

**平成13年度
セマンティックWeb技術の
調査研究報告書
～現状、課題と今後の方向性に関する調査～**

平成14年 3 月

財団法人 情報処理相互運用技術協会

はじめに

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とをマシンリーダブルなデータでその意味を記述し（このデータを、データを記述するデータという意味でメタデータと言う）、人間の代わりをするソフトウェア（これをエージェントと言う）で自動処理させることである。

セマンティック Web により、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になり、人間が簡単な指示を行うだけで、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が可能になる。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基にセマンティック Web の研究開発プロジェクトを推進しており、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。このまま放置すればセマンティック Web の国際標準規格が、米国と欧州とで我が国の事情を考慮することなく、総て決められてしまう恐れがあり、我が国も早急に取り組み、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、続々と生まれつつあり、このままだと、我が国の情報技術の決定的な遅れと情報技術産業の空洞化を招く危険がある。

このような背景を踏まえて、本調査では、Web 技術標準化団体である W3C（World Wide Web Consortium）におけるセマンティック Web の標準化動向、米国におけるセマンティック Web に関する研究開発動向、および欧州におけるセマンティック Web に関する研究開発動向を調査し、セマンティック Web の課題と今後の方向性について検討することを目的とする。

検討メンバー

[セマンティック Web タスクフォース]

委員長	清水 昇	日本電気(株)	NEC ソリューションズ
顧問	斎藤 信男	慶応義塾大学	常任理事
顧問	萩野 達也	慶応義塾大学	環境情報学部
委員	熊白 侃彦	沖電気工業(株)	システムソリューションカンパニー (SSC)
委員	後藤 龍男	日本電気(株)	政策調査部
委員	藤田 悟	日本電気(株)	NEC ソリューションズ
委員	来間 啓伸	(株)日立製作所	システム開発研究所
委員	豊内 順一	(株)日立製作所	システム開発研究所
委員	川村 直道	富士通(株)	ソフトウェア事業本部
委員	田中 省三	富士通(株)	ソフトウェア事業本部
委員	松井 くにお	(株)富士通研究所	コンピュータシステム研究所
委員	清野 正樹	松下電器産業(株)	マルチメディアシステム研究所
委員	今村 誠	三菱電機(株)	情報技術総合研究所
オブザーバ	小泉 雄介	(株)NEC 総研	調査研究グループ
オブザーバ	三田 昌弘	(株)NEC 総研	調査グループ

[セマンティック Web 委員会]

委員長	清水 昇	日本電気(株)	NEC ソリューションズ
顧問	斎藤 信男	慶応義塾大学	常任理事
顧問	萩野 達也	慶応義塾大学	環境情報学部
委員	森田 幸伯	沖電気工業(株)	研究開発本部
委員	上田 健次	九州日本電気ソフトウェア(株)	ソリューション基盤事業部
委員	布目 光生	(株)東芝	研究開発センター
委員	小倉 弘敬	日本アイ・ピー・エム(株)	ソフトウェア開発研究所
委員	細見 格	日本電気(株)	NEC ラボラトリーズ
委員	来間 啓伸	(株)日立製作所	システム開発研究所
委員	豊内 順一	(株)日立製作所	システム開発研究所
委員	津田 宏	(株)富士通研究所	コンピュータシステム研究所
委員	松井 くにお	(株)富士通研究所	コンピュータシステム研究所
委員	清野 正樹	松下電器産業(株)	マルチメディアシステム研究所
委員	今村 誠	三菱電機(株)	情報技術総合研究所
オブザーバ	小泉 雄介	(株)NEC 総研	調査チーム
オブザーバ	三田 昌弘	(株)NEC 総研	調査チーム
オブザーバ	佐藤 宏之	日本電信電話(株)	NTT 情報流通プラットフォーム研究所
事務局	神原 顕文	(財)情報処理相互運用技術協会	
事務局	小島 富彦	(財)情報処理相互運用技術協会	
事務局	石橋 博	(財)情報処理相互運用技術協会	
事務局	香取 良和	(財)情報処理相互運用技術協会	

目 次

第 1 章 セマンティック Web 技術の概要	1
1.1 セマンティック Web とは何か	1
1.2 セマンティック Web のレイヤー構造	2
1.3 従来技術との相違点	3
第 2 章 セマンティック Web を何故やるべきか	7
2.1 必要性	7
2.2 効用	8
2.3 将来性（市場性）	9
2.4 アプリケーション	11
2.4.1 セマンティック Web の応用分野	11
2.4.2 セマンティック Web のアプリケーションの分類	12
2.4.3 セマンティック Web のアプリケーション事例	13
第 3 章 セマンティック Web をやらないとどのような影響があるか	17
3.1 セマンティック Web が実現しなかったらどうなるか	17
3.2 日本でのセマンティック Web の普及が遅れたらどうなるか	17
3.3 欧米に遅れてスタートしたとしても早期にキャッチアップできるであろうか	17
第 4 章 W3C のセマンティック Web に関する標準化動向の調査	19
4.1 W3C の Semantic Web Activity	19
4.2 Semantic Web Activity 配下のグループ	24
4.3 Semantic Web Advanced Development	25
第 5 章 米国のセマンティック Web に関する研究開発動向	27
5.1 DAML (DARPA Agent Markup Language)	27
5.2 DCMI (Dublin Core Metadata Initiative)	33
5.3 RSS (RDF Site Summary)	34
第 6 章 欧州のセマンティック Web に関する研究開発動向	37
6.1 EU 委員会 IST プログラム	37
6.2 OIL (Ontology Inference Layer)	39
6.3 セマンティック Web の応用システムとツール	44
第 7 章 セマンティック Web の課題と今後の方向性	45
7.1 セマンティック Web の課題	45
7.2 セマンティック Web の今後の方向性	47
用語集	49
付録 1 良くある質問に対する回答	53

図 表 目 次

図 1.1	セマンティック Web のアーキテクチャ	2
図 1.2	メタデータの記述例	2
図 5.1	DAML(小分類)の実施計画及び予算(2001年)	29
図 6.1	On-To-Knowledge プロジェクトの全体構成	42
表 1.1	現状の Web とセマンティック Web との相違点	3
表 1.2	従来人工知能技術との相違点	4
表 1.3	Web サービスとセマンティック Web との相違点	5
表 6.1	IST プログラム予算	37
表 6.2	各パートナー組織の役割	42
表 6.3	ワークパッケージ	43

第 1 章 セマンティック Web 技術の概要

第1章セマンティック Web 技術の概要

1.1 セマンティック Web とは何か

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、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（国防高等研究計画局）を中心に米国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データ（これをメタデータという）を付加することによってその意味を記述し、人間の代わりをするソフトウェア（これをエージェントという）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、このセマンティック 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¹。」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。欧州委員会は、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

¹ 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。

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 ではフーリー大学（オランダ）、カールスルーエ大学（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。

1.2 セマンティック Web のレイヤー構造

セマンティック Web は、W3C の策定した XML（Extensive 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²において、セマンティック Web のアーキテクチャとして、図 1.1 のようなレイヤー構造を提示した。

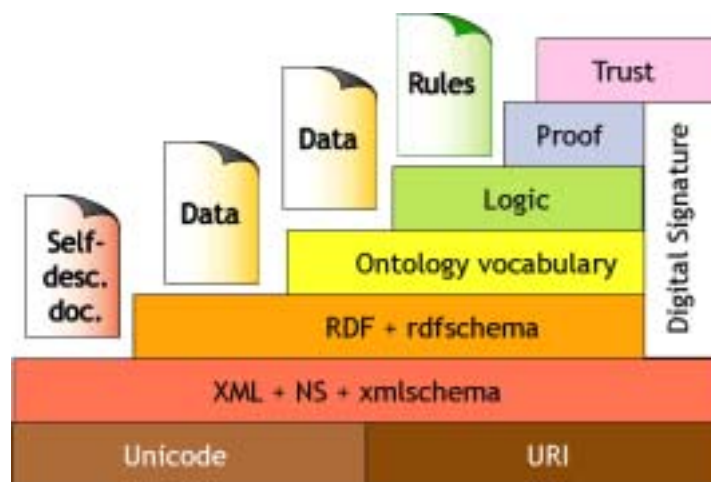


図 1.1 セマンティック Web のアーキテクチャ

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 1.2 に示すような主語・述語・目的語の 3 項関係で記述される。

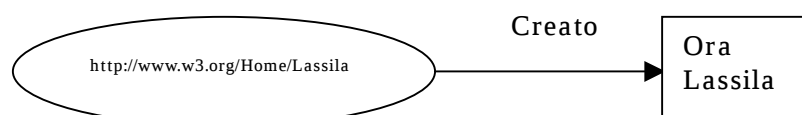


図 1.2 メタデータの記述例

² <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

- 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³。論理レイヤーは、推論を行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。

1.3 従来技術との相違点

セマンティック Web の技術特徴を明確にする目的で、現状の Web との差異、従来の人工知能技術との差異、および Web サービスとの差異について述べる。

（１）現状の Web との差異

セマンティック Web は、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ができることができるように、現状の Web を拡張したものである。セマンティック Web では、例えば、「cook」という単語で検索した場合、それが料理人を指すのか、調理法を指すのか、単に地名や人名、ビジネスあるいはその他の組織の名前を指すのかをコンピュータに理解させることができる。

表 1.1 に、目的と機械処理レベルについて、現状の Web とセマンティック Web との差異について述べる⁴。

表 1.1 現状の Web とセマンティック Web との相違点

	現状の Web	セマンティック Web
目的	人間が読むための Web コンテンツを表現。 ex. ページが A 医師のホームページであるとか、このリンクは A 博士の履歴書に飛ぶといったようなことを、理解しているわけではない。	コンピュータが解釈するための Web コンテンツを表現。Web ページに、コンテンツの意味を表す構造を付与する。 ex. 「A 先生が、その病院、月、水、金曜日に働いていること、日付の範囲を yyyy-mm-dd という形式でスクリプトに指定すると予約の時間を返答してくれる」ことがわかる。
機械処理レベル	レイアウトを解析してコンテンツを画面に表示する。	コンテンツの意味を理解し推論することにより、自動で Web 上のサービスを提供する。

³ 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）。

⁴ Tim Berners-Lee, James Hendler, Ora Lassila. "The Semantic Web." *Scientific American*, May 2001 を参考にした。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。

(2) 従来人工知能技術との差異

80 年代の人工知能(AI: Artificial Intelligence)の知識表現システムに似た高度で複雑なモデルであるとして、セマンティック Web の実現性を疑問視する意見をもつものも少なくない。しかし、セマンティック Web の推進者らは、過去の AI 技術をベースとするプロダクトが商用として成功しなかった理由の反省にたち、また、最近の Web サービス技術や EC 標準の動向とも整合性を取りながら研究開発を進めようとしているので、80 年代の AI 技術との類似性をもってセマンティック Web の実現可能性を論じるのは必ずしも適当でないといえる。

表 1.2 に、80 年代の AI とセマンティック Web の知識表現システムの差異について述べる。

表 1.2 従来人工知能技術との相違点

	80 年代 AI の知識表現システム	セマンティック Web の知識表現システム
対象世界/環境 集中か分散か	クローズドで仮想的な(頭の中の)世界 集中方式。 例えば、「親」や「乗り物」といった概念を正確に定義し、共有しておかねばならないので、システムの規模が大きくなると、管理が困難になる。	オープンな Web 環境での現実世界 分散方式。 インターネットと同様に、広域に分散化したシステム。インターネットが分散化のために、すべてのリンクの整合性を保つことはあきらめたように(リンク先がない場合に、「ページが見つからない」というメッセージを返すことを指す)、ある程度の矛盾を容認する。
オントロジの 多様性について の考え方	一つの普遍的なオントロジを想定する。	その時々や、用途、業界に応じて多様なオントロジが存在する。それらのオントロジ間の整合性は保証できないが、共通部分についてはオントロジ間の変換などにより、複数のオントロジが並存するモデルを実現しようとする。
知識の作成と 規模	少数の人が開発、ルール数も少ない	多数の人が作成。ルール数も多い。他のサイトのオントロジを参照する。
矛盾に対する 考え方	コンピュータが信頼性の高い答えをだせるように、質問を慎重に制限する。また、限定的な固有のルール体系を推論に利用する。	矛盾を受け入れることにし、Web ができるべく多様な推論をして答えをだせるように、ルールを記述する言語もできる限り自由な表現ができるようにする。また、矛盾がある場合は、「人間と対話的に問題を解決する」、あるいは「より信頼性の高いルールを採用する」等により、ルールの運用に柔軟性をもたせることで、矛盾があっても実用上困らないような頑強なシステム構築をめざす。
推論やプラン ニングの実行	単一計算機で実行。	分散環境で実行。

(3) Web サービスとの差異

最近、IBM や Microsoft らが提案している Web サービスが、インターネット上でソフトウェア機能を動的に探し出し、アプリケーションとの連携を可能にするためのしくみとして注目されている。複数の Web サービスを統合化したサービスを実現しやすくする点では、セマンティック Web と Web サービスは、その目標が似ているといえる。しかし、以下に示すように、目的やアプローチが異なるので、将来的には、セマンティック Web は Web サービス技術に対して形式的なデータモデルを提供するという関係で、両者が互いに他を補い合いながら、次世代 Web を実現していくことが期待されている。

表 1.3 に、Web サービスとセマンティック Web の差異について述べる⁵。

表 1.3 Web サービスとセマンティック Web との相違点

	Web サービス	セマンティック Web
目的	Transactional Web (情報交換の方法を改善する)	Intelligent Web (情報間のリンクをもっと知的にする)
アプローチ	今すぐ使える技術を商用に開発するというアプローチ(パズルピースからビックピクチャーへ)	目標を達成するためにはどのような技術が必要かを検討し、その技術を開発していくというアプローチ(ビックピクチャーからパズルピースへ)
両者の関係 (共存の可能性)	セマンティック Web のための技術インフラを提供する。 Web サービスの入出力インターフェイス仕様を記述するための言語である WSDL(Web Service Description Language)では、形式的なデータモデルを規定しているわけではないので、セマンティック Web のオントロジー記述言語(RDF、RDF Schema、DAML + OIL、DAML-S)とはコンパチブルであるといえる。	Web サービスの形式的なデータモデルを提供する

⁵ James Snell, Web Services -Components of the “Transactional Web”-, The First A Semantic Web Working Symposium 2001(SWWS '2001) (2001)を参考にした。

第2章 セマンティック Web を何故やるべきか

第2章セマンティック Web を何故やるべきか

2.1 必要性

今、インターネットの Web において、利用者は必要な情報を Web で検索することが年々難しくなっているばかりではなく、各種エンジンを駆使した検索においても得られた検索結果には、欲しい情報とは関係のない情報が多く含まれていることに気付いている。ガーナー社は、Web 上の情報量は 6 ヶ月で倍になっているので検索結果は悪化する一途と発表している。コンピュータに関して充分知識のある人はインターネットの情報を活用できるが、それ以外の人にはインターネットの情報を活用できなくなり、デジタルデバイドが拡大するのではないかと懸念が持たれている。情報処理の相互運用性を研究・開発する団体である INTAP に Web の増加に対応した効率的な情報流通を達成する技術の実現が期待されている。

このため、情報流通を達成する次世代 Web 技術としてセマンティック Web が注目されている。セマンティック Web のルーツは、W3C (World Wide Web Consortium)が開発したインターネット上に氾濫している有害な情報を不用意に子供たちがアクセスしないように Web に制御情報を加える仕組み、即ち、PICS(Platform for Internet Content Selection)である。1999 年には PICS におけるメタデータ（データに関するデータ）の考え方を発展させ、Web 上のメタデータを関係づけ、且つ、機械的に処理できるように構造化した汎用性の高い記述形式の Resource Description Framework (以下 RDF と称す) を規格化した。1999 年に W3C の最高責任者である Tim Berners-Lee 氏が Web にセマンティクス(意味付け)を導入することを提唱し、2001 年 2 月に W3C にセマンティック Web アクティビティが発足した。セマンティック Web は、Web にメタデータを付加する事によって現在のインターネット世界を一変させる革命的アイデア、つまり“次世代 Web”となる可能性を大いに持っている。既に、米国及び EU は調査、開発に国家予算を当てている。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。EC(欧州委員会)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）の拠出をしている。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、メリーランド大が、EU ではフーリー大（オランダ）、カールスルーエ大（ドイツ）、ブリストル大（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。このようにセマンティック Web の潜在力に期待した実用化競争が既に始まっている。

セマンティック Web は、次のような発展をすると考えている。

第 1 段階では、メタデータにより精度の高い検索を可能とし、第 2 段階では、メタデータの内容に一次情報も付加し、さらに、検索精度を高めることが期待されている。第 2 段

階までで、セマンティック Web は、(1) Web を巨大な知識データベースとして活用すること、(2) Web に散在するバラバラな情報を組み合わせると共に統合して、新たな情報の導出を可能にするものと考えられている。

第 3 段階では、人が高度な知的仕事に専念できる環境が実現されると期待されている。セマンティック Web は、世界中に分散配置されている Web の情報をそのセマンティクスを使ってエージェント・ソフトウェア（エージェントともいう）が条件にあった情報を自動収集、整理、分析してくれるものと考えられている。

Web のリンクを辿っていくと様々な情報にたどり着き、興味がある情報が見つかる喜びが湧いてくるのがインターネットである。教育現場、企業においてはインターネットに分散配置された情報を共有する効果と Web を通じた情報の流通性が認識されてきた。世界各国は、医療、地域、教育、電子図書館、電子商取引、電子政府などにおいてコンテンツの情報流通性を高めるためのプロジェクトを国家レベルで推進している。これらの情報を産業、教育レベル、地域などの区別なく、全ての人々が横通しで使える情報サービスのプラットフォームを構築する必要がある。分散配置された情報を共有し、流通させる次世代 Web 技術がセマンティック Web である。セマンティック Web では、オントロジー（ontology）で様々な分野の情報資源である Web ドキュメントが使う用語の収集や分野毎の用語の相互変換を行うため、情報の流通性が高い。セマンティック Web は、新しいサービスの創出と安全で豊かな国民生活をもたらす技術と考える。こうした広範囲なテーマを持ったセマンティック Web は、産学官の連携によって研究開発、実用化、普及推進が図られなければならない。INTAP には、セマンティック Web 推進における知識共有システム、コンテンツ利用、コミュニケーションなどのビジネスモデルの確立とさまざまな分野で応用されていくためのコスト対効果比が十分に釣り合う基盤技術開発を目指した我が国の情報拠点となる活動が期待されている。

2.2 効用

セマンティック Web では、XML 文書、図、写真等あらゆるものに対して、その構成要素にメタデータの付加記述を行う。又、XML 文書、図、写真等の関連、メタデータ間の関係や推論、オントロジーの定義等も可能である。セマンティック Web の効用はこれらの種々の付加情報を有効に利用する事で得られる。

例えば、「木曜日に診療している小児科で渋谷にある病院」を探すのに現在では、「木曜日、小児科、渋谷」のような検索条件を設定するだろう。結果として、「木曜日の新聞」や「渋谷一人歩きの手引き」等、目的とする物以外の結果が得られる事になる。条件がアンド条件でも、「木曜日が休診の渋谷先生のいる小児科」等が見つかる事になる。これをセマンティック Web で、「病院の診療科目は、内科、外科、歯科、小児科、眼科である」、「病院の診療日は月、火、水、木、金である」、「病院は存在場所がある」、「存在場所は地名の集

合である」のように意味を付加して置けば、病院に関して、「診療科目=小児科」、「診療日=木」、「存在場所=渋谷」のような検索で目的の物を探す事ができる（=の左辺はシステムが自動生成出来る）。更に、病院の診療科目、病院の診療日は、夫々一覧が出、それらから望みの物をクリックするだけで良いようにする事も可能である。又、自然言語処理と組み合わせ、最初の文をエージェントに入力し、エージェントが自動的に検索を行うようにする事も可能である（例題は、ガートナーコラム栗原氏の資料の例を流用させていただいた）。

このように、メタデータを適切に付加する事により、Web の検索を飛躍的に向上させる事ができる。

また、メタデータ間の関連を辿ることにより、膨大な Web リソースを恰も一つのデータベースのように容易に取り扱う事も可能である。

オントロジーでは、異なる言葉を同じ意味を持つ言葉として解釈させる事ができる。例えば、「マップ」という言葉で検索を行う時に「地図」も同じ意味と解釈し、「地図」を含む文書も「マップ」を含む文書と同じく検索する事ができる。

オントロジーとセマンティクスを組み合わせれば、同じ言葉でも、違う意味を持つ物と解釈させる事も可能である。文脈で異なった意味に使われる事を明確にする事が出来ると言う事である。

オントロジーとして、唯一の(標準の)オントロジーを持つのではなく、分散したオントロジーの集合のデータベースを使用する事も出来る。どのオントロジーが使い易く、十分役に立つかと言う事は競争原理でデファクトが決まっていく事になる。

また、将来は推論を記述する事で AI の推論と同じように未来の予測等も可能である。但し、AI と異なる所は、推論に矛盾があっても良く、矛盾がある場合には、推論がそれ以上進まず、人に助けを求める事になるのではないかと言う事である。即ち、エージェントは人に代わりセマンティック Web の処理を行うが矛盾点があれば人に援助を求めると言う、人との対話の余地を残しており従来の AI の完全性を求める方式とは異なっており、柔軟さがある。

セマンティック Web が一般的に使用されるようになると、ユーザは発信情報をセマンティック Web で作成する事により、十分に情報の利用を行って貰えるようになるし、セマンティック Web のメタデータを使用して、発信情報のデータベースを他の目的に構成変更して使用するような、データの加工を行う中間業者のような新しい産業の育成を促す可能性がある。

2.3 将来性（市場性）

インターネットのキラーアプリは Web である。これは今後も続くであろう。セマンティック Web の市場性は、かつてアメリカ国防省が提唱した CALS が企業間のノードをネットワークによってシームレスに繋ぎ、抜本的な経営革新、生産革新、企業競争力獲得につなげることを目指したように、企業統合、バーチャルカンパニーにあると考える。

第 1 世代 Web の限界は、人間対人間インターフェースにとどまっている点である。Web

とブラウザとの組み合わせがインターネットの普及を爆発的に拡大させたが、その効用はあくまで特定、不特定個人同士の情報共有であり、応用としても広報、宣伝、広告、閲覧、掲示板、書類送付、文書交換など、人為的活動の補助的役割にとどまっている。いわば電話、F a x など旧来型通信手段のエンハンスメントとも言える。インターネットの効用が最も期待される企業活動への応用においても、大方は広告、宣伝、伝票のやり取りなどの従来的に行われてきた周辺プロセスの延長線、もしくは補助的手段にとどまっている。最近では e ビジネスや e-マーケットプレイスなど、企業活動の中核である生産、販売プロセスにも応用が展開されつつあるが、第 1 世代 Web の限界から本質的な経営革新には繋がっていない。

一方、セマンティック Web の革新性は、マシンリーダブルな特性とインテリジェンスを生かして分散化されたコンピュータノードを相互に接続し、自律型のコンピュータネットワークに発展させ得る可能性が挙げられる。そのような自律型コンピュータネットワークが企業経営を革命的に進化させ、新たな競争力の源泉につながられる可能性がある。企業活動にとどまらず、電子政府、福祉、遠隔医療、教育などインターネットノードを介する多くの社会活動においても同様のスケールの革新を生み出せるだろう。単純なデモやアプリに拠るのではなく、セマンティック Web を採用したシステムが現実的に目覚ましい社会的効果をあげることを実証して見せるアプローチが最有力だろう。それに最も適した応用分野として、例えば C A L S が目指したような本格的企業間 e ビジネス、もしくは広域の電子政府などが挙げられる。

C A L S は、現在の企業では企業間業務を対象にした企業改革である S C M (Supply Chain Management) として議論されている。S C M の課題は、顧客満足度を最大にするために、高効率の調達、生産及び間接コスト、流通コストを削減することである。S C M では、業務を標準化し、これらの構造をモジュール化することによって最適な調達が行われると考えられている。IT(情報技術)で言えば、情報を共有し、価値ある情報に加工をすることである。コンピュータによる情報伝達、情報表現、業務フローと管理(トラブル管理も含む)、取引に関わる契約、物流、商流、金流のトータルな管理ができるかである。(J E C A L S 発行の「S C M の運営に関わる企業モデル約款(α版要約版)」から流用させていただいた。) S C M における情報にセマンティクス(意味付け)が行われれば企業は高度な業務に注力できるであろう。セマンティック Web のオープンな仕組みが S C M を導入するときの費用負担の悩みを解決すると考える。セマンティック Web をベースにした共通プラットフォームを組立て、その実用性、革新性を実証し、普及の原動力にするべきである。こまごました派生的な応用やデモンストレーションを繰り返すのではなく、より本格的な実証を目指すべきである。その意味でセマンティック Web の市場性、将来性に関しては、アカデミックな R & D だけに任せず、IT 産業界側の主体的コミットメントが求められる。I N T A P がセマンティック Web により共通基盤性の高い技術開発に早く取り組み、実用化に移行できる実証実験がセマンティック Web の市場性を高め、後に続く企業、業界に拡散すると考える。

2.4 アプリケーション

2.4.1 セマンティック Web の応用分野

これまでの Web がそうであったように、セマンティック Web の応用分野はあらゆる業種に及ぶだろう。これまでの Web では人間が読むためのデータだけを取り扱っていたために、主に情報提供の目的にしか使われなかったが、セマンティック Web は機械的に処理可能なデータをインターネット上に広めるため、トランザクション処理などの様々なデータの処理が可能になる。しかも、複数のデータ処理を組み合わせることが可能になる。また、Web サービスと異なり、異業種間でのデータの定義の仕方が異なっても、それらの間の対応を適切に与えることにより、組み合わせることが可能である。インターネット上であらゆるデータ処理を自由に行うことができる空間ができていく。

応用分野として、まず情報検索の分野があるが、これまでは、Web ページの内容を適当に処理することによりキーワードなどを抽出していたが、セマンティック Web では、Web ページにつけられたメタデータを用いることにより、それぞれの Web ページが何について書かれたものであるかを機械的に処理できるため、これまでにない正確でしかも絞り込まれた検索結果を得ることができる。また、検索結果に関して、どうして検索に引っ掛かったのかという理由が明確であるため、検索結果をさらに利用したり、そこから別の推論を行ったりすることも可能である。また、一般の検索サイトではなく、企業の製品紹介のページであったり、新聞の記事のページであったりする場合、Web ページだけでなくメタデータを付加しておくことで、利用者側から特定の情報を得ることが可能になる。

現在、利用者毎にカスタマイズできる機能を提供しているサイトもあるが、これを利用者側で行うことが可能になる。また、関連する情報を複数のサイトにまたがって検索し、利用者が一つのページを構成することも可能になる。また、ある製品について価格などを比較したページを作りサービスを行っているサイトも多数あるが、これをこれまでのように手作業で作るのではなく、自動的に作ることも可能になる。これまで、情報はデータの提供者が好き勝手な形で書いていたが、セマンティック Web のメタデータとして統一することによって、直接提供者以外がデータを加工して新しい情報を提供することが可能になる。

応用分野として次に e コマースの分野を考えてみよう。この場合も、これまでは取り引きを行うものの情報を Web ページで記述し、実際の取り引きの処理を e コマースの処理系にしたがって個々に記述する必要があったが、セマンティック Web では取り引き情報をメタデータとして表現することによって、汎用的にだれでもが処理可能になる。したがって、特定の e コマースの処理系に頼る必要がない。あるサイトで利用している e コマース情報を他のサイトでも利用可能である。また、支払い処理などもサイト間で相互に利用可能になる。電子貨幣をセマンティック Web のメタデータとして利用することができれば、あらゆる場面で利用することが可能になる。

また、セマンティック Web では異業種間での取り引きを可能にする。メタデータで使う言葉の定義などは各業種で行って良く、それらの関係を定義することによって異業種の間でも情報の解釈や交換が可能になる。世界的に考えても、日本では日本語を使ってメタデータを与え、それを海外からは英語と対応させることによって利用することが可能になる。

電子政府などでの応用はどうであろうか。電子政府において様々な文書を Web 上に人が読める形で置くだけでなく、セマンティック Web のメタデータを付加することにより、だれでもが必要な情報を検索し、それを利用することが可能になる。文書ではなく処理しやすいデータを手にいれることができるので、様々な加工が可能になる。もちろん、データのセキュリティや信頼性の問題などもあり、電子署名などの技術がセマンティック Web においても使われる。

教育の分野においても、教材や講義ノートを Web 上に公開する場合、メタデータを付加することによって、教材を組み合わせたリ、問題を自動的に生成したり、様々な利用可能になる。これまで、教材を再利用するためには、その中身をそれぞれの人が確認して使う必要があったが、メタデータが付加されていれば、機械的に選別することが可能になる。学習者にあった最適な教材を検索することも可能になる。これまでのように、提供者側が教育の中身を決めてしまうのではなく、それぞれの学習者にあった教育内容を自動的に合成することが可能になる。

セマンティック Web は情報の利用者へのメリットだけでなく、情報提供者へもメリットを与える。メタデータを付加することによって、1 つのデータを様々な用途で利用可能になり、情報提供者への負荷が軽減される。また、情報の中身を的確に表現できるために、情報を提供したい相手にすばやく伝達することが可能になる。

2.4.2 セマンティック Web のアプリケーションの分類

ビジネス適用という視点からセマンティック Web の応用アプリケーションを、情報仲介、カタログ生成、および Web ページ分類の 3 種に大きく分類した。それぞれのアプリケーションの特徴と簡単な事例を述べる。

(1) 情報仲介

e コマース (オークション)、e マーケットプレイス、人材リクルーティング、住宅・マンション仲介、中古車売買などの仲介・斡旋のための情報交換サービス。

多数の情報提供者と多数の利用者の間に介在し、適切な情報を適切な人に提供。

情報提供者は、フォーマットや用語にとらわれることなく、自由に情報を記述できる。また、特定の仲介市場に固定されることなく、自由に複数の仲介市場に参加できる。

利用者は、自分の要求に合った情報だけを選択的に受け取ることができる。

米国での事例としては、人材リクルーティングが有望。各社の HP の人材募集情報をエージェントが収集し、就職・転職希望者の条件に合った情報を提供。

(2) カタログ生成

多数の情報提供者によって記述される情報を整理し、カタログ化するサービス。

上記の「情報仲介」の前半部分（上流の処理）にも相当。

e コマースであれば、複数の企業の Web ページに記載されている商品情報から、同じ種類の商品を抽出し、比較表を提供するサービスが有効。例えば、トヨタ、日産、ホンダ等の Web ページから、1800cc クラスの自動車の比較表を生成するような場合が典型的。

このように作成されたカタログ情報は、利用者の嗜好に応じてカスタマイズされた情報の配信サービスにも適用可能。

医療分野であれば、複数の病院の Web ページから特定の科だけの比較表を作成することで、患者の選択肢の幅を広げることができる。また、複数の医者によって個別に記述された、患者の症例や退院サマリーの整理も有効であり、この事例では類似の医療ケースの発見への貢献が期待できる。

(3) Web ページ分類

同じカテゴリに属する Web ページに対し、分類の軸を与えて、自動的に分類（ディレクトリ情報の生成）を行うサービス。

それぞれの Web ページは独自の編集方針で作成されているが、それらの中から共通性のあるページだけを抽出する点が特長。

電子自治体の取り組みであれば、自治体の Web ページを対象に、文化活動情報、地域おこし情報、特産品情報などのディレクトリを抽出するような事例が典型的。

なお、アプリケーションシステムの開発では、対象とする用途における専門知識の取り扱いが重要であるため、その分野の専門家や団体との協力関係の締結が不可欠である。

2.4.3 セマンティック Web のアプリケーション事例

(1) e マーケットプレイス

現在の e マーケットプレイスは、商品の提供者（サプライヤ）および購入者（バイヤ）の間に商品の種別ごとの仮想的な市場を構築し、サプライヤとバイヤの間の商品情報や注文情報の交換、商品情報の効果的な検索や宣伝などを支援するものである。米国では CommerceOne など既に幾つかの e マーケットプレイスが運営されており多くのユーザを集めているが、以下のような幾つかの制約があり、必ずしも誰もが参入が容易なシステムとはいえない。

- ・サプライヤ、バイヤともに、e マーケットプレースの運営者と事前に契約を結ばねばならない。また多くの場合は、月々固定の参加費が必要となる（プラス売上げに応じた手数料も）
- ・e マーケットプレースの運営者からトップダウンに規定される項目が多い：商品情報の記述のフォーマット、検索手段など
- ・基本的には契約した e マーケットプレース内でしか情報の流通がなされない（注：一部 e マーケットプレース間の連携も実現されている）

このような制約下では、中小のサプライヤにとって、どの程度の売上げが期待できるかわからない e マーケットプレースに参加費を払って契約し、またそれなりの手間をかけて商品の登録を行うということは、非常に高いハードルである。また、バイヤに取っても、どの e マーケットプレースに参入すれば、最も安価で良質な商品を調達できるかは、ある種の賭けと言える。

しかし、もしセマンティック Web が普及すれば、より柔軟な e マーケットプレースが多数運営され始めるであろうことは想像に難くない。このような e マーケットプレースは、セマンティック Web に対応したロボットを用いて、ネットワーク上でサプライヤの提供する商品情報を収集し、メタデータを用いて商品情報を分類して独自の商品カタログを作成し、この商品カタログをバイヤに無料あるいは非常に低価格に公開するような運用形態を取るであろう。サプライヤは、現在の人間向けの Web ページに一手間加えるだけで、複数の e マーケットプレースで自動的に自分の商品を取り上げて貰え、バイヤは自分のニーズに合う、使い勝手のよい e マーケットプレースで商品の検索が可能となる。このような仕組みで運用される e マーケットプレースは、売買が成立した場合の手数料を徴収する点は変わらなくとも、現在のシステムより柔軟で、バイヤやサプライヤにとっても参入しやすく、また参入のためのコストは大幅に安価になるであろう。

（２）情報配信サービス

現在でもメーリングリストに代表される情報配信サービスは多々運営されており、新製品情報、経済情報から趣味の情報まで、ありとあらゆる種別の情報が配信されている。しかし、このような情報サービスの運営者にとって最大の課題は、いかにユーザの満足度が高い新鮮な情報を収集するかである。無論コストをかければ価値のある情報の収集は可能であるが、今のところ、こういった情報配信サービスに高額の料金を支払うコンセンサスは醸成されていない。現にほとんどのサービスは広告情報を入れることによって無料で運営されており、結果として BtoC ビジネスにおいても大きな市場ではない。従って安価なコストでの魅力ある、かつ他の配信サービスと差別化できる情報の収集が大きな課題となっている。

もしセマンティック Web が普及すれば、ネットワーク上に公開された Web ページから、エージェントを用いて、高いコストをかけることなく的確かつ新鮮な情報を収集することが可能になる。エージェントの収集ポリシーをチューニングすれば、メタデータをチェッ

クしつつ情報収集を行うことで、非常に対象を絞りカスタマイズされた情報（関連分野、地域など）を自動的に集めることも比較的容易となる。情報収集やカスタマイズのためのコストが安価になれば、当然配信サービス自体も安価に行える。

実例として、マンションの住人やオフィスビルの利用者を対象とした情報配信サービスを想定しよう。このような場合には、ローカル色を出した地域性の高い情報（近くにある評判の高いレストラン情報や狭い地域の天気予報、交通情報など）が好まれるであろう。また、特定のスポーツのジャンル、あるいは特定のチーム、より細かく特定の選手の情報を選択して配信することも、セマンティック Web のメタデータがその情報の細かさ（粒度）に対応してさえいれば容易である。

セマンティック Web の普及に伴い、個人で自分のエージェントをカスタマイズして情報収集を行うツワモノ（情報強者）も増加するであろう。しかし、セマンティック Web が、付加価値の高い「有料に値する」情報配信サービスビジネスの後押しをすることは疑いない。

第3章 セマンティック Web をやらないと どのような影響があるか

第3章セマンティック Web をやらないとどのような影響があるか

セマンティック Web の開発や普及がなされないとするるとどのような問題が起り得るのか、いくつかの視点で考えてみる。

3.1 セマンティック Web が実現しなかったらどうなるか

現在の Web のままだとすると 6 ヶ月で倍増しているといわれている Web 情報が有効に使われることなく情報提供者の努力が無駄になってしまう。また、情報を検索する側から言えば自分が本当に必要とする情報を見つけ出すことがますます困難になる。

3.2 日本でのセマンティック Web の普及が遅れたらどうなるか

欧米を中心とする先進国にセマンティック Web が先に普及して日本での普及が遅れたらどのようなことになるだろうか。

欧米ではセマンティック Web がもたらす利便を多くの個人・企業が享受している時に、日本ではいち早くエージェント技術を導入した英語に堪能な一部の個人・企業だけがセマンティック Web の恩恵を受ける。このような人にとっては欧米の情報を日本に紹介するビジネスチャンスが生れるが、大部分の日本人は増え行く情報の渦の中でもがき続けることになる。また、英語圏とのビジネスを狙って英語のホームページを充実したとしてもメタデータを持たない Web サイトが英語圏からアクセスされる可能性は高くはならないだろう。

3.3 欧米に遅れてスタートしたとしても早期にキャッチアップできるであろうか

現在の Web は人間が読むことを前提にしているので言語の違いはコード体系とフォントの違いだけであるのに対し、セマンティック Web は意味を扱うので状況は大きく違ってくる。従来のように欧米からソフトを導入して画面などを日本語に翻訳するだけでは使用できない。オントロジーも日本独自のものを作成しなければならなくなる。メタデータを作成するためのツール類も日本化(単なる日本語化ではない)しなければならない。また、セマンティック Web を普及させるためにはそれをサポートするための技術者を育成しなければならない。

したがって、短期間でキャッチアップすることは難しい。日本がキャッチアップを図っている間も欧米の技術は進んで行く。セマンティック Web は Web 上への知識の効率的な蓄積と機械と人間の協調動作を実現することを目指している。この技術開発・普及に遅れをとることはナレッジがもっとも重要になると言われている 21 世紀に日本がグローバル競争に打勝つための産業構造改革(知価革命)にも重大な悪影響を及ぼす懸念がある。

第 4 章 W3C のセマンティックに関する標準化動向の調査

第 4 章 W3C のセマンティック Web に関する標準化動向の調査

4.1 W3C の Semantic Web Activity

(1) セマンティック Web の定義

- Semantic Web Activity ホームページ⁶の冒頭には、次の言葉がある。

「セマンティック Web は現在の Web を拡張したものである。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⁷」
- セマンティック Web の目標 (Goal)

「WWW は、ドキュメントをリンクで繋いだハイパーテキスト構造によって、社会に大きなインパクトをもたらした。しかし、現在のリンク構造は暗黙の意味関係に依存するため、ドキュメントの内容を理解できる人間には豊かな世界を提供するが、そうではない計算機には平板で単調な世界に過ぎない。セマンティック Web は Web のシンプルな拡張であり、Web 上の情報に well-defined な意味を与えてコンピュータにも理解できるようにするものである。

 - ・ W3C のセマンティック Web の目標は以下のものである。
 - ・ Web が持つ潜在的な可能性を現実化すること
 - ・ Web 上への知識の効率的な蓄積を実現すること
 - ・ 情報の機械処理を可能にしてコンピュータと人間の協働を実現すること

そして究極の目標は、有効で効率の良い知識共有を世界的な広さで可能にすることにある。⁸」

(2) W3C における Semantic Web Activity の位置付け

- Semantic Web Activity ホームページの Introduction には、次の言葉がある。

「Web がその完全な可能性を開花するためには、自動化されたツールによって Web 上のデータを共有し処理することが必要である。Web が先へ進むためには、今後のプログラムは、それらが完全に独立に設計されたプログラム間でも、データを共有し、処理することが可能でなければならない。セマンティック Web は、一つのヴィジョンである。すなわち、Web 上で明確にデータを定義し、それをリンク付けすることによって、機械がそれらのデータを、通常の表示のみならず、アプリケーション間でのデータの自動化、統合、及び再利用のために利用できるようにしようとする考え方である」
- W3C トップページ⁹には、次の言葉がある。

「W3C は、相互接続可能な (interoperable) 技術 (仕様、ガイドライン、ソフトウェア、ツール) を開発することによって、Web のその完全な可能性を開花させようとするものである」
- W3C の組織概要ページ¹⁰には、W3C の目標 (Goal) として以下の 3 つが挙げられている。

⁶ <http://www.w3.org/2001/sw/>

⁷ Tim Berners-Lee, James Hendler, 及び Ora Lassila の言葉である。出典は、Tim Berners-Lee, James Hendler, Ora Lassila. "The Semantic Web." *Scientific American*, May 2001.

⁸ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

⁹ <http://www.w3.org/>

¹⁰ <http://www.w3.org/Consortium>

- ・ユニバーサル・アクセス
- ・セマンティック Web¹¹
- ・信頼の Web

これらから、Semantic Web Activity が W3C の掲げる目標にとって、いかに重要な取り組みであるかが分かる。

上述の Introduction では「セマンティック Web は、一つのヴィジョンである」と言われているが、ヴィジョンとは W3C が従事する 3 つの役割 (Role) のうちの一つである。W3C の 3 つの役割とは以下のようなものである¹²。

- ・ヴィジョン：

W3C は WWW の将来のヴィジョンの構想と推進を行う。W3C の会員組織や W3C チーム、全世界の Web コミュニティの研究者・開発者からの寄与によって、W3C は Web が真にユニバーサルな情報スペースになるための技術的な必要条件を見きわめることができる。

- ・設計：

W3C は上記のヴィジョンを実現するために、既存の技術と将来の技術を考慮に入れながら、Web 技術の設計を行う。

- ・標準化：

W3C は仕様 ("Recommendation") 作成によって、Web 技術の標準化活動を行う。

(3) Semantic Web Activity の活動概要

- 2001 年 2 月 9 日に Semantic Web Activity グループが設立。従来の Metadata Activity グループ¹³からの名称変更である。2003 年 2 月に終了予定である。
- 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受けている¹⁴。
- Technology & Society Domain¹⁵に属する。

Technology & Society Domain：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ならない。このような内容に関する活動が行われている¹⁶。

- Semantic Web Activity Statement¹⁷によると、Semantic Web Activity の目標は次の通りである。「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

¹¹ 「各利用者が、Web 上で利用可能な資源を最大限活用できるようなソフトウェア環境を開発すること」という説明が続く。

¹² <http://www.w3.org/Consortium>

¹³ <http://www.w3.org/Metadata/>

¹⁴ <http://www.w3.org/Consortium/>

¹⁵ <http://www.w3.org/TandS/>

¹⁶ 斎藤信男「次世代 Web コンピュータにおけるセマンティック Web 等関連技術動向」(2001 年 3 月) より。

¹⁷ <http://www.w3.org/2001/sw/Activity>

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。この目標（Goal）を実現させるために、Semantic Web Activity は、以下の具体的な目標（Objective）を掲げている。これらは、W3C の Metadata Activity によって開発された既存の技術をベースにしている。

- ・ RDF インタレストグループの活動の継続：
RDF インタレストグループは RDF の実装と普及を促進し、W3C やその他のコミュニティにおける RDF 関連活動との連携を取る。
- ・ “RDF Model and Syntax Specification: W3C Recommendation 22 February 1999¹⁸”の改版
- ・ “RDF Schema Specification 1.0: W3C Candidate Recommendation 27 March 2000¹⁹”の仕様策定：
RDF の試験的な実装の結果を仕様書にフィードバックし、また” XML Schema Part 2: Datatypes: W3C Candidate Recommendation 24 October 2000²⁰”で示された方向性を吟味する。
- ・ P3P、CC/PP、XML Protocol、WAI 等の、他の W3C の活動との協調
- ・ Semantic Web を実現しようとしている他の団体の活動との協調：
DCMI、DAML、OIL、SHOE を含む。W3C としては、DAML+OIL の活動が W3C の Semantic Web Activity に参画することを期待している。このような協調の目標は、各仕様が分散することを避け、ソリューションの一般性を保証することである。
- ・ XML と RDF をサポートする先進開発プロジェクト（Semantic Web Advanced Development）の実施」
- 他の W3C の活動との関係としては、Semantic Web Activity は以下の W3C の活動に対して情報提供と支援を行う²¹。
 - ・ “Harvesting RDF Statements from Xlinks: W3C Note 29 September 2000”
 - ・ XML Signature を用いる際の署名者の意図を記述する
 - ・ P3P におけるプライバシーポリシーや CC/PP におけるプロファイルと関連づけられたセマンティクス
 - ・ WAI における、障害者の Web ページへのアクセシビリティを改善するためのあのテーション
 - ・ 非ツリー構造データ向けの XML DOM の利用についての調査
 - ・ XSLT を用いたドキュメント構造変換
 - ・ 他のジェネリック・データモデルやタイプ階層と XML Schema との統合
 - ・ 一つの XML Schema の中に RDF セマンティクスを組み込むこと
 - ・ URI Activity によって定義された情報資源の識別をサポートするような、情報資源の特徴の表現

¹⁸ <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

¹⁹ <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>

²⁰ <http://www.w3.org/TR/2000/CR-xmlschema-2-20001024/>

²¹ <http://www.w3.org/2001/sw/Activity - Context>

(4) セマンティック Web に対するアプローチ

- セマンティック Web の展開に向けた W3C のアプローチは、以下のものである²²。
 - ・ 支配的な原則の確立
 - ・ 標準化や技術開発の促進
 - ・ 短期間の展開への集中
 - ・ 長期的な研究課題への注目
 - ・ 政策的含意の理解
 - ・ 効果的な協力と協働のための環境の育成
 - ・ 多くのハードワーク
 - ・ 少しの運
- W3C にとっての、セマンティック Web の原則は以下のとおりである²³。
 - ・ 物理世界の全てのものは URI によって識別されるオンライン表現を持ち、セマンティック Web の中で関連付けられる。
 - ・ セマンティック Web は述べられている内容、対象、場所に基本的な制約を加えるべきではない。Web 上の情報は部分的である（完全ではない、不整合を含み得る）ので、誰でもあらゆることについてあらゆることを言うことができる。
 - ・ セマンティック Web は Web 上の記述が「真実」であることを主張してはいけない。真実あるいは情報の信頼性は、Web 条の情報 W 於処理するアプリケーションが評価する。
 - ・ セマンティック Web は分散したコミュニティが同一の領域に異なる表現の情報を独立に与えることを許さなければならない。
 - ・ 最小限の設計とし、簡単なものは簡単に、複雑なものは可能な限り簡単にする。
 - ・ セマンティック Web はシステムやアプリケーションに依存しない、一般性のある共通モデルに基づく必要がある。
- セマンティック Web は、以下の既に検証されたアイデアの上に構築される²⁴。
 - ・ RDF、XML、Xlink、ハイパーテキスト、メタデータを結合
 - ・ 現在はシンプルで将来は複雑になるアプリケーションを確立する（例えば、Dublin Core、RSS、PRISM、MusicBrainz）
 - ・ Web 自動化とデータ集約のための一般原理に集中する。
- セマンティック Web のアーキテクチャが目指すものは、次の通りである²⁵。
 - ・ Web 上でメタデータ交換を行うアプリケーションのための規約の定義
 - ・ “serialization syntax”としての XML の利用
 - ・ 専門家のコミュニティによって定義される語彙の意味の確立
 - ・ 多様なメタデータの細粒度での融合

²² <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

²³ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

²⁴ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

²⁵ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

(5) セマンティック Web の課題 (W3C の立場から)

- ツール、格納方法、処理方法等、開発や策定すべきことが沢山ある。また、本当の関係者 (セマンティック Web 上のエージェントの開発者) が、いまだ関与していない²⁶。
- セマンティック Web の実現における困難として、以下のものが挙げられる²⁷。
 - ・ 電子署名のためのビルトイン
 - ・ 証明の生成
 - ・ 古典的なセマンティクスと、極端に地域的社会的な背景を持つ現実の人間同士の意味との基本的な不整合
 - ・ RDF / セマンティック Web 技術を理解しやすくすること
 - ・ 多数のためのデータモデリング
 - ・ レガシーアプリケーションのデータ利用のための、意味の明確化
 - ・ また、オープンイシューとして、URI の問題と、スケーラビリティの問題がある。
 - ・ 標準化とテクノロジーの面では、Web オントロジーや Query、Rules が課題である。
- Semantic Web Advanced Development に関しては、目標とミッションの透明性、タスクフォースの拡大、分散開発を効率化するツールとの統合、類似する問題を扱うグループ (例えば open source、.NET、DotGNU) との協力が課題である²⁸。

(6) セマンティック Web の将来の方向性 (W3C の立場から)

- Semantic Web Activity については、他の技術 (例えば、Topicmaps) とのさらなる関連、RDF と XML ファミリーの間のより明確な結びつき、セマンティック Web と Web サービスの間のより良い理解、EU のセマンティック Web イニシアティブとの協調などが求められる²⁹。
- Tim Berners-Lee の意見では、研究レベルでは企業もセマンティック Web にかなり積極的に参加するようになったが、ビジネスレベルでは未だ二の足を踏んでいる (これも商用プロダクトが世に出てないがためであろう)。今後 1~2 年でツールが出てくると、ソフトウェアベンダも商用ベースで本分野に入ってくるだろう³⁰。
- W3C としては、やがては IBM やマイクロソフトといった Web サービス陣営が「セマンティック Web の弱点を補強するための Web サービス」という風に考えるように持っていきたい。セマンティック Web は Web サービスのように企業間のデータ交換を目的とすることに留まらず、インターネット上にある全ての Web を対象としたもっと大きな問題を扱うことを目的としている³¹。
- セマンティック Web のサービスやアプリケーションの例としては、Netscape の RSS を

²⁶ INTAP 「WWW10 調査報告書」より。

²⁷ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

²⁸ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

²⁹ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

³⁰ INTAP 「W3C 出張報告」より。

³¹ INTAP 「W3C 出張報告」より。

使った例がある。また、RDF ベースの DB も開発されている。電力業界のコンソーシアムでは、電力事業の情報交換に RDF を適用した例もある³²。

- メタデータの半自動生成ツールとしては、MIT の David R. Karger が Haystack というオートマティック・クラスタリング・データ（個人情報からメタデータを作成するツール）を開発している。また、OCLC のメタデータ変換や Netscape の RSS もある³³。
- W3C としては、DAML+OIL の技術を利用し、オントロジーファイルの定義を行う方向である。

4.2 Semantic Web Activity 配下のグループ

（１）RDF Core ワーキンググループ

RDF Core ワーキンググループ³⁴は RDF Model and Syntax Specification（W3C 勧告）の改版について検討するために、また RDF Schema Specification 1.0（W3C 勧告候補）についていくつかの見直しを図るために設立された。その他の目的としては、上記 RDF Schema Specification 1.0（W3C 勧告候補）へのフィードバックに対して、レスポンスを行うことである。

新たなワーキンググループを立ち上げるための提案もなされており、この中には query 言語およびサービス、RDF グラフのための代替 XML シンタクス、相互運用性のテストベッドといったテーマが含まれている。RDF query、RDF logic、DAML、OIL、XML スキーマといった問題に関してワーキンググループの立ち上げが提案されるかもしれない。

（２）RDF インタレストグループ

RDF インタレストグループ³⁵は、RDF の革新的なアプリケーションについて議論するための、W3C 会員及び非会員向けのフォーラムである。同グループはまた、RDF に関連した将来の潜在的なワークアイテムや、それらのワークアイテムと W3C の他の活動との関係性、Web を巡る広範な社会的・法的文脈との関係性について議論を行っている。同グループは 1999 年 8 月 20 日に立ち上げられた。

同グループの憲章（charter）によると、「RDF インタレストグループは、インターネットコンテンツの選択、フィルタリング、ラベリング、署名、品質保証等をめぐる社会的・法的・技術的問題の文脈において、W3C のメタデータ技術の適用を探求する。また、セマンティック Web、知的財産権、P3P、電子商取引の文脈における RDF の適用も扱う」という。

RDF インタレストグループの下には 2 つのサブグループ（ RDF+Logic と RDF カレンダー（スケジュールに特化））がある。

（３）Semantic Web コーディネーショングループ

Semantic Web Activity のアドバイザー・フォーラムである Semantic Web コーディネ

³² INTAP「W3C 出張報告」より。

³³ INTAP「W3C 出張報告」より。

³⁴ <http://www.w3.org/2001/sw/RDFCore/>

³⁵ <http://www.w3.org/RDF/Interest/>

ーショングループ³⁶は、セマンティック Web に関連した標準や技術に注力する他のグループとの間の関係や相互依存性をマネージしている。

Semantic Web コーディネーショングループは 2001 年 7 月現在、Web Ontology ワーキンググループの憲章 (charter) をレビューしている。

(4) Web オントロジー・ワーキンググループ

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。Web オントロジー・ワーキンググループの活動目的は、RDF Core ワーキンググループの成果に基づいて、各研究開発コミュニティ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。

(5) W3C のアクティビティについて

W3C のアクティビティは一般に、以下に挙げるグループに組織される³⁷。

- ワーキンググループ：

技術開発を行うグループ。仕様書やソフトウェアの作成、他グループの成果物のレビュー等を行う。電子メール、電話会議、face-to-face 会議により運営される。ワーキンググループのメンバーには、会議参加に関する一定の義務がある。作業を迅速に進めるために、メンバーの総数は 15 名以下程度が目安とされ、特定分野のエキスパートから構成される。ワーキンググループの人数が大きくなりすぎた場合、インタレストグループと、より少数のワーキンググループ (コアなグループ) とに分割することができる。

- インタレストグループ：

より一般的な活動を行うグループ。仕様書の作成は行わない。Web 技術と将来動向について評価を行う人々に参加してもらい、アイデア交換を行う。その分野での様々な意見を集めたり、関連するワーキンググループに対するフィードバックなどを求めたりする。インタレストグループのメンバーには、会議参加に関する一般的な義務はない。原則として、メンバーの総数に制限はない。

- コーディネーショングループ：

関連グループとの間の連携を行うグループ。コーディネーショングループは W3C 内外のグループの間の連携を促進する。

4.3 Semantic Web Advanced Development

(1) 概要

- Semantic Web Advanced Development³⁸では、Semantic Web Activityを支援するための共同作業と技術開発を促進する努力を行っている。セマンティック Web の基盤となる

³⁶ <http://www.w3.org/2001/sw/CG/>

³⁷ <http://www.w3.org/Consortium/Process-20010719/groups.html>

³⁸ <http://www.w3.org/2000/01/sw/>

ようなコンポーネントの開発のためにリソースを提供している。

- Semantic Web Advanced Development は、プロトタイプアイデアの調査、研究コミュニティとの連絡、ワーキングドラフトの実装試験のためのテストベッドの提供、共同開発環境の提供、さらなるセマンティック Web インフラストラクチャ開発の促進を行っている³⁹。
- Semantic Web Advanced Development のハイライトとして、日程管理とワークフローを挙げることができる。RDF Calendar タスクフォースは、日程管理に関心のある個人によるインフォーマルなグループで、行事のための RDF スキーマのドラフト作成と、日程管理における RDF の利用のデモンストレーションを目指す予定である⁴⁰。
- Semantic Web Advanced Development の目標は、実地で試されていないアイデアにワーキンググループが従事しないようにすることであり、また仕様書の開発の際に関連するツールを提供するということである。その目標には、パーサ、データ貯蔵庫、ブルーフ・チェッカー及びその他のセマンティック Web 処理モジュールから構成される複数のコンポーネントの間の相互通信パイプラインのプロトタイプの開発も含まれる。
- W3C は、セマンティック Web 技術の商業利用に先立った試験ツールを開発する機会を提供している。しかし、これらのツールの開発における究極の目標は、セマンティック Web を支援するために必要な Web サービス及びエージェントのより効果的な記述、管理、統合及び自動化の基盤を提供することである。
- Semantic Web Advanced Development はワーキンググループ及びインタレストグループと緊密な連携のもとに実行される。

(2) 開発システム

● Annotea

Web ドキュメントに注釈のメタデータを付加するツールである。付加された注釈を共有することも可能にする。Annotea は Annotea クライアントと Annotea サーバから構成され、その間の通信には HTTP+RDF のプロトコルが用いられる。クライアントとしては W3C が開発中のブラウザ兼エディタである Amaya (クライアント独立であるため、Annozilla、Bookmarklet 等のクライアント実装も存在している) を用い、サーバとしては W3C Generic Server (Apache+CGI+Perl+..) を用いる。その他、Xpointer、Xlink といった W3C の技術が可能な限り再利用されている。

課題としては、誤った箇所に注釈がついてしまうことや、一度注釈をつけた箇所の前後に修正が入ると修正によって文書中の位置がずれてしまうために、Xpointer による指定が正しくないものになってしまうことがある。また、注釈をつけた箇所を削除すると、指し示すものがない注釈情報ができてしまう⁴¹。

³⁹ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

⁴⁰ <http://www.w3.org/Talks/2001/07/30-swws/slide1-0.html>

⁴¹ INTAP「WWW10 調査報告書」より。

第 5 章 米国のセマンティック Web に関する研究開発動向

第5章米国のセマンティック Web に関する研究開発動向

5.1 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)⁴²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) およびソフトウェア (知的エージェント) 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⁴³。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⁴⁴。
- DAML 仕様書
 - ・ 2000 年 10 月 10 日に "DAML-ONT" 仕様書の初版⁴⁵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)⁴⁶が DARPA の Jim 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)⁴⁷という仕様書をリリースし、DAML と OIL との統一化を行った。2001 年 3 月 27 日には "DAML+OIL (March 2001)"⁴⁸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 プロジェクトのアプローチ⁴⁹
 - ・ W3C、OIL、SHOE と連携
 - ・ W3C と共同で言語の定義を行う
 - ・ プロトタイプツールの開発
 - ・ オントロジーの作成と編集
 - ・ オントロジーの翻訳とマッピング

⁴² INTAP 「WWW10 調査報告書」より。

⁴³ <http://www.daml.org/researchers>

⁴⁴ INTAP 「BBN Technologies 出張報告」より。

⁴⁵ <http://www.daml.org/2000/10/daml-ont.html>

⁴⁶ <http://www.daml.org/committee/>

⁴⁷ <http://www.daml.org/2000/12/daml+oil-index.html>

⁴⁸ <http://www.daml.org/2001/03/daml+oil-index>

⁴⁹ <http://www.daml.org/2001/06/swday-daml/Overview.html>

- ・分散化された知識ベース
- ・マークアップ・エディター
- ・Validation サービス
- ・DAML APIs
- ・DAML 対応のブラウザ
- ・その他アプリケーション
- DAML 言語の特徴⁵⁰
 - ・ WWW (URI、XML シンタックス、HTTP) を基盤としている。多様性、スケーラビリティ、セキュリティ等に対する強い関連性を持つ。
 - ・その他のクラスに関するプロパティの付加
 - ・タグ付け / reification / 引用・・・ステートメントに関するステートメント
 - ・形式セマンティクス：モデル理論的、公理的。
- DAML 言語の将来的作業⁵¹
 - ・フィードバックの統合
 - ・W3C Semantic Web Activity への提案
 - ・rules
 - ・可能なレイヤリング
 - ・ロジック、proof-checking
- DAML プログラムの研究領域⁵²
 - ・言語設計
 - ・オントロジーの開発：UML、パワーポイント、コンポジション
 - ・サービス：WWW ゲートウェイ、記述
 - ・オントロジー翻訳
 - ・分散化された query 処理 / 推論
 - ・レポート・ジェネレーション
 - ・セキュリティ：署名された XML、信用
- DAML-S version0.5

2001 年 5 月 30 日に DAML-S version0.5 がリリースされた⁵³。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) とサービス間操作 (Interoperation)、および実行モニタリング (Execution Monitoring) な

⁵⁰ <http://www.daml.org/2001/06/swday-daml/Overview.html>

⁵¹ <http://www.daml.org/2001/06/swday-daml/Overview.html>

⁵² <http://www.daml.org/2001/06/swday-daml/Overview.html>

⁵³ <http://www.daml.org/services/daml-s/2001/05/>

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⁵⁴。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）。

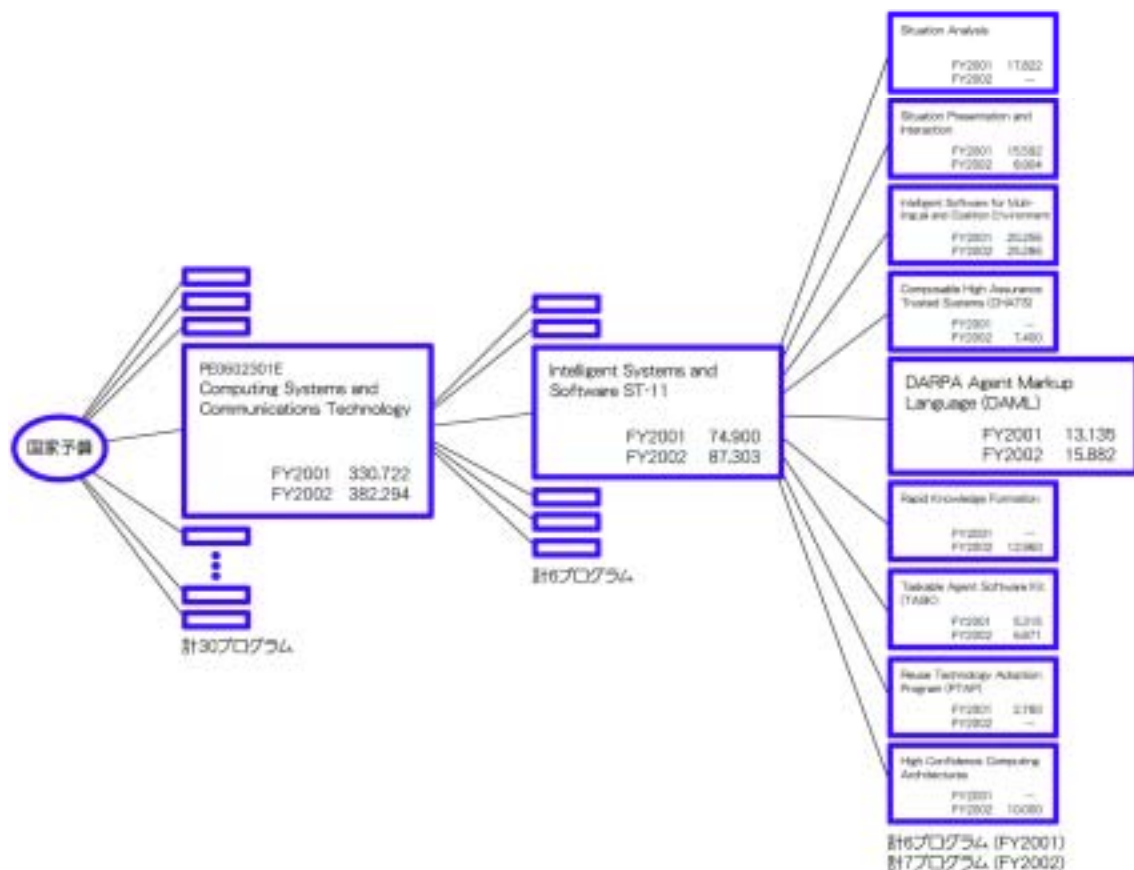


図 5.1 DAML (小分類) の実施計画及び予算 (2001 年)

- DAML 言語仕様の完成

⁵⁴ INTAP 「SWWS の聴講メモ (暫定版)」より。

- Intelink⁵⁵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)：

- 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 万件以上に上る DAML ステートメントのレポジトリ作成
 - dynamic prioritization management 向け技術の開発
 - 帯域状況が変化する場合のインテリジェントな情報配信技術の開発
-

⁵⁵ 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。

計 9 項目 \$M15.882 (約 19 億円)

(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⁵⁶。

- スタンフォード大学及びカールスルーエ大学 (ドイツ)

Web 上の知的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。
<http://www.daml.org/ontologies/>を参照のこと。上部集合としては、Cyc が無料で公開しているものを使うとのことである⁵⁷。

(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⁵⁸。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⁵⁹。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。

- DAML Crawler

2001 年 5 月 4 日には DAML Crawler がリリースされた⁶⁰。同ツールでは、DAML で書

⁵⁶ INTAP「BBN Technologies 出張報告」より。

⁵⁷ 浦本「Semantic Web Working Symposium 参加報告メモ」より。

⁵⁸ <http://www.daml.org/viewer/>

⁵⁹ <http://www.ontoprise.de/>

⁶⁰ <http://www.daml.org/crawler/>

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⁶¹。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⁶²

- ・ 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 パーサおよびシリアライザー
- ・ 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、PalmDAML⁶³、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⁶⁴

- ・ 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、個人のプロフィール、カレンダー、地図情報（物理的にその講演を聴きにいくか否か）と組み合わせ、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⁶⁵。

（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⁶⁶。
- 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 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

⁶¹ <http://www.daml.org/validator/>

⁶² <http://www.daml.org/2001/06/swday-daml/Overview.html>

⁶³ PalmDAML とは、PalmOS 向けに DAML Viewer を実装したもの。

⁶⁴ <http://www.daml.org/2001/06/swday-daml/Overview.html>

⁶⁵ 浦本「Semantic Web Working Symposium 参加報告メモ」より。

⁶⁶ INTAP「BBN Technologies 出張報告」より。

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⁶⁷。

5.2 DCMI (Dublin Core Metadata Initiative)

(1) 概要

- 1994 年にシカゴで開催された WWW5 から本格的活動開始。
- 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⁶⁸。
- 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 といった基本要素の集合) 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。
- Dublin Core の歴史⁶⁹
 - 94 年 : Web 記述のための簡単なタグを開発
 - 95 年 : Dublin Core は、必要とされた沢山の語彙 ("Warwick Framework") の一つ
 - 96 年 : テキストとイメージの記述のため、Dublin Core を 13 要素から 15 要素に拡張
 - 97 年 : Warwick Framework は、RDF のため、形式記述を必要とする
 - 2000 年 : DCMI は修飾子の推奨を行った。また活動領域を Dublin Core 以外にも拡大した。
- Dublin Core とは、情報資源を記述するための以下の 15 の特別な属性である。多くの人たちが共同に使うためのポリティカルなアグリーメントの結果 15 個のメタデータエレメントとなった。Core とは、メタデータ記述のベーシックパートである。用途ごとに拡張可能である。共通になる部分として 30 ヶ国 (25 言語) 間でコンセンサスを取れたものが 15 個であり、それで有効だと考えられている。

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

- Dublin Core の拡張性としては、各 AP ごとに拡張部分を作成。例えば、「インターナ

⁶⁷ INTAP 「OCLC (DCMI) 出張報告」より。

⁶⁸ INTAP 「OCLC (DCMI) 出張報告」より。

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LEMENTの上
にそれぞれが拡張を行っている。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⁷⁰。

- 2000 年 10 月、DCMI は活動領域の拡大を行った。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簡単にすることとなった。
 - ・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
- Dublin Core の 2001 年の Workshop は 10 月 22 日～26 日に東京で開催された。Web
サービス、チュートリアル、カンファレンスの 3 トラックから成る。

（２）今後の展望

- Dublin Core に基づいた技術適用の例としては、RDF を使用したアドビのアクロバット
製品がある。この製品が最初の大型製品になるのではないか。また、Dublin Core 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⁷¹。

5.3 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。
- 確立されたボキャブラリ（Dublin Core）の上に構築される。RDF シンタクスを使用。
- アプリケーション独自のセマンティクスのレイヤー：
シンディケーション・ボキャブラリ、注釈ボキャブラリ等
- RSS は、Netscape が 1999 年に自社ポータル（My Netscape⁷²）のために開発した。こ
れが RSS0.9 であり、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 Winer⁷³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

⁶⁹ INTAP「WWW10 調査報告書（別紙）」より。

⁷⁰ INTAP「OCLC（DCMI）出張報告」より。

⁷¹ INTAP「OCLC（DCMI）出張報告」より。

⁷² 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。

⁷³ 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）でホストされた WebLogs からコンテンツを取り入れ
ている。

モジュールによる拡張機能を備えた RSS1.0 仕様書⁷⁴が 2000 年 12 月に発表された。また、RSS1.0 とともに、Dublin Core 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⁷⁵。

(2) RSS の応用システム：O'Reilly Network

● O'Reilly Network

オリリー・ネットワーク⁷⁶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し、また提携サイトとコミュニケーションする機会を提供する。第一の目的は、技術者が新たな技術コミュニティに横断してくることをサポートすることである。

● O'Reilly Network の Meerkat⁷⁷

- ・ 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、議論スレッド、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 は時間的な順番を重視する。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 のフロントエンドは、利用者が指定したプリファレンスを用いて一つの query を作成し、マッチングするものをストーリーのデータベースの中から探し、それらを表示する。
- ・ Meerkat のバックエンドは、Perl で記述され、以下のモジュールを用いている：
LWP::UserAgent DBI, DBD::mysql, XML::Simple
- ・ フロントエンドは PHP で記述されている。Meerkat は Apache Web サーバで提供され、データは MySQL サーバで保存されている。
- ・ Meerkat 以外にもリーダーはあるが、Meerkat の特徴は開発者、プログラマー、Web デザイナー、イントラネット/エクストラネット管理者の興味のあるストーリーを時間的に配列するところにある。
- ・ 利用者が自分のチャンネルを登録する場合は、

⁷⁴ <http://groups.yahoo.com/group/rss-dev/files/namespace.html>

⁷⁵ INTAP「WWW10 調査報告書」より。

⁷⁶ <http://www.oreillynet.com/>

⁷⁷ <http://www.oreillynet.com/meerkat/>

xmlTree: directory of content

Aggregator.Userland.Com

というアグリゲータを用いて登録する。どちらを用いても、RSS または ScriptingNews でフォーマットされたチャンネルを登録することができる。

- Meerkat は実際にはこれらのアグリゲータを毎朝訪問し、新たに見つけたチャンネルを書きとめる。それらの中から、テーマ的に適切なもの（通常は、技術 / コンピュータ / ギーク / 科学関連）が O'Reilly Network の中にピックアップされる。
- 提案やバグの報告、賞賛、その他のフィードバックは O'Reilly Network RSS フォーラムに寄せることができる。
- 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。

第 6 章 欧州のセマンティック Web に関する研究開発動向

第 6 章欧州のセマンティック Web に関する研究開発動向

6.1 EU 委員会 IST プログラム

(1) EU 委員会 IST (Information Society Technologies) プログラム⁷⁸

- 「EU 第 5 次 RTD (research and technological development) フレームワーク・プログラム—ユーザフレンドリーな情報社会の創造」 (1998-2002) の一部。
- 情報処理技術、通信技術およびメディア技術の集積の上に築かれる、統合された調査プログラム。
- IST プログラムは総額 36 億ユーロの予算を持ち、EU 委員会の情報社会総合理事会 (Information Society Directorate-General) によって運営されている。

表 6.1 IST プログラム予算

EU 第 5 次 RTD フレームワーク・プログラム テーマ 2 : ユーザフレンドリーな情報社会の創造 (IST プログラム)	
アクティビティ	予算額 (百万ユーロ)
a. 主要なアクション	
i. 市民向けのシステムとサービス	646
ii. 仕事と電子商取引の新たなメソッド	547
iii. マルチメディア・コンテンツとツール	564
iv. 基盤技術とインフラストラクチャ	1363
b. 一般的性格の調査及び技術開発活動： 将来の技術及び出現しつつある技術	319
c. 調査インフラの支援： 調査ネットワーキング	161
総計	3600

(2) IST プログラム内の Semantic Web Technologies

- 欧州委員会は、Semantic Web technologies に関する 2001 年 IST ワークプログラムにおいて、アクションラインを発表した⁷⁹。
 - ・コールナンバー：IST コール 7
 - ・領域：情報のアクセスとフィルタリング
 - ・アクションライン： 4.1-Semantic Web technologies
 - ・発表日：2001 年 6 月を予定
- コンテンツの処理を自動化するための Semantic Web の概念は、W3C や他の国際組織・国内組織において、重要性を増してきている。このグローバルなイニシアティブと、情報へのアクセスとフィルタリングにおける EU のマルチメディア・アクティビティとの関係性を構築するために、100 人以上の専門家がルクセンブルグに召集され、欧州に

⁷⁸ <http://www.cordis.lu/ist/home.html>

⁷⁹ <http://www.cordis.lu/ist/ka3/iaf/iii41obj.htm>

おける更なる調査開発と、以下のようなマルチメディア・アプリケーションのための展望と必要条件とを見極めようとした。

- ・ Web ナビゲーションと情報検索
- ・ ナレッジマネジメント
- ・ コンテンツにおける商取引 (Commerce-in-content)
- Call for Proposal は Web 向けのセマンティックベースの技術とアプリケーションにおける R&D のための協働的な提案を提出する具体的な機会である。ここで Web とは公共の WWW、企業内の Web 及びイントラネット、及びモバイル向けコンテンツの Web を意味する。提案に際しては、以下の Web サイトが参考になる。
 - ・ the Workprogramme 2001 themes,
 - ・ the 'Semantic Web Technologies' Workshop Report,
 - ・ answers to frequently asked questions,
 - ・ links to ongoing projects (especially ONTOWeb, ONTO-KNOWLEDGE), to other Key Actions, especially KA II (E-Commerce) and KAIV (Software for ontologies),
 - ・ links to W3C and related bodies.
- コンタクトポイントは、DG Information Society / Unit D5, Management of Content and Information の Hans-George Stork である。
- アクションライン⁸⁰ :
 - 「 .4 情報に関するアクセス、フィルタリング、分析および処理」

WWW は、固定プラットフォームと移動プラットフォームからのアクセスの両面で、ビデオ・オーディオを含むマルチメディア・コンテンツの提供と配信のための主要な手段になりつつある。しかし、情報資源や知識の発見、情報のフィルタリング、及び知的ブラウジング等を容易にするためにセマンティックな構造を Web コンテンツに課すことが必要である。すなわち、Web コンテンツをマシン・リーダブルにすることが必要である。Web が完全なサイズとダイナミクスを与えられた際には、このことは高度に自動化された処理を通じてのみ達成されうる。
 - 「 .4.1-Semantic Web technologies」
 - ・ 目的：利用者の関心とニーズに関連し、またクオリティに対する利用者の期待に適合させながら、利用者が Web から情報へアクセスし、情報を抽出し、またフィルタリングすることを可能にすること。このことは、Web ベースのコンテンツのアクセス、抽出、フィルタリングに関連したサービス（とりわけ、既存の、また出現しつつある標準に適合した新たな検索システムと検索マシン）のための、新しくかつ先端の方法、モデル、ツール及びシステムを必要とする。これらのサービスは通常、知的エージェントとして提供されるであろう。
 - ・ 焦点：このアクションラインは、セマンティック Web の作成と活用に寄与するようなコ

⁸⁰ <http://www.cordis.lu/ist/ka3/iaf/iii41text.htm>

ンテンツ技術に焦点を当てている。Web のコンテキストにおいてこれらの技術を開発し、適用し、ベンチマークするプロジェクトを歓迎する。

- ・デジタルコンテンツのセマンティクスを定義し宣言するための、デジタルコンテンツをコード化したり、構築したりするメソッドとツール。これらには通常は、セマンティックな相互運用性のために、また、ある分野特定のアプリケーションのオントロジー等を推論 (reason) するために、XML、RDF およびその他の技法が用いられるだろう。
- ・自動特徴探知、ビデオセグメンテーション、ポスト処理等を通じた、Web ベースのコンテンツ (とりわけビデオ、オーディオ、画像) のセマンティックな属性を抽出するためのメソッドとツール。これらは、コンテンツ分析や Web 情報資源の自動カテゴリ化に基づいたマルチメディア目録化 (indexing) を容易にするだろう。
- ・情報エージェントや特定の query 言語等の、知識発見および知的フィルタリング・プロファイリングのためのセマンティクス・ベースのツール。協調フィルタリングや、特定の、または一般のユーザコミュニティにおいて知識共有するためのセマンティクス・ベースのツール。
- ・情報のビジュアル化：利用者が精通していない複雑な情報空間の中を自然にナビゲートし、検索するためのラディカルな新しい方法を提供するために、セマンティックな情報を利用するような、知的でビジュアルなインターフェイス。

このアクションラインの下プロジェクトによって対処される問題は、メタデータ・レポジトリやオントロジー等に関する提案されたアプローチの維持力 (sustainability) とスケラビリティとを含むものである。プロジェクトは、Web アクセスのための家庭内プラットフォーム (WebTV 等) やモバイルプラットフォーム (WAP や UMTS 等) の制約と特別な要求事項、およびそれらの利用者のコミュニティに特に焦点を置いてよい。スケラビリティと同様にアクセスの容易さと、コンテンツの表示の容易さは、主な課題となるであろう。さらに、Web ベースのサービス環境に地理的な情報を統合することは、モバイルプラットフォームのユーザビリティに大きなベネフィットをもたらすであろう。」

- 各プロジェクトのコンソーシアムとしては、これまでの RTD プロジェクトの平均的なサイズは 3 ~ 4 ヶ国間の 4 ~ 6 のパートナー組織から構成されているものである。「情報に関するアクセス、フィルタリング (IAF)」のプロジェクトに対する平均的な EU からの資金提供額 (プロジェクト予算総額の 50%) は、100 万ユーロから 250 万ユーロとなっている。

6.2 OIL (Ontology Inference Layer)

(1) On-To-Knowledge⁸¹

- EU 委員会 IST プログラムの一環であり、EU 委員会から資金提供を受けている。予算規模は総額 250 万ユーロで、うち EU 委員会からは 134 万ユーロである。

⁸¹ <http://www.ontoknowledge.org/>

- 2000 年 1 月から第 5 回欧州フレームワーク・プログラムの下で開始された。プロジェクトの継続期間は 2002 年 6 月まで。(プロジェクト番号 IST-1999-10132)
- プロジェクトの成果の一つが OIL (Ontology Inference Layer)⁸²の開発。OIL 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なされている。
- On-To-Knowledge プロジェクトでは、共有可能かつ再利用可能なオントロジーを基盤とするナレッジマネジメントをサポートするためのツールと方法論 (Methodology) の開発を行う。
- On-To-Knowledge プロジェクトの目標 (Objective) は、
 第一に、オンラインの情報サービスにアクセスする敷居を低くすること。
 第二に、異なるユーザグループのニーズに適合し、整合的かつタイムリーな情報を抵抗するためのコストを削減すること。
 第三に、3 つのケーススタディが On-To-Knowledge において開発されたメソッドとツールがユーザ・フレンドリーであることを保証すること。
- On-To-Knowledge プロジェクトの目標 (Goal) :
 大規模かつ分散した企業におけるナレッジマネジメントの品質を高めるために、オントロジーを電子情報に適用する。オントロジーは知識の共有と再利用を支援するための形式的な理論である。半構造化された情報のセマンティクスを明示的に表現するために利用される。このことによって、情報の獲得と維持、アクセスを、高度な方法で自動的に支援することを可能となる。このために、下記の 2 つのパースペクティブから、ナレッジマネジメントのサポートにおけるオントロジーの利用を可能とするような、インターネット (及びイントラネット、エクストラネット) 環境における半構造化され、莫大な量のテキスト情報の資源に知的にアクセスするための方法論とツールを開発する。
 - ・ 情報クライアント・・・知識へのアクセスはシンプルかつ効果的でなければならない。知識にアクセスする際のコストと障壁は低減されるべきであり、利用者は既存の知識ソースについて教えられる必要がある。既存のキーワードベースの検索技術は明らかにこれらの要求事項を満たしていない。情報ソースへのアクセスを改善することが、On-To-Knowledge の第一の主要なゴールである。
 - ・ 情報プロバイダ・・・現状の技術において莫大な量のテキスト化され、半構造化された情報を提供し維持することは、労力を要し、かつコストがかかる活動である。これらのコストを低減することが、同プロジェクトの第二の主要なゴールである。
 On-To-Knowledge プロジェクトの目標 (Goal) は、効率的で効果的なナレッジマネジメントを支援することである。緩やかに構造化されたオンラインの情報ソースを獲得し、維持し、アクセスするという 3 つのプロセスに焦点を当てる。
 - ・ 獲得 (Acquiring) : テキスト情報からセマンティック情報を抽出するために、テキストマイニングと抽出技術が用いられる。
 - ・ 維持 (Maintaining) : RDF と XML は、半-構造化された情報ソースのシンタクスと

⁸² <http://www.ontoknowledge.org/oil/index.shtml>

セマンティクスを記述するために用いられる。ツールサポートは、このような知識に関する自動的なメンテナンスと、定義の閲覧を可能にする。

- ・アクセス（Accessing）：プッシュサービスとエージェント技術は、利用者が情報にアクセスすることを支援する。
- On-To-Knowledge プロジェクトのアプローチ方法は上記の 3 つのプロセスのそれぞれについてオントロジーを利用することである。これらの目標（Goal）に到達するために、3 層から成るツール環境を開発する。
 - ・最下層（情報レベル、抽出レイヤー）においては、弱く構造化された情報ソースが処理され、それらからマシン・プロセッサブルなメタ情報を抽出される。
 - ・中間層（表現レベル、表現レイヤー）においては、これらの情報ソースへの自動アクセスや、自動作成、自動メンテナンスを提供するために、メタ情報が利用される。
 - ・最上層（アクセスレベル、アクセスレイヤー）においては、これらの情報へのアクセスの敷居を低くするために、更なるプッシュ及びプル技術が利用される。
- On-To-Knowledge プロジェクトの主要な成果物（Outcome）は、ナレッジマネジメントのための方法論、ツール及びベストプラクティスである。
 - ・方法論：方法論はナレッジマネジメントの概念とツールを企業に導入するためのガイドラインを提供する。その際、知識プロバイダが自分たちの知識を効率的かつ効果的に表現することを支援する。方法論には、ナレッジマネジメントのツールによって得られるべきゴールの同定が含まれる。方法論は、ビジネスプロセスの分析及び組織においてナレッジワーカーが演じる様々な役割の分析に基づいている。
 - ・インテリジェントサーチツール：インテリジェントサーチツールは、利用者が情報ソースのメンテナンス、集積及び獲得のために、情報及びツール環境にアクセスすることをサポートする。これらのツールは 3 層のアーキテクチャに基づいており、この 3 層アーキテクチャは、利用者の情報ニーズと利用可能な情報ソースとの間の大きなギャップの橋渡しを支援する。
 - ・3 つの大規模なケーススタディ：ケーススタディは大規模 / バーチャル規模の実際のニーズに適合した技術の開発の手助けとなる。また、方法論とツールの評価のための理想的なテストベッドを提供するだろう。
- On-To-Knowledge プロジェクトの全体構成
ツール環境のアーキテクチャは、異なる機能を持つ 3 つの層から構成される。抽出レイヤーは CognIT、表現レイヤーは AIdministrator、アクセスレイヤーは BT が担当している。BT は同システムのバックボーン（3 つの層において利用されるオントロジー的な側面）を担当しており、CognIT はあるアプリケーションコンテキストにおけるツールの適用の方法に関する方法論的な側面を開発している。3 つのケーススタディはそれぞれ、スイス・ライフ、BT、Enersearch が担当している。図 6.1 は、プロジェクトチームの構成を、ツール、方法論、およびケーススタディの観点から示している。

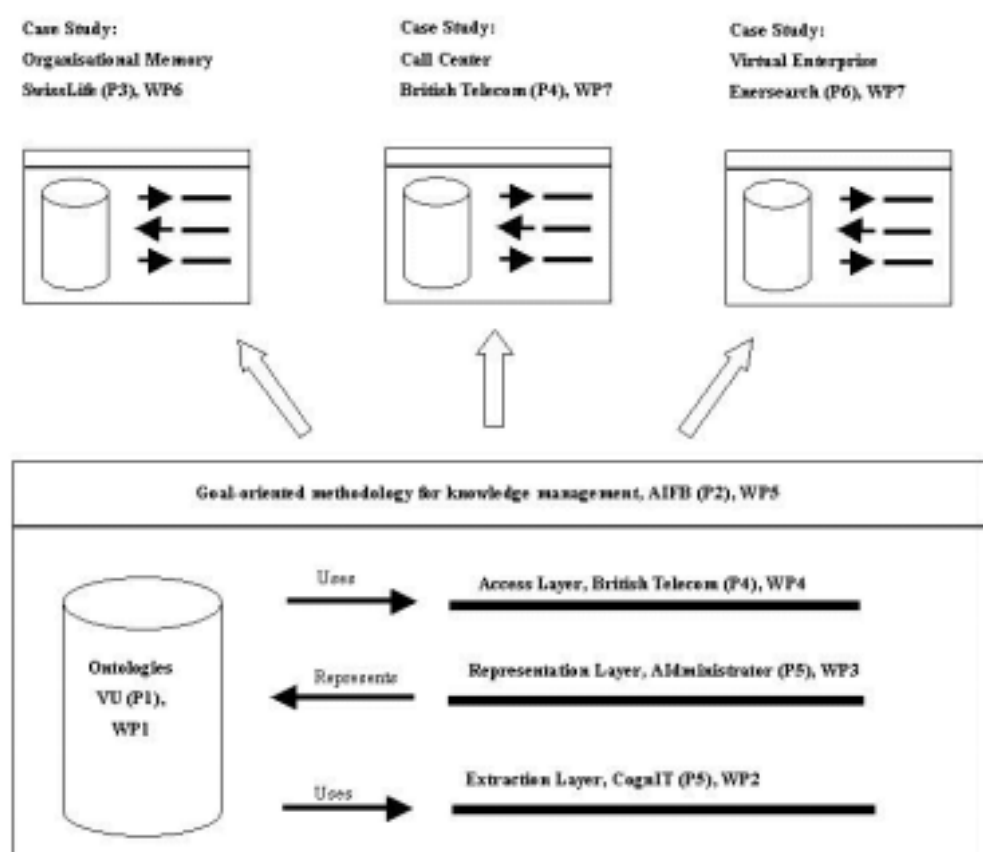


図 6.1 On-To-Knowledge プロジェクトの全体構成

- 表 6.2 は、各パートナー組織の役割を示すものである。フーリー大学（オランダ）、AIFB 研究所（ドイツ）、カールスルーエ大学（ドイツ）、Administrator（オランダ）、BT 研究所（英国）、スイス・ライフ、CognIT（ノルウェイ）、Enersearch（スウェーデン）が参加している。

表 6.2 各パートナー組織の役割

Partner	Role	Task
P1. VU	Methodology Provider, Coordinator	Method Development
		Tool Development
P2. AIFB	Methodology Provider	Method Development
		Tool Development
P3. Swiss Life	User	Evaluation
P4. BT	Technology Developer, User	Tool Development
		Evaluation and Learning
		Exploitation
P5. CognIT	Technology Developer	Tool Development
		Exploitation
P6. Enersearch	User	Evaluation
P7. Administrator	Technology Developer	Tool Development
		Exploitation

- ワークパッケージは、表 6.3 に示す通り。

表 6.3 ワークパッケージ

	タイトル	主担当
WP1	Developing tools for Ontology Construction and Interchange	VU
WP2	Developing support in Information extraction	CognIT
WP3	Developing an infrastructure for Information representation	AIdministrator
WP4	Developing various means for Information access	BT
WP5	Methodology and Management Guidelines	AIFB
WP6	A case study on Organisational memory	Swiss Life
WP7	A case study on call centres	BT
WP8	A case study on knowledge management in a virtual enterprise	Enersearch
WP9	Information dissemination and take-up of results	VU
WP10	Project Management and Exploitation Management	VU& BT

(2) IBROW⁸³

- EU Information Society Technologies (IST) プログラムの一環であり、EU 委員会から資金提供を受けている。予算規模は総額 153 万ユーロで、うち EU 委員会からは 110 万ユーロである。
- 第 4 回欧州フレームワークの下でプレ・フェーズが始まり、2000 年 2 月から第 5 回欧州フレームワーク・プロジェクトの下で完全な IST プロジェクトとなった。プロジェクトの継続期間は 2003 年 2 月まで。(プロジェクト番号 IST-1999-19005)
- 目的は、複数のデジタル・ライブラリから、利用者の要求に基づいてナレッジの検索を可能とする知的ブローカ・サービスの開発である。
- アムステルダム大学、スペイン科学研究協議会、AIFB 研究所(ドイツ)、スタンフォード大学、インテリジェント・ソフトウェア・コンポーネンツ社、フーリー大学(オランダ)が参加している。また、業界諮問委員会のメンバーは以下の企業から送られている。ダイムラー・クライスラー、ドイツ・テレコム、Bolesian、ユニリーバ、IBM、英国人工知能アプリケーション研究所である。
- プロジェクトの概要は、以下の通りである。
 - ・ IBROW のコンパチブルなライブラリの開発
 - ・ UPML とともに知識要素の注釈手続きをサポートするためのツールの開発
 - ・ 知識要素のキャパビリティを利用者の要求に適合させるための、また包括的な知識要素の導入と統合のためのフレームワークの開発
 - ・ インターネットサービスの開発

⁸³ <http://www.swi.psy.uva.nl/projects/ibrow/home.html>

- ・ 包括的な知識要素を記述するための注釈言語の実験と検証
- ・ 国際的な情報インフラを通して分散した知識要素の相互運用性のためのインフラ
- ・ 標準的なブラウザによってアクセス可能な、包括的な知的推論サービスおよび特定の知的推論サービスの一組
- ・ IT ワーカーが事業特有の問題に対するカスタマイズ可能な問題解決およびオントロジーを閲覧・選択・修正・設定するためのインターネットサービス
- ・ 特定のタイプの問題または業務向けの特殊なソフトウェア・ブローカの構築方法
- ・ 分子生物学における標準アルゴリズムの容易な検索と適用をサポートするためのシステム
- ・ Web / イン트라ネット上で情報検索を行うためのフィルタリング・エージェントの設定とカスタマイズを可能とするシステム
- プロジェクトのマイルストーンは、以下の通りである。
 - ・ 分散された知識要素の相互運用性のためのインフラの開発
 - ・ 標準的なブラウザによってアクセス可能な知的推論サービスの開発
 - ・ Web / イン트라ネット上の情報検索のためのフィルタリング・エージェントを作成・カスタマイズするためのシステムの開発
 - ・ 問題解決およびオントロジーを選択・修正・作成するためのイン트라ネット / Web サービスの開発
 - ・ 知識要素を記述するための注釈言語およびツールの開発

6.3 セマンティック Web の応用システムとツール

● Sesame⁸⁴

オランダの Aidadministrator 社で開発された、RDF を格納するスキーマベースのレポジトリとその検索機能から成るツール。EU 委員会 IST プログラムの中の On-To-Knowledge プロジェクトの主要配布物の一つ。検索言語として、OQL スタイルの RQL を使用しており、データベースの内容をオントロジーとして取り出すことが可能である⁸⁵。

⁸⁴ <http://sesame.aidadministrator.nl/>

⁸⁵ INTAP「WWW10 調査報告書」より。

第7章 セマンティック Web の課題と今後の方向性

第7章セマンティック Web の課題と今後の方向性

7.1 セマンティック Web の課題

セマンティック Web を実現する課題として、以下のものがある。

- ・ オントロジーの設計・作成に関する課題
- ・ メタデータ付与に関する課題
- ・ 応用システムの開発・運用に関する課題

(1) オントロジーの設計・作成に関する課題

セマンティック Web において、オントロジーは、記述における情報の粒度を決め、それによってできるアプリケーションの範囲も決めるという意味で極めて重要な位置を占める。しかし、その構築における方法論の確立が第一の課題となる。

かつての CYC プロジェクトのように世界的に共通なオントロジーを構築するのは、非常に困難を極め、またできあがったものも汎用的すぎ直接アプリケーションで使えるというレベルにならない。また、Dublin Core が書誌情報における十数のエレメントを決めるだけでもあれだけ長い時間をかける必要があったことから見て、例え狭い業界であっても統一的なオントロジーの合意を得るとするのは難しい。そこで、おそらく、セマンティック Web では、以下の方針に基づいてオントロジーを構築するということが現実的になると考えられる。

- ・ ボトムアップアプローチ
- ・ 想定するアプリケーションに合わせたオントロジー
- ・ それぞれの文化に対応させたオントロジー

これは、一つのアプリで閉じている間は良いが、いざ他との「統合」をやろうと考えるとオントロジー変換が次なる課題となる。例えば、 n 個のボトムアップのオントロジー間のインターオペラビリティを実現するには、 nC_2 のオントロジー変換(コミュニケーションオントロジー)を行わなければならないのか、それとも機械翻訳におけるピボットのように共通オントロジーを作るのか、さらには別々の観点で作られた粒度の異なるオントロジーを対応させるのか、などといった変換における種々の問題が生じる。こうしたオントロジー変換における方法論の確立も重要な課題である。

また、グローバルな視点で見ると、各国の文化を反映させることも重要な課題である。現在、セマンティック Web の開発は欧米中心に行われ、日本はセマンティック Web の開発に参加していないし、何ら影響力を有していない。この為、セマンティック Web 関連の仕様は、日本文化を考慮したものになっていない。オントロジーの定義には、我が国の文化的背景を反映させることも必要である。今後は、日本もセマンティック Web の開発に積極的に関与することで、セマンティック Web の仕様に、我が国の事情を反映させる努力が必要である。

(2) メタデータ付与に関する問題

現在世界中の Web ページは 20 億～40 億と言われる。その全てにメタデータを付与する必

要はないとは言え、Web においては書誌情報に比して桁違いの数のメタデータをどうやって付与するかの方法論の確立が課題となる。対処方法としては、以下の方法が考えられる。

- ・自動付与
- ・利用者（とツール）による手動付与
- ・半自動付与

これらの方法は、どれを取っても決め手になるものではなく、一長一短がある。一番目の自動付与は、AI や自然言語処理技術における自動分類/抽出/翻訳といった技術をベースに行うことが考えられる。これは確かに数はこなせるが、精度を上げることは困難であろう。使える領域もあるかもしれないが、極めて限られたものとなろう。

二番目の利用者による手動付与は、数には数で対抗できる可能性のある方法である。知的作業の雇用促進を促す面もあるが、これまで以上にユーザインターフェイスの良いオーサリングツールの研究開発や、複数人の結果をマージすることでミスを減らすなどの管理手法が必要となる。

三番目の機械処理で半自動的に付与し、それを専門家がチェックして手を入れるというのは、おそらく最も可能性のある方法である。今後セマンティック Web 世界が充実されると、このような専門家はいくらでも足りなくなっていくかもしれない。教育体系や雇用体制の充実をはかるなどの方策が必要となる。また、分野の専門家と言っても、必ずしもセマンティック Web や RDF に堪能である必要はない。半自動付与システムの研究をさらに進めていく必要がある。

（３）応用システムの開発・運用に関する課題

セマンティック Web は手段であって目的ではない。それによってどれだけ素晴らしい世界になるかを、誰でもわかるような形で具体化して見せることが急務の課題である。

基本的にメタデータというのは、付与するのにそれなりの手間がかかるものだけに放っておいては普及するものではない。また、自分だけが付与することで何かが変わるという性格のものでもない。皆が付けることで皆が利益を得るという状態でないと、だれもが進んで付けるようにはならない。一方 HTML が短時間で流行したのは、HTML で簡単に綺麗なページを作ることができ、さらに多くの人に見て貰えて情報改訂の意欲を得るというポジティブなフィードバックができていたためである。セマンティック Web も、体制、運用としてポジティブなフィードバックをどう実現するかがまず課題である。

例えば、電子政府プロジェクトで、まずセマンティック Web データを役所などの公共情報に付与し、電子自治体などの基本的なアプリケーションを実現し、それによって誰でもが効果がわかるような仕組みを実現するというのも一案である。十分な効果があればメタデータを生成したり維持したりするために雇用創出効果が生まれ、セマンティック Web の普及にはずみがつくだろう。セマンティック Web の普及に向けては、メタデータをどうするか等を議論するよりも、まずこういった便利なサービスが実現されるのかから考えてい

く必要がある。これらの実証実験を行うことがまず必要である。

同時にこれらのシステム開発をする上でのツールや API の整備も急務となる。現状ではセマンティック Web を使う上で必要となる使い易いツールが少ない。使い易いツールが充分提供されていない場合、セマンティック Web のアプリケーションを開発しようとする基本ツールから手作りしなければならず非常に効率が悪い。例えば、RDF データ生成、RDF データ管理、RDF 検索、RDF 収集、オントロジー定義、オントロジー・ビジュアライゼーション、オントロジー・インポート、オントロジー自動生成などのツールが必要である。

さらに、API(アプリケーションプログラムインターフェイス)の仕様が決まっていないことも課題である。セマンティック Web のツールやアプリケーションを効率的かつ簡素に作るためには、セマンティック Web の各階層単位に標準 API を定義し、それに則って作ることが望ましい。セマンティック Web の標準 API として、RDF API、RDF スキーマ API、DAML+OIL API、DAML-S API、DAML-L API が必要である。しかし、現時点においては、セマンティック Web の API 仕様を開発する動きは無いように見える。

運用に関しては、セマンティック Web とメタデータ収集の課題を指摘しておきたい。現在、Web 検索エンジンの多くは、クローラー(Web ロボット)と呼ばれる収集プログラムが、継続的に Web 世界を巡回することで成立している。しかし、これは Web 文書量の増加と共に、巡回に要する時間の増加(リソースの更新頻度に巡回頻度が追いつかない)、完全収集の放棄といった方向に向かっている。セマンティック Web のメタデータについても、ボトムアップのものである限りは同様のことが起る。Web サービスについては、UDDI のようなディレクトリがある。セマンティック Web についても、メタデータ、スキーマ(オントロジー)を検索できる統一的なディレクトリが構築されることが望まれる。

最後に、信用と情報保護の課題について触れる。セマンティック Web は、メタデータをソフトウェア(このソフトウェアをエージェントと言う)で自動処理する。この為、偽のメタデータを検知する機構、適正なメタデータを適正なエージェントでのみ処理させる仕組み、エージェントによる不正アクセス防止、メタデータで記述された個人情報の漏洩防止などが必須である。しかし、これらの検討は、後回しになっている。セマンティック Web を実用的なものとするには、これらの検討を早急に行う必要がある。

7.2 セマンティック Web の今後の方向性

現在のインターネットの市場性は Web とその応用が中核となっている。この“第1世代 Web”によって確立された人間対人間インターフェイスが現在のインターネットの市場性を引き出し、有用性を確立したことになぞらえれば、次世代 Web となるべきセマンティック Web は、インテリジェンスとマシンリーダブルな特性を生かした応用で、第1世代 Web 以上のスケールの大きい革新をもたらすことが期待される。すなわちインターネット空間上で第1世代にはない知能的な機械対機械インターフェイスを実現することで、インターネットの有用性を革命的に高められる可能性がある。その意味で、セマンティック

Web の市場性を考察する場合、現 Web 応用の漸次的改良、強化にとらえず、より本質的な革新にとらえるべきである。

第 1 世代 Web の限界は、人間対人間インターフェイスにとどまっている点である。Web とブラウザとの組み合わせがインターネットの普及を爆発的に拡大させたが、その効用はあくまで特定、不特定個人同士の情報共有であり、応用としても広報、宣伝、広告、閲覧、掲示板、書類送付、文書交換など、人為的活動の補助的役割にとどまっている。いわば電話、Fax など旧来型通信手段のエンハンスメントとも言える。インターネットの効用が最も期待される企業活動への応用においても、大方は広告、宣伝、伝票のやり取りなどの従来的に行われてきた周辺プロセスの延長線、もしくは補助的手段にとどまっている。最近では e ビジネスや e-マーケットプレイスなど、企業活動の中核である生産、販売プロセスにも応用が展開されつつあるが、第 1 世代 Web の限界から本質的な経営革新には繋がっていない。

一方、セマンティック Web の革新性は、マシンリーダブルな特性とインテリジェンスを生かして分散化されたコンピュータノードを相互に接続し、自律型のコンピュータネットワークに発展させ得る可能性が挙げられる。そのような自律型コンピュータネットワークが企業経営を革命的に進化させ、新たな競争力の源泉につながられる可能性がある。企業活動にとどまらず、電子政府、福祉、遠隔医療、教育などインターネットノードを介する多くの社会活動においても同様のスケールの革新を生み出せるだろう。単純なデモンストレーションやアプリケーションに拠るのではなく、セマンティック Web を採用したシステムが現実目覚ましい社会的効果をあげることを実証して見せるアプローチが最有力だろう。それに最も適した応用分野として、例えば企業間 e ビジネス、もしくは広域の電子政府などが挙げられる。そう言った分野でセマンティック Web をベースにした共通プラットフォームを組み立て、その実用性、革新性を実証し、普及の原動力にするべきである。派生的な応用やデモンストレーションを繰り返すのではなく、より本格的な実証を目指すべきである。その意味でセマンティック Web の市場性、将来性に関しては、アカデミックな R&D だけに任せず、IT 産業界側の主体的コミットメントが求められる。

用語集

用語集

エージェント (Agent) : 場面に応じて意図を理解して自立的な判断に基づいた処理を実行する機能。次世代のマンマシン・インタフェースとして提案されている概念である。インターネットにおいても、様々なサービスを有効利用するための技術として提案されるようになった。ネットワーク上のサービスを分散配置されたエージェントととらえ、より高度なサービスを提供するというのがこの技術の目的である。

オントロジー (Ontology) : 概念どうしの関係や、それらを解釈するための規則の集合。規則は、RDF のような言語でかかれている。コンピュータプログラムは、リンクで示されたオントロジーを参照することにより、Web コンテンツの意味を理解できるようになる。また、典型的なオントロジーは、「分類体系」と「推論規則集」である。

デジタルデバイド (Digital Divide) : 情報格差。何らかの原因でインターネットの利用が阻害されたり、又は、何らかの原因でインターネットからの情報取得が出来ないために生ずる情報格差。

プラットフォーム (Platform) : コンピュータのオペレーティングシステム (OS) や仮想マシン言語 (例、Java) など基盤となるソフトウェアを示す。また、コンピュータハード及び OS が一体化された装置を示すこともある。

メタデータ (Meta Data) : データに関するデータ。例えば、PICS、CC/PP、RDF などがある。

CC/PP : CC/PP (Composite Capability/Preference Profiles) のプロフィールはその装置に提示されたコンテンツの適応をガイドするために使用することができる装置能力およびユーザプリファレンスの記述です。インターネットに接続された装置の数および種類が増大するとともに、異なる装置の能力に適合するコンテンツを伝える必要に対応する増加があります。HTTP「受理」ヘッダーおよび HTML<中高音>タグのようないくつかの制限のある技術は、既に存在します。コンテンツ適応などのフレームワークの一部として、ユーザエージェントの能力およびそのユーザのプリファレンスについて記述することができる一般的な目的プロフィール・フォーマットであることを目指しています。

DAML+OIL : 米国 DARPA(Defense Advanced Research Projects Agency)の資金提供を受けて開発された DAML(DARPA Agent Markup Language)と、ヨーロッパの研究機関を中心に開発された OIL(Ontology Inference Layer)とを統合したオントロジー記述言語。

RDF Schema をベースとして、論理的な規則の表現能力を向上させている。

DAML-S : Web 上で提供されるサービスのオントロジーを記述するための言語。オントロジー記述言語 DAML(DARPA Agent Markup Language)をベースとして、サービスの機能、振る舞い、およびアクセス方法などを記述することができる。コンピュータプログラムは、DAML-S で記述されたオントロジーを参照することにより、複数の Web 上のサービスを統合化したサービスを提供できるようになる。

DARPA (Defense Advanced Research Project Agency) : 米国国防省高等研究計画局。DARPA は米国国防省 (DoD) の中央研究開発機関であり、DoD 特定の基礎および応用研究開発プロジェクトを管理・監督している。

HTML : ハイパーテキスト・マークアップ言語。WWW のページを記述するための言語。画像や音声、ビデオなどを含むページを表現できる。SGML を基に策定された。現在の共通基盤の仕様は WWW の標準化団体 WWW コンソーシアム (W3C) が 1997 年 1 月に標準化した HTML 3.2。それ以前に、米 Microsoft 社や米 Netscape Communications 社などが個別に自社の WWW ブラウザ向けに仕様を独自拡張していた。

W3C は新版の HTML 4.0 を 97 年 12 月に標準化した。フレーム関連の仕様を追加したほか、タグを整理した。ISO (国際標準化機構) は、HTML 4.0 をベースに、より厳格なサブセットという形で、ISO/IEC 15545 の標準化を進めている。

HTML で記述されたファイルはテキスト形式。“<”と“>”で挟んだ予約語を使って、テキストの整形や、画像ファイルなどの存在位置、リンク先などを記述する。WWW サーバーが蓄積している HTML 形式のファイルを、WWW ブラウザが読み出して解釈・表示する。

また、SGML が XML に発展したのに合わせ、HTML 4.0 を XML 1.0 準拠のタグセットとして再定義した XHTML 1.0 が 99 年 2 月にワーキングドラフトとして W3C で公開された。

P3P : Platform for Privacy Preferences Project。W3C が開発中の、インターネットを含むネットワーク上のプライバシー保護を目的とした技術標準である。P3P を用いて、Web サイトはプライバシーポリシーを標準化された形式で、かつマシンリーダブルな形式 (XML 形式) で記述することができる。サイト利用者側では、P3P 対応のクライアントツールまたはブラウザによって、個人情報収集ページやクッキー利用ページにおいて Web サイトのプライバシーポリシーを参照したり、予め登録しておいたユーザプロフィールとポリシーとを照合して、個人情報を開示するか否かの判断を半自動で行ったりすることが可能になる。2000 年 12 月には P3P1.0 の Candidate Recommendation (勧告候補) が公開された。

PICS(Platform for Internet Content Selection) : W3C はインターネットの社会的責任を技術的に解決するために、1995 年夏から、規制なしでインターネットアクセスをコントロールするための技術標準の開発を進めてきた。PICS の特徴は、インターネットにおける情報発信を制限することなく、受信者が設定するレベルに合わせて、選択的に情報を受信 (フ

フィルタリング)できるようにするところにある。マイクロソフトのインターネットエクスプローラ 3.0 以降、またネットスケープのコミュニケーター 4.5 以降 (6.0 未満) には PICS 準拠のフィルタリング機能が組み込まれている。フィルタリングを実現するためには、そのようなソフトウェアやレイティング基準の他に、発信者が自分の発信情報に対して付加したレイティングデータか、または第三者が Web 上の様々な情報に対して付加したレイティングデータから成るレイティングデータベースが必要である。これらのレイティングデータを、フィルタリングソフトウェアが設定に基づいて参照することによって、受信者が受信する情報、または親や教師が子供たちに与える情報をコントロールすることを可能にする。

RDF : リソース・ディスクリプション・フレームワーク。インターネットのプッシュ型向けのデータ・フォーマットの標準。WWW コンテンツの配信、プログラムの配信を体系化し、統一された形で操作・管理できるようになる。

WWW コンテンツの配信には CDF (channel definition format) があり、プログラムの配信には OSD (open software description) があったものを、WWW コンソーシアムはネットスケープ・コミュニケーションズ社が提案した RDF を優先させて標準化を進めている。

RDF スキーマ (RDF Schema) : リソース・ディスクリプション・フレームワーク スキーマ。メタデータを処理するための基礎。ウェブについてのマシンリーダブルな情報を交換する相互運用を提供します。RDF は、ウェブ資源の記述を交換するために XML を使用します。また、ウェブ資源の自動処理を可能にする設備を持っている。例えば、よりよい検索エンジン能力を提供する資源発見で、コンテンツとコンテンツの関係についての記述のためにカタログを作る際、利用可能な特別のウェブサイトで、ページ、あるいはデジタル図書館、知識共有および交換を促進する知的なエージェントによって、単一の論理的な「ドキュメント」を表わすウェブ・ページの記述する知的所有権のために、またウェブサイトのプライバシーポリシーと同様にユーザのプライバシープリファレンスを示すことのために、ウェブサイトの デジタル署名を備えた RDF は、エレクトロニック・コマーс、共同作業および他の適用のために「信頼のウェブ」を構築することの鍵になります。

UDDI : Universal Description、Discovery、 and Integration。米 IBM、米アリバ、米マイクロソフトなどが共同で設立した、Web サービス・ディレクトリの推進団体、または、その標準仕様。前述の 3 社は、UDDI のディレクトリ・サービスを無償で提供している。ディレクトリ・サービスは WSDL。

URI : Universal Resource Identifiers。ネットワーク上のリソースを特定するためのコンパクトな文字列の仕様です。仕様を定めている RFC 2396 が、URI の文法と基本的な機能、利用のガイドラインを示しています。URL との違いは、識別名をもつリソースの永続的な標識化を第一の目的に掲げている点です。

URL : Uniform Resource Locator。インターネット上の各種情報リソースにアクセスする手段（使用する通信プロトコル）とリソースの名前をどのように指定するかを定めた規格。ある通信プロトコルを使って、インターネット上のサーバーが蓄積しているファイルを指定するには「プロトコル名://サーバー名/ファイル名」と記述する。指定できるプロトコルとしては、HTTP（http）、FTP（ftp）、NNTP（C_blanks）、Telnet（telnet）、およびローカル・ディスクへのアクセス（file）などがある。

WWW ブラウザは URL を指定する機能を備えている。例えば、WWW サーバー（www.nikkeibp.co.jp）のファイル（test.html）を読み出すには「http://www.nikkeibp.co.jp/test.html」と入力する。

W3C : W3C(World Wide Web Consortium)は、Web の相互運用性を向上させる共通のプロトコルを開発し、標準規格を提供する。MIT/LCS (マサチューセッツ工科大学コンピュータサイエンス研究所)、INRIA (フランス国立情報処理自動化研究所)、および慶應義塾大学 SFC 研究所がホスト機関として三極共同で運営している。

Web サービス : ウェブテクノロジーを使ったサービスの総称。分散アプリケーション連携技術の一つです。分散アプリケーション連携技術とは、あるアプリケーションからネットワーク上にある他のアプリケーションを呼び出して実行結果を返してもらう、という形で複数のアプリケーションを連携させる技術を示す。例えば、米 OMG の CORBA(Common Object Request Broker Architecture)などがある。

XML : エックテンシブル・マークアップ言語。HTML (hypertext markup language) に代わるものとして標準化作業が進むページ記述言語。HTML で普及したリンク（関連づけ）機能などを拡張するとともに、SGML (standard generalized markup language) をインターネット向けに最適化した。HTML と SGML の長所を併せ持つように定めたと考える。97 年 12 月に WWW コンソーシアムで標準化作業が完了した。

米マイクロソフトは XML をベースに、プッシュ型ソフト向けの情報配信フォーマット CDF (channel definition format) を規定し、Internet Explorer 4.0 で実現している。また、EC (エレクトロニック・コマース) などでのデータ交換の新たな枠組みとして「BizTalk」を 99 年 3 月に提唱、その基盤技術に XML を採用した。

放送分野では、2000 年末から放送衛星 (BS) を利用してはじまる BS デジタル放送で予定されているデータ放送サービスのコンテンツにテキスト形式の XML をベースにする方向で、99 年 10 月をメドに最終仕様が確定する。

また、HTML 4.0 を XML 1.0 準拠のタグセットとして再定義した XHTML 1.0 が公開されている。

XML 言語に関連して、XSL (extensible style language)、XML ハイパーリンク機能、XML DOM (document object model) などの仕様がある。

付録 1 良くある質問にたいする回答

付録 1 良くある質問に対する回答

Q1：セマンティック Web と AI とは、どう異なるのか？

A1：AI は、閉じた仮想的な世界での自動処理を指向し、その応用範囲は限られ、特定の人しか関与していないのに対し、セマンティック Web は、現実存在するオープンなウェブ世界での自動処理を対象にし、誰もが自由にあらゆる局面に活用できる。

Q2：セマンティック記述は、XML を使えば出来るので、RDF は必要ないのでは？

A2：XML は言語定義の為の言語、メタ言語であり、セマンティック記述を行うことも可能である。しかし、その記述の仕方は複数考えられ、その複数の記述を解釈し処理するパーサは複雑になる。当然、複数有るならば、どれか 1 つに決めようという事になるが、その結果、生み出されたものが RDF である。

RDF により、セマンティックの効率的な記述が可能になると共に、それを処理するパーサも簡単になる。

Q3：セマンティック Web はトップダウンアプローチでは？

A3：セマンティック Web はボトムアップアプローチです。セマンティック記述はそれぞれの人が好きなように行えば良く、それがどの位使われるかは、オープンな場での評価によって決定される。

Q4：オントロジーとは何？

A4：語彙の定義です。語彙の指し示す事(例：実体)が何か定義するものである。

[例] 風邪薬 = 総合感冒薬 = 新ルル・A錠

Q5：セマンティック Web は画一的なオントロジー定義を押し付けるのでは？

A5：それぞれの人がそれぞれのオントロジー定義を行う事ができる。同じ定義であっても、相反する定義であってもかまわない、また、異なるオントロジー定義を如何に変換するかは、利用者に任される。どのオントロジー定義が支配的になるかは、オープンな場での淘汰に任される。従来の標準化と異なり、用語を統一しなくても実体が同じか異なるか認識することが可能となる。

[例] 太郎さんの風邪薬 = 総合感冒薬 = 新ルル・A錠

花子さんの風邪薬 = 解熱鎮痛薬 = ナロンエース

Q6：それぞれの人がそれぞれのオントロジー定義を行うとすると、発散するのでは？

A6：ある 1 つのものに対し、色々なオントロジー定義が生ずる可能性がある。しかし、それらがウェブ上で公開された場合、実際に使われるのは、それらの中の限定されたものだけに成る。

この点は、実社会において、人それぞれが、色々な価値観やイメージを有していても、社会全体に影響を及ぼす価値観やイメージが、ある特定の人のものに限定されていることと同じである。

Q7：セマンティック Web は機械(ソフトウェア)にとって理解しやすくなるということですが、機械が読むのはメタデータだけで本文の方は人間が読まないといけないのか？

A7：メタデータでない HTML など記述されている本文を機械が読んで処理する必要はない、もっとも、機械がそれを読んで処理してはいけないとは、規定されていないので、其れが可能な機械は、読んで処理してもかまわない。

Q8：メタデータに本文の要点が記述されていれば、かなりのレベルまで機械が処理出来るようになる可能性があると思うが、メタデータに書く内容には制限があるのか？

A8：メタデータに書く内容に制限はない。何を、どれだけ書くかは、利用者の任意である。

Q9：本文の方も機械が読めるようにすることはできますか。その場合現在のウェブ情報のままだと可能か？それとも機械が読みやすいように本文の方も変更する必要があるのか？

A9：本文を変更する必要はない。本文を変更するか否かは、利用者の任意であり、メタデータの付与とは無関係の問題である。因みに、メタデータを Web 情報に付与する方法には、次の2つがある。

HTML データの中に、メタデータ識別タグ(例えば、<rdf：RDF >・・・</rdf：RDF >など)を付加して、そこにメタデータを記述する方法。

対象 Web 情報とは別にメタデータを作成しておき、何らかのポインター情報(例えば、URI)によってメタデータと Web 情報を対応付ける方法。

の方法の場合、全く Web 情報を変更する必要がない。

Q10：企業間取引のように特定された範囲ではセマンティック Web を導入しなくてもウェブの機械化が可能になるのでは？

A10：機械が読み、その意味を理解し、処理できるのなら、そのように言えると思う。しかし、問題は、今のままでは、機械がその意味を理解するのが難しいということである。従って、企業間取引のように限定された範囲で意味理解を可能にするに当たっては、何らかの意味表現方法の約束事か、標準を決め、更に、必要なツールを開発することになるであろう。そのような努力をするよりは、セマンティック Web を導入すれば、意味表現には、既の開発済みで基礎もしっかりしている RDF を用いる事ができ、色々なツールを流用できる。故に、セマンティック Web を用いた方が格段に効率的である。

Q11：セマンティック Web は 4～5 年先の技術では？

A11：セマンティック Web は、直ぐ使える技術である。

セマンティック Web の活用フェーズには、次の 3 フェーズが有る。

- ・フェーズ 1：メタデータに基づいて処理(例；PICS)
- ・フェーズ 2：オントロジーを用いた意味処理
- ・フェーズ 3：エージェントによる自動処理や推論

フェーズ 3 に到達するのは、かなり先だと思うが、フェーズ 1 は直ぐ活用でき、既に色々な分野で利用されている。

Q12：あいまいなオントロジーで問題ないのか？

A12：セマンティック Web は、あいまいなオントロジーを推奨しているわけではない。どんなオントロジーを定義するかは、利用者の任意なので、数あるオントロジー定義の中には、あいまいなオントロジーが生ずるかもしれない。

そのようなあいまいなオントロジーを利用していると、何らかの混乱が生じたり、矛盾が生じたりする可能性が、高いので、オープンな場で評価を行うと、使われなくなり、廃れてしまう。

この点は、実社会において、色々な理論が、提唱されたとしても、あいまいさの少ない、しっかりした理論が、評価されることと同じである。

Q13：セマンティック Web がそれ程素晴らしい技術なら、なぜ皆が使わないのか？

A13：次の 4 つの理由が考えられる。

セマンティック Web の認知度が低い。

特に我が国では、セマンティック Web 技術は全く知られていないに等しい。しかし、INTAP の普及活動により、今後、普及と活動とが、進むであろう。

セマンティック Web 関連の仕様が、固定されていない。

現時点では、RDF の基本仕様書のみが、確定しているにすぎないが、来年には、セマンティック Web 関連の仕様が、固定される予定である。

セマンティック Web のツールが整備されていない。

セマンティック Web 関連の仕様が、固定されていないことと相まって、セマンティック Web のツールの商用化が、進んでいない、来年、セマンティック Web 関連の仕様が、固定されれば、欧米においては、商用化されたツールが多数出現する見込みである。

RDF の形式に準拠したメタデータが、少ない。

セマンティック Web は、メタデータを処理することが前提となっている。

セマンティック Web は、メタデータの数が少ないとその効果を発揮する事が出来ない。

従って、メタデータを如何に生成するかがポイントとなる。

メタデータを生成するには、一般のインターネット利用者に自発的にメタデータを付けて貰うことが望ましいが、その為には、セマンティック Web の有効性を示し、メタデータを付けることのメリットを理解してもらう必要がある。

ここには、一種のデッドロックがあり、セマンティック Web の有効性実証実験等による、このデッドロックのループを解き放つ施策が必要となる。

これらの理由があるものの、行政、INTAP 及び W3C の支援や努力により、今後は、セマンティック Web の普及が進むであろう。

発行日 平成14年3月

編集・発行 財団法人 情報処理相互運用技術協会

〒113-6591 東京都文京区本駒込2-28-8

文京グリーンコートセンターオフィス 13F

TEL(03)5977-1301 FAX(03)5977-1302

印刷・製本 ホクエツ印刷株式会社

All rights reserved, Copyright © INTAP 2002

KEIRIN



この事業は、競輪の補助金を受けて実施したものです。