

セマンティック相互互換実用化コミュニティ
(SICoP : Semantic Interoperability Community of Practice)

セマンティック技術の概要とセマンティック Web の ビジョン

白書シリーズ モジュール 1
2005年2月16日 改定
バージョン 5.4

セマンティック技術の概要とセマンティック Web の ビジョン

エグゼクティブエディタと共同議長

Dr. Brand Niemann, U.S. EPA, Office of the CIO (SICoP Co-Chair)

Dr. Rick (Rodler F.) Morris, U.S. Army, Office of the CIO (SICoP Co-Chair)

Harriet J. Riofrio, Senior Staff Officer for Knowledge Management, Office of Assistant Secretary of Defense for Networks and Information Management, Deputy Chief Information Officer, Information Management (OASD NII DCIOIM), U.S. Department of Defense (KM.Gov Co-Chair)

Earl Carnes, Nuclear Industry Liaison, Environment, Safety & Health, Office of Regulatory Liaison, U.S. Department of Energy (KM.Gov Co-Chair)

エディタ管理

Jie-hong Morrison, Computer Technologies Consultants, Inc.

エディタ

Kenneth R. Fromm, Loomia Inc.

コピーエディタ

Michael J. Novak, Senior Analyst, Headquarters Office of Research, Internal Revenue Service

最貢献者

Kenneth R. Fromm, Loomia Inc.

Irene Polikoff, TopQuadrant, Inc.

Dr. Leo Obrst, The MITRE Corporation

Michael C. Daconta, Metadata Program Manager, Department of Homeland Security

Richard Murphy, U.S. General Services Administration

Jie-hong Morrison, Computer Technologies Consultants, Inc.

貢献者

Jeffrey T. Pollock, Network Inference Inc.

Ralph Hodgson, TopQuadrant, Inc.

Joram Borenstein, Unicorn Solutions, Inc.

Norma Draper, Northrop Grumman Mission Systems

Loren Osborn, Unicorn Solutions, Inc.

Adam Pease, Articulate Software Inc.

査閲者

Irene Polikoff, TopQuadrant, Inc.

Jeffrey T. Pollock, Network Inference, Inc.

Adam Pease, Articulate Software, Inc.

Dr. Yaser Bishr, ImageMatters LLC

Kathy M. Romero, U.S. Army Training and Doctrine Command, Futures Center

David Wood, Tucana Technologies, Inc.

注:この文書中の見解は貢献者個人のものであり、貢献者が属している組織の公式な見解や役職としての見解を示したものではない。

この文書の翻訳に当って

この文書は <http://colab.cim3.net/file/work/SICoP/WhitePaper/>を清水昇が翻訳したものです。当該文書は、本来米国の電子政府プロジェクト向けに作られたものですが、セマンティックWebの全貌を知るのに非常に有益だと思い、微力ながら翻訳する事を決意いたしました。本翻訳文書の中で、不明な点、或いは、おかしい点等あると思いますが、その様な場合、原文を読まれることをお勧めいたします。

謝辞

この文書の翻訳を快く許可して頂いた U.S. EPA の Dr. Brand Niemann に感謝申し上げます。また、翻訳許可を頂くに当って、仲介の労を取っていただいた米国メリーランド大学の Jim Hendler 教授と Mindswap Projects(及び Tucana Technologies, Inc.)の David Wood,氏に感謝致します。この翻訳文書を作るに当って協力して頂いた(株)サイバーエッジの足立和夫さんにも感謝致します。

注意事項

この翻訳文書の著作権は清水昇が所有します。
この文書を読んだことによるいかなる損害についても清水昇は責任を負いません。

2005年11月8日

以上、

目次

1.0	エクゼクティブサマリ	7
2.0	セマンティックコンピューティング序論.....	9
2.1	企業内に於ける意味矛盾	9
2.2	World Wide Web内に於ける意味問題	11
2.3	セマンティック技術の主な機能	11
2.4	セマンティック技術 対 セマンティックWeb技術	14
3.0	セマンティックWebのビジョン.....	14
3.1	何がセマンティックWebであり、何がセマンティックWebでないか.....	15
3.2	直近の利益	17
4.0	キー概念.....	18
4.1	より豊富なデータ、より柔軟な関係付け及び進化可能なスキーマ.....	18
4.2	データの形式	20
4.3	メタデータ	21
4.3.1	メタデータ標準.....	22
4.4	セマンティックモデル(タクソノミとオントロジ).....	23
4.4.1	オントロジ記述標準.....	27
5.0	コアビルディングブロック	28
5.1	セマンティックWebのウェディングケーキモデル	28
5.2	言語	29
5.2.1	XML(eXtensible markup Language).....	29
5.2.2	RDF(Resource Description Framework).....	29
5.2.3	OWL(Web Ontology Language).....	30
5.2.4	その他の言語開発の為の努力	31
6.0	セマンティックツールとコンポーネント	32
6.1	メタデータ生成と管理ツール.....	32
6.2	モデリングツール(オントロジの作成と変更).....	33
6.3	オントロジ.....	34
6.4	マッピングツール(オントロジの移植).....	35
6.5	データ格納	36
6.6	仲介エンジン	37
6.7	推論エンジン(Inference Engines).....	37
6.8	その他のコンポーネント	38
7.0	セマンティック技術のアプリケーション.....	38
7.1	セマンティックWebサービス.....	38
7.2	意味相互互換	39
7.3	知的検索.....	40

8.0 本資料以降の話題.....	41
9.0 参考資料.....	42
10.0 注釈.....	44
付録A: SICoPのチャーター.....	48
付録B: 用語集.....	49
付録C: 意味矛盾の種類.....	56

白書シリーズの概要説明

この白書は、連邦CIO評議会に於ける二つの作業グループである知識政府(KM.GOV <http://km.gov>)とセマンティックス相互互換実用化コミュニティ(SICoP)との活動成果である。この白書の目的は、セマンティック技術の概要とセマンティックWebのビジョンを理解してもらう事にある。これ等の技術は情報理論として進歩しつつあるが、未だ総てのIT問題に対する万能薬となるものではない。

この文書は、連邦政府、州政府及び地方政府に於けるデータ管理、情報管理および知識管理に責任を持っている部門エクゼクティブ、CIO、エンタープライズアーキテクト、IT 専門家及びプログラム管理者向けに作られたものである。

モジュール 1: セマンティック技術の概要とセマンティック Web のビジョン

この白書はセマンティック技術の原理および能力とセマンティック Web のゴールとを読者に理解してもらう事を狙いとしている。これは、セマンティックと言う概念を研究段階から実世界へと移らせつつある新しい標準、スキーマおよびツールに関する情報に沿って述べる事により意味処理分野に関する基礎知識を提供する。更に、より効率的な機械処理や最終的には、多くの新しく、且つ、強力な自動計算能力を実現可能ならしめる特定のシステムやアプリケーションに依存しないより豊富な用語を用いたデータ記述の方法に付いても説明する。

この白書は、政府機関や政府パートナーに対して直ぐ効果があると思われるセマンティック技術アプリケーションに焦点を絞っている。それらにはセマンティック Web サービス、情報相互互換および知的検索が含まれている。更に、セマンティック Web に至る道を成すプロトコル、スキーマおよびツールの状況や現在の利用状況に付いても説明する。

[コメント]: 我々は読者にセマンティック技術に対して深く理解する事、次世代のワールドワイドウェブに対して感謝する事、デジタル情報を処理する為のこれ等の新しいアプローチが、難しい情報共有問題を解決する為にどの様に用いられるのか分かる事を欲する。

1.0 エクゼクティブサマリ

セマンティック技術は次世代の Web すなわちセマンティック Web の推進源であり、スマートデータのマシンリーダーウェブと自動化されたサービスは、Web の現在の能力を遥かに超越して拡大する。

eGov コンファレンスに於けるセマンティック技術より(2003年9月8日)

セマンティック技術は連邦政府におけるより困難な情報技術課題の多くを解決できる可能性を有している。その一つの例は環境保護局であり、この局の主たる役割は、子供の生活環境をより良く改善する為に公共の健康データと環境データとを調和させる事である。子供は環境汚染に対して大人より非常に感受性が高いので、公共の組織は子供の環境品質と子供への影響に正しく関与する必要がある。インターネットへのアクセスビリティとその他の情報技術の進展に伴い、生活環境における危険に対する世論の懸念の増大は、公共組織や政府の組織が公衆衛生と環境に関する日々の最新の情報に対処する事が要求される事態になっている。しかし、これ等の期待は連邦政府の持っている現状の情報技術ツールやアーキテクチャでは充分対処できない。

この問題はリソースの問題ではない。現在の重要なリソースは子供に対する環境汚染の健康に対する危険情報をアクセスする為のデータの収集と分析に費やされている。政府の部局、機構および第三者の間の情報共有の現状に加えて蓄積されたデータに対する知的問合せ、推論および原因究明の為のツールの技術レベルは、これ等の期待に応えられるものではない。

公衆衛生と環境に関するデータは色々なソースから送られてくるが、そのソースの多くは互いに連携していない。語彙やデータの形式が良く分からないものであったり、不統一であったりする。特に組織を跨る場合(公衆衛生局 vs. 環境団体)、これが顕著である。データ値間の構造や関係性の場合、あるデータ集合から他のあるデータ集合へと一致させる事は難しい。これ等のデータセットを再評価したり、再編成したり、標準化したりする事は、時間が掛かり又誤りを生じる元凶となる。現在、この種のデータを横断的に知的検索したり、合理的に推論したりするツールは存在しない。

公正に見て、物理的なコンピュータの連動技術や大量データの高信頼且つ高度に安全な方法での交換する技術に関しては非常に大きな進歩が見られ、多くの有名ベンダが色々なデータベース、アプリケーション、ネットワークおよびコンピュータを結合する為の実績あるミドルウェアソリューションを提供している。しかし、これ等の技術はアプリケーションや色々な情報の塊を結合し、データを移動したり出来るけれども、情報システムを結合する場合における現実の難間に対処しているものではない。すなわち、IT システムを全面的に見直したり、もしくは、当該組織の運用方法の根本的変更をすること無しに、或るシステムが他のシステムの中の情報を、透過的、タイムリー且つ独立に利用したりする事を可能にするものではない。

これを可能にする為には、情報の論理的な形式変換が必要である。すなわち、或るシステムでは当該情報が何を意味し、そして、どの様に利用されるか、又、他のシステムの中で、当該情報が何を意味し、如何に利用されるか理解する事が必要とされる。これが、公衆衛生と環境とに関する利用可能なデータを合理的に利用する上での大きな障害である。

目標はシステムを結合させるだけではなく、それらのシステムの中のデータと情報とを相互に

機械的に処理可能にする事と人間が理解する事を可能にする為に、相互互換とアクセスとを可能にする事である。

この問題を解決する為の試みとして、米国環境保護庁(EPA : Environment Protection Agency)の中で、セマンティック技術を用いて、米国疾病予防管理センター(CDC : Disease Control and Prevention)のセンターの情報を EPA の情報と州のパートナーからの情報とを結合する一つのパイロットプログラムが進行中である。それは、EPA が「私の子供は環境毒素から守られているの?」と言った市民の質問に答えられるようにする事を目差すものである。(Sonntag 2003) このパイロットシステムの目的は本質的に、膨大な人間の考察知識を有するより拡張性の高い機能を首尾良く装備すると言う技術開発であるけれども、全国の数百万の子供の健康と自立とを改善する大いなる可能性を有している。毒素排出源の可能性を迅速に特定し、それによって影響を受ける場所と影響の程度とを知り、そして、効果的な環境浄化作業の優先付けの 3 つが、政府機関と産業界とこれ等の活動の支援者(子供、親、及び、社会のその他の総てのメンバ)との基本的優先事項である。

このストーリーは、連邦政府が直面している膨大な IT チャレンジの構想の一つである。連邦政府の複雑さ、連邦政府のデータ装置のサイズ、及び、州や地方や部族政府機関に結び付けられた連邦政府の組織、更には、増えつつある私企業や非政府組織(NGOs)に結び付けられた連邦政府の組織は、システム、アプリケーション及びデータを連携させた中で、必要なものをより早く見つけ、安価で且つより信頼性の高い方法を実現するよう圧力を受けている。政府部門及び第三者間およびそれらの中の情報の離れ小島を結合する事は、政府サービスの改善、財政と流通とを合理化し、複雑な機械の運用の信頼性の増大、人々の健康と福祉の推進、ネットワーク中心の防衛能力(net-centric defense capabilities)の実現、更には、我が国の安全を保障する為のキーステップであると見られている。

広範囲に亘る情報の相互互換性を実現できる事は、セマンティック技術に対して多くの研究者、思考リーダー(thought-leaders)及び実践技術者が抱いている利点の一つである。しかし、手段が無ければ、単なる利点でしかない。より表現力のある記述法、よりアクセスがし易く且つ自律的な情報の上部に構築される知的検索、知的推論及び真に適合可能なコンピューティングの実現が近づいていると見られている。

セマンティックコンピューティングの分野に於ける開拓者達は、既に数年に亘って仕事をしてきたが、2004 年の初めに W3C により二つの新しいプロトコルの承認がされたことは、セマンティック技術の商用化の為の重要なマイルストーンである。更に、セマンティック Web の目標に向って幾つかの開発が行なわれている。W3C の言葉に従うと、”セマンティック Web イニシャチブの目標は、データ交換の為の普遍的な媒体を形成する事で、現在の Web と同様な広がりを持つ。セマンティック Web とは一つの構想であり、機械により処理される事が出来る Web 上で定義されリンクされたデータを持つ事である。これは、表示の為だけのデータではなく、色々なアプリケーションに跨って自動化、統合化及びデータの再利用を可能にする為のものであり、それによって、情報の意味の力をフルに利用可能となる。”

情報技術に於けるこれ等の新しい機能は、初期の開拓者による重要な仕事と投資無しに獲られたものではない。セマンティックコンピューティングは、階層型データベースからリレーシ

ナルデータベースに、また、手続き型プログラミング技法からオブジェクト指向アプローチに移行した時と同様に、セマンティックベースのアプローチの意味や機構を人々が理解するには多少時間を要するであろう。しかし、何れ人々はこれ等の新しい技術やアプローチの力を理解するに違いなく、技術革新の第一段階で既存の IT 問題領域の多くに対して目覚ましい効果をもたらすであろう。技術革新の継続は、ユーザ間、システム間及びネットワーク間での情報共有と交換の為の基盤に於ける劇的な能力向上をもたらすであろう。(Fromm and Pollock 2004)

数年のレンジで見ると、これ等の技術革新は、メインフレーム、パーソナルコンピュータ、イーサネット及びワールドワイドウェブの初期のものをもたらしたのと同様なコンピューティングに於ける新たな波を起こすことが約束されている。

2.0 セマンティックコンピューティング序論

人々は情報の方がソフトウェアより長生きする事を認識し始めている。 **Tim Berners-Lee**
情報の持つ意味は、その情報の最初のユーザもしくはアプリケーションに結び付き過ぎている。従って、(a) マシンによる情報の再利用、もしくは、(b) 人間による概念(単なる用語ではない)問合せ、の何れも非常に難しい。 **Jeffrey T. Pollock**

政府が直面しているデータ管理課題に対するより良い情報技術ソリューションの必要事項を明示する事は難しい事ではない。情報共有は一つの例である。連邦、州及び地方機関で(法の執行、海上輸送、環境保護、青少年支援、公衆衛生または本土防衛の何れかの為に)持っている情報の共有と有効化の課題は気の遠くなるものである。それらの機関は共通語彙標準を作成する為に多くの時間と費用とを支出しており、その結果、システムインテグレータはこれ等の標準を作ったり、付加したりする為に、データの持ち主に対して骨の折れる作業を行なっている。不幸にも、このアプローチは(例え、標準語彙作成段階になったとしても)、実装の遅延や問題を直ぐ生じる。本質的に異なるソース間での情報共有の真の課題は、共通言語を作る事ではなく、特別の語彙標準を作ったり付加したりする事を殆ど総ての場合に阻害している「組織と文化に於ける違いを処理すること」である。(Fromm and Pollock, 2004)

2.1 企業内に於ける意味矛盾

組織に於ける IT システムに内在する構造的違いや文化的違いは、その独自のミッション、階層、語彙、作業手順及び作業形態を反映している。あるシステムでは”価格”とされているものが他のシステムでは”コスト”となっていたり、大尉の事を陸軍では”キャプテン”と呼び、海軍では”ルーテナント”と呼び、海軍の”キャプテン”は陸軍の大佐(コロネル)に対応している。(この種の違いは、軍事関係に留まらず、多くの州警察組織は海兵隊をモデルにした階級を使っており、多くの公衆衛生局は海軍をモデルにした階級制度を用い、多くの警察や調査部門は独自の命令構造を有している。) 同様に、法執行組織に於ける”情報提供者”を諜報機関では”情報源”と言う用語で呼ばれるかも知れない。(後者は人間以外のソースを含む可能性がある。) これ等は名称付けにおける比較的簡単な違いの例である。

より複雑且つ抽象的な概念のものがあり、それらは、構文や構造の違いであり、より重大なものに意味の違いがある。システム開発者及び(また)情報モデラに関係する問題に名前の違いが概念の違いに起因するものか、意味の違いに起因するものか決めなければならない問題があ

る。名前の違いは、ルックアップテーブルやシソーリ等の既存のツールを用いて比較的簡単に処理する事ができるが、概念の違いや定義の違いの場合は、その意味のより深い洞察が必要である。

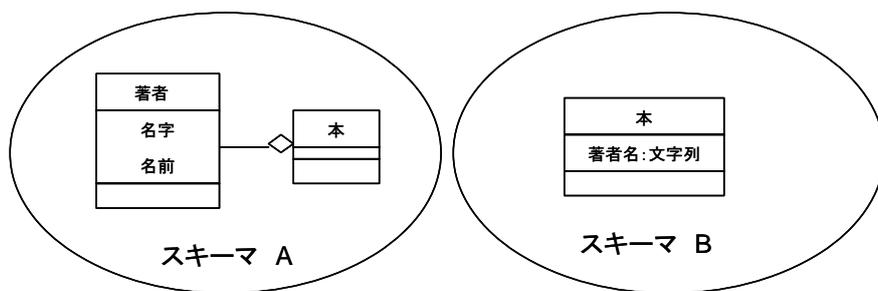
例えば、異なるシステムが異なる概念またはバリーチェーンに於ける異なるステージを示すのに同一の用語を用いる事がある。多くのシステムでは”コスト”と言う用語は、消費者が購入する場合の”価格”を示し、同時に、他のシステムの中で用いられる用語”コスト”は、サプライヤが卸売業者にある品物を販売する場合の”価格”を指し示している事がある。

意味は時間と共に変化するものもある。個人的な変化、組織の履歴、組織の方針/文化及び法人運営の為の命令などが時間と共に意味の変更を来たす契機と成り得る。(そのような変化が生じる事を明示する事無く使われている。) この様な複雑な矛盾に対処する為には、より大規模なセマンティックベースのソリューションを必要とする。

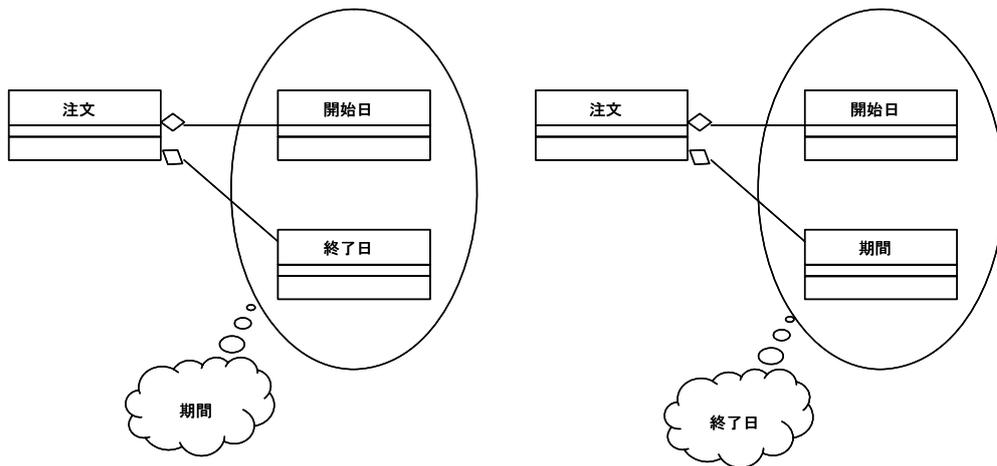
図1は色々なデータセットを連携させた場合に於ける意味矛盾の二つの例を示している。

(付録Cには意味矛盾の表がある。)

この様な矛盾は複数のデータセットを連携させる場合に通常存在し、多くの場合、別々の開発、ニーズの変更、組織や構造の違い等々に起因してデータモデリングの副産物として発生する。



データ収集に於ける矛盾 - 構造の違い



値表現の矛盾

図1 意味矛盾のタイプ

(Pollock and Hodgson,2004 より引用)

構文的、構造的及び意味的な矛盾問題は、企業及び政府機関の両方の中で大きな問題になって来ている。

メッセージングや転送ソリューションは益々一般化し商品化され、XMLがデータ交換の為の

基本構築要素となっており、これらは部分的に完璧な為、多くのものが容易に実現可能な様に見える。更なる技術が、組織と個人的意図との間の処理と情報セットを、点から点への個別データや用語のマッピングや同時プロセスなどを必要としないで効率的かつ合理的に処理する為が必要とされる。セマンティック技術で実現を保証しているものに論理的な言語を使える事がある。

それにより、より明確にデータの意味と構造とを明らかにする事ができるので、用語や定義に違いや矛盾があっても、ソフトウェアがうまく処理できるようになる。

2.2 World Wide Web 内に於ける意味問題

意味問題は、色々な組織やそれらのビジネスや運営協力者の内外にのみ存在するのでは無く、World Wide Web 上に色々な形式で存在している。

ウェブ上の情報は、日々断片化が進み、適切さ、時宜の度合い及び信頼度は変化する。

サーチエンジンは、素晴らしいツールであるが、正しく検索できないケースがどんどん増加している。これらの失敗したケースの存在は検索の完全性に問題のある事を証明し、“filetype”の様なサーチコマンドライクのスクリプトの利用の増大をもたらし、RSS を用いた送付、イメージ又は音楽などの特定のタイプのデータやメディアに特化したサーチエンジンを出現させた。

連邦政府や州政府は、情報を公共のオンラインに利用できるように巨額の予算を費やしたが、それでもまだ World Wide Web の現状は、これ等の情報の利用や応用に於ける制約要因となっている。

2.3 セマンティック技術の主な機能

幸運にも、丁度インターネットや World Wide Web のプロトコルが人間の(利用の)為に、膨大な情報を結び付ける手助けをしたように、今度はマシン操作やマシン処理の為に同等かそれ以上の情報を結び付ける新しいアプローチが現れつつある。これらの技術革新は情報の相互互換を簡単にし、企業内や World Wide Web でのよりよい情報連携と信頼性を提供する。徐々にこの技術革新は、新しい知的仲介機能や収集された情報のフィールドに跨った知的推論機能への道を拓くであろう。図 2 にセマンティックコンピューティングの主要能力とステークホルダに及ぼす影響とを示す。

能力	狙い	ステークホルダ	影響	もたらすもの
短期				
セマンティック Web サービス	柔軟なサービス検索、サービス発見及びスキーマ変換	システム開発者及びシステムインテグレータ	Web サービスの付加と配備に於ける摩擦の軽減	より自動化された柔軟なデータの連携
情報統合及び/または相互互換	情報統合数を n^2 個から n 個へ削減	データ及びメタデータアーキテクト	ヘテロなデータソースを統合するコストの削減	素早い相互互換性の増大とコスト削減
知的検索	文脈を踏まえた検索、概念問合せ及び利用者フィルタリング	ビジネス及び技術管理者、アナリスト及び個人	検索結果の人間フィルタリングの減少、より意味のある検索	検索精度の向上はより高い生産性をもたらす。

能力	狙い	ステークホルダ	影響	もたらすもの
長期				
モデルドリブンアプリケーション	アクションナブルモデルからドメインロジックを処理する事を可能にする。	ソフトウェア開発者	ドメインロジックの素早い変更とコーディングの削減	コードメンテの減少と要求に対する迅速な変更
アダプティブ及び自動コンピューティング	診断アプリケーション及び予測システム管理に必要な機能の提供	システム管理者	複雑なシステムの自己診断とプランニングによる信頼性向上、コスト削減	システム保守及び人間の介入を少なくする事によるコスト低減
知的推論	豊富なデータと進化可能なスキーマとに基づく機械推論のサポート	アプリケーション及びコグニティブエージェント	埋め込みロジックに対する要求の削減とドメインモデルから制約の分離	アプリケーション開発コストの低減

図 2: コンピューティング能力の調査

(Richard Murphy による)

最近、W3C で承認された二つの新しいデータと論理構造は、情報をより豊富にし、より自律処理可能にし、そして最終的には、さらによりアクセスしやすく且つ利用しやすくすることを可能にしている。

この新たな構成要素—**Resource Description Framework (RDF)**と **Web Ontology Language (OWL)**—は、既存の W3C のマークアップランゲージと互換性を持たせつつ機能を付加し、知識表現規則の為の利用拡張を行っている。RDF はデータ間の関係を確立する為のフレームワークを提供し、一方、OWL は RDF を拡張して、異なったデータ要素間の特別な制約や他のデータとの関係を指定する事を可能にしている。これらの標準(その標準を支えるために作られた新しいツールや基盤要素を含む)は、企業内のアダプティブコンピューティングの開発を牽引するだけでなく次世代のウェブの発展をも牽引し、セマンティック Web と呼ばれる

セマンティック Web のビジョンは、インターネット上の転送情報やアクセス情報に良く定義された意味を付加する事により現在のウェブを拡張する事である。従って、コンピュータは、データが高度に関連した情報や知識になる様に、必要なデータ処理や組み立て処理をより多く出来るようになる。言い換えると、セマンティック Web の独創的な点は、データの関係とカテゴリーとを改善する為に必要なプロトコルと技術とを確立している事であり、それにより、異なったシステム間やデータ間の関係を生成したり推論をしたりする能力を強化している。

例えば、図書館員に南北戦争当時のゲティスバーグの地図を要求すると、おそらくその時代の

地図を含んだ本が出てくるだろう。しかし、検索エンジンで検索すると、「ゲティスバーグの地図」と言うテキスト情報を含む多くの結果を得られるであろう。しかし、これらの結果には実際の地図を含んでいるかもしれないし、そうでないかもしれない。加えて、検索文字列の中に指定された日付の形式と一致していない引用資料は得られないかも知れない。

同様に、ワシントン DC 地区のネットワークセキュリティイベントについての検索を行った場合、バージニア州(VA)の McLean で行われたアンチスパムメールに関する講演は得る事ができない。何故なら、ネットワークセキュリティとアンチスパムメールの関係とワシントン DC 地区在るバージニア州(VA)の McLean の概念とは、World Wide Web の領域の中では、まだ基本的に結び付けられていないからである。W3C による取り組みは、このようなデータの関連性と集合全体の理解とのギャップを埋めることを目指している。

XML と XML に関連して開発されたデータ記法とを基盤としている RDF と OWL とは、ウェブの諸々のツールや World Wide Web の中に組み入れられ始めている。図 3 は、W3C のどの標準が構文、構造及びデータや情報の意味記述に対応しており、そして、どの様にセマンティックコンピューティングのスペクトラムをカバーしているか示している。

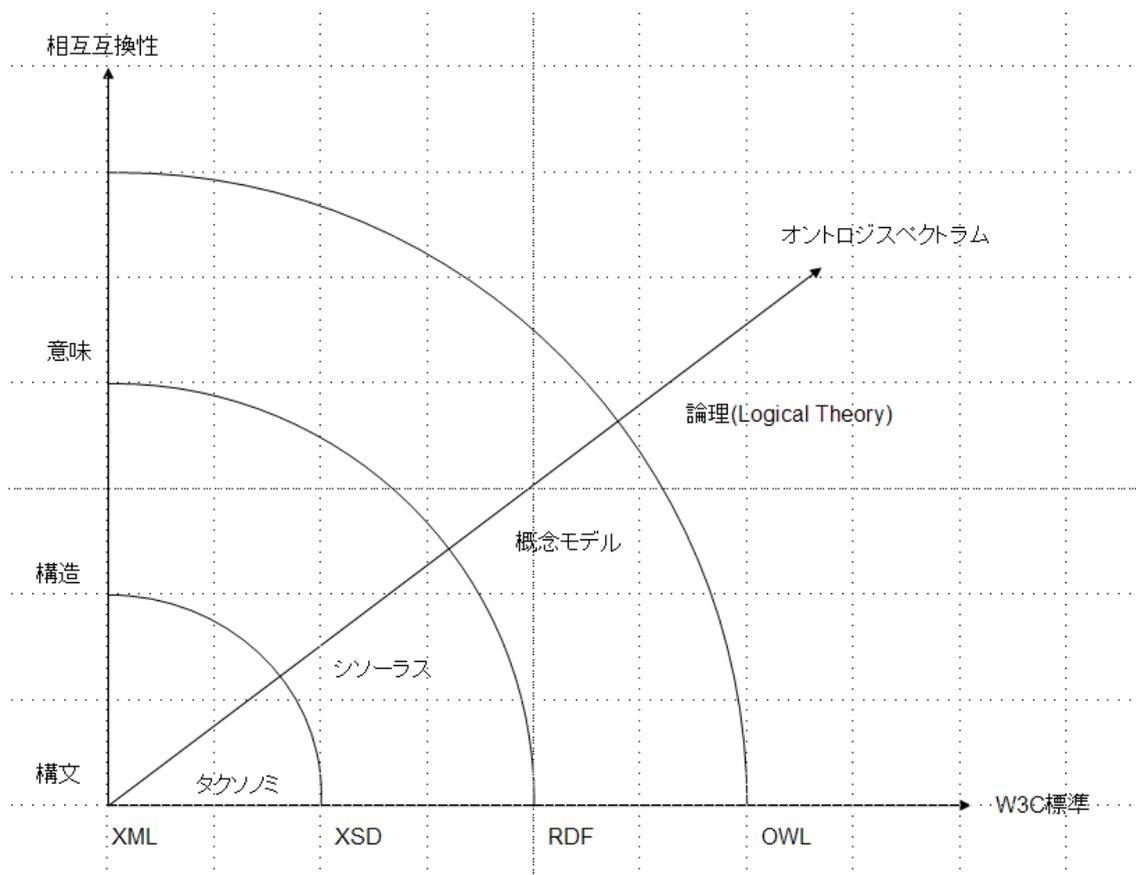


図 3: セマンティックコンピューティングの三つの次元

(Richard Murphy により、Daconta, Obrst, Smith 2003 から引用)

これらのデータと論理構造の定義に加え、W3C は既存アプリケーションやデータセットと連携したセマンティックソリューションを実装する為に必要とされる基礎的なアーキテクチャと論理とを定義している。多くの会社が既にセマンティックアプローチとセマンティック Web の

ビジョンを採用しており、さらにその分野を発展させる技術戦略を積極的に迫及している。
(Pollock and Hodgson,2004)

これらの技術やアプローチは、先進的な利用者に使われつつあり、そして、その様な利用者は増えつつある。初期のアプリケーション開発事例は、徐々に実装可能である事や投資収益率(ROD)が期待に沿うものである事をはっきり証明している。いくつかの政府機関では、W3Cのこれ等の新しい標準を用いて対象領域を絞った挑戦的な複雑問題を扱うパイロットプログラムを計画したり実行に着手したりしている。本白書とセットになっている他の報告書にこれらの機関とプログラムに関する詳細が紹介されている。

2.4 セマンティック技術 対 セマンティック Web 技術

この資料ではセマンティック Web の説明に先立って多くのセマンティック技術の説明を行っている。(時には、これ等の技術は「セマンティック」と言う用語を使っていないこともある。)

それらの幾つかは、人工知能に起源を有し、またあるものはマークアップ言語の拡張であり、またあるものは、EAI(Enterprise Application Integration)から論理的に派生したものであったりする。データの集合と論理言語との形成に関する W3C の努力は、多くの分野や技術を共通の基盤に乗せる際の重要な触媒になっている。しかし、全てのセマンティック技術が、W3C の承認した言語とフレームワークを利用している訳ではないので本書では、本質的にセマンティックの概念を利用する技術である”セマンティック技術”と W3C の勧告に完全に準拠した技術である“セマンティック Web 技術”とを区別している。本資料の中でしばしば用いられている「セマンティック技術」と言う用語は、セマンティックに関する学問分野の広義の議論の為にのみ用いるのではなく、数年の内に実用化されるであろうセマンティック Web 技術とを分かり易く区別する為にも用いられている。

3.0 セマンティック Web のビジョン

セマンティック Web は現在の Web の発展形であり、そこでは情報はわかりやすく定義された意味を持ち、コンピュータや人間の共同作業を可能にする。

Tim Berners-Lee, James Hendler, Ora Lassila(2001)

W3C によれば、データが、人間だけでなく自動化されたツールによっても共有でき、処理でき更には理解可能になった時に初めて、Web はその潜在能力の極限に到達できる。将来のプログラムは、ウェブのスケールで、たとえ夫々独立して設計されたものであってもデータを共有でき、処理でき、また、理解できなければならない。

定義段階であるにもかかわらず、セマンティック Web という用語は、IT 業界の人のみならず多くの人々にとって多分まだ聞き慣れない言葉であろう。しかし、それが対象としている問題は、この数十年の間解決するのに苦闘していたものである。

その問題の例として、情報過多問題、ストーブパイプシステム問題、低品質収集(poor content aggregation)問題などがある。(Daconta, Orbst, and Smith, 2003)

これらの問題の根本的原因は、個々のシステムでのセマンティック定義の欠如、データセット間のセマンティック統合の欠如、本質的に異なるシステム間における意味の相互互換性の欠如で

ある。セマンティック Web は、現在の Web の能力及び既存の情報技術を超えた領域を有し、それによって、より効率的な共同作業やより賢い意思決定が可能となる。セマンティック Web は、セマンティック技術や概念基盤やより良く理解可能な相互操作の約束規則を用いてアクセス可能な知的ウェブサイトとデータストアとの集合であり、それにより、サービス要求に対して応答する為により多くの仕事を機械が行う事(機械的な検索手順、より関連性があり信頼度の高い情報の提供、知的推論の実行、又は、知的仲介(brokering)の実行等)を可能にする。

図4はセマンティック Web の概念図であり、セマンティック技術を既存の Web にどのように付加したか示している。

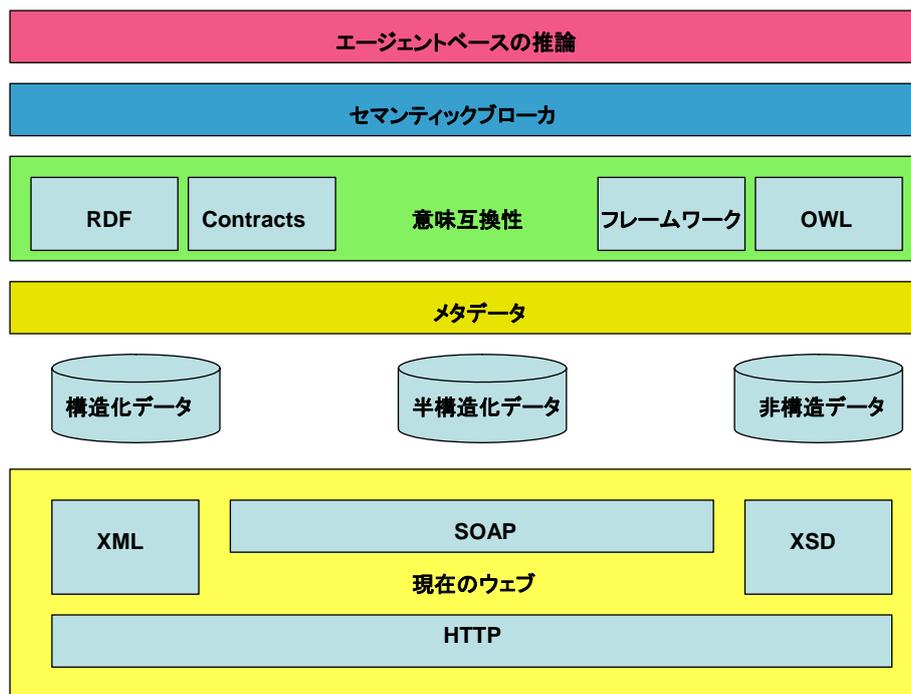


図4: セマンティック Web の概念階層

セマンティック Web のこのような構想は、数年で実現できるものではないが、確かな事は、低コストあるいは無料のウェブサーバ及びグラフィカルなブラウザが利用できるようになった後、沢山のウェブサイトが出現した様に何等かの要因で急速な進展が起こるのであろう。しかし、そのウェブサイトの開発の進行状況は、6年に亘って比較的混沌としたものであった。スクリプト言語の付加から始まり、末端の利用者用ツール、カスタムビルドのサーバコンポーネントそして比較的整然と進化したコア言語の統一セット、アプリケーションサーバ、コンテンツ管理システム、e-コマースエンジン、Web サービス及びその他の企業にとって価値のあるコンポーネントや提供物などがあった。

セマンティック Web の成長は、市場ダイナミクスの中の進化を通じて行われるであろう。セマンティック Web の互いに結合された世界のビジネスモデルは、より理解され、先端ウェブ技術の認知度は高くなるだろうが、その構想の多くの部分が組み立てられるまでには、タイムラグがあるであろう。

3.1 何がセマンティック Web であり、何がセマンティック Web でないか

1.セマンティック Web は新しいものではなく、また、別のウェブサイトの集合でもない。

セマンティック Web は、現在の World Wide Web の拡張(発展形)であり、新しい別のウェブサイトの集合ではない。セマンティック Web は、現在の World Wide Web の機構とトポロジとの上に構築されており、より豊富な意味の関係を表す為の標準と機械処理可能なデータとを定義することにより、より高い能力を実現している。検索エンジンやスパイダサーチ技術や解析スクリプトのような外部処理がより容易にアクセスできるようにする為にウェブページに情報を記述するのに、既存のサイトはこれらの仕組みを利用するかも知れない。さらに、新しいデータ格納機構(多くのデータベースを含むが)がマシン処理をしたり、されたりする事が可能になり、種々の構文、構造および意味を統合した質問や結果を生成する様に高度化出来る。セマンティック Web で使われるプロトコルは、現在の World Wide Web を実現している既存の色々な技術に対して透過な手段である。

2.セマンティック Web は人間からのアクセスだけを考慮して作られたものではない。

現在の Web は、非常に粗雑なレベルの情報を構造化し相互結合する為のテキストマークアップとリンクプロトコルとに本質的に関連している。そのプロトコルは主として、人間が利用するのに適した形式で文書を記述したりリンクしたりするために使用される(しかし、それは基本的な機械検索と収集とに対する有用な手掛かりとなっている)。セマンティック Web のプロトコルは、より洗練されたレベルで情報を定義し結合している。意味が機械により、容易に理解や処理し易い形式で記述される事により、データ格納機構間の構造や意味の違いを超えてブリッジ可能となる。この抽象化とアクセス性の増大は、現在の Web 能力が増し、また、拡張される事を意味している—そして、新しく強力なものになるであろう。

3.セマンティック Web は、急進的な未実証の情報理論の上に構築されたものではない。

セマンティック Web の出現は、広く認知された情報理論の中の自然な成り行きであり、その情報理論は、知識表現や知識管理からのコンセプトと World Wide Web コミュニティにおける改訂構想に起因するコンセプトとを借用しているものである。新たに承認されたプロトコルは、コンピュータ言語、情報理論、データベース管理、モデル指向の設計技法及び論理学における多くの熟達した人々のアイデアを数年にわたって具体化したものである。これらのコンセプトは、多くの現実世界の場面で実証がなされているものであり、W3C の統一的な標準は、企業の中や Web 上で広く採用され、また、その利用が加速する事を約束している。

知識表現上の問題と未だ実現されてない構想に関して、歴史を振り返ると、あるコンセプトが受け入れられる為には、大勢を決定付けるのに統一的な標準が必要である事を多くの例が示している。HTML は SGML から派生したものであり、唯一のある程度普及したテキストマークアップ言語であるが、まだ情報技術の利用において大変化をもたらすに至っていない。この分野の多くは、オブジェクト指向プログラミングと物理概念プログラミングモデル (conceptual-to-physical programming models) とが受け入れられるのに長い時間を要した事を示している。Ralph Hodgson によれば、“知識表現は基盤的学問分野であり、現在それは、研究所内から現実世界の利用へ移す為に必要なインフラとサポート標準のセットを有している”。³

4.セマンティック Web は現在のデータモデリングの概念から大きく逸脱するものではない。

Tim Berners-Lee によると、セマンティック Web のデータモデルは関係データベースモデル

と似ている。“関係データベースは列もしくはレコードから構成される表である。各レコードはフィールドの集合から成っている。RDF がプロパティ値間のコネクションだけの様にそのレコードはフィールドのみからなる。その間のマッピングは大変容易であり、一つのレコードは RDF の一つのノードとなり、フィールド(カラム)名は RDF のプロパティタイプとなり、且つ、レコードフィールド(テーブルセル)は一つの値となる。実際、セマンティック Web の主たるドライビングフォースの一つは、マシン処理可能な方法で関係データベースの膨大な情報をウェブ上で表現する事である”(Berners-Lee,1998) セマンティック Web がより高い表現力を持ち、包括的で、また、データモデリングの強力な形式を有する事を意味し、それは伝統的なデータモデリング技法(実体関連モデル[ER モデル]等)の上に構築され、より厳密に理解可能な方法で色々な関係を表現する為のより強力な方式にそれらを転換する。

5.セマンティック Web は人工知能(AI)を用いた魔法ではない。

機械が解読可能な文書というコンセプトは、人間がももごと話す事を機械に理解可能にさせる魔法のような人工知能を意味するものではない。それはただ、実在する良く定義されたデータを用いて良く定義された操作を行なう事により、良く定義された問題を解く機械の機能を示しているに過ぎない。(Berners-Lee, Handler, and Lassila, 2001) 現在の検索エンジンは 20 年前には魔法の様に見られた機能を実行している、そして我々は、IP プロトコル,HTML,ウェブサイトの概念、ウェブページ、リンク、画像対応ブラウザ、革新的検索、ランキングアルゴリズムそして勿論非常に多くの信じ難いほどの高速なサーバと巨大で高速のディスクストレージ等も同じ様に認識している。セマンティック Web の機能も、標準とアプローチとの共通の基盤の上に構築された知識表現と情報技術とにおける継続的な進化の論理的な結果と同様なものとなるであろう。

6.セマンティック Web の利用可能な実物は未だ存在しない。

現在セマンティック Web は、有望的かつ魅惑的なものだが、一つのビジョンである。現在の Web と同様に、セマンティック Web は、オープンな標準と固有のプロトコル、フレームワーク、技術及びサービスの組み合わせにより形成されるであろう。W3C が承認した標準である XML、RDF 及び OWL はセマンティック Web の基本プロトコルを構成する。これらの新しいプロトコルを使って構築された新しいデータスキーマや連携メカニズムは、同じ興味や産業や目的のコミュニティの周辺で開発され、それらの内のあるものは、熟練したデータアーキテクトによって注意深く設計され、そして、権威ある標準化機関により公式なものと承認され、またあるものは、どこからともなく現れて、突然広く受け入れられるであろう。新技術や新サービスの採用は、意味を意識したコンテンツ発信ツール、コンテンツモデル化ツール、仲介し推論する評判の良いエンジン、データの浄化やシソーラスサービス、新しい監査及び検証コンポーネントなどで行われるであろう。しかし、セマンティック Web のビジョンの色々な要素は既に存在しているが、これ等の技術の拡散や競合機能との調整やそのビジョンを全体的にカバーするには数年を要するであろう。

3.2 直近の利益

セマンティック Web のビジョンの完全な実現は少し先になるかも知れないが、直ぐにでも、その機能を活用する事により、エンタープライズソフトウェアをより連携し易くし、相互互換を

容易にし、色々なものに付加し易くし、劇的に保守費用を低減できる。既存あるいは新しいセマンティックベースのスキーマやツールと組み合わせたセマンティックアプローチの利用は、ITを率先して導入している多くの企業や政府機関に直ぐもしくは近い将来に利益をもたらす事ができる。

例えば、意味相互互換性は、セマンティック Web のビジョンの部分または限定されたものであるが、情報共有及び情報の相互互換の領域で特定の目的の為に比較的限定され且つ明白な領域の中で、構造や意味や文脈の仲介や調整にセマンティックベースのツールを用いれば、目覚しい成果を得る事ができるであろう。言い換えると、セマンティックインターオペラビリティとは、より明白に定義されたエンドポイントでより離散的な問題集合を扱うものである。(Pollock and Hodgson, 2004)

セマンティック技術は、既存の Web サービスと XML フレームワークの上に、より疎結合な被いを提供でき、現在利用可能なものよりも遥に色々なものを付加する機能を提供可能である。そして、それらを、サービス発見や調整や異なる語彙を持つレスポンスや要求のネゴシエーションを容易にする為に直ぐに取り込む事ができる。問題の深さや難しさを考えた場合、連邦、州及び地方の機関はこれらを注視すべきであり、セマンティック技術は、初めて柔軟で、オープンで且つ包括的な手段を提供できるかも知れない。

4.0 キー概念

少量のセマンティックスであつても長い道のりを歩んでいる。

James Hendler

セマンティックコンピューティングは、ここで述べられているように形成され整形されつつある新たな学問分野である。多くの異なる定義と解釈とが存在し、低強度哲学論争(low-intensity philosophical wars)が思考リーダーや実装者の間で多少行われているが、今年の初めに W3C の推奨仕様として RDF 及び OWL をリリースした事は、非常に多大な一般化をもたらしたと言われている。

セマンティックコンピューティングとは、色々な形態の抽象的且つ論理的な表現の利用を可能にするものなので、RDF と OWL とが、前節の中で説明した強力な諸機能をどの様に提供されるべきか示すのは難しい。しかし、インターネットと WWW が、プロトコル層と諸技術の上に構築されているとすると、セマンティック Web も同じである。幾つかのキー概念を理解する事及びセマンティック Web のコアとなる構築ブロックを良く知る事は、どの様に、より高次のツール、コンポーネント及び技術がより柔軟で且つ豊富な機械処理可能なデータの利点をもたらす事が出来るかを見通す為の基礎となる。土台を形成する概念の幾つかを理解する事は、読者がこの技術の現状をより良く理解する事にもなり、セマンティック Web のビジョンを完全に実現する為に未だ改善すべき領域が何かを理解する事にもなる。

4.1 より豊富なデータ、より柔軟な関係付け及び進化可能なスキーマ

セマンティック技術は、データベーススキーマやデータ辞書や統制語とはある重要な点で異なる。それは、異なる概念領域が一つのネットワークとして一緒に働くように、関連性を持って設計されている事である。図 5 に示された地下鉄の地図は、どのようにして、各概念が結ばれているか、あるいは、関係又は関係しない概念と構成されているかを示す正当なセマンティック Web ダイアグラムの一つである。

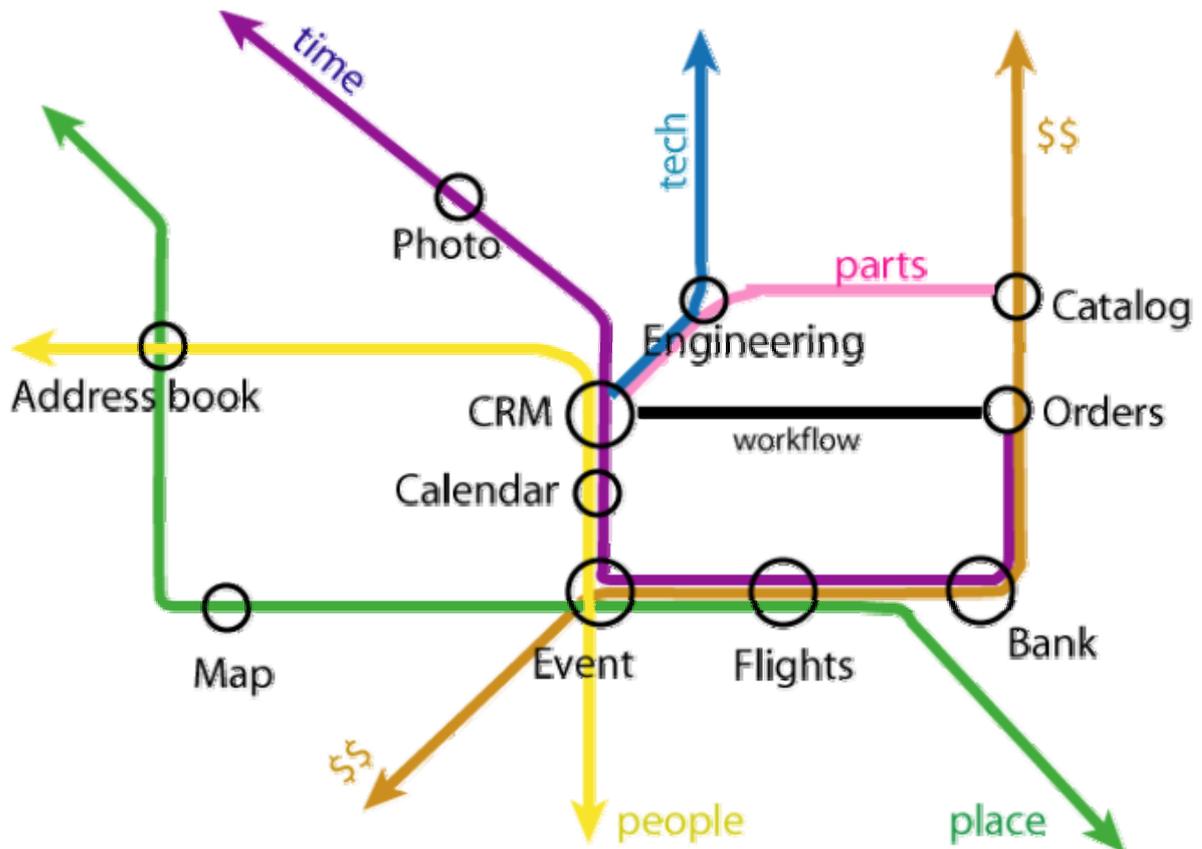


図 5: セマンティック Web 地下鉄マップ i

(From Berners-Lee, ISWC, 2003)

基礎概念がこのダイアグラムの中の線として示されており、人や価格や時刻や場所の様なデータを特別な形式でそれぞれを識別できる。これ等の基礎概念の交点は、意味の高位レベルの何らかの実体を形成する。例えば、アドレスブックの概念は、人と住所(及び連絡先)との組合せであり、カタログの概念は、部品と価格との集合である。

本質的に単純だが、このダイアグラムは、データがより豊富になった時、例えば、日付または位置が追加された場合、特別な形式又はデータの表現より遥にうまくそれらに関係付ける事が出来る事を示している。言い換えると、例として、1863年のゲティスバーグの地図をサーチするのに用いると、もしも、地図を構成するデータがより”リッチャー”(あるデータが日付としてタグ付けられ、あるいは、識別されている事を意味する)なら、知的サーチ機能は、柔軟な日付表現を用いて検索を行うことができる。例えば、色々な日付(例えば、1863年7月、1863年、1860年代)更には、色々な概念(“ゲティスバーグの戦い”または“南北戦争”で示される期間、それぞれ、1863年の7月1日から3日及び1861年4月から1865年4月を示す)が考えられる。

同様に、バージニア州 McLean でのアンチスパムメールのセミナーの例の場合、場所の概念

が、より柔軟な境界領域(ワシントン DC 地域)、隣接都市、郵便番号、公園、記念碑及びその他の横断的概念領域の場所情報と関係付けられているならば、セマンティックアプローチを用いているインテリジェント検索は、構文、構造、表現又は文脈の違いをブリッジする柔軟な構成モデルに従ってデータの型(日付または場所の様な)を認識できる。

“非集中で結合可能”(更に、“進化可能な”を付け加えるかも知れないが)なアイデアは、セマンティック Web のビジョンの根幹を成すものである。

4.2 データの形式

少なくとも、セマンティック Web 技術の進化に於けるデータ構造の観点において、如何なるアプローチがネイティブでない処理の為にデータを記述しデータを提供するのに用いられるべきかを考える上で、データの構造は非常に重要な問題である。企業のデータセットは、多くの様式と構造を持っている。企業のデータは、テキストファイルとオブジェクトの間の差異のように、内部のバイナリ形式が異なっているのみならず、その情報は特定の構造と表現の中に組み込まれている。データ構造は、極度に非構造なものから高度に構造化されたものまで色々である。

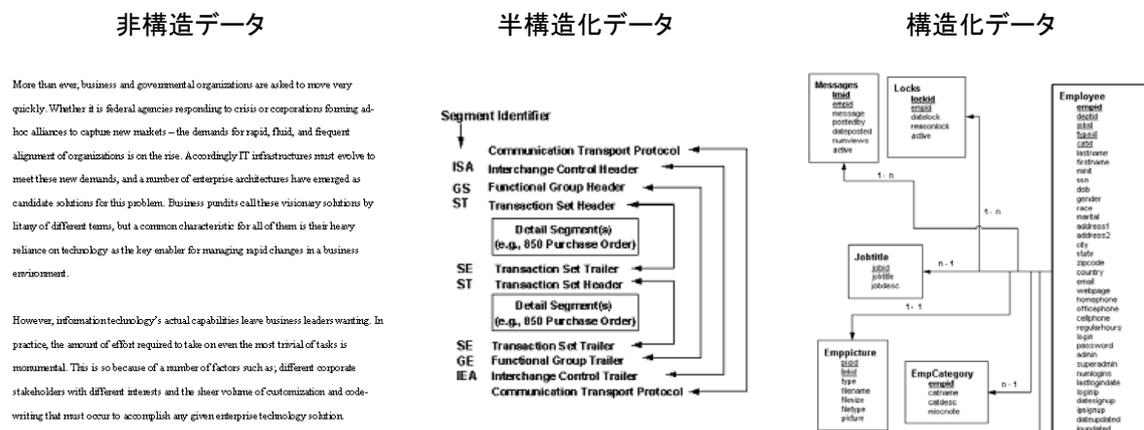


図 6: データ構造の種類
(From Pollock and Hodgson, 2004)

構造化データはこれらのデータの中で最も組織化されたものである。一般に、構造化データはタイプ、長さ、テーブル、制約のようなメタデータ定義を有する。構造化データの例として関係データベースモデル、オブジェクトモデル、XML 文書などがある。通常、構造化データは機械処理や機械利用の為に生成される。

非構造化データとは、既存の知識の延長線上もしくは組織の普遍形式によって意味を交換可能なデータ固有の構造または(明確なレイアウト規則に則った)関係を有しないデータを言う。その良い例が、自由形式の文字列から成る文書であり、その記述方法は任意であり、題名や著者名を除けば、他の文書との関連を定義する如何なるメタデータも構造も無い。非構造化データは、主に人間向けに作られ、機械の場合(特にインターネット時代において)、強力な機能を用いなければ非構造化データの意味を理解する事ができない。

半構造化データは、構造化データと非構造化データとの中間のものであるが、その境界は曖昧

で、それを明示的に示すことはできない。多くの場合、半構造化データは、組織化されてはいるが、高度な連携手法では明示的に定義されていないデータだと思われる。既存の半構造化データの例には、EDI、CSV(comma separated value)プロパティデータファイル、PostScript ファイル、HTML ファイルなどの位置で意味を示しているテキストメッセージがある。半構造化データは、一時的に使われるデータ(あるシステムから他のシステムに譲渡されるデータ)や特別な処理目的で且つ蓄積する事を想定していないデータに使われる事が多い。(Pollock and Hodgson, 2004)

データの形式は、ある特定のデータアイテム若しくはデータ自体を記述するのに、そのデータの構造及び意味に付いての情報を公開するのに、また、他のデータ要素もしくはデータと連携するのに、どの様なアプローチを用いるべきかに直接関わっている。セマンティック技術の本質は、色々な形式のものに横断的に作用できることである。例えば、米国本土の安全情報共有環境の場合、用語の矛盾、定義、電子メールの中のデータ、RSS のデータ、メモ若しくは報告書が、格納されたデータや、または、構造化データベース及び他の高度に構造化された形式から取り出されたデータと一致させる事ができる必要がある。

4.3 メタデータ

補助的データ記述の最も原始的な形式の一つがメタデータである。メタデータは、まさしく文字通り、”データについてのデータ”である。メタデータの最も単純なものはデータフィールドのラベルである。メタデータの細目として、データタイプやデータ長を持つ事ができる。XML は、データフィールドを記述する用語機能や構成データ要素に関する論理スキーマ機能やネームスペースを作成する機能の為にプロトコルを定義する事により、メタデータの利用している。

高度に構造化されたデータセットの中には、相当な量のメタデータが存在するであろう。しかし、より構造化されていないデータセットの中には、より少ないメタデータしか存在しない、故に、それをより扱い易くする方法は、それ自身に関する情報のタグを付けることによりメタデータを作ることである。その様なタグは通常 XML で記述される。写真に付ける”写真”または地図に付けられる”地図”などのタグ付けは、イメージファイルの集合を検索する時、非常に大きな意味を持つ。更に言うと、ある写真に対して、それが撮影された場所、日付、その対象などに関する情報を附加できる。地図の場合、日付や場所などを附加し、街路図、トポロジカルマップまたは戦闘指令図等の地図の種類で分類できる。

メタタグと呼ばれるメタデータの種類は、正規のウェブページの仕様書の一部として定義されており、ウェブページの内容に関するより良い情報を提供する事を目的としている。しかし、検索サイトや信頼性に関する懸念を欺く為にメタタグが用いられた為、検索サイトがメタタグを使う事をやめたので主流となれなかった。しかし、メタタグ以外のメタデータの利用は、企業の中やウェブ上の基本的なデータ統合の方法として別の形で、復帰しつつある。データソース(資源)の評判や信頼性を評価するための新しい方法も開発されつつあり、それによって検索能力を高めたり、信用を改善したりするだろう。

メタデータの一般化しつつある代表的な例は、2003年と2004年のRSS送出の劇的な増加である。RSSは、Real Simple Syndicationの略称であり、ニュース若しくはニュースの様なコンテンツを配給するためのフォーマットである。

要するに、RSSは(XMLで記述された)メタデータ標準の一種であり、ニュース配給チャンネルの中で、ニュースの見出しや(著者や作成時刻のような)アイテム情報を記述するのに用いられる。RSSは、比較的軽量のメタデータ記述フォームであるが多目的で拡張性がある。その標準は数年前から存在していたが、ほんの2003年から特にブログの世界で利用が広まった。(2004年9月時点で)ウェブの中に90万を越えるRSSのチャンネルが存在し、毎日1000のオーダーで増加している。RSSの利用者にロイターやW3CニュースやスラッシュドットやXMLニュースなどがある。そのメディア領域の拡大及びその技術領域に於けるRSS開発戦略の新たな出現は、この技術の可能性を明示している。

XMLの出現と共通メタデータスキーマとを原動力として、企業内のメタデータの利用は、ここ数年着実に増加している。メタデータの形式と意味との合意がなされた時、XMLは、情報を生成し、また、格納するシステムやアプリケーションに依存しない情報を作る為に単純ではあるが強力なツールとなる。しかし、組織や個人が、ドキュメントを作らず、あるいは、他人が理解するのに記述が不十分なまま、独自のやり方でメタデータを実装したら、幾つかの問題が起きるのであろう。この様な独自のアプローチは、しばしばデータストアの所有者以外の人に、基本的な情報が分からず、また、アクセスすら出来ないといった状態を来す。

メタデータの中の用語、構造及び意味の矛盾や不調和を取り扱う事は、セマンティック技術の初期のアプリケーションの一つである。例えば、ある連邦機関が、色々なソースからある一つのWebポータルに統合されている健康と環境汚染物質との情報を共有する為の最適な方式を決めようとした時、色々な形式のデータを結合しなければならなくなり、メタデータの重要性を明確に理解した。このプロジェクトの最重要の挑戦事項は、形式とソース(その機関の管理下以外のもの及び影響下以外のものをソースとして含む)が本質的に異なる情報をどうするか検討する事であった。その問題を分析した後、システム設計者は発散的な用語矛盾と意味の不一致との調和は、このシナリオを処理するためのメタデータ管理ツールを活用することで達成する事が最良であると結論付けた。

このツールは、コアの機能として、色々なデータソースの意味衝突を調整し、正規且つ一貫したクエリを行い、更には、表示する機能を有する。セマンティック技術は、色々な理由で企業の中のメタデータの活用を加速する。この技術は、メタデータを(a)有用にし、(b)容易に管理可能にし、(c)再利用可能にする。開発者により再利用され、利用者により二度以上アクセスされ、アナリストにより正確であることが保証されたメタデータは、性能と生産性を向上させるデータでもある。加えて、ある特別なニーズ(及び当該組織にとって有意義である何か)に関連すると見做されるメタデータは、従業員やその他の人々により保全されるべきメタデータでもある。

4.3.1 メタデータ標準

メタデータ標準にはDublin Core Metadata Initiative(DCMI)等がある。DCMIは、“相互互換可能なメタデータ標準の広範な採用を促進し、且つ、より知的情報発見を可能にするリソース

記述を行うための特殊なメタデータ語彙を開発するための”ものである。ISO 16642 (Computer applications in terminology -- Terminological markup framework)は、用語データ集合の中に記録されたデータを記述するための基本原則に関するガイディングフレームワークを定義している。このフレームワークは、メタモデルと XML を用いた特殊な用語のマークアップ言語を記述するための方法とを含んでいる。ISO/IEC 11179 は、データ要素の標準化に関する 6 編から成る標準である。その標準はデータ要素の構造定義を行なうための規則とガイドラインとを規定している。PRISM は、出版業界の団体が開発している標準メタデータ語彙である。OMG(Object Management Group)は、アプリケーション開発及びシステム開発の為の沢山のモデル化仕様書とメタデータ仕様書とを提供している。注目に値する標準は MOF(Meta-Object Facility)である。MOF は、プラットフォームに依存しない方法で、メタデータとデータを定義したり操作したり統合したりするための拡張性のあるモデルドリブンの統合フレームワークの一種である。MOF ベースの標準は、ツールやアプリケーションやデータを統合するために使用されている。その他のメタデータに関連する例には、メタデータ交換のための OMG の CWM (Common Warehouse Metamodel)がある。(OMG と W3C とは現在、より密接に活動する方法を模索している)

メタデータ標準、そして/または、標準化努力は、地理空間的 (geospatial) な情報及び健康管理から一般の消費者市場に至る迄多くの産業で見られる。地理空間的 (geospatial) な分野における注目すべき成果には ISO 19115 と Digital Geospatial Metadata と連携している Federal Geographic Data Committee がある。商用で市場に受け入れつつある標準に XMP(Extensible Metadata Platform)と呼ばれるものがあり、これはアドビシステム社によって開発されたものである。XMP は、RDF のサブセットを使ってファイルの中にメタデータを埋め込むことを可能にする。特に注目すべき事は、XMP は PDF と沢山のイメージ形式をサポートしている事であり、それにも拘らず、殆ど総てのファイルタイプをサポート出来る様設計されている事である。

多くのアドビのアプリケーションは、XMP スキーマを書く事ができ、加えて、アドビは XMP ソフトウェア開発キットを提供している。

ウェブコンテンツのデジタル著作権管理を目的とする非営利団体の Creative Commons は、機械処理可能な形式でデジタル著作権管理情報を埋め込むために XMP(及び、ネイティブの RDF、SMIL、複数のオーディオ形式など幾つかの他のメタデータ形式)を利用している。この機能は、デジタル著作権の自動的な管理やネゴシエーションを行うのに役に立つ。例えば、イメージファイルの集合に対して、特定の主題でないもので、且つ、非商用利用なら無料のイメージを検索するために問合せを行う事が可能になる。

4.4 セマンティックモデル(タクソノミとオントロジ)

このごろ、'ontology'という言葉を使うことができるようになり、人々はそれが何を意味しているか知っている。

Michael Daconta

組織、情報実体もしくはその他の知識領域空間の中に於ける関係の膨大な配列を充分かつ正確に記述可能にするデータモデルの研究が現在進行中である。その挑戦は、機械計算処理可能であり、且つ、意味モデルを用いてコンピュータにより大規模に自律的且つ決定可能に成った時、よ

り高度なものとなる。これまで、多くの知識記述技術が考案されてきた。そのあるものは成功し、また、あるものは成功しなかった。これらの努力の結果として、コンピュータ科学者は、ビジネス環境、組織的相互関係、更に言えば、日常社会の中に於ける関係概念及び論理的概念を高度に記述するのに最も適切な方法を見つけ出すという点において重大な進歩を得た。

沢山の語彙群の信頼性に起因するコミュニケーションギャップの克服が挑戦すべきものとして残されている。最近まで重複や冗長な用語上の一貫性の欠如を克服する事が技術的挑戦であった。その様な問題を認識する事無しに、ビジネスユニット、個人あるいはその他のものは、異なる色々な用語や異なる関係モデルを用いて同一の要素を参照することに乏しい資源を費やし、その結果、混乱を惹き起こしコミュニケーションの可能性を限定していた。

このような意味の違いを識別したり、調和させたりする事が、セマンティックモデルを使う基本的な理由である。図7は、一般的に使われているセマンティックモデルのスペクトルである。

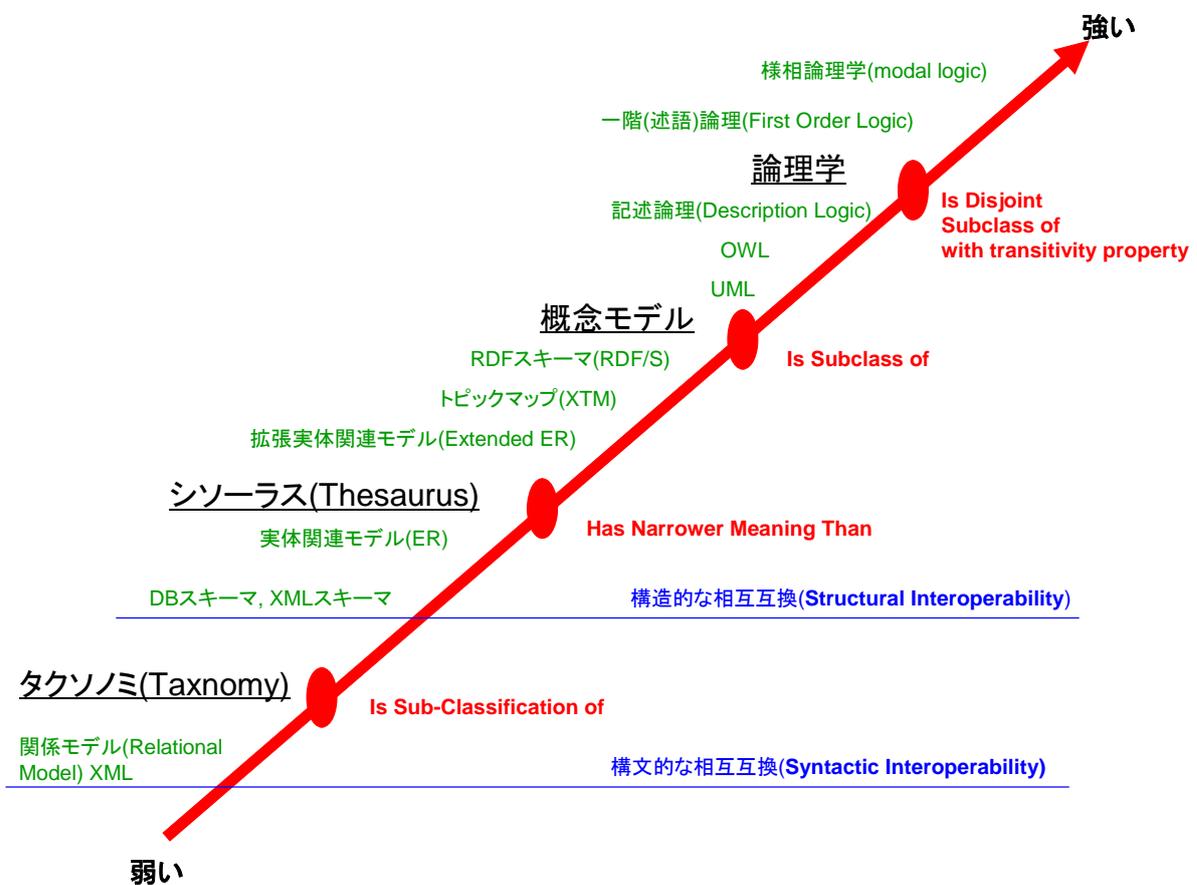


図 7: オントロジスペクトラム

(From Daconta, Orbst, and Smith, 2003)

この図は、左下の表現力の乏しい或は意味の弱いモデルから、右上に向ってより表現力の豊かな或は意味の強いモデルの位置を示している。概略的には、左下から右上への並びは、それらのモデルが表現する構造の複雑さが増加する事を示している。すなわち、最も表現力の豊富な意味モデルは最も複雑な構造を有している。この図にはモデルと言語とが示されており、それらは、左下に示されている関係データベースモデルや XML 等で読者は良くご存知であろう。以下、XML スキーマ、実体関連モデル(ER)、XTM(The XML Topic Map standard)、RDF スキーマ(RDF/S)、UML(Unified Modeling Language)、OWL(Web Ontology Language)そして、一階

述語論理(First Order Logic)まで続き、更に、上に続いている。実際には、このスペクトラムは、様相論理(Modal Logic)より上に広がっているが、それらは未だ高度に理論上の話題であり、本書の対象範囲外である。

セマンティックモデルの中で最も簡単なものはタクソノミである。タクソノミとは、合理的に綺麗に定義された関係構造を用いた情報の分類或はクラス分けの一つの方法と考える事ができるであろう。二つのアイテム間の関係の形は、それらアイテム間の構造やコネクションに固有なものである。タクソノミは用語間の結合関係を捉えているが、その結合関係の性質は定義していない。総ての関係は”親-子”リンクから成る階層構造である。しばしば、この様な階層構造はツリーと呼ばれ、最上位の根から下に向かって枝分かれする。階層構造の中には、特定のアイテム若しくは下位のアイテムと各アイテムとの間に向きの揃った結合関係が存在する。タクソノミの良くある例は、生物学における動物相や植物相を記述するのに用いられる階層構造である。

図8は連邦エンタープライズアーキテクチャ(Federal Enterprise Architecture (FEA))の一部で行政概念タクソノミの一部分を示している。タクソノミの階層構造的により、幾つかの概念は、一つ以上のカテゴリ下にグループ分けされている。例えば、”プログラム”は二回現れ、一つは”部局(Agencies)”の下にあり、他の一つは”協力関係(Partnerships)”の下にある。タクソノミは物事をクラス分けするのに有効であるが、しかし、物事の意味をモデル化するには有効ではない。

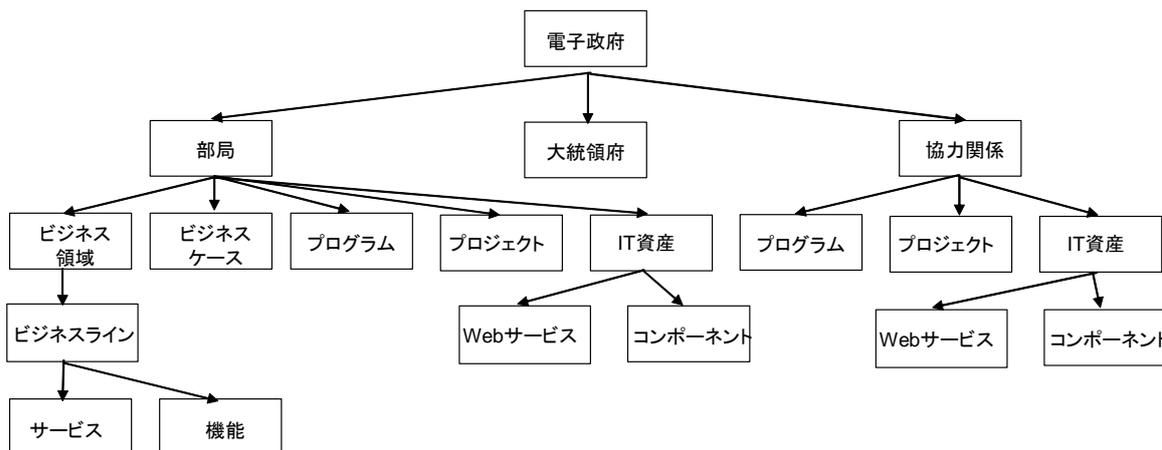


図8：電子政府に於けるタクソノミの例

シソーラスは、固有の意味をその構成の中に含み得るのでタクソノミより高次のセマンティックモデルである。言い換えると、シソーラスとは、タクソノミに規定された語彙を用いて何等かの関係の意味を付け加えたものである。シソーラスのノードは言葉若しくは文章を意味する”用語(terms)”である。これ等の用語は、他のものに対して、”より広い”又は”より狭い”関係を有する。シソーラスは、同義語(synonyms)の様に、用語間の前記以外の関係の意味を持たせる事ができる。

タクソノミ及びシソーラスの意味表現力は限られており、関係定義における一次元の公理だけしか提供できない。それ故、これ等は、通常クラス分け体系を作成するのに用いられるが、多次元概念領域や異なる概念領域を表現するには有効ではない。概念とは、意味の替りを示すものではなく意味を帯びている物である。その概念は非常に抽象的であり、そのモデル化はより複雑

である。或る概念群と他の概念群、そのプロパティ、属性及び概念間のルールとの関係性は、タクソノミを用いてモデル化することは出来ない。しかし、より洗練された形式のモデルを用いればこれ等の関係を表わす事が出来るであろう。

(アイテム間の関係)関係に明示的に名前が付けられ、また、区別するセマンティックモデルは、オントロジと呼ばれる。(図7の中の概念モデルと論理学とはオントロジと考える事ができ、前者は弱いオントロジであり、後者は強いオントロジである。) 何故なら、関係性が指定されるので、関係性を定義若しくは包含する厳密な構造を最早必要としないからである。モデルは本質的にコネクシオンのネットワークになり、各コネクシオンは、任意の他のコネクシオンとは無関係の関係を有する。通常ツリー構造で示されるタクソノミとは異なり、オントロジは、一般的にグラフの形式で表される。すなわち、ネットワークでは、ノードに対する交差する枝(他の関係性を表す)を持つ事も、また、幾つかの子ノードは複数の親からリンクされる事もある。このような豊富な結合性は、観念を表すのに非常に高度な柔軟性を提供する。何故なら、多くの概念領域はタクソノミやシソーラスで適切に表現する事はできないからである。維持不可能な辻褃合わせを強引に行う事により非常に多くの特異性や矛盾が起きる。更に、似ていない概念間の移動は、しばしば保守や拡張を行う事が難しいもろい結合機構を必要とする。

前に述べたゲティスバーグの地図を例にとると、日付の範囲を推論する為に“ゲティスバーグの戦い”や“南北戦争”といった概念を使うアイデアは、タクソノミを使って実現する事が不可能ではないが難しい。関係によって規定される関係は、“ゲティスバーグの戦い”と“1863年7月1-3日”との間の”日付”または”日付の範囲”関係を包含する事を可能にする順序だった関係性構造とは無関係に定義する事ができる。結果的に、日付に関する概念のオントロジの中の任意の関係を辿る事が出来るならば、サーチエンジンは、日付の範囲に関し推論を実行する事が出来る。前に述べた様に、これは、人間がもごもご言っているのを理解する様なマジック的な人工知能を意味しているものではない。それは、ただ単に存在している良く定義されたデータに対して良く定義された操作を行なう事により良く定義された問題を解く機械の機能を示しているに過ぎない。

図9は、TopQuadrant社によって作られた FEA Capabilities Manager の一部の為のオントロジを示している。下記のモデルは或る特別な IT コンポーネントが大統領府の支援の為に開発されている事を推論する事を可能にする。更に、このモデルはある特別なコンポーネントを開発する協力関係にある部門を特定できる。

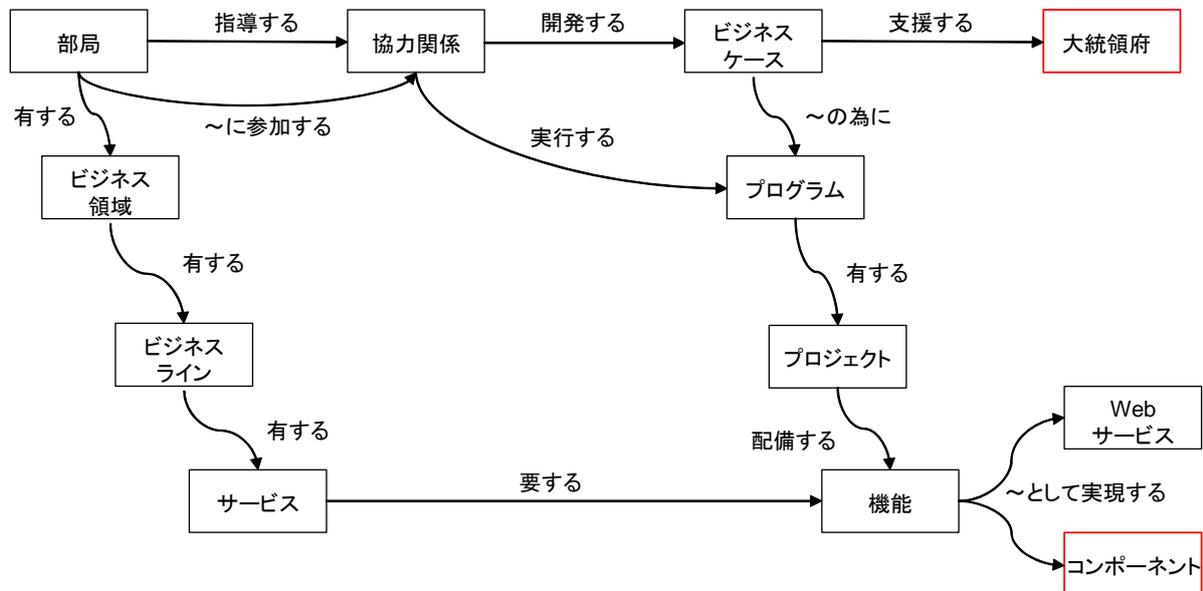


図 9: FEA Capabilities Manager のオントロジモデルの一部

(TopQuadrant 提供)

単純なオントロジは、単なるコネクションのネットワークである、例えば、より豊富なオントロジは、これらの結合を規定するルールと束縛を含み得る。言語とモデルベースプログラミングのアプローチとの改良が、人間によるコーディングステップを必要とすること無しに概念モデルから実用モデルに移行する能力を強化させたのと同様な事が、オントロジ開発の中にも起こり得る。かつて、オントロジは人間の為にだけ作られたが、その様なモデルをサポートする基盤の発展に伴ったオントロジを表現する為の強化プロトコルの開発は、これ等の関係と規則とに基づいて基礎を成す文脈の演繹や論理的な帰結を得る為のモデルの能力強化をもたらした。

4.4.1 オントロジ記述標準

オントロジ記述と利用とに関する現在の状況は1980年代に始まった色々な努力が実を結んだ結果である。初期のセマンティックシステムは、ユビキタスネットワーク基盤が無かった事と知識記述の為の標準の欠如とから当初は苦闘した。WWW の出現とウェブ上の情報交換用のデファクト標準としてのXMLの普及とで、オントロジの研究は収斂し、そして、実を結び始めた。RDF、OWL 及びトピックマップ(コンテンツリソース上に重ね合わせられるべき概念網記述の為のISO標準の一つ)は、総てXMLを用いている。(標準プロパティと順次形式の中のフィールドを持つ)強く型付けされた記述であるRDF、OWL 及びトピックマップは、ウェブ上でこれ等のモデルを転送したり、蓄積したり、また、Webサービスの様な他のウェブ標準とそれらのモデルを統合する事を容易にした。

知識管理コミュニティの幾人かから表明されている警告がある。それは、同じ様な競合するオントロジが沢山出現するのではないかと言う事である。それはシステム横断的な構造や意味のシームレスな共有を達成する阻害要因に成るかも知れない。異なるオントロジがあるモデルから他のモデルに自動形式変換する用に作られ得るけれども、通常、その様な状態に至るには多量の人間によるモデリングが必要となる。(あるサイズのオントロジの整備は、大規模なデータベースの整備に例える事ができ、その様な仕事は、しばしば多くの計画の立案と開発努力とを必要とす

る。) 「広く共有されたオントロジを用いる事で大いなる成果が得られる」と言う知識管理の専門家の警告の問題は後の節で議論される。

5.0 コアビルディングブロック

5.1 セマンティック Web のウェディングケーキモデル

Tim Berners-Lee は、XML2000 コンファレンスにおいて、セマンティック Web ウェディングケーキ(レイヤーケーキ)の説明を行なった。この説明は、セマンティック Web 社会の中で、広く興味を持たれ、他の多くの作者やアナリストたちによって引用されている。

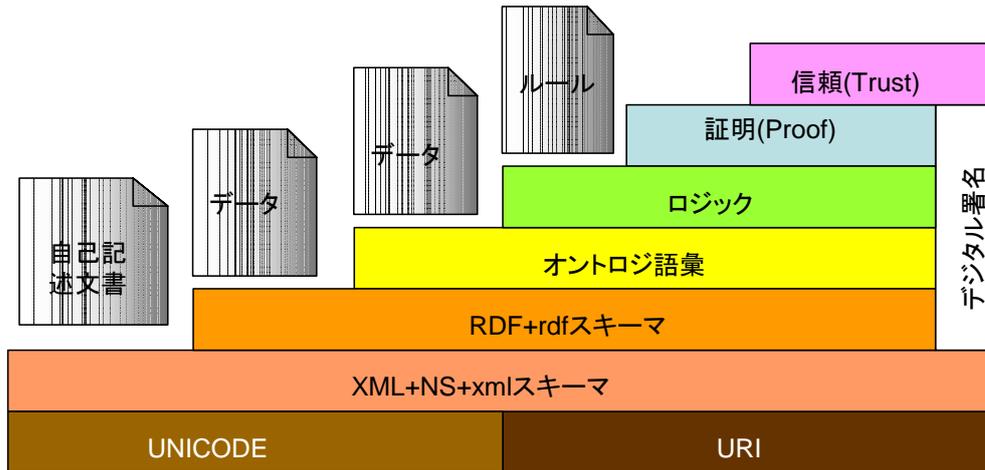


図 10: セマンティック Web ウェディングケーキ

(Berners-Lee, XML 2000 Conference より)

図 10 は、前に示した図 4 のセマンティック Web の概念階層図を体系化したものである。ここでは、技術要素に対する基盤として用いられるプロトコルと言語とを強調している。このウェディングケーキの底に良く定義され且つ広く受け入れられている諸標準を示している。Unicode は、16 ビットの文字セットであり、ASCII コードの後継として殆ど総てで受け入れられている。URI は Universal Resource Identifier の略であり、インターネットの中の現在及び将来のオブジェクトの名前と位置とを示す為に W3C が規定したものである。URI は、ウェブサイトやウェブページに対する識別子として通常知られている Universal Resource Locator 又は URL の拡張概念である。

図 10 は、どちらかと言うと非公式且つ幾つかの問題を図示している事に注意しなければならない。例えば、このオントロジは用語で且つ或る特殊な言語で定義されたものである。このロジックは、関係付けられたデータアイテムの集合を用いた論理的な推論を可能にするものを示している。この証明(Proof)は、論理的な推論のステップを追跡する機能を指し、これに反して、信頼(Trust)はデータの素性を示す。それは、オリジナルのデータか、又は、そのデータを操作したメソッドは信頼できるものか否かを示す。これ等総てのアイテムは、相対的に異なる概念であり、他のものの上に構築される必要がない、若しくは、この様に順序立てて図示し、正確に現れるものではない。或る人はオントロジ無しでロジック記述が出来、同様に、他の誰かは、ロジック無しで信頼(Trust)を実現する事ができる。すなわち、この図は、情報技術の専門化に対して、真に付加的なコンピューティングを実現する為の機能拡張の準備ができていた事を示すプロ

トコロと言語との青写真を提供しているに過ぎない。

5.2 言語

5.2.1 XML(eXtensible markup Language)

XMLはeXtensible Markup Languageの略称で、1990年代後半にW3Cによって開発されたデータ記述、データ転送及びデータ交換の標準的な方式である。XMLは、データを記述する方法として、カスタマイズされた”タグ”を用いてマークアップデータを作り上げる為の仕組みを提供する。XMLはHTMLに関する限り必要では無い、事実、これら二つは全く異なる目的の為に設計されている。このような事実にも拘らず、この二つは、利用者のニーズに則して、色々な局面でお互いを補完できる関係にある。

タグは、通常、”姓名”や”街路番地”のようなデータに対するラベルである。ある標準的な交換形式を定義するのにXMLを用いる場合、タグに関する合意を得る事が重要である。例えば、二つの出版社がデータ交換を含むパートナーシップを形成する事を望むかも知れない、この場合、出版社Aの”著者”の定義は出版社Bの”ライター”の定義と同一である事を最初に定義し、その後、XMLを用いて、その他の合意した部分を定義する。重複した用語や同じ意味の用語が、通常DTDやXMLスキーマ等を用いて定義される。(XMLスキーマは形式的な方法でXML文書を定義する仕組みであり、正確な情報交換を可能にする。)

XMLスキーマの実例を沢山の政府や実業界の中に見出す事ができる。U.S. CIO 諮問委員会のXML作業部会によれば、”組織が同一要素の定義を行ない、且つ、これ等の定義をパートナーが検索やサービス発見に使える様になって、XMLの最高の利益を獲得する事ができる。レジストリ/リポジトリは、インターネットを介して文書、テンプレート及びソフトウェア(すなわち、オブジェクト及びリソース)を検索したり見つけ出したりする為の手段である。レジストリはオブジェクト発見の為に用いられ、場所等の発見したオブジェクトに関する情報を提供する。リポジトリはそのオブジェクトの存在する場所である。”

セマンティックス及びセマンティック Webの文脈の中では、XMLは特定の領域の中で意味的に豊富なマークアップ言語を創るための構文規則のセットである。XMLは文書の任意の構造を規定する事を可能にするが、その構造の持つ意味に関しては何も述べていない。(Berners-Lee, Hendler, and Lassila, 2001) 言い換えると、ITシステム、データベース及びコンテンツ管理システムは物事を記述するには良くなったけれども、関係を記述するにはそれ程良いものとは言えない。より強固且つ忠実な記述方式が語句、用語及び領域の意味を表すのに必要である。

5.2.2 RDF(Resource Description Framework)

RDFはResource Description Frameworkの略称で、特にこの種の関係情報を記述するために設計されたものである。RDFは、豊かかつ柔軟なデータを作成する手段を提供する、故に、システムプログラマやデータモデル設計者により明示的に定義されたものを、当該環境の外に置く事が可能である。RDFは、主語、述語、目的語からなるトリプルの集合を用いて情報をコード化する。(このモデルはリソース、プロパティ、バリューの構造でも表せる)

RDFは、分散されたメタデータをリンクする為の基盤を提供し、OWLと共にオントロジを

記述したり表現したりするコア言語でもある。

データの関係記述に RDF を用いる第一の利点は、スケーラビリティと柔軟性とである。データベーステーブルを用いて、RDF と同じ様なものを作る事が出来る事は明白であるが、RDF のユニークな点は、高度な関係記述機能の為の柔軟な仕組みを提供している事である。それを用いて、問合せの機能の高度化を図ったり、また、テーブルの中に明示的に記述されていないトピックの様なものを推論する事ができる。この利点は、新たなデータソースの統合を図る時にのみ増大する。特に、それらが異なる構造若しくは意味を持つ時、又は、より重要な事には、概念領域をまたがった時、特に顕著となる。(環境データと健康データとの統合、又は、法執行データと諜報データとの統合等の場合)

RDF トリプルは XML 形式で記述され、XML タグ、又は、機械処理し易い形式の構文を用いてデータ要素間の関係を記述する一つの方法を提供する。疎結合及び/又は仮想アーキテクチャをサポートする為に、そのトリプルの要素を特定するのに URI が用いられる。URI の目的は、主語、動詞、若しくは、目的語の形式の中で、夫々に該当する概念を定義している大元の場所にリンクする事により対応する概念を曖昧無く識別する事である。

RDF スキーマ(RDFS 又は RDF-S と記述されることもある)は、RDF の意味の記述や RDF の拡張を行う為のものである。関連するリソースのグループやリソース間の関係を記述する仕組みも提供する。RDF スキーマは、XML に対して DTD や XML スキーマが行なうのと同様な事を RDF に対して行なう。

RDF データに対する問合せ言語が、研究者や産業界で沢山開発された。2004 年 10 月に W3C の Data Access Working Group は、SPARQL(“スパークル”と発音する)の仕様書の草稿をリリースした。これは、RDF の問合せ言語の一つであり、開発者やエンドユーザが書くべき言語の統一化を模索したものであり、且つ、広範囲の情報からの検索結果を使える様にする事を目指したものである。

5.2.3 OWL(Web Ontology Language)

OWL は、Web Ontology Language の略である。(この頭字語は、Winnie the Pooh の本に出てくる owl (フクロウ) の名前を連想するように、本当の名前 WOL から意図的に OWL へ変更された) RDF の主要な価値が分散データの結合と統合を可能にする事であるのに対して、OWL の主要な価値は、分散データの推論を可能にする事である。

OWL は高度な表現力を有するモデリング言語であり、既存のデータ格納機能や XML や関係及びオブジェクト指向アプローチ等のモデル構築機能との互換性を有している。更に、OWL は、簡単に構築でき且つ発展可能な複数の知識ベースから成るデータの疎結合ビューを提供する。最も重要な事は、OWL が機械処理可能なセマンティックス(machine-actionable semantics)を持っていることである。実行時および設計時に、ソフトウェアツールは、人間の助けや高度に特殊なアプリケーション命令を必要とせずにモデル、データ、メタデータ、規則(rules)或はロジックを用いて”ある事”を行なう事が可能となる。(Pollock, 2004)

OWL は、柔軟且つ計算可能な論理構造のセットを開発しようとした数多くの開発努力から発展したものである。その開発努力の多くものは数年前に始まったものである。OWL は、DAML+OIL と呼ばれるオントロジ言語の次の世代のものであり、DAML+OIL は、米国の DAML(the DARPA Markup Language)とヨーロッパの OIL(the Ontology Inference Layer (又は Language))と言う二つの開発努力を統合したものである。また、OWL は SHOE(Simple HTML Ontology Extensions)にもルーツを持っている。SHOE は知的エージェント機能を実現する事を目的として、マシンリーダブルな知識を Web 文書と連携させる為に、メリーランド大学の James Hendler をリーダーとして開発されたものである。OWL には、(OWL Lite, OWL DL 及び OWL Full) 3 つのレベルが定義されており、漸次、表現力と推論の能力とが高度になる。これらのレベル分けは、ツールベンダが OWL の特定機能範囲のツールを開発するのをより容易にするために作られた。

RDF と OWL とは一緒に使うことも出来るし、別々に使うことも出来る。ある場合、分散データをサポートすることが主な目的かも知れない、その様な場合は、RDF のみが使われるかも知れない。また他の場合、分散と推論機能とが必要とされるかも知れない。この場合、RDF と OWL との両方が利用されるかもしれない。他の場合、推論機能だけが要求とされるなら、この様な場合、OWL だけで充分であろう。

5.2.4 その他の言語開発の為の努力

前記以外の言語の開発が、セマンティック Web ビジョンの中の上位層に対応するために現在進行中である。例えば、ルール言語は、機械処理に適した形式である論理関係を表現するための機能を提供するであろう。この言語はビジネスルールの表現を可能にし、より高度な推理や推論能力を提供するであろう。RuleML が最初ルール言語として提案されたが、Semantic Web Rules Language(SWRL)を形成する努力が現在 W3C で進行中である。ロジック言語は、単調論理を表現したり、証明を検証したりするための普遍的な機構を提供するものと考えられている。

長期的な野望は、最終的にウェブなどから新しい知識を引き出して何等かの事実を利用可能にする事である。(ここでの問題は、演繹システムが恐ろしく相互互換でなければならないと言う事ではない。何でも出来る唯一の推論システムを設計するというより、むしろ、現在の活動は、証明を記述するための普遍的言語を定義する事に集約されている。それが実現すれば、色々なシステムが、他のシステムで利用したり組み込んだりする為に、これ等の証明をデジタル的に署名したりこれらの証明をエクスポートしたりする事が可能になる。)

同様に、評判や信用を推論する為の構造、スキーマ及びアーキテクチャが、W3C の中や、より大きな Web コミュニティの中で開発されつつある。これらのアプローチは、個人や人々との間の評判や信頼を推論するだけでなく、人々のグループ(会社、メディアソース、NGO 及び政治的動向など)や無生物(本、映画、音楽、研究論文及び消費物資など)やアイデア(信用システム、政策上のアイデア及び政策提案など)を推論するための手段として注目されている。(Masum and Zhang, 2004)

この分野の技術者が直面した 1 つの問題は、あいまいな人間の言葉で述べる事が可能な知識を把握するのに十分な記述力を有するフレームワークと言語とを開発する事である。問題は、複雑な事を表現するのを可能にしつつ、単純な事を簡単に表現する事を可能にする言語、ツール及び

システムを如何にして作るかである。もう1つの問題は、構文に対する人間の可読性や扱い易さを勘案しながら HTML、XML 及び RDF の様な既存の構文標準といかに互換性を保持するかである。最終的には、これらの問題を極小化する様なより良いツールが開発されるであろう、しかし、暫くすると、何等かのより高次の言語における複雑さが、現在のエディットツールやモデリングツールを用いた完全準拠の実装開発をより困難にするかも知れない。

6.0 セマンティックツールとコンポーネント

セマンティック Web のツールは日々良くなり、新しい会社が参入し始め、大会社は動き始めている。

James Hendler

技術のライフサイクルや成熟の段階を表すモデルが幾つか存在する。一般に、これらのモデルは、参入(または定義)、成長(または検証)、成熟(または改善)、衰退(または合併)の4段階を有する。大方の尺度では、セマンティック Web は、公開されているものを見る限りでは、未だ参入/定義フェーズにある。しかし、セマンティック技術の幾つかは成長/検証フェーズに入っていると言える。(成熟過程におけるフェーズの移行点は大体分かり難く、その転換点は変革の後で分かるものである。また、時には、疑いや不適切な期待により見逃してしまうものである。)

政府や私企業に於いて新技术を採用する先駆者は、新たな道を切り開き、彼らのセマンティック実装プロジェクトから成果を得ている。既に、多くの政府部門でセマンティック研究プロジェクトを実施している。セマンティック製品は、Adobe、Hewlett Packard 及び IBM などの有名な大企業から提供されると共に、Unicorn、Network Inference 及び Semagix の様な小さいが先駆的な企業からも提供されている。加えて、数多くのオープンソースや誰でも使えるツールが公的或は私的研究所や組織などにより開発されている。

以下の章では、セマンティック技術を活用したアプリケーションを作るのに利用できる商用ツール若しくはオープンソースツールとしてどの様なものがあるのか調べた結果を記述する。これ等のツールを理解するには、それらが設計時のツールか実行時のツールか見分ける事が必要である。設計時ツールは、文書の著者、システム設計者及びその他のもの作りや設計や監査をする人達により使われるツールである。例えば、メタデータ作成ツールやオントロジの作成ツールやオントロジ移植ツールなどが含まれる。実行時ツールは、問い合わせ処理、データ変換又は処理結果生成に用いられるソフトウェアコンポーネントである。例えば、仲介サーバや推論エンジンなどが含まれる。多くのツールは、実装過程で一纏めにして利用される。例えば、事業連携の設計時に用いられるモデリングツールやマッピングツール、実行時の問い合わせ機能や仲介サーバなどをワンセットで利用する。

6.1 メタデータ生成と管理ツール

或る文書若しくはデータのアイテムが作成又は著述される時、それをコンテンツ管理システムまたはウェブサイトにインポートする時、または、それらを利用者に対して表示或いは利用者が読む時、当該文書若しくは当該データに関するメタデータを作る作業を行う事が可能である。更

に、該当するデータアイテムが存在する間は、その作成作業は、何時でも何等かの明示的或いは暗黙的アクションにより行われ得る。言い換えると、メタデータの生成は、ある一点で行われるのでは無いと言うことである。メタデータは継続的に積み重ねられ、任意の時点で任意の人々により修正され得る。

コンテンツ作成時に通常著者は、文書に主題、作成者、場所、言語或いは著作権状況などの情報を付加する。この種の情報は該当する文書をより検索し易くする。RSS は基本的にこの種の情報から構成され、情報を検索したり選別したりする為に特化された機能を持つニュース読み取り用アプリケーションを提供している。SixApart と言う会社の Moveable Type は、RSS 準拠の文書を作成するのに使われているブログコミュニティの中で最も普及しているツールの一つである。RSS の普及の拡大と簡単さとは、ブログコミュニティを超えて一般メディアに広がり、企業の中にも広がりつつある。デスクトップやウェブオーサリングツールの他のベンダは、RSS 作成機能を提供すべく対処を急いでいる。

メタデータの生成作業と言うのは、諸々のステップの中の一つに過ぎない。メタデータ管理ツールは、メタデータ語彙の保守機能、メタデータを用いた問い合わせ機能、更には、対象領域に於ける変化を監視する為の可視化機能等を必要とする。

連携とデータ共用システムとを実現する為にメタデータを活用しているウェブサイトの例として、Flickr がある。このサイトは、人々が簡単に、デジタル写真をアップロードし、共有する為のサイトの一つである。何が他のデジタル写真サイトと異なるかと言うと、このサイトは、写真に対するタグ機能を提供していると共に、写真のカタログを見る為の画期的なインターフェースを提供している事である。(そのタグは、ある種の地図の中に含まれており、データ格納装置内でのタグの利用頻度に応じてサイズが変化する。) そのサイトが初期のメタデータ実装と異なる点は、フィードバックが非常に厳密に行われていることである。それは、タグの割付が、その利用に応じてなされている事を意味する。写真及び写真の集合は、タグ付けされるや否や直ぐに、同一のタグを持つアイテムの集団を見る事ができるので、利用者はその集団を再定義する為にタグを容易に変更することができる。

メタデータ問い合わせ用のツールの観点では、それらのコンポーネントは、現在の検索エンジンと大差はないが、メタデータを包含させる事により、より情報が豊富なデータを作る事が可能になり、その結果、より正確且つ関係性の深い検索が可能となる。問い合わせ用スクリプト及び言語は、利用者がより精緻な機能を付加する事を可能にするであろう、しかし、機能の高度化と単純さとのバランスは、相変わらず流動的であり、特に、広く利用可能な検索エンジンでそれが顕著である。前述の Flickr で見られるように、新たな可視化ツールが関連データの複雑なフィールドを利用者が分かり易く見る事のできるツールとして開発されて来ている。

6.2 モデリングツール(オントロジの作成と変更)

モデリングツールは、オントロジの作成や変更に使われる。知識をモデル化する人(知識モデル)は、クラス構造やモデル領域の作成及び編集にモデリングツールを用いる。これ等のツールは、ファイルシステムのディレクトリ構造に似たインターフェース若しくはブックマークフォルダのインターフェースを有する。更に、これ等は、データベーススキーマ、商品カタログ及び

イエローページリスティングなどの既存のオントロジをインポートする機能、形式変換する機能及び流用する機能などを部分的機能或いは主機能として持つ傾向にある。その他の特筆すべき機能として、類似の用語や概念に対する編成、マッチング及び連携の為の進んだ仕組みを有している。

その上、一つの大きなモデルを作るよりは小さなオントロジを連結するのが、モデラにとっては普通の感覚なので、(何故なら、扱い易いし、再利用し易い)分割、併合及び連結の為の機能の提供が、オントロジエディタの重要な機能となっている。幾つかのエディタは協調作業メソッドやより高度な可視化機能やグラフィカルな操作モードを具備しているものさえある。

Protégé-2000 は、大勢且つアクティブな利用者のコミュニティと一緒にスタンフォード大学で開発されたフリーのオントロジエディタであり、色々な開発者が Protégé の機能を拡張可能にするプラグインの仕組みを有するオープンアーキテクチャを備えている。これ等以外にも商用のモデリングツールが沢山のベンタから提供され利用可能である。それらの会社には、Network Inference、Language and Computing 及び Intelligent Views などがある。

IBM のオントロジ管理システム(SNOBASE として知られ、セマンティックネットワークオントロジの基盤用)は、ファイルからインターネットを介してオントロジをローディングし、且つ、ローカルで作成し、変更し、問い合わせをし、そして、オントロジを格納する為のフレームワークの一つである。内部的には、SNOBASE は、推論エンジン、オントロジ格納維持機能、オントロジディレクトリ及びオントロジソースコネクタを使っている。アプリケーションは、作成されたオントロジモデルに対する問い合わせを行うことができ、そして、その推論エンジンは、答えを演繹し、JDBC(Java Data Base Connectivity)のリザルトセットに似たりザルトセットを返す事が出来る。しかし、この資料を作成していた時点では、SNOBASE は OWL と互換性は無かった。Sigma ontology development and reasoning system は、完全な設計時および実行時のオントロジ管理システムを装備しており、フリーライセンスで利用可能であるが、SNOBASE と同様、OWL との互換性が無い。

6.3 オントロジ

正しいオントロジを実現する事が、しばしばセマンティックを基盤とするプロジェクトの隘路要素となる。データベース設計の時以上に、オントロジ作成は、高度に特殊化された分野の一つである。熟練した技術者が充分いるか否かだけでなく、オントロジ作成には、ある概念領域を把握するのに必要なオントロジを実現するのにかなりの時間を必要とする。結局、最初から何かを作る代わりに既に使われている既存のものを見直し、活用を図る事が重要である。流用可能な既存のオントロジは、前に述べた名前のオントロジモデリングツールに密接に関係している事が多い。故に、それらのオントロジの活用は、該当するモデリングツールのライセンス状況次第かも知れないが、その努力をする事は無駄ではなく、オントロジ開発の可能性を広げる事になる。

その他のオントロジは、商用目的及び非商用目的に利用可能なフリー且つオープンなものであろう。それらは、Linux、JBoss、Wikipedia、Musicbrainz 及びその他のオープンソースソフトウェアやオープンなデータリポジトリと同様な性質のものである。

現在のオントロジ開発の努力は、スコープや規模において変化に富んでいる。決算報告書又は医療記録の整合の様なローカルな実装の為に開発されたオントロジの主要な役割は、情報の相互互換(構文、構造及び意味の橋渡し)を実現する事であり、ロジックプログラミングはあまり必要とされない。他のオントロジ開発の場合は、広い知識対象領域でビューを共有する事が、計算能力の付加及び知的推論機能を広範囲で増加させる為には重要なので、よりトップダウン的なアプローチが採用されるかも知れない。後者の人々に取っては、任意の数の領域オントロジの為の基盤を提供すると言う構想の基に企業規模の共通上位オントロジを作ると言うのは重要な考え方である。この場合、新たな領域オントロジは、その上位オントロジの拡張であり、且つ、その上位オントロジに準拠する事が可能である。既存のオントロジや昔のデータモデルは、この上位オントロジに対応するものである。理論的には、領域横断的な高度なセマンティック相互互換を達成するのに最小ステップ構成でそれを行うことができる。(しかし、依然として付加的な開発や作業が、このアプローチの実現可能性やスケーラビリティを実証するために必要とされるという事に注意されたい。)

現在幾つかの上位オントロジの候補が存在し、例えば、DOLCE(Gangemi, et al., 2002), Upper Cyc(Lenat, 1995)及び SUMO(Niles and Pease, 2001)があるが、しかし、これ等の何れもマーケットでの確固たる地位を獲得してはいない。これ等の上位オントロジアプローチの支持者達は、米国国防総省及び/又は連邦政府でこれら候補の一つを採択すると信じている。今は産業界がこの技術をフォローする良いチャンスであり、米国はそれが出来たあかつきには、それを標準として ISO に提案するであろう。

領域特有のオントロジが存在しない場所であっても、既存のタクソノミ、XML標準若しくは他の低次のデータモデルを活用することにより、オントロジ開発をジャンプスタートさせることは可能である。連邦政府レベルで、知識管理の作業グループ(<http://km.gov>)は、機関横断型のタクソノミプロジェクトに関する情報共有化に著しい進展を見た。XML.Gov(<http://xml.gov>)は、利用可能文書及びデータのシームレスな共有を可能にする為、機関横断のXMLの効率的且つ効果的な利用機能の実現をミッションとしているが、多くの政府機関が既存のタクソノミを持っているか、あるいは、自分たちの情報領域向けのタクソノミの開発を始めている。例えば、JusticeXMLは興味深いプロジェクトであり、より柔軟なデータモデルを提供するためにRDFやOWLによって拡張され強化されており、連邦政府、州及び地方の法の執行の情報を他の機関がより簡単にアクセスできるようにする道を拓く試みである。

6.4 マッピングツール(オントロジの移植)

一度或るオントロジモデルを作ると、データ(“オントロジ用語”で言うクラスインスタンス)をその中に組み込む事が必要となる。この様なデータを組み込む処理は、通常マッピングツールを用いて色々なデータソースを或るオントロジの中の特定の概念にリンクする事で達成される。”マップ”を作成すると、一つのデータソースの中の或る問い合わせは、それをそのオントロジにマップする事により形式変換され、そして、そのオントロジから他のデータソースにそのデータソース用のマップを用いて形式変換する事が可能となる。従って、その問い合わせの通信データを、他のデータソースに付いて知る事や配慮するデータを持つ事を必要とせずに同様の方法で返す事ができる。すなわち、各データソースは、色々なソースやターゲット間で中心となる支配的

な一つのオントロジに対する独自の”マップ”を持つ。この種の抽象階層を提供する事は、オントロジ作成やデータマップの作成の作業の部分に多少骨の折れる作業を必要とする。しかし、これは、一度実施すれば、各データソースは間違いなく他のデータソースとの実行時の相互互換を実現できる。また、新たなデータソースが組み込まれた場合でも、大部分のケースでは、既にあるデータソースへの影響が全く無いか、影響を少なくする事ができる。

この様な方式は、現在の企業アプリケーションのアプローチとして通常採用されているデータ値のマッピングや意味矛盾の解決処理の数を劇的に減少させる。そのアプローチとは、これまで一般に n の二乗個のマッピング(各データソースから総てのデータソースへのマッピングの数)を必要とするか、そうでなければ、硬直し、柔軟性が無い明示的な標準にエクスポートするアプローチを採用しなければならない。このモデリング及びマッピング方式は、政治的意図を排し、より柔軟にし、且つ、色々な機能を付加し易くする。例えば、ある一つのデータソースに特化された特異なものを、ほぼ透過的に扱う事が可能であるが、その様な特異なものを典型的な標準処理で取り扱おうとすると大変な時間とエネルギーを必要とする。構造化データを処理するのに用いられる多くのツールは、データベースのフィールドからオントロジへのマッピング処理を自動的に行なう機能を持っている。Network Inference と Unicorn とはこの種のツールを提供しているベンダである。収集し、正規化し、そして、非構造化データをオントロジへとマップするツールは、Word、RTF、テキストファイル及び HTML 等色々な非構造化データを通常処理可能である。Semagix は非構造化データに対するツールのリーディングベンダである。

6.5 データ格納

オントロジ及びその他の RDF データモデルは、RDF 専用データストア、若しくは、連想データ技法をサポートする為にカスタマイズされた関係データベースの中に蓄積され得る。RDF 専用のデータストアは、トリプルズ概念をサポートする為のみに設計され、オントロジの格納に効率的且つ独創的な方式を提供できる。RDF 専用データベースは、Tucana Technologies や Intellidimensions 等の会社から提供されている。高品質のオープンソースの RDF データストアも幾つか存在し、Kowari、Redland、Sesame、3Store 等の会社から提供されている。関係データベースを利用する場合、該当するデータベースを多少伝統的でない方法で設計しなければならない。各主要な概念を記述するテーブルの代わりに、一般に、そのデータベースは 4 個のカラムを含んだ単一のテーブルを使用することによってトリプルズ概念を模倣するように設計される。そのカラムの内の 3 個にはトリプル要素(すなわち、主語と述語と目的語)を格納し、4 番目のカラムには、その識別タグを格納するために使用される。(”セマンティックデータと RDBMS とのマッピング”と題した研究報告は、関係データベースへのトリプル格納の実装方法に関し多くのことを知ることのできる優れた情報源である。)

[トリプル(すなわち RDF)の記述、格納及び問い合わせ]対[伝統的な関係手法](更には、実装時にこれ等二つのタイプのどちらを用いたら良いか、また、共用するには如何にしたら良いのか)に関する問題は、業界や市場の中で未だ結論の出していない問題である。格納機能と問い合わせ機能とのそれぞれは、今のところ他では出来ないユニークな機能を提供する。RDF は、将来実行される問い合わせのタイプを予測する事が難しい場合の様な状況にうってつけである。更に、RDF はメタデータの処理や不正確若しくは本質的に異なるデータの横断検索を必要とする問い

合わせを行うのに非常に適している。例えば、“今年のエネルギー生産者状態の‘緑’は、幾つ？”という行の問い合わせは、SQL を使うよりは（一度、色々なデータストアと結びつけたモデルを作れば）RDF の問合せ言語を用いて実行する方が簡単である。反面、SQL を用いると非常に簡単な”今年、最も CO2 の排出削減を行ったエネルギー生産者は誰”と言うような問い合わせは、RDF 問い合わせ言語を用いた場合、非常に複雑になることもある。

RDF 問い合わせ言語は、未だ、発展を続けている言語であり、上記の様な制約を超えて広がるかも知れない事に注意されたい。RDF の他の制約は性能に関する問題である。問い合わせは用語だけでなく概念を含むように拡大するので、検索空間が劇的に増大する。RDF データの格納技術は比較的新しく、また、実装事例も少ないので、システム開発者は、性能にネガティブな影響を与える問い合わせと機能とに対して特別な注意を設計時に繰り返し払う必要がある。業界成長の観点から見ると、RDF がデータベース業界に如何なる影響を及ぼすかを見通すことは難しい。RDF データ格納機能は、独自のデータ格納カテゴリとして存続するかも知れないし、または、オブジェクト指向データベースが関係データベースの中に吸収された様に、その機能が関係データベースの中に吸収されてしまうかも知れない。

6.6 仲介エンジン

仲介エンジンとは、形式変換コードを書く代わりにモデルを用いる事により、異なる構文、構造および意味のデータの動的な形式変換を行う自動化ツールである。仲介エンジンは如何なる相互交換アーキテクチャに於いても重要な要素である。仲介エンジンは、データマップ、オントロジや概念モデルの他の形式を利用する、多様なデータセット間の抽象階層となる実行時プロセスであり、その抽象階層により、色々な組織が、データ及び情報に関する統一標準を作らなくてもすむ事を可能にする。通常、仲介エンジンは、高度に構造化されたデータを使って実行される。非構造化データ又は半構造化データの場合は、先ず、仲介マップを作るのに必要な或るスキーマに連結されなければならない。(Pollock, 2004)

6.7 推論エンジン(Inference Engines)

推論エンジン（時々、リーズナ(reasoners)と呼ばれる）は、既存の情報から新たな事実や関係を導き出すソフトウェアツールである。推論エンジンは、推理によってある結果に到達する人間の能力を模倣しているとしばしば言われる。実際には、推論とは、神秘的な人工知能の能力などではなく、むしろ、データ処理における極めて普通のアプローチである。複雑なデータマイニングを行なう事は、推論の1つの形と考えることができる。情報と関係とのモデルを作ることによって、推論エンジンは、そのモデルに基づく論理的結果を導き出す事ができるようになる。推論の良くある例として、人々や他の人々との関係のモデルを用いる事で、新しい知識を得る事がある。この種のネットワークグラフ(モデル)の探求は、明示的に定義されていなかった関係を知る事を可能にするであろう。

RDF と OWL とを用いた推論は、該当するモデルの中に表現された関係に限定されていること(主に、遷移的関連を推論する手段であること)に注意されたい。しかし、ルール言語やロジック言語を付け加える事により、概念理解、学習及び適応に於ける大いなる飛躍を行うことが可能であるが、しかし、この種の能力の実装事例は未だ少なく、実現からは程遠い。

フリーや商用の推論エンジンが幾つか存在する。例えば、Jena は、HP の研究所で開発されたセマンティック Web アプリケーションを開発する為のオープンソースの Java フレームワークであり、推論エンジンサブシステムを有している。Jena の推論エンジンは、OWL Full のサブセットである OWL Lite 用及び RDFS 用のルール集合と一緒に構成された推論エンジンをベースとする汎用のルールを有している。JESS は、カーネギーメロン大学で開発されたポピュラーな OWL 推論エンジンである。また、Network Inference 社は、記述論理(OWL-DL)に基づく商用の推論エンジンを提供している。

6.8 その他のコンポーネント

普通のウェブページは、実体情報の良い情報源の一つである。オントロジ生成の為の多くのツールは、ウェブページに注釈を付ける事でオントロジを生成している。W3C の Annotea プロジェクトは無料の注釈ツールを提供している。Semagix, Siderian Software, Entopia 等の商用ベンダは、情報を分類したり、進んだ検索機能や案内機能を提供したりする為に、オントロジを使った商品を提供している。

7.0 セマンティック技術のアプリケーション

セマンティック技術は、既存技術ではどんな犠牲を払っても解けない問題を解くことができる。

Don Hall

セマンティック技術が、主要な役割を果たす事の出来る多種多様なアプリケーションが存在する。それらの核心は、セマンティックアプローチを他のキー技術と連携させた時、それが計算処理の次の波を演じる基盤機能であることである。数年先の視点で見ると、今まで捉え所の無かった真の適応コンピューティングを実現するのに、セマンティック技術が IT 産業の救世主になることが約束されている。注意して見ると、その未来は既に起きつつある。私企業や政府機関は、既存のセマンティックデータ格納機能、オントロジ、ツールセット及びアプリケーションを用いながら実用レベルのプログラムを実装しつつある。これ等の短期的プロジェクト領域の幾つかは、セマンティック Web サービスや意味相互互換や知的検索を包含している。

7.1 セマンティック Web サービス

Web サービスとは、ネットワークを介してマシンツーマシン相互連携の相互互換を実現する為に設計された一つのソフトウェアシステムである。Web サービスは、Web サービス記述言語(WSDL)を用いて機械処理可能な形式で記述されたインターフェースを有する。WSDL、UDDI 及び SOAP の組合せは、マーケット全体をサービス指向アーキテクチャ(SOA)に向ってシフトさせる為の記述の三つ組み(triad)を形作っている。これ等の技術は、一体となって、HTTP 又は SMTP ネットワーク上でディレクトリ、コンポーネント検索及びサービス交換プロトコルを実現する。

マイクロソフト、IBM 及び他の大手ソフトウェアベンダは、Web サービスのモデルの根底をなす概念と言語とを採用し、更に、サービス指向アーキテクチャを採用する事の利益を解説する

書籍や業界記事が増えている。しかし、Web サービスに欠点が無い訳ではない。セキュリティ問題は、長い間懸念されていたが、これ等の問題に対処する為の幾つかのソリューションは、ここ数年の間に導入されたばかりである。多分、Web サービスに関して、最も重要な改善すべき領域は、(a)柔軟な検索機能や発見機能と(b)情報管理機能やスキーマ変換機能とに関するものである。基本的に、Web サービス技術とは、疎結合方式でメッセージを操作するものであり、記述用語の違いをブリッジするものではなく、また、送信されて来たメッセージを受信者が個別に理解する事を可能にするものでも無い。Web サービスでは、その交換に関する部分は、利用者のコードソリューションに依存、及び/又は、大きなコミュニティの中のある種の文書交換標準の合意に依存している。(後者の目的が達成されたケースは、稀である)

柔軟な発見機能やサービス開始機能の実現の難しさ(更には、Web サービスでの情報交換機能のシームレスな運用)は、W3C にセマンティック技術を取り込んだセマンティック Web サービスイニシアティブを作る事を余儀なくさせた。セマンティック Web サービスは、OWL-S(Web Ontology Language Service specification)により強化された Web サービスであり、Web サービスに於けるサービス開始やサービス記述を行う為の柔軟なフレームワークを実現できる。OWL-S は、Web サービスプロバイダが Web サービスの特徴や機能を明白、且つ、コンピュータ理解可能な形式で記述する為のコアのマークアップ言語である。Web サービスを OWL-S によりマークアップする事は、Web サービスの発見や実行や相互操作や合成や実行監視の自動化などの Web サービスの自動実行を可能にする。次節以降、マークアップ言語に於ける階層アプローチについて説明する。現在の OWL-S は、W3C 標準である OWL の上に構築されたものである。

7.2 意味相互互換

行政管理予算局(OMB : Office of Management and Budget)は、連邦政府全体に於けるビジネスプロセスと情報技術との統一及び簡素化の指示と指導の呼びかけとを行った。これを達成する為には、各機関は、各機関が持つ情報を連邦政府内で共有できるようになっていなければならない。しかし、各機関が共に働く環境には、多種多様な用語や定義があるので、この情報共有の為には、異なる文脈の中のデータの意味を理解できる必要がある。セマンティック技術は、分散データを結び付ける事や文脈依存の中でデータを記述する為のフレームワークを提供する事でこの問題を解決できる。

公式な説明では、セマンティック技術は交換されている情報の論理的性質や文脈を記述する事ができ、且つ、通信者間の最大限の独立性を保証し得ると言う。その結果、ビジネスロジック、プロセス及びワークフローに関係無く情報ドメイン間の高度な透明性やよりダイナミックな通信が可能となる。(Pollock and Hodgson, 2004)

その技術構想は、柔軟でないプログラム又はコードで無い柔軟な情報モデルを実現する事であり、膨大な量のミッションクリティカルなデータの共有の為の動的且つ自己修復できる新たな基盤を実現するのに用いる事である。タクソノミ技術やシソーラス技術や文脈モデル化アプローチ技術や推論技術やオントロジドリブンの相互互換技術の最近の進歩により、それらの技術を、複数の知識コミュニティに分散している情報の分散管理方式の動的変化に対応するフレームワー

クに適用できる様になった。(Pollock and Hodgson, 2004)

NASA は、意味相互互換技術を特定の形式或いは語彙の標準化や唯一つのモデルに準拠する中核データベースを作る事無く総ての関係者に必要な情報を共有させるのに極めて有望な技術と見ている。一つの例として、NASA は、スペースシャトルに於ける配線劣化システムに関する重要且つ稼働中の保守問題に対処するのに、この概念を使用している。既存の配線システムデータベースは、部品の仕様、伝票、設計図、注文変更、標準手順、テスト手順、テスト報告書、調査報告書、故障原因と報告情報、作業依頼及び修復処理文書に関する情報を含んでいる。数 10 の雑多なデータベース(技術的作業を支援する)と雑多なシステム(設計作業を支援する)とが、スペースシャトルプログラムを支援している幾つかの契約企業間に分散したデータを処理する為に NASA の中で使われている。配線問題の処理には、沢山の組織横断システム、データベース及び知識リポジトリにタイムリーにアクセスする事が必要であり、その範囲は広大である。フライト異常時の診断やトラブル処理は、特にクリティカルであり、タイムリーな解決はミッションクリティカルと言うだけでなく命に関わる問題でもある。多くの関係者が必要なデータにアクセスし易くし、且つ、より有用なデータとする為の作業は、未だ、始まったばかりであるが、この章の最初の方で強調した様にセマンティック技術は、既存のデータ統合アプローチでは解くことのできない問題の多くを処理する為の有望な方法の一つであると言われている。

その様なプロジェクトで如何にセマンティック技術を活用できるか要点を以下に述べる。

設計時ツールは、或る特殊な領域を含む RDF 及び OWL モデルの開発に用いられる。これ等のモデルは、既存の XML 標準か他の方法で定義されたものに基づいているかも知れない。そして、他の設計時ツールを、特定のデータ表現からこれ等のモデルに柔軟にマップするのに使う事で、或るデータ標準にコンバートするアプリケーションをわざわざ作る必要性を無くする。故に、実行時処理は、これ等のモデルを使う事で、ソースからターゲットへデータの形式変換を行ったり、又は、独自の問い合わせステートメントを連携用問い合わせとして実行したりする中心軸に変換できる。このタイプの意味相互互換フレームワークは、構文、構造及び意味の差異をよりうまく解決する為の堅固な基盤を提供可能であり、将来、色々な組織が、現状の操作メソッドを変更する事無くデータを共有したり相互連携したりする事を可能にする道を拓くであろう。

意味相互互換アプローチの重要な利点の一つは、既存の統合技術やデータベースやソフトウェアアプリケーションを置換する必要が無い事である。種々のセマンティックベースのコンポーネントとアプリケーションインターフェース(APIs)とから構成されるセマンティックフレームワークは、Web サービス若しくは伝統的なミドルウェア APIs と共に配備する事ができ、既存の基盤投資を強化できる。その様な場合でも、仮想的な集中問い合わせや変換やビジネス規則メタデータにより大きな利益を得る事ができる。そのビジネス規則メタデータはネットワーク基盤のパイプを流れる事が出来る。同様に、そのソフトウェアは、設置作業の少なさ、コーディングの最小化及び最大限の再利用性により、顧客の既存 IT エコシステムに適合する事ができるであろう。(Pollock and Hodgson, 2004)

7.3 知的検索

意味相互互換性に非常に関係の深いものに知的検索の領域がある。前に述べた様に、意味相互互換技法は、或るシステム固有の問い合わせを他の異なるシステムと連携する為の問い合わせに

マップする事を可能にする。この場合、普遍的な問い合わせ言語へ変換するシステムの必要性を無くし、そして、現在のデータ形式を使い続ける事を可能にする。データソースの上を仮想的な層で覆う事により、問い合わせを普遍的な方法で定義でき、マップ先の総ての資産にアクセスする事が可能となる。この様に連携した検索は、より意味的に正確な検索が可能になることで、より賢い検索と成り得る。すなわち、検索機能に概念を包含する様に拡張でき、そして、特定のキーワードのみ含む場合よりも検索範囲を絞る事ができる。この様な検索の深さ、若しくは、木目の細かさは、利用者の欲する検索を可能ならしめる。

知的検索の他の側面は、情報の識別と関連性とを利用する事によって、人の検索に関して、より関係を意識した検索を可能に出来る事である。人々の間及び人々に付いての情報間の関係性は、より高度な適切性及び信頼性を判断する鍵となる。知識管理システムにお金を投じているにも拘らず、多くの人々は、エキスパートを探したり或いは信頼性の高い情報を見つけたりするのに、友達、身近な人、同僚等の人のネットワークに依存している。個人間の関係性は、販売活動や多くの組織間の相互連携を行う場合にも有用である。社会ネットワーク及びソフトウェアは、この種の広域利用を可能にするものである。

この様な情報が大規模にどの様に用いられるかの事例が、ある電話会社に於いて見られる。この電話会社はより知的な電話番号検索サービスを提供する為の技術として研究している。そこでは、その電話会社は、一般的な名前の表を提供する代わりとして、より知的な一致サービスを提供する為に、人と検索対象の人に関係する可能性のある名前の表とから成る混合情報の検索を行っている。例えば、社会ネットワークの関係性の推論は、(FOAF(friend of a friend)をよく調べる事により)或る人が他の或る人を知っているか、又は、知っていることを期待できるか否かの情報を明らかにできる。位置や学校の出席状況や元の職業や現在の職業といった他の情報は、一致を推論するのに利用する事ができる。確かに、多くの人が信じている様に個人情報に関する重大な問題を内在させているが、しかし、個人識別可能な情報のハッシングや暗号化の技術と漸進的な告示技術とにより多くの個人情報に関する懸念を解決できるであろう。

知的検索を可能にするセマンティックアプローチは、知識管理システムへと進むべき道を見つけ始めている。現在の知識管理システムは、彼らの領域内にのみ止まり、組織境界を横断する事は難しい傾向にあるが、知的検索技法を既存の情報基盤に覆いを掛ける形で付け加えることにより、物理的なデータ形式、知識領域及び組織構造の違いを橋渡しする事ができる。

8.0 本資料以降の話題

セマンティック技術及びセマンティック Web に関する追加の話題を探求する為の追加モジュールを作成中である。この白書の第二モジュールでは、連邦政府内の意味相互互換の為のビジネスケースを詳しく調べるであろう。何故、そして、どの様に意味相互互換が ROI を生み出すかを記述するのにビジネスシナリオとストーリーボードアプローチとを用いるであろう。これ等のビジネスアプローチは、ビジネスとアーキテクチャ用語の中に書き表されたビジネス問題の詳細な記述を含むであろう。

第三モジュールでは、セマンティック技術の利益を得、そして、セマンティック Web の実装開発を始める為の政府機関用のロードマップを提供する。新しい技術、アプリケーション及びサービスは、これ等の新しい進歩の利益を得る為に開発されている。このモジュールは、手順と実装推奨とを提供するであろう、これを参考する事により、政府機関は、彼らの進行状況を照らし合わせながら、且つ、採択時間を最適化すると共に開発摩擦を最小化しながら将来のプロジェクトを立案できる。

9.0 参考資料

Bedford, Denise. "Charter Statement of Taxonomy and Semantics Special Interest Group." 2004.

<<http://www.km.gov>>

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." ScientificAmerican.Com, May 17, 2001.

<

http://www.scientificamerican.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 >

Berners-Lee, Tim. "Semantic Web: Where to Direct Our Energy?" Invited Talk at the International Semantic Web Conference (ISWC), October 2003.

<<http://iswc2003.semanticweb.org/>>

Berners-Lee, Tim. "Semantic Web Road Map." W3C. September 1998.

<<http://www.w3.org/2003/Talks/0521-www-keynote-tbl/> >

Berners-Lee, Tim. "Web Services – Semantic Web." W3C. 2003.

<<http://www.w3.org/2003/Talks/0521-www-keynote-tbl/> >

Berners-Lee, Tim. "What the Semantic Web Can Represent." W3C. September 1998.

<<http://www.w3.org/DesignIssues/RDFnot.html>>

Daconta, Michael C., Leo J. Obrst, and Kevin T. Smith. The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. Indianapolis: Wiley, 2003.

Daconta, Michael C. "10 Practical Reasons Why You Need an Ontology."

<<http://www.daconta.net/briefs/why-ontology-general.pdf>>

Daconta, Michael C. "The Semantic Web Foundations of Net-Centric Warfare."

<<http://www.daconta.net/Articles/Net-centricWarfareWhitepa.html>>

Gangemi, Aldo, Nicola Guarino Claudio Masolo, Alessandro Oltramari, and Luc Schneider. "Sweetening Ontologies with DOLCE." EKAW 2002, Siguenza, Spain, October 1-4, 2002, Proceedings. Lecture Notes in Computer Science 2473. Springer, 2002.
<<http://www.loa-cnr.it/Papers/DOLCE-EKAW.pdf>>

Fromm, Kenneth. "The Semantic Web in the Enterprise - EAS Speakers Shed Some Light on Semantic Technologies and Their Roles in Web Service and XML Frameworks." Enterprise Architect, Summer 2004.
<<http://www.ftponline.com/ea/default.aspx>>

Fromm, Kenneth, and Jeffrey T. Pollock. "Semantic Computing's Building Blocks." Enterprise Architect, April 15, 2004 (Vol.2, No.8).
<<http://www.ftponline.com/ea/default.aspx>>

Lenat, Douglas B. "CYC: A Large-Scale Investment in Knowledge Infrastructure." Communications of the ACM, November 1995 (Vol.38, No.11).

Masum, Hassan, and Yi-Cheng Zhang. "Manifesto for the Reputation Society." First Monday, July 2004 (Vol.9, No.7).
<http://www.firstmonday.org/issues/issue9_7/masum/index.html>

Niles, I., and A. Pease. "Towards A Standard Upper Ontology." Proceedings of Formal Ontology in Information Systems (FOIS 2001), October 17-19, Ogunquit, Maine, USA, pp 2-9. See also <<http://www.ontologyportal.org>>.

Obrst, Leo. "Ontologies for Semantically Interoperable Systems." MITRE, Center for Innovative Computing & Informatics. Presentation to the KM.Gov Semantics Interoperability Community of Practice. April 14, 2004.

Obrst, Leo. "Ontologies and the Semantic Web: An Overview." MITRE, Center for Innovative Computing & Informatics. July 13, 2004.

Obrst, Leo J., Howard Liu, and Robert Wray. "Ontologies for Corporate Web Applications." AI Magazine (Fall 2003): pp. 49-62.
<http://www.findarticles.com/p/articles/mi_m2483/is_3_24/ai_110575583/>

Pease, A. "The Sigma Ontology Development Environment." In working notes of the IJCAI-2003 Workshop on Ontologies and Distributed System, Heiner Stuckenschmidt, ed, Acapulco, Mexico, August 2003.

Pollock, Jeffrey T, and Ralph Hodgson. Adaptive Information: Improving Business Through

Semantic Interoperability, Grid Computing, and Enterprise Integration. Wiley-Interscience, September 2004.

Pollock, Jeffrey T. "The Evolution of Integration to EAI, EII, and Onward to Semantics." ManTech EII Workshop, July 2004

<www.meiim.com/files/Jeffrey_Pollock.ppt>

Sonntag, William. Submission to "Problem Statements for Semantic Technology Panels – Interactive Discussion Session" in the One-Day Conference on "Semantic Technologies for eGov" White House Conference Center, Monday, September 8th, 2003. United States Environmental Protection Agency, Office of Environmental Information.

<<http://www.topquadrant.com/documents/Sept%208th.%20-%20Collected%20Problem%20Wner%20Statements-less%20DIA,%20DCMA.pdf>>

TopQuadrant. "Harnessing the Value of Semantic Integration For Your Business." TopQuadrant Whitepaper. June 15, 2004.

TopQuadrant. "Semantic Technology, Version 1.2." TopQuadrant Technology Briefing. March 2004. <http://www.topquadrant.com/documents/TQ04_Semantic_Technology_Briefing.PDF>

Udell, Jon. "Collaborative Knowledge Gardening." Infoworld.com, August 20, 2004.

<http://www.infoworld.com/article/04/08/20/34OPstrategic_1.html>

W3C. "Resource Description Framework." World Wide Web Consortium. August 2004.

<<http://www.w3.org/RDF/>>

W3C. "Web Ontology Language (OWL)." World Wide Web Consortium. August 2004.

<<http://www.w3.org/2004/OWL/>>

10.0 注釈

¹ The World Wide Web Consortium's Semantic Web Activity Statement:

<http://www.w3.org/2001/sw/Activity_intro>

² The World Wide Web Consortium's Semantic Web Activity page:

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

³ Ralph Hodgson, Semantic Web in the Enterprise Panel, Enterprise Architect Summit, June 2004.

⁴ 用語“リッチデータ”は、その用語が存在するシステム又はアプリケーションに対して独立で且つ正確な記述を行っているデータに対して用いられる。用語“スマートデータ”は、Michael Dacont やその他の人々により普及させられた用語である。“スマートデータ”は、時々“リッ

チデータ”と同様に使用されるが、その定義には異なる条件を持ち、そして、推論と高次処理を可能にする論理構造を包含している。それらの定義の微妙な点を理解するよりは、読者は、それらにより実現されるより高度な自律性や自己記述性に注意すれば良い。それらの多くは、それらのデータのネイティブなシステムやアプリケーションの外で利用される。

- 5 Tim Berners-Lee, Semantic Web Status and Direction ISWC2003 Keynote, October 2003.
<<http://www.w3.org/2003/Talks/1023-iswc-tbl/slide10-0.html>>
- 6 Information about Dublin Core:
<<http://www.dublincore.org/about/>>
- 7 ISO 16642:2003: Computer Applications in Terminology – Terminological Markup Framework
<<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=32347&ICS1=1&ICS2=20&ICS3=>>>
- 8 ISO/IEC 11179: Information Technology – Metadata Registries
<<http://metadata-standards.org/11179/>>
- 9 Information about PRISM:
<<http://www.prismstandard.org/>>
- 10 Information on OMG’s Meta-Object Facility (MOF):
<<http://www.omg.org/technology/documents/formal/mof.htm>>
- 11 Information on OMG’s Common Warehouse Metamodel (CWM):
<<http://www.omg.org/cwm/>>
- 12 ISO 19115:2003: Geographic Information – Metadata
<<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70>>
- 13 Information on the Federal Geographic Data Committee (FGDC) and the Content Standard for Digital Geospatial Metadata (CSDGM):
<<http://geology.usgs.gov/tools/metadata/tools/doc/faq.html>>
<<http://www.fgdc.gov/metadata/contstan.html>>
- 14 Information about Creative Commons and Creative Commons metadata-supported formats:
<<http://www.creativecommons.org/>> and
<<http://creativecommons.org/technology/usingmarkup>>
- 15 ワールドワイドウェブ及びセマンティック Web の中に於ける利用者やデータと言う文脈の中で信頼及び信頼性は、急速に発展している概念である。伝統的な IT 定義では、しばしば、保全性(完全性)、タイムラインレス(流通、更新)、信頼性又は正確性の様なデータの性能的性質であると述べられてきた。現実世界での実装では、次第にデータプロバイダに対するお役立ち度(カルマポイント)の評価点(Slashdot)や利用者による評価(eBay)やデータプロバイダ若しくはデータ利用者の公開あるいは非公開の調査による評判を考慮するようになってきた。社会及び組織の仕組みは、大規模分散協業開発で高品質のアウトプットが得られるよう発展

を続けている。Linux に於けるオープンソース開発や Wikipedia に於けるオープンコンテンツ作成が、この仕組みの実例として先ず挙げられる。

- 16 An XML registry for the exchange of environmental data can be found at:
<[http://oaspub.epa.gov/emg/xmlsearch\\$.startup](http://oaspub.epa.gov/emg/xmlsearch$.startup)>
- 17 Information about XML.gov XML Registry/Repository efforts:
<<http://xml.gov/registries.asp>>
- 18 SPARQL Query Language for RDF:
<<http://www.w3.org/TR/2004/WD-rdf-sparql-query-20041012/>>
- 19 James Hendler, Closing Keynote, Semantic Technologies for E-Government, September 2003.
- 20 友達や家族に説明している Flickr とは、デジタル写真を簡単にアップロードし、そして、共有する方法である。更に、del.icio.us もウェブのブックマークに限定した同様のサービスを行っている。CTO に理解して欲しい事は、両者とも或るアイテム群に対するメタデータ語彙を開発し、メタデータドリブンの問い合わせを実行し、調べたい部分にモニタリングを変更可能な共有データベースの為の協業システムであると言う事である。Flickr の場合、そのアイテムは写真であり、del.icio.us の場合、そのアイテムとは URL である。しかし、そのメソッドは我々が日常作成し、見つけ、利用するデジタルな物を共有する為にも流用できる物である。(Udell, 2004)
Flickr に関する詳細な情報は、次のページを参照されたい。:
<<http://www.flickr.com/>>
- 21 この白書は、オントロジ開発に必要なツールに言及しているが、しかし、オントロジ及びその他の論理モデルと一緒にそれらがどのような処理を行なうか、また、どの様に使いこなしたら良かに付いては言及していない。これ等に言及する事は、モデリングツールやマッピングツールの現状は、人間により要求されるものより劣る思考過程のオントロジの生成や利用の大自動化を図っているとの間違った印象を与えるかもしれない。これは正しくない。この白書の最初の方で述べた様に、セマンティックモデルの利用技術は、フラットファイルのデータベースから関係データベースへ進化した、或いは、手続き型プログラミング技法からオブジェクト指向アプローチへと進化したのと同様に進化するものである。この様な言い回しとセマンティックベースのアプローチのアーキテクチャとを人々が理解するには多少時間が掛かるかもしれない。また、特にモデリングやオントロジのマッピングやデータ構造に関するツールが円熟するには時間が掛かるであろう。
- 22 Mapping Semantic Web Data with RDBMSes
<http://www.w3.org/2001/sw/Europe/reports/scalable_rdbms_mapping_report/>
- 23 Don Hall, Program Director, Logistics Enterprise Support Program, in support of the Assistant Deputy Under Secretary of Defense for Logistics Systems Management, in numerous conversations and presentations.
- 24 From W3C Working Note on Web Services Architecture:
<<http://www.w3.org/TR/ws-arch/#whatis>>

- ²⁵ Universal Description, Discovery, and Integration (UDDI):
<<http://www.uddi.org/>>
- ²⁶ Simple Object Access Protocol (SOAP):
<<http://www.w3.org/TR/soap/>>
- ²⁷ For additional background, see “Semantic Discovery for Web Services” in Web Services Journal, April 2003.
<<http://www.sys-con.com/webservices/article.cfm?id=507>>
- ²⁸ Extracted from NASA presentations and reports.
- ²⁹ ハッシングが、或る文字列(例えば、ユーザ名及び認証情報等)を数学的サマリーに変換するのに使うある種のアゴリズムを用いている事を意味する。従って、そのサマリーは、色々な公開鍵暗号システムを用いて暗号化される。

付録 A: SICoP のチャーター

SICoP(The Semantic Interoperability Community of Practice)は、政府分野の中の意味相互互換及び意味データ統合の実現を目指す人々のグループにより設立された。SICoP は、オンライン対話、会議、チュートリアル、コンファレンス、パイロットプロジェクト及び最適事例の開発と普及を目的とするその他の活動を通じて意味相互互換性を可能にする為(特に、これらの技術とアプローチとの運用化)の方法を探究する。参加者は、広範囲にわたる政府組織及び業界更にはこれ等の活動を支援する大学の協力者からなる実施コミュニティを結成する。しかし、SICoP は如何なる組織の代表に依って公式或いは非公式に認可されたもので無い事を宣言する。

SICoP は、XML 作業グループと協力関係を持つ CIOC(Chief Information Officers Council)の最良実践諮問委員会(the Best Practices Committee)をスポンサーとした知識管理作業グループ(KMWG: Knowledge Management Working Group)の SIG の一つである。他の組織との関係で見ると、SICoP 及びその親である KMWG は、総ての連邦組織、利用者及びパートナーに対して政府の知的資産の恩恵をもたらす為に政府省庁間の仲介役を果たす。SICoP は、KMWG を通じて最良実践諮問委員会に、SICoP の行動並びに決議を伝えるであろう、その伝える目的は、CIOC 及びそのメンバ機関が、セマンティック技術を導入する場合に SICoP のメンバを支援する事及び各機関内のセマンティック Web 能力を発展させる為である。

付録 B: 用語集

用語	定義/説明	出典
アプリケーション プログラムインタ ーフェース(API)	一つのアプリケーションプログラムインターフェース (API)とは、或るコンピュータソフトウェアの部分が他のものと通信する為の方法の定義集合である。それは、通常(しかし、必須条件では無い)、より低位レベルソフトウェアやより高位レベルソフトウェアを抽象化するための手段の一つである。	Wikipedia, 自由コンテンツ百科事典 < http://www.wikipedia.com/ >
ブログ(Blog)	<p>ウェブログ(weblog)又は略してブログとは、ウェブアプリケーションの一種であり、普通のウェブページ上に定期的に更新され、新しい記事を上にしたコンテンツを有する。この種のウェブサイトは通常、任意のインターネット利用者がアクセスできるようになっている。ブログという言葉は、サーバログと誤解されるのを避けるためにだんだん普通に使われるようになった。</p> <p>ブログは個人の日記から政治宣伝の手段、メディアプログラム及び企業にまで広がり、更に、たった一人の時々記事を書く人から大勢のライターコミュニティまで広がっている。ウェブログ若しくはブログ関連ウェブを総称して、通常、ブログスフェア(ブログ天体)と呼んでいる。</p> <p>ブログの形式は、単純なハイパーリンクの表から、利用者が付加したコメントや格付け情報を持った記事の要約まで色々ある。個々のウェブログのエントリは、通常ほぼ総てに日付と時刻とがあり、ページのトップには最新の記事が来るようになっている。ウェブログではリンクが重要なので、大部分のブログは、古い記事を保存し、個々のエントリに対して固定のアドレスを付けている。この固定アドレスのリンクは永久リンクとして参照に使われている。最新の(ハイパーリンクと要約と一緒の)ヘッドラインは、RSS XML形式でウェブログに投稿され、RSS フィードリーダーで読まれる。</p> <p>ウェブログは、しばしばコンテンツ管理システム若しくは CMSにより実行される。</p>	Wikipedia < http://www.wikipedia.com/ >
統制語彙 (Controlled Vocabulary)	タクソノミのカテゴリの中で用いられる標準用語の有限集合である。或る組織の中で、複数の統制語彙が存在し得る。例えば、組織全体の為のコア統制語彙とその組織の中の各ビジネスユニット固有のサブ統制語彙。統制語彙はタクソノミのカテゴリ、情報カ	“Taxonomy Analytical Brief”, Department of State, IRM

	タログ化及びタグ付けに記述に使う事ができる。更には、ウェブサイトの案内インターフェースの為のラベルの記述にも使う事ができる。	Business Center, May 27, 2003
データ	生の事実(raw facts)、指示又は孤立した記述の集まり。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
情報	関連する事実、指示若しくは与えられた文脈(例えば、特定の場所や時刻)の中で何かに付いての記述の集合であり、その真/為又は価値の判明していないもの。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
情報アーキテクチャ	広く理解されている情報アーキテクチャとは、簡単に言えば、情報ニーズを情報リソースに合致させるのを助ける手段の集合である。良く作られたアーキテクチャ設計は、特定の形式、カテゴリ及び関連性を通じて組織内の情報を構造化する。その為には、ビジネスコンテキスト、コンテンツ(情報)及び利用者を勘案する必要がある。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
知識	関連する事実、指示若しくは与えられた文脈の中で何かに付いての記述の集合であり、その真/為及び価値の確定しているもの。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
メタデータ	メタデータの簡単な定義は、データに付いての”構造化されたデータ”である。 メタデータとは、物理的か電子的かに関係なく物或いはリソースに付いての記述情報である。メタデータは比較的新しいものであるが、メタデータの背景となる概念は、組織化された情報の集まりとして昔から使われてきたものである。図書カードカタログは、よく使われているメタデータの例であり、過去数十年の間コレクション管理やリソース発見のツールとして使われてきた。メタデータは、人手またはソフトウェアを使って自動的に作る事ができる。	Dublin Core Metadata Initiative, Frequently Asked Questions < http://dublincore.org/resources/faq/whatismetadata >
名前空間	多くのプログラミング言語に於いて、名前空間は識別子に対する背景規則となっている。	Wikipedia < http://www.wikipedia.com/ >

	<p>一般に、名前空間とは、名前若しくは技術用語若しくは語句を使う事が有効となる抽象領域である。一つの名前空間は名前の集合をユニークに識別し、それにより、名前は同じだが出所の異なるものを使った時、曖昧なく区別できる様にする。ある名前空間の中では、各々の名前はユニークでなければならない。</p> <p>名前空間は背景規則となり、且つ、名前空間の中の各々の語句は実世界の概念をユニークに示す(写像する)ことが可能となる。</p> <p>自然(民族)言語、人工言語、ある職業の技術用語集、方言、社会方言若しくはコンピュータ言語(例えば、プログラミング言語)であるか否かに拘らず、各言語は一つの名前空間である。</p>	
<p>オントロジ</p>	<p>オントロジとは概念化の仕様である。</p> <p>知識共有の文脈の中では、オントロジと言う用語を概念化の仕様と言う意味で用いる。オントロジは、或るエージェント或いはエージェント群の或るコミュニティの為に存在し得る概念及び関係性の記述(プログラムの形式仕様と同様)である。この定義は概念定義の集合としてのオントロジの活用法と矛盾していない、しかし、より一般的には、哲学の中で用いられているオントロジと言う言葉とは、その意味が異なる事は確かである。</p> <p>重要な事はオントロジが何の為のものかである。私の同僚や私は、知識共有と知識の再利用とを可能にする為に幾つかのオントロジを設計して来た。その様な状況の中では、オントロジとは、オントロジ的な約束事を作る為に使われる一つの仕様である。オントロジ的な約束事の形式定義とは以下の様なものである。</p> <p>実用的理由により、我々は形式語彙の定義集合としてオントロジを書く事を選択した。であるが、これは概念化を規定する方法だけでなく、AI ソフトウェア間で知識共有するのに適した特性を有している。(例えば、読者や文脈からの意味的独立性)</p> <p>実際、オントロジ的な約束事とは、オントロジによって指定された理論に沿った(しかし、完全にでは無い)方法での語彙の利用法に関する合意(例えば、問い合わせを出すとかアサーションを作るとか)である。</p> <p>我々はオントロジに束縛されるエージェントを作るかも知れない。我々はオントロジを設計する事で、我々はこれ等のエージェント間またはそれらと知識を共有する事が可能となる。</p>	<p>Tom Gruber, Stanford University <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html></p>
<p>スキーマ</p>	<p>スキーマと言う言葉は、ギリシャ語の"σχῆμα" (schema)に由来し、それは、"形"、若しくは、より汎用的には、"計画"を意味する。スキーマと言う言葉は、以下の様な異なるものを表現する事ができ</p>	<p>Wikipedia <http://www.wikipedia.com/></p>

	<p>る。</p> <ol style="list-style-type: none"> 1. コンピュータ科学においては、スキーマとはモデルである。 2. 形式論理においては、ステートメントの集合(通常、無限集合)を描写する規則(通常、再帰的に定義可能)である。 3. データ構造の記述、又は、データベースの定義された部分。ソフトウェアアーキテクチャを参照すると、概念スキーマ、Sowa の概念グラフ、意味ネットワーク、Berners-Lee のセマンティック Web。 4. XML スキーマは、構造、内容及び拡張部分、XML 文書の意味等の定義手段を提供する。 5. Z 形式仕様言語で書かれた形式仕様の部分。 6. 或る小さく且つ特化されたオントロジ(すなわち、質問に対する表)は、どれが、当該対象世界の中に何が存在するかを記述しているかに答え得る。この場合、幾つかの狭い範囲のアクションに何が要求されているかのみを記述している。例えば、図書カードカタログのスキーマは、図書館利用者が、書籍を読みたいか否か、そして、もしも読みたいならば、その書籍をどの様に見つけられるかを補助する為に必要な当該書籍に関する十分な情報の提供のみを必要とする。これとは逆に、或るオントロジは、広い範囲のアクションを包含する。これ等の総ては、通常、取引や専門領域に関係している。 	
セマンティックス	<p>セマンティックスとは、元々タクソノミを利用する、若しくは、作成する手順である。セマンティックスが無ければ、タクソノミは単純又は凝ったものであっても構造の無いものとなる。公式には、セマンティックスは、意味の研究や意味の変更や文又は語句とそれらの意味との間の関係を律する原理を扱う言語学の一分野である。単純なタクソノミの意味生成に必要なセマンティックスは、ネットワーク或いはファセットタクソノミの意味生成に必要なセマンティックスとは異なる。セマンティックスは、サインやシンボルの間の関係性の研究を含む。情報と言う観点では、セマンティックスは、更に、言語や情報のすり替えや情報組織や情報発見に於ける及びそれらの間の効果的な情報通信を包含している。</p>	<p>タクソノミとセマンティック SIG のミッションステートメントより抜粋</p> <p><http://km.gov/></p>
意味統合(Semantic Integration)	<p>意味統合は、しばしば意味相互互換の同義語として用いられて来た。しかし、幾つかのベンダは既存の XML 統合の上に構築する包括性の少ないソリューションを指すのに用いて来た。</p>	<p>編集者</p>
意味相互互換(Semantic Interoperability)	<p>意味相互互換とは、推論し、関係し、翻訳し、そして、デジタルコンテンツの暗示的意味を分類する特殊な技術のアプリケーションから引き出される能力の一種である。それにより、ビジネスプロセス、企業知識、企業規規則そしてソフトウェアアプリケーション</p>	<p>“適応情報: 意味互換によりビジネス、グリッドコンピューティング及び企業統合を</p>

	ヨンの相互互換性を実現する。	改善する” Jeff Pollock と Ralph Hodgson の本より, Wiley Publishing 2004
セマンティック Web	セマンティック Web は、既存のウェブの拡張であり、ここでは、情報には良く定義された意味が付与され、コンピュータ処理がより容易になり、そして、人々は共同して作業する事が容易になる。	“The Semantic Web”, By Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, May 2001
セマンティック Web サービス	セマンティック Web サービスは、OWL-S(Web Ontology Language Service)により強化された Web サービスの実現形の一つである。OWL-S は、Web サービスに於ける曖昧さが無く、且つ、コンピュータ理解可能な形式で Web サービスの特性と機能とを記述するのに必要なマークアップ言語のコアセットを供給する。Web サービスの OWL-S によるマークアップは、自動的な Web サービスの発見、実行、相互操作、構築及び実行監視を含む Web サービスの自動化を可能にする。マークアップ言語の階層化アプローチに従うならば、現在の OWL-S は、W3C 標準である OWL の上に作られたものである。	“適応情報 : 意味互換によりビジネス、グリッドコンピューティング及び企業統合を改善する” Jeff Pollock と Ralph Hodgson の本より, Wiley Publishing 2004
タクソノミ (Taxonomy)	タクソノミを簡単に定義すると、情報を組織化するのに用いる構造である。タクソノミと言うと人々は、通常、生物科学で用いられるような階層構造であると理解する。情報科学の視点では、タクソノミは幾つかのタイプの構造の一つか、又は、それらを組み合わせたものである。すなわち、単純な平面構造、階層、ネットワーク/複合の構造若しくはファセットタクソノミかも知れない。これ等の構造の各々は、夫々異なる種類の情報管理やアクセス目的に用いられる。これ等総ては、今日の複雑な情報ソリューションをサポートする為には重要であり、且つ、今日の複雑な情報システムの統合要素でもある。	タクソノミとセマンティックSIGのミッションステートメントより抜粋 < http://km.gov/ >
タクソノミ構造	タクソノミ構造は、米国議会図書館の分類大系の様な、定義された範囲と文脈との中の概念の基礎をなす階層的構造を言い、コンテンツ管理ワークフロー手順の中で情報のカテゴリ化の為にコンテンツ管理者により利用される。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
タクソノミビュー (Taxonomy View)	タクソノミビューは、エンドユーザに提示されるタクソノミ構造の可視的な見え方である。それは、タクソノミ構造と同一、又は、全く異なる事もある。タクソノミビューは、ウェブのコンテンツ	“Taxonomy Analytical Brief”, Department of

	を、Yahoo の階層型ディレクトリ表の様に論理的にグループ化する。タクソノミビューの出力には、例えば、概念案内モデル、情報アクセス点、広範囲に亘る情報カテゴリ及び関係する標準やガイドラインなどを含む。	State, IRM Business Center, May 27, 2003
シソーラス (Thesaurus)	文書の或る集合を記述するのに関係する用語の集合である。これば階層的では無い：統制語を用いて概念の標準用語を記述する。複数のシソーラスは、同義語、及び、広い若しくは狭い範囲の用語、関係のある用語やその他の語句等、より複雑な関係性を包含する。	“Taxonomy Analytical Brief”, Department of State, IRM Business Center, May 27, 2003
トピックマップ (Topic Maps)	<p>これは、知識を表現したり交換したりする為の ISO 標準の一つであり、情報を見つけ易くするという事に重点を置いている。この標準は、ISO/IEC 13250:2003 である。</p> <p>トピックマップは、トピック(topics)と関連(associations)と出来事(occurrences)とを用いて情報を表す。トピックは、人々、国々そして組織群からソフトウェアモジュール、個々のファイルそして事象に至るまで任意の概念を表現し、関連は、トピックの間の関係性を表現し、出来事は、トピックとそのトピックに関する情報資源との間の関係性を表現する。</p> <p>トピック、関連及び出来事は、タイプ付け可能であるが、そのタイプは、該当するトピックマップの作成者により定義されたものでなければならない。このタイプは、トピックマップのオントロジとして知られている。更に、マーキング(merging)やスコープ(scope)等幾つかの付加的な機能がある。マーキング及び識別(identity)の概念は、バラバラなソースから首尾一貫した新しいトピックマップへ自動的に統合する為のものである。</p> <p>トピックマップは、XML に基づいた交換構文の標準を有し、更に、デファクト標準の API 及び問い合わせ言語とスキーマ言語とが ISO で開発されている。</p>	Wikipedia < http://www.wikipedia.com/ >
Weblog	ブログを参照。	
Web サービス	Web サービスとは、アプリケーション間でデータ交換する為に用いられるプロトコル及び標準の集まりである。色々なプログラミング言語で書かれ、そして、色々なプラットフォーム上で実行されるソフトウェアアプリケーションは、インターネットのようなコンピュータネットワークを介してデータを交換するのに Web サービスを使う事が可能である。この相互互換性はオープン標準に起因するものである。OASIS と W3C とは、Web サービスに関する	Wikipedia < http://www.wikipedia.com/ >

	<p>るアーキテクチャ及び標準化の為に責任を持つ複数の専門委員会 (steering committees) を設けている。Web サービス実装間の相互互換性を改善する為、WS-I という組織は、複雑な標準をより分かり易く定義する為、一連のプロファイルを開発している。</p>	
--	--	--

付録 C: 意味矛盾の種類

矛盾	説明	例
データ タイプ	同一の情報に対して、異なる基本要素又は抽象タイプを使用。	社会保障番号(SSN)として、文字列使った場合と数字を使った場合
ラベル付け	同義語/反意語が異なるテキストラベルを持つ	ORG_NAMEとCOMPANYテーブルとが、同じもの(例えば、会社)を意味するデータの場合
集合体	類似のデータセットの概念間の関係性に関して異なる概念作用。集まり若しくは制約が同一の情報に対して違う様にモデル化されている。	「オートバイが1、2、3、4又は5以上の車輪を持つとした場合、貴方のスキーマではどの様に、この制約モデルを作れるか？」を考えると良い。
汎用化	同一の領域のモデルに異なる抽象化が用いられる。	車及びトラックは乗り物の一種か？又は、それらはトップレベルのクラスか？
値表現	何らかの概念を明示的記述にするにあたって、異なる選択が可能である。	開始時刻を期間を加えるとは、終了時刻と等しい、又、終了時刻から開始時刻を引くと期間と等しいか？これらをどの様にモデル化するか？
インピーダンス不一致	基本的に異なるデータ表現方式が使われた	オブジェクトマッピング関連(キーの移行の場合、多様性、等)
ネーミング	同義語/反意語が同一/類似の概念値として存在する。	“会社”の名前が、“ダイムラーベンツ”、“メルセデス”、“クライスラー”等として記述されているが同じものを指している。
尺度と単位	非互換の異なる計測単位	4段階尺度 対 5段階尺度
混同	異なる定義の類似概念	NASDのEPS 対 NYSEのEPS
ドメイン	背景となるドメインが基本的に非互換	“食品生産システムの主組み立て設備 対 ブレーキ生産システムの主組み立て設備
同一性	同一性制約に於ける相違	航空券は旅客認識IDとしての主キーとなるか？(大部分はそうならないが)

(注)

NASD : 全米証券業協会

EPS : EarningsPerShare=1株当たり利益

NYSE : ニューヨーク証券取引所

Jeff Pollockにより提供, 2004年12月.

<http://jtpollock.us/semanticconflicts/chart_semantic_conflicts.pdf>