カテゴリの説明は、以下の通りである。

### <physical/>

実社会における連絡先情報:実社会において個人に問い合わせを行なったり、所在を突き 止めたりすることを可能にするような情報。電話番号や住所など。

### <online/>

オンライン連絡先情報:インターネット上で個人に問い合せを行なったり、所在を突き止めたりすることを可能にするような情報。電子メールアドレスなど。この情報は、ネットワークにアクセスするときに使用される特定のコンピュータには依存しないことが多い。

# <uniqueid/>

ユニークな識別子: 個人を整合的に特定したり、認識したりするために発行された識別子。 金融機関の ID 番号を除く。また、政府発行の識別子を除く。Web サイトやサービスから 発行された識別子を含む。

# <purchase/>

購買情報:商品やサービスを購入することによって積極的に生成される情報。支払方法の 情報を含む。

### <financial/>

金融情報:口座、残高、支払い、借越し、購入、クレジットカード、デビットカードなどの個人の金融情報。 個人による個々の購買に関する情報は、それ単体では「金融情報」には属さない。

### <computer/>

コンピュータ情報:個人がネットワークにアクセスするときに使用しているコンピュータシステムに関する情報。 IP アドレスやドメインネーム、ブラウザの種類、OS など。

#### <navigation/>

ナビゲーションとクリックストリームのデータ:Web サイトを閲覧することによって受動的に生じるデータ。訪問したページやページごとの滞在時間など。

### <interactive/>

インタラクティブデータ: Web サイトを通じた、サービス提供者との明示的なやりとりから積極的に生じるデータ。また、そのようなやりとりを反映したデータ。検索エンジンでの検索事項やアカウント活動のログなど。

### <demographic/>

人口統計学的・社会経済学的データ:個人の特徴に関する情報。性別や年齢、収入など。 <content/> 文章の内容:通信活動に含まれる言葉や表現。電子メールの文章や掲示板への掲示内容、 またチャットルームでの通信内容など。

#### <state/>

状態管理メカニズム:利用者とのセッションを維持したり、また以前に特定サイトを訪問したことや特定コンテンツにアクセスしたことがある利用者を自動的に特定したりするメカニズム。クッキーなど。

# <political/>

市民情報:宗教団体、労働組合、専門的な協会、政党などの会員、または所属。

#### <health/>

健康情報:個人の肉体的または精神的健康、性的志向、ヘルスケアサービスや製品の使用 または調査、ヘルスケアサービスや製品の購入等に関する情報。

# 

プリファレンス(嗜好)データ:個人の好みや嫌いなものに関するデータ。好きな色や音楽の好みなど。

#### <location/>

位置データ:個人の現在の物理的な位置情報と変更した場合にその位置の追跡のために使うことができる。GPS 位置データなど。

#### <government/>

政府発行の識別子:個人を識別するための政府が発行した識別子。

#### <other-category>

その他:上記の定義にあてはまらないその他のデータ。

上記のデータ要素のうち、利用者プロファイルとしての利用が可能なのは、<physical/>, <online/>, <uniqueid/>, <purchase/>, <financial/>, <computer/>, <demographic/>, <state/>, <political/>, <health/>, <preference/>, <government/>, <location/>である。ただし、金融情報<financial/>や健康情報<health/>、市民情報<political/>など、秘密性の高いデータも含まれるため、プロファイルに含まれる個々のデータについて、それを利用できる機関や、利用目的などに制限をかけることが望ましい。

その他の<navigation/>, <interactive/>, <content/>データ要素については、個人の行動履歴に関する情報であったり、一時的な情報であったりするため、利用者プロファイルの中に含めるのは適切ではない。

# (2)データ要素

P3P1.0 勧告案では上記のカテゴリの他に、個々の個人データまたは個人の行動に関連したデータとして、いくつかのデータ要素が規定されている。基本データ要素としては、以下のものが挙げられている。

#### 1. 日付

	• _ • •	<del></del>
∥date	∥カテブリ	簡旦某記夕
uaie	カノコン	間勿衣記古

ymd.year	(可変カテゴリ)	年	
ymd.month	(可変カテゴリ)	月	
ymd.day	(可変カテゴリ)	-ゴリ) 日	
hms.hour	(可変カテゴリ)	時	
hms.minute	(可変カテゴリ)	分	
hms.second	(可変カテゴリ)	秒	
fractionsecond	(可変カテゴリ)	秒(小数点以下)	
timezone	(カテゴリ)	タイムゾーン	

# 2. 名前

personname	カテゴリ	簡易表記名
prefix	人口統計学的・社会経済学的データ	敬称
given	実社会における連絡先情報	名(Given Name)
family	実社会における連絡先情報	姓
middle	実社会における連絡先情報	ミドルネーム
suffix	人口統計学的・社会経済学的データ	名前の接尾語(Name Suffix)
nickname	人口統計学的・社会経済学的データ	愛称

# 3. ログイン

login	カテゴリ	簡易表記名
id	ユニークな識別子	ログイン ID
password	ユニークな識別子	ログインパスワード

# 4. 認証

certificate	カテゴリ	簡易表記名
key	ユニークな識別子	認証鍵
format	ユニークな識別子	認証フォーマット

# 5. 電話

telephonenum	カテゴリ	簡易表記名
intcode	実社会における連絡先情報	国番号
loccode	実社会における連絡先情報	局番
number	実社会における連絡先情報	電話番号
ext	実社会における連絡先情報	内線
comment	実社会における連絡先情報	注釈

# 6. 連絡先情報

contact	カテゴリ	簡易表記名
postal	実社会における連絡先情報, 人口統計 学的・社会経済学的データ	郵便情報
telecom	実社会における連絡先情報	テレコミュニケーション情報
online	オンライン連絡先情報	オンラインアドレス

# 6.1 郵便

postal	カテゴリ	簡易表記名
name	実社会における連絡先情報, 人口統計学的・ 社会経済学的データ	氏名
street	実社会における連絡先情報	町・番地
city	人口統計学的・社会経済学的データ	市・区
stateprov	人口統計学的・社会経済学的データ	都道府県
postalcode	人口統計学的・社会経済学的データ	郵便番号
country	人口統計学的・社会経済学的データ	国
organization	人口統計学的・社会経済学的データ	団体

# 6.2 テレコミュニケーション

telecom	カテゴリ	簡易表記名
telephone	実社会における連絡先情報	電話番号
Fax	実社会における連絡先情報	ファックス番号
mobile	実社会における連絡先情報	携帯電話番号
pager	実社会における連絡先情報	ポケットベル番号

# 6.3 オンライン

online	カテゴリ	簡易表記名
email	オンライン連絡先情報	電子メールアドレス
uri	オンライン連絡先情報	ホームページアドレス

# 7. アクセスログとインターネットアドレス

# 7.1URI

Uri	カテゴリ	簡易表記名
authority	(可変カテゴリ)	URI 権限
stem	(可変カテゴリ)	URI ステム
querystring	(可変カテゴリ)	URI の照会列部分

# 7.2ipaddr

ipaddr	カテゴリ	簡易表記名
hostname	コンピュータ情報	完全なホストとドメイン名
partialhostname	人口統計学的・社会経済学的データ	ホスト名の一部
fullip	人口統計学的・社会経済学的データ	全 IP アドレス
partialip	人口統計学的	IP アドレスの一部

### 7.3 アクセスログ情報

loginfo	カテゴリ	簡易表記名
uri	ナビゲーションとクリックストリームデータ	要求されたリソースの URI
timestamp	ナビゲーションとクリックストリームデータ	要求とタイムスタンプ
clientip	コンピュータ情報、人口統計学的・社会経済学的データ	クライアントの IP アドレ スまたはホスト名
other.httpmethod	ナビゲーションとクリックストリームデータ	HTTP 要求方式
other.bytes	ナビゲーションとクリックストリームデータ	レスポンスのデータバイト
other.statuscode	ナビゲーションとクリックストリームデータ	レスポンスステータスコー ド

# 7.4 その他 HTTP プロトコル情報

httpinfo	カテゴリ	簡易表記名
referer	ナビゲーションとクリックストリームデータ	ユーザーが要求した最後の URI
useragent	コンピュータ情報	ユーザーエージェント情報

上記のデータ要素のうち、利用者プロファイルとしての利用が可能なのは、「2.4名前」、「3.0グイン」、「4.8認証」、「5.0電話」、「6.2 是絡先情報」、「6.1 郵便」、「6.2 テレコミュニケーション」、「6.3 オンライン」、「7.3 アクセスログ」の中の「clientip」、「7.4 その他 HTTP プロトコル情報」の中の「useragent」である。

その他のデータ要素については、個人の行動履歴に関する情報であったり、サーバ側の 情報であったりするため、利用者プロファイルの中に含めるのは適切ではない。

# 3.1.5.2 OECD プライバシーステートメントジェネレーター調査

OECD のプライバシーステートメントジェネレーター<sup>94</sup>の調査を行なった。OECD プライバシーステートメントジェネレーターとは、Web サイト上に掲載された規定フォームの質問に答えていくことにより、自社サイトのプライバシーポリシーを作成、ダウンロードできるオンラインツールであり、2000 年 7 月頃に OECD が Web サイト上で公開した。OECD プライバシー・ガイドライン 8 原則をインターネット上で普及させることを狙いと

91

<sup>94</sup> http://cs3-hq.oecd.org/scripts/pwv3/pwhome.htm

# するものである。

OECD プライバシーステートメントジェネレーターでは、Web サイトがサイト訪問者から収集する個人データとして、以下のような個人データの種別を規定している。

### 1. 主要な個人データ

- ・ 氏名
- 性別
- 住所
- ・ 電子メールアドレス
- ・ 電話番号 / FAX 番号
- ・ その他

### 2. ビジネス情報

- · 勤務先/組織
- 役職
- · 勤務先住所
- ・ 勤務先電子メールアドレス
- · 勤務先電話番号 / FAX 番号
- ・ その他

# 3. 詳細な個人データとプロファイルデータ

- ・ 個人の詳細:ニックネーム、生年月日/年齢、出生地、国籍など.
- ・ 物理的な詳細:慎重、体重、その他の身体的特徴など
- ・ 家族構成:婚姻、パートナーシップ、扶養家族など
- ・ 教育とスキル:学歴、専門的興味など
- ・ ライフスタイルまたは嗜好:商品・サービスの購入情報、レジャー、スポーツ、個人または家族のふるまい、喫煙、飲酒、好きな色、好きな食べ物など
- ・ 金融資産:収入、財産など
- ・ その他

### 4. 識別子

- ・ オンライン識別子: Web サイトのパスワード、クッキー、訪問者の身分証明書(個人認証)など
- ・ 金融識別子:クレジットカード番号、銀行口座番号など
- ・ 公共機関に割り当てられた識別子:社会保障番号、個人識別番号など
- ・ バイオメトリック識別子:DNA、虹彩認識、指紋など
- ・ その他

# 5. 特殊なデータ

・ 民族または人種

- · 政治的見解
- ・ 宗教的信仰または哲学的心情
- ・ 労働組合への加盟
- · 医療健康情報
- 性生活
- ・ 警察・司法データ
- ・ その他

また、個人データ(個人を識別できるデータ)でないデータとして、以下のものを挙げている。

- ・ IP アドレス、選好する言語、セッション(番号、キー)、Web ページへの滞在時間、閲覧したバナー広告、訪問した Web ページ
- ・ ただし、上記データが特定個人と結び付けられている場合は、個人データとみなされる。

上記の個人データ種別のうち、利用者プロファイルとしての利用が可能なのは、「1.主要な個人データ」「2.ビジネス情報」「3.詳細な個人データとプロファイルデータ」「4. 識別子」である。

「5.特殊なデータ」については、個人情報保護の観点から、利用者プロファイルの中に 含めるのは適切ではない。

# 3.2 情報のセマンティック検索に関する調査

電子政府関連の情報に関するセマンティック検索システムを実現するための要素技術 として、下記の調査を行なった。

- ・メタデータ管理システム調査
- ・半自動メタデータ生成システム調査
- ・対話型メタデータ生成システム調査
- ・利用者プロファイル定義システム調査
- ・セマンティック検索システム調査

Web 情報にメタデータを誰が付けるのかという問題はセマンティック Web の実用化・普及に取って一番大きな課題であるとなると言われている。有効なメタデータが広く付けられることでセマンティック Web の有効性は加速度的に高まる。そこでメタデータを生成するツールが重要になってくる。自動的に生成できることが最も望ましいが、現在のWeb ページはコンピュータが理解できるようには書かれていないので完全には自動生成できない。コンピュータが生成したメタデータを人手で手直しするとか、対話型メタデータ生成システムを用いて人の労力を削減するなどの半自動的なメタデータ生成ツールに期待が掛けられる。

### 3.2.1 メタデータ管理システム調査

メタデータ管理システムとしてどのようなものが現在存在しているか、また今後どのようなツールがあり得るかについて調査を行なった。

メタデータもデータの一種であることに変わりはないため、その管理に既存の DBMS (データベース管理システム)を利用することは十分可能である。実際、いくつかの研究では、インターフェイス・レイヤを介してリポジトリにアクセスすることで、リポジトリの実現によらないメタデータ管理システムを構築するアプローチがとられている。リレーショナルデータベース(RDB)は、スケーラビリティの点からリポジトリの有力な候補であるが、一般にはインターフェイスとの間の対応にギャップが存在する。メタデータの管理をオブジェクト指向データベースで行なうものなどもある。

またメタデータの記述は XML を用いるケースが増えているため、XML のデータ管理システムそのものがリポジトリとして有効である。 XML データの管理システムとしては、 XMLデータ中のテキストを RDBのデータ構造にマッピングしてデータベース化するものと、 DOM のインターフェイスを用いてツリー形式の構造をそのままバイナリ化してデータベース化するものが存在する。 大量のメタデータに複数のエージェントがアクセスする可能性があるセマンティック Web の世界では、高速性の観点から後者の XML データをオブジェクトとして管理する技術が、鍵を握る要素技術の一つとなると考えられる(既に商用製品は登場している)。

以下、既存の DBMS よりももう一段階上のレイヤーで行われているメタデータ管理の試みを紹介する。(1)ではメタデータ間のオントロジーの違いを吸収するアーキテクチャも検討されている管理システムとして、Multi-schema Metadata Format(MMF)を説明する。(2)ではセマンティック Web のメタデータとして、第一に挙げられる RDF に特

化したメタデータ管理の取り組みの事例を紹介する。(3)では、メタデータの設置や収集方法に焦点をあてて、その事例を紹介する。

# (1) Multi-schema Metadata Format(MMF)

セマンティック Web が話題になる以前から、データベース毎にそれぞれ[生産県]、[生産地方」など、商品に関して異なるスキーマに基づくメタデータの差異をどのように吸収するかが問題になっている。これを解決するため、ディジタル・ビジョン・ラボラトリーズ(DVL)によって開発された、Multi-schema Metadata Format(MMF) 95がある。商品特徴情報をメタデータ化し、WWW 上で提供するためのメタデータ形式である。MMFでは、メタデータをメタデータインスタンス、スキーマ定義、スキーマオントロジー、標準属性辞書の4階層の構成とすることで、様々な人間が定義したメタデータ間のオントロジーの違いを吸収している。同時に、Metadata Mediator というエージェントを介してMMFの検索や、Metadata Mediator間で交換するためのプロトコル Metadata Mediation Protocol (MMP) を規定し、MMFの運用を支援する。メタデータインスタンスは XMLで記述され、Metadata Mediator内のデータベースに保持されるが、どのような実装形式(DBMS も含め)でメタデータを管理するかは規定していない。MMFと MMPは、オントロジーの管理や運用までを考慮したメタデータ管理システム体系ということで、セマンティック Web に近いアプローチと言える。

### (2) RDF の蓄積・検索に特化したデータベース

また、RDFに特化したデータベースの開発が研究レベルで進められている。

European Commission Joint Research Center (JRC)で開発されている RDFStore (Perl API for RDFStorage)%は、RDFの構造を解析、格納、管理する Perl のライブラリーである。このプロジェクトは 2001 年の 10 月に始まっている。RDF の 3 つ組(リソース、プロパティ、値)の検索を効率的に行なうことを目的としている。ここでは、RDQL/SQUISH と呼ばれる RDFStore で蓄積されたデータに対して SQL のような形式で検索するための表記法などについても開発している。

同様なシステムとして、 $RDFdb^{97}$ がある。(RDFStore よりも先に開発されている模様)。 こちらも Algae と呼ばれる SQL のような形式で RDF のステートメントを検索する言語が 開発されている。

Sesame<sup>98</sup>は、EU 委員会 IST プログラムの 1 プロジェクトである On-To-Knowledge で Aidministrator Nederland によって開発された、RDF スキーマベースの Repository and Querying facility である。リポジトリには RDF データと RDF スキーマ情報を格納することができ、RQL と呼ばれる問合せ言語を用いて検索を行なう。リポジトリとして現在は Object-Relational DBMS である PostgreSQL が使われている (Repository Abstraction

 $<sup>^{95}</sup>$  坂田毅、多田浩之、大竹智久(1997)「メタデータの WWW 上への実装と電子商取引への応用」画像電子学会第 9 回メディア統合技術研究会

<sup>96</sup> http://rdfstore.sourceforge.net/

<sup>97</sup> http://www.w3org/2001/Talks/0505-perl-RDF-lib/

<sup>98</sup> http://sesame.aidministrator.nl

Layer を通じて様々なタイプのリポジトリを使用可能とされる<sup>99</sup>)。

RDFの取り扱いに特化したデータベースでは、RDFの3つ組を中心とした処理が行われると考えられる。RDFの3つ組によって記述されたリソース同士は柔軟に結合することができるが、それらをRDFデータモデルのままで参照することが可能である。従来のデータベースは構造化が進み柔軟性を失っているのに対して、RDFデータベースでは半構造的なデータ管理が可能であり、さまざまなアプリケーションからの要求にあわせてデータベースを変更するコストの削減などの効果も期待できる。

# (3) セマンティック Web におけるメタデータの設置・収集・管理方法

ここまでは、データをシステムがハンドリングするための要素技術について述べたが、 アプリケーションによる利用や、サービスという側面から考えると、メタデータをどこに おいて管理するかが問題となる。

# (a) 既存の Web サイトへのメタデータの設置

セマンティック Web では、Web の特性としてコンピュータネットワーク上のさまざまなユーザーがメタデータの生成や管理に携わることになる。個々のメタデータの置き場所が、各 Web サイトにあるというケースがまず考えられる。セマンティック Web のファーストステップとして、HTML 中に RDF を記述することが唱えられている。そして、RSS (RDF Site Summary) 100はライトウェイトな RDF 準拠のメタデータフォーマットであるが、これも基本的には各 Web サイトがサイトの情報の要約などを提供するのに使われている。これらのケースは Web 全体が巨大なメタデータ管理システムという位置付けになる。

### (b) メタデータの集中管理

RDF のリポジトリを構成する試みもある。W3C の Annotea プロジェクト<sup>101</sup>のように Annotea サーバをインターネット上に置いておくことによって、RDF で記述されたアノテーション情報を集中的に管理してユーザーに提供できる。ユーザーが Amaya などの Web ブラウザ上で表示している情報の URI に対応したアノテーション情報をサーバに問い合わせる形になる。

# (c) メタデータのインデックス管理及び収集ツール

RDFWeb  $^{102}$ のように、メタデータ自体は個々のユーザーが作成・設置するが、そのインデックスのようなものを管理する機構も存在する。 DARPA による DAML プロジェクトにおいても DAML によって記述された Web 上のステートメントを収集する試みも  $^{2001}$  年  $^{5}$  月から DAML Crawler  $^{103}$ によって開始されている。

また、RDF Crawler<sup>104</sup>もメタデータ収集ツールであり、インターネットから RDF フラグメントをダウンロードして知識ベースを構築する。RDF Crawler は Java API を提供す

103 http://www.daml.org/crawler/

<sup>99</sup> http://www.On-To-Knowledge.org/downl/sesame.ppt

<sup>100</sup> http://groups.yahoo.com/group/rss-dev/files/specification.html

<sup>101</sup> http://www.w3.org/2001/Annotea/

<sup>102</sup> http://rdfWeb.org/

<sup>104</sup> http://ontobroker.semanticWeb.org/rdfcrawl/

るので、単独のアプリケーションとして利用する以外に、他のツールの中に埋め込むことも可能となっている。

SHOE<sup>105</sup>はWebのためのオントロジー言語であり、Webページの作者が機械処理可能な形式でドキュメントにアノテーションを付ける事ができる。SHOEでは、作者が付けた情報をWeb上に置くダイレクト・アクセスと、Web-crawlerを使って集めて知識ベースに格納するリポジトリ・ベースド・アクセスの2つのアプローチが考慮されている<sup>106</sup>。後者に関しては、SHOEデータの格納と検索エンジンへのアクセスのためのJava APIが提供されており、Web-crawlerのExposéなど、このAPIを使うツールが各種提供されている。

<sup>105</sup> http://www.cs.umb.edu/projects/plus/SHOE/

<sup>&</sup>lt;sup>106</sup> Jeff Heflin, Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment, Ph.D. Thesis, University of Maryland

# 3.2.2 半自動メタデータ生成システム調査

半自動メタデータ生成システムとしてどのようなものが現在存在しているか、また今後 どのようなツールがあり得るかについて調査を行なった。半自動メタデータ生成システム の一部としてのメタデータ生成規則定義システムの調査(メタデータ生成規則の調査を含む)も行なった。

セマンティックWeb社会の実現においては、少なくともあるコミュニティー(電子政府や業界団体など)でメタデータの枠組み(属性、エレメントリスト)を決め、それに従ってメタデータを付与するという実際の作業が必要である。

特に、メタデータの付与については、Web時代でリソースが爆発的に増えている状況にあり、人手工数の削減は大きな課題となる。膨大で良質なメタデータを継続的に付与、更新していくことができなければ、セマンティックWeb社会は絵に描いた餅にすぎない。

こうした状況を背景に、コンピュータによる半自動メタデータ付与によって、膨大な人手工数の 削減だけでなく、質のばらつきの少ないメタデータの作成が期待できる。本節では、半自動メタデ ータ生成について、関連する技術を調査した。

# 3.2.2.1 メタデータ自動付与における背景技術 メタデータ自動付与のためには、以下のような技術が必要である。

# (1) HTMLパーザ

メタデータ付与対象がHTMLの場合、HTMLソース内の特定タグにメタデータとなるべき情報が含まれていることがある。その場合は、HTMLパーザによりソースのタグ構造を解析し、必要な部分を取出せば良い。例えば、以下のようなタグがある。

### METAタグ

METAタグはHTMLの規格としてフォーマットが決まっているだけで、エレメントは作者が自由に付けて良いことになっている。おおむね、以下の2つは多くのページで良く見られる。

- KEYWORD: キーワードを記入する (DC:Subject)
- DESCRIPTION: ページの説明文を記入 (DC:Description)

TITLEタグ (DC:Title)

(例)実際には、これらのメタデータは次のようにHTML内に埋めこまれる。

<head>

- <META NAME="AUTHOR" CONTENT="H. Tsuda">
- <META NAME="KEYWORD" CONTENT="情報検索,テキストマイニング,Webマイニング,自然言語処理">
- <title>テキストマイニング研究のページ</title>

</head>

#### (2)情報抽出技術

情報抽出 (Information Extraction ) とは、文書から特定のイベント情報 (あるイベント5W1H情報)などを自動で抜き出す技術である。自然言語処理研究では、従来、新聞などのテキスト群を対象に、特定パターンの文字列から情報を抜き出すなど、多くの研究が行なわれてきた。Web文書を対象としての最近の動きとしては、Wrapper(ラッパー)と呼ばれる、タグを手掛りにした手法が多くなっている。ラッパーを使った応用としては、複数のショッピングサイトから商品の価格情報を取出して横断検索をできるようにするとかのように、Web情報をデータベース化するのが一般的である。

例えば、坂元論文<sup>107</sup>には、以下のようなHTML文書から国別の統計情報を抜き出しレコード化する例と、その自動化手法が解説されている。

#### (例)

元の HTML文書:

<HTML><TITLE></TITLE><BODY>

<B>Congo</B><I>242</I><BR>

<B>Egypt</B><I>20</I><BR>

<B>Belize</B><I>501</I><BR>

<B>Spain</B><I>34</I><BR>

</BODY></HTML>

抽出されるレコード情報:

((Congo, 242), (Egypt, 20), (Belize, 501), (Spain, 34))

情報抽出技術はメタデータの特定エレメント抽出にも利用可能である。例えば、DC:Rights (copyright表記を手掛り)、Location (住所表現を手掛り)などが考えられる。

### (3)文書自動分類技術

文書自動分類技術は大別すると、以下の2種類に大別できる。

クラスタリング (clutering): 文書群を与え、文書の類似度により複数のグループに分けるカテゴライゼーション (categorization): 文書群と既存クラスを与え、文書群をクラスに割り当てる。

メタデータ自動付与としては、カテゴライゼーションの方が関連が深い。というのも、例えば DC:Subject のようなエレメントは、統制されたシソーラス語彙を付与することが推奨されている。 統制語によるクラスへの割当てと考えれば、カテゴライゼーションの問題に帰着する。

カテゴライゼーションは以下のような技術の組合せで実現される。

・ 形態素解析:文を単語に分割する。日本語の場合は単語間にデリミタがないので、辞書やルールを用いた処理になる。英語の場合は、動名詞の活用の統一(stemmingと呼ばれる)などが必要。

<sup>107</sup> 坂元,有村「Webマイニング」人工知能学会論文誌, Vol.16, No.2,pp.233-238, 2001

- ・ キーワード抽出: 単語の中から、文書やクラスに特徴的なキーワードを取出す。特徴的なキーワードとは、例えば、その文書に多く現われ、他にはあまり出現しないものというものが考えられる。 TFIDFと呼ばれる有名な単語の重みづけ手法は、その一例である。
- ・ 特徴素選択: 文書やクラスの特徴となる情報を選別する。たとえば、キーワードの列(ベクトル) とかが一般的。Web文書の場合には、URL特徴や、リンク元のアンカー文字列の組込みも有効 とされる。
- ・ 類似度マッチ: 文書やクラスの特徴素間の類似度を定義し、その近さを計算する。例えばキーワードベクトルに対しては、その方向の近さをベクトルの内積(二本のベクトルの作る角度のコサイン)で表すなどの手法がある。

### (4)その他

例えば、Dublin Coreエレメントのうち、Date(日付)、Format(フォーマット)、Identifier (Webの場合はURL)などは、WebにおけるHTTPプロトコルの結果などから得ることができる。

### 3.2.2.2 メタデータ自動生成ツール108

# (1)メタデータエディタ

メタデータエディタとは、メタデータの中身を入力するのを補助するツールやテンプレートのことである。メタデータの中身は基本的に作業者が付与し、出力フォーマット(RDF、HTML内のMETAタグ埋め込みなど)への変換はエディタが自動的に行なうことで、人手によるミスと労力を減らす。

Dublin Core Metadata template

[開発元] Nordic Metadata Project

[URL] http://www.ub.lu.se/metadata/DC\_creator.html

[対応メタデータ] ダブリンコア

[出力フォーマット] HTML (3,4) Metaタグ埋め込み

Webのフォームになっており、ダブリンコアの15エレメントのうち必要な情報を入れると、HTMLのMETAタグ埋め込み形式のダブリンコアメタデータを作成する。利用者は、その部分を自分のページにカット&ペーストする。出力形式はHTML(3,4)のみ。

Reggie Metadata Editor

[開発元] Distributed Systems Technology Centre

[URL] http://metadata.net/dstc

[対応メタデータ] ダブリンコア、スキーマファイルを与えれば他のAGLSなどにも対応可 [出力フォーマット] HTML3.2(METAタグ)、HTML4.0 (METAタグ)、RDF、RDF Abbreviatied ダブリンコアだけでなく、スキーマファイルを定義することで、独自のメタデータ形式にも対応で きる。

<sup>108</sup> http://www.dstc.edu.au/Research/Projects/metaWeb/toolpost.html を参考にした。

MetaWeb Generic Edit Tool

[開発] State Library of Tasmania

[URL] http://www.dstc.edu.au/RDU/MetaWeb/generic\_tool.html

[対応メタデータ] ダブリンコア、AGLS,ADMIN Core

[出力フォーマット] HTML (Metaタグ)

Tasmania Online serviceで用いられている。出力されるファイルをHTMLやテキストに埋め込む。

# (2) メタデータジェネレータ

エディタは作業者が全ての属性値を入力する必要があるのに対して、ジェネレータは、いくつかの属性値はシステムが自動で作成してくれるものである。メタデータのエレメントにより、自動化の精度はまちまちである(詳細は(2)参照)。したがって、ジェネレータは、実際はツールとしてはエディタ機能も持ち、誤って付与された値は簡単に修正できるようになっている。

DC-dot

[開発] UKOLN, University of Bath

[URL] http://www.ukoln.ac.uk/metadata/dcdot/

[対応メタデータ] ダブリンコア, 他の形式 (USMARC, SOIF, IAFA/ROADS, TEI headers, GILS, IMS)にも変換可能

[出力フォーマット] HTML (Metaタグ), RDF

URLを与えると、そのページ本文からダブリンコアメタデータエレメントを自動的に作成する。いくつかデータを与えた結果から類推するに、大体次のような動作をしているようである。

DC.Title: <title>タグを参照する。

DC.Subject: <meta>タグのKEYWORDを参照する。無い場合には、アンカー文字列から抜き出したキーワードを入れる場合もある模様。

DC.Description: <meta>タグのDescriptionを参照する。

DC.Date: 最終更新日を参照する。

DC.Format: htmlの場合は text/htmlとする。

Dc.Identifier: URLを参照する。

HiSoftware TagGen

[開発] HiSoftware社

[URL] http://www.hisoftware.com/taggen.htm

[対応メタデータ] GILS, WAGILS, ダブリンコア

HTML(XHTMLにも対応)ファイル群を選ぶと、ウィザード形式でメタデータを付与してくれるツール。

- ・キーワードは人手付与または本文からの自動抽出が可。スペルチェックや、シソーラスによる統制語への変換にも対応
- ・自動的にメタデータつきのページに変換し、サーバにアップする。(形式はHTML内METAタグで、RDFなどはない様子)

- ・検索エンジンコントロール用メタデータにも対応。例えば、ロボットがリンクを辿って良いかどうか、インデクシングして良いかどうかなど。
- 3.2.2.3 実例 1: オーストラリアの政府系メタデータ AGLS と半自動生成ツールオーストラリアは非ヨーロッパ圏ではメタデータに関して最も進んでいる国と言われる。オーストラリアにおける政府系メタデータとして有名なものがAGLS (Australian Government Locator Service)である。AGLSは、オーストラリアのNational Archives of Australis (NAA)により、政府系公開情報リソース(文書、サービス)に対して付与されているメタデータエレメント標準である。NAAでは、メタデータを付与した政府系リソースをアーカイブとして管理している。AGLSはダブリンコアを拡張した19のエレメントからなり、1997年から1998年にかけて策定された109。AGLSメタデータは基本的には、リソースの作者(Creator)または出版人(Publisher)が

例えば、Subjectフィールドは、リソースの主題キーワードを記述する。シソーラスによる統制キーワードを付与することが推奨されている。例えば、

付与するものである。その作業において、メタデータ付与ソフトウェアが利用できる。

# Library of Congress Subject Headingsの例 [LCSH] Biographies - Government - Australia

# MeSH - Medical Subject Headingの例 [MeSH] Gene Expression Regulation, Bacterial

AGLSのメタデータ付与ツールには以下のようなものがある。

Reg: 前述のReggieの軽量版

メタデータエディタである。対応メタデータは、ダブリンコア、AGLS (ver. 1.0, 1.1, 1.2), EdNA, QLDGOV。他にもスキーマを定義すれば他メタデータにも拡張可能である。

Klarity

[URL] www.klarity.com.au

URLを入れると、ページ内テキストを解析して、TITLE, KEYWORDS, DESCRIPTIONを入れて

 $^{109}$  AGLS のエレメントは以下の 19 個である。ダブリンコアと同様に、qualifier も許される。 (NAA のホームページ http://www.naa.gov.au/ よりアクセスできる)

Creator Publisher Contributor Rights Title Subject Description Source Language Relation Coverage Function Date Type Format Identifier Availability Audience Mandate

AGLS はオーストラリアにおける各種メタデータの基本と言っても良く、実際 Australia and New Zealand Land Information Council (ANZLIC) 以外の次のようなメタデータは、全てダブリンコアまたは AGLS 準拠している。

- Educational Network Australis (EdNA)
- Environmental Resoruces Information Network (ERIN)
- Business Entry Point (BEP)
- HealthInsite

102

くれる。文書自動分類を行なっていると思われる。メタデータ形式はダブリンコアである。出力形式はHTML, RDFに対応している。

### 3.2.2.4 実例 2: 富士通研究所による Web ディレクトリ自動生成ツール

津田論文<sup>110</sup>では、Yahooのようなディレクトリを自動的に作成することを目的として、いくつかのページメタデータの自動付与を試みている。一部は商用に使われている。

Webディレクトリは、人気がありかつ情報の入口となるメニュー的なページがコンテンツとして望まれる。そのため、そのようなページを選別するためのメタデータを付与しているのが特徴的である。技術的には、Webのリンク関係の解析により、そうしたメタデータを付与しているのが特徴的である。ダブリンコアを拡張した、以下のエレメントの自動付与を行なう。

- Description: METAタグ内またはページ本文の最初

- Publisher: ホスト名

- Date: 最終更新日またはcrawlerによる最終取得日

- Type:ページ(ナビゲーション)タイプ。
メニュー、リンク集、コンテンツを、ページ間のリンク関係を元に判別している。例えば、メニューページはローカルリンク数や被リンク数が多い、リンク集はサーバ外へのリンク数が多いなどの特徴を使っている。

- Format: HTMLなど

- Identifier: URL。ミラーサイトは同一のものとして縮退。末尾のindex.htmlなどは適宜削除

- Source: リンク元URL集合

- Language: コード種別からヒューリスティックスで判定

- Relation:特定ドメインとの関連度 リンクの近さで定義する。社内向けディレクトリであれば、社内ページからのリンクトポロジの 近さにより、業務に関連した情報かそうでないかを判別する。

- Coverage (Location):情報抽出により、ページ本文中の関連地域を取出す。
- Rank:ページ人気度。そのページがどのような他ページからリンクされているかを基準に算出し、 時系列的動きの情報も加える。

このようにして、地域/ジャンル/時間と様々な観点を組み合わせた、多観点のディレクトリインターフェイスを実現している。図 3-9は、経済産業省に関係するカテゴリのページ画面である。 Rankの情報は、ダブリンコアにはないが、情報の新鮮度を与えるため、特にWebのように動きの激しい世界では重要となる。例えば、図 3-10に、厚生労働省の狂牛病関連のページの動きを示す。 上に行くほど、そのページへのリンクが増え、世間から着目されていることがわかる(数字は、同ページの注目度順位を表し、小さいほど上)。世間で話題になった2001年10月中旬から、同ページへのリンクが増えているということがわかる。

 $^{110}$  津田,鵜飼,三末「Web ディレクトリのためのページメタデータの自動付与の試み」情報学シンポジウム 2002, pp.17-24, 2002

Webの場合、アクセスログがサーバにしか残らないこともあり、外部からページの人気度を測定するのは難しいが、リンク関係の解析によりある程度人気度を測定することが可能である。

Webリソースは雑多なので、著者がつけたメタデータだけでなく、外部からの評価指標も必要と考えられる。そのようなメタデータ付与には、ここで紹介したような技術が必要と考えられる。



図 3-9 経済産業省のカテゴリ



- [タイトル] 「牛海綿状脳症(BSE)関係」ホームページ
   説明] (なし)
   (キーワート] (なし)
   [ティレクトリ登録済] (厚生労働省)(人)
   収集日最終更新日] 20020116,
   ジャナル」厚生労働省,
   地域」,
   [順位] 21348
   [一次格納ディレクトリ] 20020112/185/18545#7 (mache)
   [内部[0] 128304472
   [ランク重み] 5594, [参照数] 53(うちリモート 7), [被参照数] 57(うちリモート 57)

# 1. 人気度ランキングの変動履歴

# 10月末から 伸び

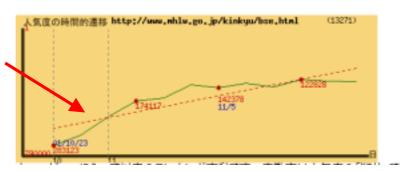


図 3-10 厚生労働省の狂牛病関連情報ページの動き

# 3.2.3 対話型メタデータ生成システム調査

セマンティック Web 技術を利用するにあたって、ひとつの大きな課題はオントロジー (Ontology)と呼ばれる実世界の知識の記述を如何におこなうかということである。RDF による記述は、明確ではあるものの、定義された内容の相互関係を直感的に把握するのは困難である。複雑な知識を記述するためには、内容を図を用いて可視化したり、対話的に編集したりできる編集システムが必要である。このために開発されているいくつかオントロジーの作成・編集システムについて説明する。

# (1)オントロジーの作成・編集システム

Amsterdam 大学 Social Science Informatics (SWI)では、WonderTools<sup>111,112</sup>というプロジェクトで、利用者が適切なツールを選択するのを支援するという目的で、オントロジーを作成するための以下のツールの評価をおこなっている。

- · Ontolingua
- · WebOnto
- · ProtégéWin
- · OntoSaurus
- · ODE
- · KADS22

この中からスタンフォード大学 KSL( Knowledge Systems Laboratory )の Ontolingua システム<sup>113</sup>を紹介する。

Knowledge Acquisition, Modeling and Management (KAW '99), 1999

<sup>111</sup> WonderTools, http://www.swi.psy.uva.nl/wondertools/

Duineveld, A. J., Stoter, R., Weiden, M. R., Kenepa, B., Benjamins, V. R. Wonder Tools? A comparative study of ontological engineering tools, In Proc. the 12th Internatinal Workshop on

<sup>113</sup> Ontolingua, http://www-ksl-svc.stanford.edu/

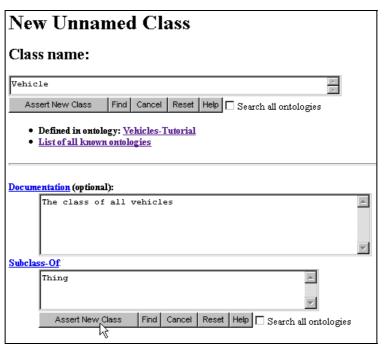


図 3-11 Ontolingua におけるクラスの定義画面 (Vehicle クラスの作成)

Ontolingua システムは、サーバと記述言語からなり、Web ブラウザからサーバにアクセスしてオントロジーの作成、編集、挿入や統合を行なえる。また、複数の利用者が協同して設計できるという特長を持っている。図 3-11及び図 3-12に Ontolingua のオントロジー作成 Web 画面を示す。

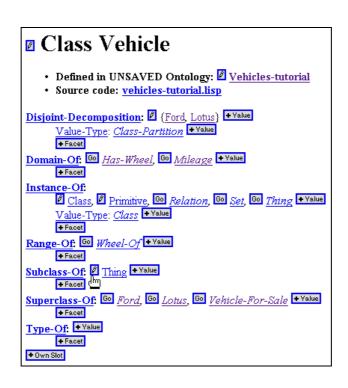


図 3-12 Ontolingua におけるクラスの定義画面(Vehicle クラスの属性を定義)

オントロジーは、一般的に概念とその関係性の記述として表現されるが、Ontolinguaでは、それぞれクラス及びスロットとして表現している。

図 3-11は、Vehicle-Tutorial という新しいオントロジーを定義したのち、Vehicle-Tutorial に「車」を表す新しいクラス Vehicle を定義している画面である。クラス Vehicle は、"the class of all vehicle"を記述したもので、既に定義されている「物」を表すクラス Thing のサブクラスとして定義している。図 3-12は、クラス Vehicle にいくつかのスロットを定義した後の画面である。

Ontolingua は、公理(Axiom)と呼ばれる規則を定義する機能も有している。例えば、「2匹の動物が兄弟ならば、彼ら両方の母であるものが(1匹)存在する」という規則(公理)によって、兄弟という概念と母子という概念の関係を記述している。

サーバは、スタンフォード大学のサイト<sup>113</sup>から使用できるようになっており、上述した 機能のガイドツアーも用意されているので、詳細についてはそちらを参考にされたい。

上記 WonderTools によって挙げられたツール以外に、商用システムとして OntoEdit システムがある<sup>114</sup>。OntoEdit はカールスルーエ大学の AIFB が開発し、独 Ontoprise 社 が商用化したもので、Ontolingua と同様にオントロジーの生成、処理等の機能を有し、さらに W3C 標準である RDF( Resource Description Framework )や、DAML+OIL、F-Logic の各形式に対する入出力機能を有している。OntoEdit のオントロジー作成画面を図 3-13 及び図 3-14 に示す。

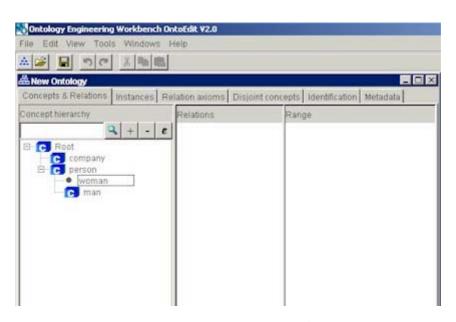


図 3-13 OntoEdit におけるコンセプトの定義画面

-

<sup>114</sup> OntoEdit, http://www.ontoprise.de/

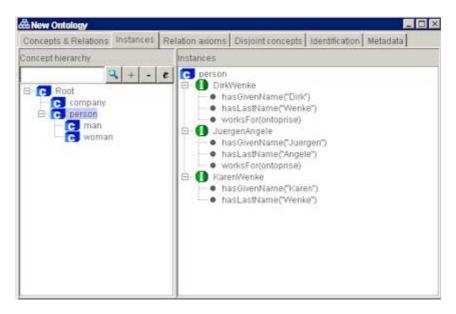


図 3-14 OntoEdit におけるインスタンスの作成画面

OntoEdit では、情報を階層的に表示できる。このため、概念階層を辿りながら必要な情報を編集できる。図 3-13は、人(person)という概念の下位概念として女性(woman)や男性(man)を定義している画面であり、図 3-14は、人(person)という概念の実体(インスタンス)として3名を定義している画面である。

また、OntoEditでは、公理を記述し、それを適用するために、F-Logic という述語論理の処理系を組み込んでおり、述語論理により規則(公理)を記述できる。しかし、F-Logic による公理の記述は、利用者には難解である場合が多いため、公理を設定するたに、「背反 (disjoint)の概念  $^{115}$ 」や「対称的 (symmetric)  $^{116}$ 」「反対 (inverse)  $^{117}$ 」「 $^{118}$ 推移的 (transitive)」な「関係」を指定できるようになっている。例えば「女性」と「男性」は背反の概念であるという公理を F-Logic で記述せずに、「女性」と「男性」のコンセプト間は「背反」であると設定することにより公理とすることができる。これにより実体が「女性」かつ「男性」を継承することはできなくなる(即ち、女性でありかつ男性であることはない)。

<sup>115</sup> 背反の概念:概念 A と概念 B の両方の下位概念になることはないこと。上記、女性・男性の例参照。

 $<sup>^{116}</sup>$  対称の関係:一つの関係 R に対して A と B とがある関係 R にあるとすると、B と A とも同じ関係 R にあること。例えば、A と B とが兄弟であれば、B と A とも兄弟であると言って良い。

 $<sup>^{117}</sup>$  反対の関係:二つの関係 R と S について、A と B とが関係 R にあるとき、B と A とは関係 S にあること。R が会員を表し、S が所属団体を表すとすると、A が B の会員であれば、B の所属団体に A があるというのは真である。

 $<sup>^{118}</sup>$  推移的な関係:一つの関係 R に対して、A と B が関係 R にあり、B と C とも関係 R にあるとき、A と C も関係 R であること。例えば、R が到達可能の関係を表すとき、A から B へ到達可能で、B から C へ到達可能であることが成り立つ。

## 3.2.4 利用者プロファイル定義システム調査

利用者プロファイルとは、サービスの提供者がその利用者各々に対して適切なサービスを提供するために参照するデータのことであり、いわゆる情報サービスのパーソナライゼイションやワン・トゥ・ワン・マーケティングに用いられる。電子政府においては、国民に対して年齢層や地域特性に応じた情報サービスを行なうために重要な役割を果たすものと考えられる。

利用者プロファイルの身近な例としては、ユーザーエージェント(Web ブラウザなど)で管理され Web サーバ側で参照される Cookie がある。Cookie の典型的な用途は、ある Web ページに対して各訪問者が何回訪れたかを記録しその回数をページ上に表示したり、会員登録した訪問者の登録内容(の一部)を記録して次回訪問時に自動的に認証を行なったりすることである。

- (1)プライバシー情報取扱いに関連した利用者プロファイルの定義 利用者プロファイルは次のように分類できる。
- (1) 利用者個人に関する情報(主に認証やサービスの個人向け最適化に利用):氏名、性別、年齢、住所、電話番号、Eメールアドレス、公開鍵証明書など
- (2) 利用者の嗜好情報(主にサービスの個人向け最適化に利用):閲覧したコンテンツや利用したサービスの一覧(履歴)など
- (3) 利用者の端末デバイス関連情報(主に端末向けのコンテンツ最適化に利用):画面の 解像度、カラー表示の可否、音声出力の可否、OS の種類など

これらは利用者個人のプライバシーに関わる情報であり、特に(1)や(2)に該当する情報をサービス提供者側が無断で収集したり、利用したりすると問題である。そこで、W3CではP3P<sup>119</sup>という技術標準を策定中であり、サービスを提供するサイトがプライバシーポリシーを利用者に明示したり、サイトとユーザーエージェントとの間でプライバシーポリシーに関する不一致を自動的に検出したりできるようにしている。また、サイトが収集する利用者の個人情報について、データの分類を行なっている(3.1.5.1節を参照のこと)。

P3P 1.0 勧告案ではプライバシーポリシーなどを XML で記述するようになっており、テキストエディタで直接記述できる。しかし、サイト管理者や Web ブラウザの利用者が全てのタグとその意味を理解して記述することは困難であるため、様々な設定支援ツールが開発されている。

Microsoft 社は現在最もポピュラーな Web ブラウザである Internet Explorer のバージョン 6 で P3P に対応し、Cookie のフィルタリング処理について利用者に簡易な設定手段を提供している<sup>120</sup>。また、財団法人ニューメディア開発協会では、P3P 準拠のプライバシーポリシーの作成を支援する各種ツールを開発、無償提供しており、サーバ(Web サイト)側とクライアント(ユーザーエージェント)側のそれぞれについて詳細な設定が可能になっている<sup>121</sup>。

.

<sup>119</sup> http://www.w3.org/P3P/

<sup>120</sup> Internet Explorer 6 のプライバシー機能:

http://www.microsoft.com/japan/developer/articles/dnpriv/html/ie6privacyfeature.asp

<sup>121</sup> プライバシー情報管理システム: http://www.nmda.or.jp/enc/privacy/

## (2) Web サイト構築システムにおける利用者プロファイル定義

Web サイトから提供されるコンテンツやサービスを各利用者の個人情報や嗜好に応じて最適化するサーバ側のパーソナライゼイション機能は、多くの商用 Web サイト構築システムに搭載されている。IBM 社の WebSphere には WebSphere Personalization というオプション製品が用意されており、専用のルールエディタによって利用者プロファイルのルール定義とコンテンツ定義が可能になっている<sup>122</sup>。BEA Systems 社の WebLogic にも同様の機能を持つ WebLogic Personalization Server が用意されている<sup>123</sup>。

## (3)利用者が用いる端末デバイスのプロファイル定義

一方、前述の"(3) 利用者の端末デバイス関連情報"を定義するための規格としてW3C では Conposite Capabilities/Preference Profiles (CC/PP)の策定が進められている[6]。端末デバイスは各利用者がコンテンツや情報サービスにアクセスする手段であり、最近はデスクトップ PC やノート PC ばかりでなく PDA や携帯電話、セットトップボックスなど多様なデバイスを利用できる。端末デバイスの選択には利用者の嗜好や状況が反映されうることから、これも利用者プロファイルの一部と考えることができる。

CC/PP は RDF で記述するメタデータであり、将来的には P3P とも密接に関係してくるものとみられる。CC/PP Working Group のホームページ<sup>124</sup>にも記載されているように、HP Labs の DELI や W3C の Jigsaw など CC/PP に対応したサーバはいくつか発表されているが、定義ツールやクライアント端末側のサポートはまだこれからという状況である。

# (4)電子政府における利用者プロファイル定義

現在、日本国内全ての都道府県庁公式サイトで、情報公開もしくは各種申請書用の文書フォーマットの1つとして Adobe Systems 社が開発した PDF が採用されている $^{125}$ 。 PDF 文書のオーサリングツールである Acrobat では、PDF 文書にその作成者のプロファイルをRDF 形式で自動的に埋め込むようになっている(図 3-15、図 3-16参照のこと)。

http://www.beasys.co.jp/products/Weblogic/commerce/index.html

<sup>122</sup> IBM WebSphere Personalization: http://www-6.ibm.com/jp/software/Websphere/personalization/

 $<sup>^{\</sup>rm 123}\,$  BEA WebLogic Personalization Server :

<sup>124</sup> CC/PP Working Group: http://www.w3.org/Mobile/CCPP/

<sup>&</sup>lt;sup>125</sup> アドビシステムズ(株)、全都道府県庁の公式サイトが Adobe PDF を活用と発表

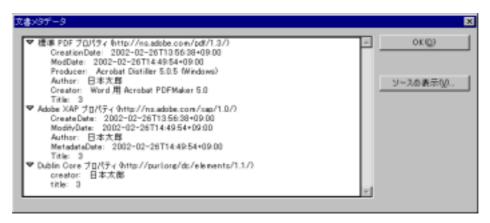


図 3-15 Acrobat で自動生成されたプロファイルの例(一覧表示)

図 3-16 上記プロファイルのソース(一部)

Acrobat では、このようなメタデータに加え、PDF 文書上に設定されたフォームに必要事項を入力し、電子署名を付加した電子申請書としてオンラインで提出することが可能となっている。(図 3-17参照のこと。)



図 3-17 電子署名(右上部分)を付加した PDF 文書の例

このほか、特に電子政府においては、利用者プロファイルに登録すべき情報が手書きやプリントされた紙の帳票の形で提出、管理されている場合も多いと考えられる。既に紙の文書として蓄積された情報は、人手によるキーボード入力とスキャン画像を組み合わせて電子化されるほか、OCR(光学的文字認識)を用いて自動的にテキスト情報を入力し、全文検索が可能な製品もある。

さらには、定型的な紙面を OCR で認識し、プロファイルとして適切な情報を抽出する技術も開発されている。実用化している典型例としては名刺管理システムがある。専用のスキャナにより OCR を通じて氏名、会社名、住所、電話番号などを自動的に認識し、データベースへ登録することができる<sup>126</sup>。今後は PDF や Web ページを用いた電子的帳票入力と OCR を用いた紙帳票の電子化を統合したプロファイル入力・管理システムが、各電子政府・電子自治体に導入されてくるものと予想される。

<sup>126</sup> 名刺管理ソフトの例:http://www.mediadrive.co.jp/products/package/win/econ/econ40/

# 3.2.5 セマンティック検索システム調査

セマンティック Web の電子政府への応用として、政府や自治体が収集・発行・管理する情報に対する、より人間に近い知的なセマンティック検索が期待されている。本節では、セマンティック検索を必要とする応用事例と、そこで使用あるいは要求される技術について述べ、セマンティック検索の可能性・将来性を示す。

#### 3.2.5.1 応用事例

## (1) ホームページ検索

政府には多くの省庁や関連機関が存在し、それぞれが独自のHPを立ち上げて情報を提供しているが、一般利用者にとっては、求める情報の管轄省庁がわからないとなかなか情報にたどり着けない。このため、電子政府全体を統合的に検索する仕組みが必要である。現状では、例えば、総務省のHP<sup>127</sup>では、「電子政府の総合窓口」として政府系の全HPに対する検索がかけられるようになっているが、まだキーワード検索の域を出ないため、必ずしも使い勝手の良いものではない。(図3-18参照のこと。)



図 3-18 総務省「電子政府の総合窓口」

-

<sup>127</sup> http://www.e-gov.go.jp/

また、地方自治体による情報提供においても同様の問題が存在する。欲しい情報はわかっていても、それがどこの場所(自治体)にあるか知らない場合には、情報にたどり着くのは容易ではない。このため、財団法人地方自治情報センター(総務省関連機関)では、NIPPON-Net<sup>128</sup>と呼ばれる、地方自治体のHP検索のためのポータルサイトを運営しているが、キーワード入力の仕組みに工夫は見られるものの、まだ使い勝手は十分ではない。(図 3-19参照のこと。)



図 3-19 地方自治情報センターの「NIPPON-Net」

## (2) デジタル・ミュージアム

日本には国や地方自治体が運営する美術館・博物館等が多く、電子政府の一環として、文化財の保護と国民への情報公開サービスの両方の観点からデジタル・ミュージアムが構築されている。代表的なものとして、デジタル・ミュージアム推進協議会(総務省関連組織)が運営する、全国のミュージアムが保有している美術品を横断的に検索することのできるデジタル・ミュージアム<sup>129</sup>がある。(図 3・20参照のこと。)

-

<sup>128</sup> http://www.nippon-net.ne.jp/

<sup>129</sup> http://www.digital-museum.gr.jp/

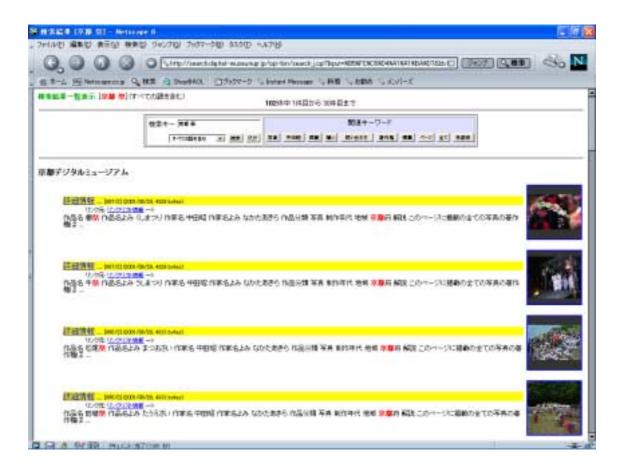


図 3-20 デジタル・ミュージアム推進協議会の「デジタル・ミュージアム」

同様の取組みはヨーロッパにおいても数多くのプロジェクトが進められている。例えば、英国では、大英博物館<sup>130</sup>の所蔵品のオンラインデータベース(図 3-21参照のこと) National Trust<sup>131</sup>の管理する歴史的資産のデータベース(図 3-22参照のこと) Wildscreen Trust による絶滅危機生物のデータベース構築を目的とした ARKive プロジェクト<sup>132</sup>(図 3-23参照のこと) などがある。

<sup>130</sup> http://www.thebritishmuseum.ac.uk/

<sup>131</sup> http://www.nationaltrust.org.uk/main/

<sup>132</sup> http://www.arkive.org.uk/default.asp



図 3-21 大英博物館の所蔵品のオンラインデータベース



図 3-22 National Trust の管理する歴史的資産のデータベース

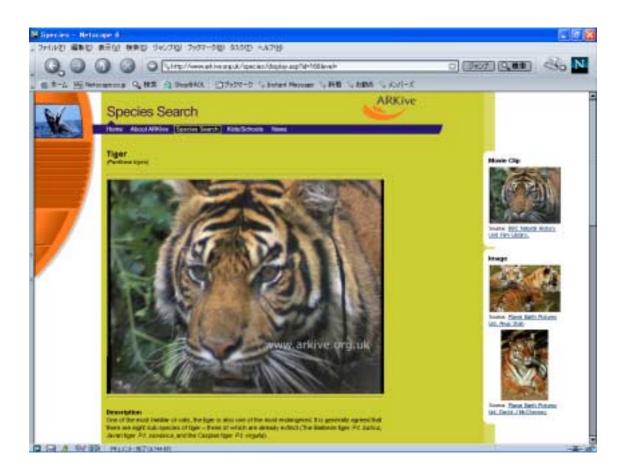


図 3-23 Wildscreen Trust による ARKive プロジェクト

# 3.2.5.2 必要な技術

## (1)知的検索

上記の応用事例のホームページ検索やデジタル・ミュージアムの多くは、インターネット上に分散した情報源を統合的に検索するものである。それぞれの情報源は独立して設計・構築されるものであるため、横断的に検索することは容易ではない。この問題を解決するため必要な要素技術としては次のようなものがある。

## オントロジー変換

分散型のデジタル・ミュージアムでは、それぞれの情報源ではシソーラスを使用しているかもしれないが、横断的に見た場合、情報アイテムの属性の名前ですら統一されている保障はない。例えば、美術品の製作された年は、"製作年""年代""時代"などの表現のゆれが見られる。このため、表現の同義性を判断する必要があり、あらかじめ一つの統合的なオントロジーにマッピングしておくか、オントロジー(表現)間の同義関係を表す変換規則を用意しておき検索時に変換するなどの手法が要求される。

自然言語表現からの情報抽出(専門用語抽出、テキストマイニング) 検索の対象となるホームページやデータベース・エントリは自然言語で表現されているため、検 索要求に適合しているかを判断するには、自然言語表現からの専門用語抽出、さらには、意味解析に基づくテキストマイニングによる情報抽出技術が必要となる。

#### 検索要求の自由記述入力

現状の検索要求入力はキーワード入力が基本で、AND/OR の組合せは可能なものの、ユーザーの 意図が十分に表現できないという問題がある。このため、検索要求の自然言語による自由記述入力 が必要である。検索システム側は、入力された検索要求を解析し、検索に必要な用語と意図を抽出 し、検索式に変換する機能が求められる。

#### 意味的マッチング

現状の検索アルゴリズムは、入力されたキーワードに対する全文検索が基本である。しかし、単純なキーワードマッチングでは、表現のゆれに対応できないため、オントロジーによってサポートされる表現間の意味的な関係に基づいて、検索対象を展開あるいは絞込みする機能が求められる。

## (2)知的問合せ(クエリー・サポート)

データベースに格納されている情報に対して、知的な問い合わせを行なう手法について、セマンティック Web では研究が進んでいる。近年、コンピュータネットワーク上でやり取りされるさまざまな情報を XML で取り扱う動きが盛んだが、セマンティック Web でも知的な問い合わせの対象となる情報を XML で記述し、相互運用性を向上させようとしている。しかし、XML は任意のデータ構造を表現できるが、構造が何を意味するかについては規定していない。そこでセマンティック Web では、XML をベースに情報の意味を記述するためのデータモデルを規定した RDF をメタデータの記述に活用しようとしている。RDF のデータモデルに従って、既存のデータベースのデータを変換したり、新たに登録する情報やデータに意味を付与したりしておけば、全文検索に頼らずに、よりユーザーの質問の意図に沿った回答を返すことができる。

例えば、官公庁のホームページを利用して「ある種の申請書の提出をどの窓口が今日受け付けているか」調べたいとする。官公庁のホームページ内を全文検索しても、たまたま自分が知りたいことがすべて載っているホームページが見つからない限り、検索結果を基にいくつものホームページを人間が読んで答えを探す必要がある。申請を受け付けない日があることを書いてあるページや、各窓口の取り扱い申請書類一覧の表などを見つけてやっと目的の知識を得ることがきるかもしれない。

ここで、各「窓口」のメタデータとして、RDFによる「取扱申請」、「受付日時」、「場所」に関する記述がされていると仮定する。また、「取扱」と「受け付け」は似た意味であることが記述されたメタデータ(オントロジー)があったとする。そのようなデータベースが整った条件下で、知的な問い合わせを行なうと、ユーザーが申請したいと考えているものが、まずどの窓口で取り扱われているかがすぐにわかる。そして、「今日」という言葉は時間の一種で、現在の日にちをシステムで調べて、その日にちが「受付日時」の範囲内に収まっているかを調べることができるかもしれない。そしてユーザーに対しては、ユーザーの住所などの利用者プロファイルを参照して、「あなたの現在いる場所から一番近い窓口はどこにあって、本日何時までなら申請書の提出が可能です」と表示されるかもしれない。(さらに、実際に窓口にいかなくて済む電子的な申請方法もあわせて勧めるかもしれない。)

上記のような世界を実現するシステムにおいて最もコアとなる要素技術の1つは、RDF専用のデータの取得・変更・挿入・削除などを行なうクエリー(Query)言語の開発とそれを利用した検索システムであると考えられる。

オランダの Aldministarator 社で開発された Sesame<sup>133</sup>と呼ばれる RDF スキーマベースのリポジトリとその検索機能を備えたツールがある (これは将来的には DAML+OIL ベースの検索もサポートする予定がある)。 EU 委員会の IST プログラムにおいて活動している On-To-Knowledge プロジェクトの主要な発表成果の 1 つである。 検索言語に OQL ベースのギリシャの ICS-FORTH で開発された RQL (RDF Schema Query Language)を用いている。

この他にも 3.2.1 で述べたように、RDFStore の RDQL/SQUISH、RDFdb の Algae、さらに HP の RDQL<sup>134</sup>など、SQL のような形式で RDF のステートメントを検索する言語が開発されている。 今後は、クエリー機能の向上と、このクエリーを操作するための扱い易いユーザーインターフェイスの開発が必要である。

現在世の中に提供されているセマンティック検索デモシステムの多くは、ある程度閉じたサイト内に蓄積されたRDFデータや、検索できるRDFデータのボキャブラリに対して制約を課している。今後は、オントロジーの相互変換や統合などを複数のサイト間などで動的に行なうフレームワークの開発が必要になると考えられる。

DAML プロジェクトの ITTALKS<sup>135</sup>は、IT 分野の講演の Web ページにメタデータを付与し、さらに個人のプロファイルやスケジュール、住所などの情報と組み合わせて、エージェントが推論を行ない、参加が推奨される講演をユーザーに提供するポータルサービスである。ITTALKS はいくつかのエージェントを持ち、その中の MapQuest エージェントと呼ばれるものは、別のサービスに対するオントロジーの翻訳機能を提供し、ユーザーの興味のある講演を探すエージェントと協調して動作する。ここでは、ACM のトピックと UMBC (University of Maryland Baltimore Country)のトピック間でオントロジー交換を実現している。このような技術が任意のオントロジーの相互変換フレームワークの開発の進展につながると考えられる。

さまざまなオントロジーが存在し、それらを相互に参照して動作するシステムの場合は、従来の AI システムでは許されなかった矛盾などについても処理できる仕組みが必要になる。現在の Web の検索結果に誤り(ユーザーの意図と結果が異なっているという意味で)があっても、検索エンジンがそれなりにユーザーの役に立っているように、まずは柔軟なルール運用による実用性の検証が求められている。また、オントロジーの信頼性に関してもデジタル署名や複数ユーザーのレイティングなどによってフィルタリングするアイデアがあるが、実用性の検証が求められている。

<sup>133</sup> http://sesame.aidministrator.nl/doc/sesame-dissemination.ppt

<sup>134</sup> http://www.hpl.hp.com/semWeb/rdgl.html

<sup>135</sup> http://www.semanticWeb.org/SWWS/program/full/paper41.pdf

# セマンティック Web 技術と次世代電子政府での 活用方法に関する調査研究

調査報告書

平成14年3月

発行 財団法人ニューメディア開発協会

〒108-0073 東京都港区三田1-4-28

TEL 03-3457-0672