「米国の産学官で利用が進むセマンティック技術」

渡辺弘美@JETRO/IPA NY

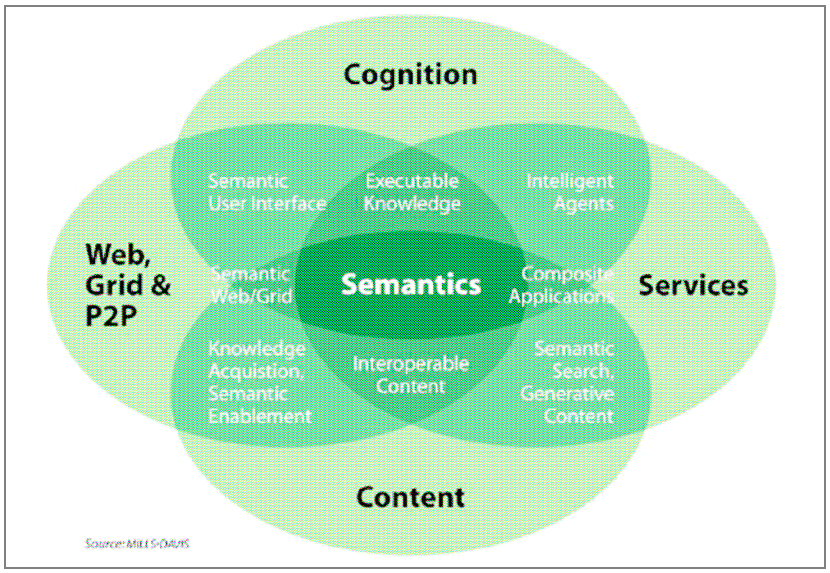
1. セマンティック技術の市場規模と応用

セマンティック技術とは、ユーザの入力した指示に従って、コンピュータがその指示を実行するだけではなく、コンピュータ自らが、入力されたデータや情報などの事象、関係、制約などの意味を理解し、その理解に基づいて様々な処理を実行する技術を指している。これまでのコンピュータの処理能力を進化させ、新しいパラダイムを築くものとして高い関心を集めている。セマンティック技術は、主にマッシュアップ技術として利用されており、米国では、政府機関および民間企業において、データの統合、情報検索、情報共有などをはじめ、様々な目的に応用されている。

(1) セマンティック技術の研究開発とビジネスへの応用

様々な分野での応用が期待されるセマンティック技術は、現在、広範な情報通信技術（ICT）の研究開発分野の中心的テーマのひとつとなっている。中でも、ネットワーキング（ウェブ、グリッド、P2P）、コンテンツ、サービス、認知（Cognition）の 4 つの主要分野において、関連する研究開発が集中している。

ICT 研究開発の中心であるセマンティック技術



セマンティック技術の主要研究開発分野

分野	主な研究開発内容
ネットワーキング（ウェブ／グリッド／P2P）	セマンティック技術を使うことでコンピュータが、ウェブ、グリッド、ピアツーピア（P2P）・ネットワーク上のバーチャル・システムの設定管理を行うことを可能にする。
コンテンツ	セマンティック技術によって情報を相互運用することで、組織やシステムなどの枠を超えて、コンテンツへの検索・アクセスを可能にする。コンテンツに対する理解を向上させるだけではなく、情報を有効に活用できるようにすることで、コンテンツの経済性を高める。
サービス	セマンティック技術を利用することにより、コンピュータ自らによるサービスの発見、開発、統合、管理ができるほか、複合アプリケーション上で異なるアプリケーションを使った情報をリンクさせることも可能となる。
認知	セマンティック技術を通じて、適応性が高く、自覚的で、自立したコンピュータによる判断および知識処理を可能とする。

このような分野において、セマンティック技術の研究開発が進められている一方で、米国では、すでに政府や産業界において、具体的なビジネス・ソリューションとしての利用も始められている。セマンティック技術は、機関・企業におけるソフトウェア開発、インフラストラクチャ、情報システム、ナレッジ・システム、行動システムの 5 つの IT 関連分野において、特に事業価値を高めることが可能とされている。

セマンティック技術によって事業価値の上がる主な分野と利点

企業における重要分野	セマンティック技術による主な利点
ソフトウェア開発（Development）	ソフトウェア開発にセマンティック技術を利用することで、ソフトウェアの開発プロセスが自動化され、通常のソフトウェア開発における複雑さ、労働力、リスク、コストなどといった問題が解決されるほか、ソフトウェアが開発されるまでの時間を短縮することができる。
インフラストラクチャ（Infrastructure）	セマンティック技術を利用し、情報やデータの輸送、保管、計算のために重要な資源を統合させ、インフラストラクチャの規模、複雑性、セキュリティに関する問題解決に役立てている。
情報システム（Information）	セマンティック・モデルを利用したセマンティック・ウェブ用のコラボレーションや複合アプリケーションでは、情報やアプリケーションの相互運用ができるほか、自動的なナレッジ処理およびナレッジ処理能力の拡張を可能とする。
ナレッジ・システム（Knowledge）	計算可能なナレッジ資産を基に、自動的なナレッジ処理やナレッジ処理能力を拡張することで、人件費削減となるほか、ナレッジワークの生産性が向上し、企業のナレッジ優位性につながる。
行動システム（Behavior）	セマンティック技術によって、コンピュータが大規模なナレッジ・ベースを使って人間のような理解、学習、認識する能力を持ち、論理的に不確実性や数値を推測することが可能となる。

(2) セマンティック技術分野の市場規模

セマンティック技術は、情報システム市場における成長株として注目されている技術であり、今後大きな伸びが期待されている。

セマンティック技術を使った製品やサービスに関するコンサルティング会社 Project10X 社の創業者で、代表取締役を務める Mills Davis は、セマンティック技術分野の市場規模は現在、情報通信技術（ICT）市場のほんの一部分にすぎないが、同市場は今後も順調に市場規模を拡大し、近い将来、数 100 億ドルを超える市場に成長していくと見ている。同氏は、ICT 市場における現在のセマンティック技術は、セマンティック・ウェブの関連技術を中心としており、2006 年のセマンティック技術全体の市場規模は 21 億ドル程度であるが、4 年後の 2010 年には 524 億ドルに達し、2015 年には 2006 年の 200 倍以上まで成長し、5,020 億ドルになると予測している。

世界におけるセマンティック技術市場規模の推移（2006-2015 年）

市場	2006 年	2010 年	2015 年
セマンティック技術開発	5,000 万ドル	4 億ドル	20 億ドル
セマンティック・インフラ	5 億ドル	170 億ドル	2,000 億ドル
ナレッジワークの自動化			
情報集約型	11 億ドル	300 億ドル	2,500 億ドル
知識集約型	3 億 5,000 万ドル	45 億ドル	400 億ドル
知的システム	1 億ドル	5 億ドル	100 億ドル
合計	21 億ドル	524 億ドル	5,020 億ドル

(注１)セマンティック技術開発市場には、オントロジー・モデリングやライフサイクル管理環境など、セマンティック技術開発におけるメソドロジー、トレーニング、ツールおよびプラットフォームなどの開発、プロトタイピング、実行、展開、管理などのための製品・サービスが含まれている。
(注２)知識集約型ナレッジワーク自動化市場では、研究、分析、計画、設計、シミュレーション、実験、決定管理のための知識資産開発や大規模なノレッジベース、知識共有および知識ツールなどによるノレッジワークの自動化に関連した製品やサービスを提供している。
(注３)知識システム（Systems that know）市場には、判断・理解能力を備えた認知システム、知的エージェント、ロボットやシステムなどの関連製品・サービスが含まれている。

世界におけるセマンティック技術市場において最も市場規模が大きいのは、情報集約型ナレッジワーク自動化の分野で、2006 年の 11 億ドルから 2015 年には 2,500 億ドルまで成長すると予測されている。情報集約型ナレッジワークの自動化市場は、セマンティック・コラボレーションやセマンティック検索など、セマンティック技術を使ったシステムと情報の統合および相互運用によるナレッジワーク自動化の関連製品やサービスを提供している。ナレッジワークの自動化を図るために、今後もセマンティック技術が積極的に採用されていくと見られる。

また、これに続くセマンティック・インフラ分野については、2015 年には 2006 年の 400 倍となる 2,000 億ドルに達するとされている。セマンティック・インフラ市場では、セマンティック・ウェブ、セマンティック・グリッド・コンピューティング、セマンティック P2P など、ネットを中心に統合、相互運用、セキュリティを構成していくためのセマンティック・エンタープライズ・アーキテクチャやセマンティック技術の関連製品やサービスが提供されている。セマンティック・インフラ市場において、現在最も注目されているセマンティック・ウェブは、コンピュータによるデータの意味やデータ間の関連の解釈を可能とし、情報を自動的に収集、整理、処理する能力を持つといわれており、ウェブ 2.0 を取って代わる次世代ウェブになる可能性が高いと期待されている。

2. セマンティック技術ソリューション提供する IT ベンダによる取り組み

このように将来の市場成長が期待される分野だけに、すでに多くの IT ベンダがセマンティック技術ソリューション分野に参入している。Project10X の調べによると、今後の ICT 市場を牽引する技術になると注目されているセマンティック技術を利用したソリューションの研究開発に取り組んでいる IT ベンダは約 200 社を超えており、既に 70 社はセマンティック技術を使った製品やサービスを提供しているとしている。

IT ベンダが提供しているセマンティック技術ソリューションは、オントロジーを利用して複数のアプリケーションによる情報共有を可能にした情報統合システムからセマンティック・ウェブをイントラネット上に実現するためのソリューションまで様々である。

セマンティック技術を利用した製品やサービスを開発している IT ベンダ

Active Navigation	CheckMi	Empolis	Intellisophic	Noetix	SilkRoad
Adobe	Cisco	Endeca	Interwoven	Northrop Grumman	Software AG
Aduna	ClearForest	Engenium	Inxight	nStein	Sony
Agilense	CoeTruman Technologies	Enigmatec	iSOCO	NuTech	SRA International
AKT Triple Store	Cogito	EnLeague Systems	ISX Software	Ontologent	SRI International
Amblit Technologies	CognIT	Entopia	ISYS Search Software	Ontology Works	Stanford University
Anteon	Cognos	Entrieva	JARG	Ontopia	Stellent
Apelon	Composite	Epistemics Ltd.	Jayna	Ontoprise	Stratify
APR Smartlogik	Compoze Software	Factiva	Kalido	OpenText	Sun Microsystems
Arbortext	Computer Associates	Fair Isaac	Kanisa Software	Oracle	Sybase
Ask Jeeves	Conformative Systems	FAST	Knowledge Foundations	Profium	Synomos
AskMe	Connecterra	FileNet	Knowledge Media Institute	Radar Networks	SYS Technologies
Aspasia	Connotate	Fujitsu	Kofax	Raytheon	Tacit
Astoria Software	Content Analyst	GeoReference Online	Kowari	Readware	Taxonomywarehouse
AT&T	Contextware	Global360	L&C	RuleBurst	TEMIS
ATG	Contivo	Gnowsis	Lockheed Martin	Reed Elsivier	The Brain
Attensity	Convera	Google	Logic Library	SAIC	Thetus
Autonomy	Copernic	Grand Central	Mark Logic	Sandpiper Software	Thomson
Axontologic	Correlate	Groxis	McDonald Bradley	SAP	TopQuadrant
BBN	Cougaar Software	H5 Technology	Metacarta	SAS	Triple Hop
BEA	Coveo Solutions	Hewlett Packard	MetaIntegration	SchemaLogic	Troux
BioWisdom	Crystal Semantics	Hummingbird	Metallect	Semagix	Ultimus
Black Pearl	Cycorp	Hyperion	Metamatrix	Semandex Networks	Unicorn
Blue Oxide	Dassault Systems	i2 Inc	Metatomix	Semantic Light	Verity
BrandSoft	DAY	IBM	Microsoft	Semantic Research	Versatile Info Sys
Broadvision	Digital Harbor	iLog	Mind Alliance	Semantic Sciences	VerticalNet
Business Objects	Discovery Machine	Image Matters	Miosoft	Semansys	Vignette
C24 Solutions	Dynamic Digital Media	Informatica	Modulant	Semaview	Visual Knowledge
Capraro Technologies	Dream Factory	InforSense	Mondeca	Semtation GmbH	Vitria
Captiva	EasyAsk	Infosys	Moresophy	Serena	Vivisimo
Celcorp	Ektron	Innodata (ISOGEN)	NCR Teradata	SiberLogic	WiredReach
Cerebra	EMC/Documentum	Intellidimension	NetMap Analytics	Siderean	XSB
		Intelliseek	Neurok		

Source: ARLLS-OMAIS—1/10/2006

こうした IT ベンダの中には、Cisco 社、IBM 社、Oracle 社、SAP 社などの大手ソフトウェア会社はもちろんのこと、セマンティック技術を専門に取り扱っている新興企業も多く見られる。

以下に、セマンティック技術を利用したソリューションを提供している中で、比較的新しい企業である Metatomix 社および TopQuadrant 社の製品・サービスを紹介する。

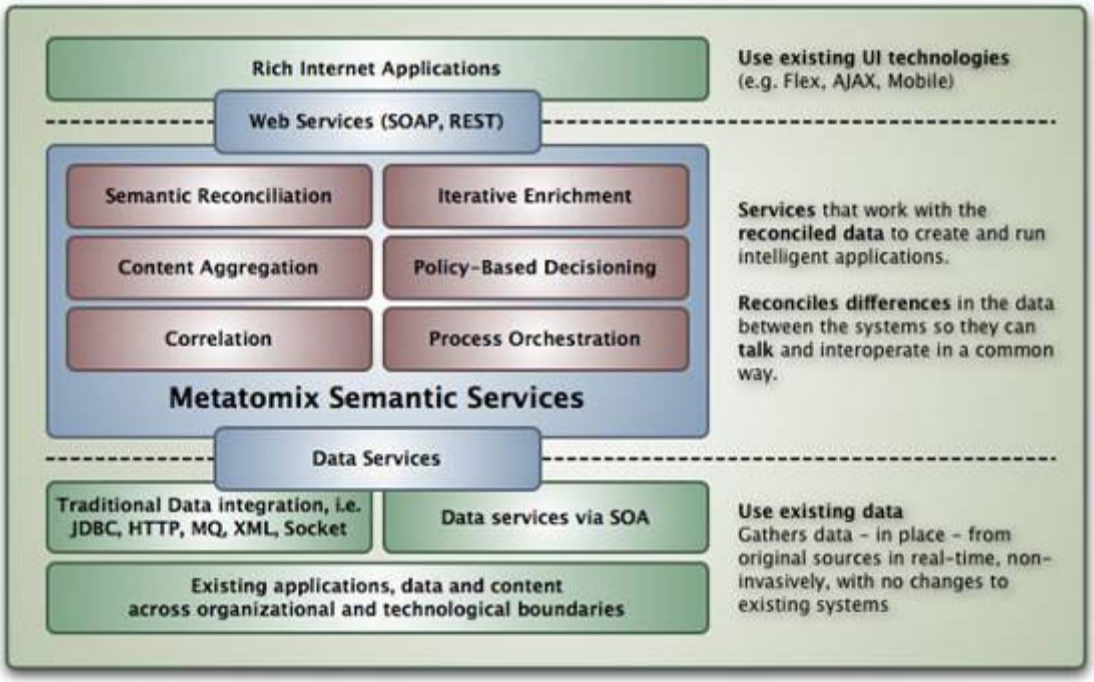
(1) Metatomix 社

マサチューセッツ州ボストンに本社を置く Metatomix 社は、2000 年に設立されたセマンティック・ソリューション・プロバイダである。司法、ライフ・サイエンス、金融、製造業など 4 つの業界を中心に、データ統合を可能とするセマンティック・ミドルウェア・ソリューションを提供している。同社は、顧客および Unisys 社や Oracle 社などの技術パートナーと緊密に協力し、同社の提供するソリューションで既存システムを統合し、顧客の意思決定や事業効率化の実施およびコスト削減などを図るための支援を行っている。

Metatomix 社のセマンティック技術ソリューションは、ウェブを利用して、異なるデータを統合させるためのミドルウェアを提供している。同社が提供しているセマンティック・サービスは、以下 6 つの機能によって、リアルタイムで異なるシステム間が所有する知識の統合を図り、企業における決断プロセスを支援するものとなっている。

①. セマンティック技術によるデーター致（Semantic Reconciliation）：異なるデータにおける違いを調整し、データの相互運用性を促す。
②. コンテンツの集約（Content Aggregation）：企業全体にわたって統合された単一ビューをリアルタイムで作成する。
③. 相互関係（Correlation）：異なるデータを分類し、位置付けおよび点数付けをすることでデータ間の関係を評価することで、新しい見解を導き出すほか、意思決定に役立つような事業知識を生み出す。
④. 相互作用の向上（Interactive Enrichment）：決断が下されるまで、ルールや方針によってやデータをダイナミックに評価・分析する。
⑤. 方針を基にした決断（Policy-Based Decisioning）：ルールや方針を基に結論に達し、決断を下す。
⑥. プロセスの組織化：下された決断を実行するために、ルールや方針を基に全ての既存システムにおけるプロセスを調整する。

Metatomix 社の提供するセマンティック技術ミドルウェア・ソリューション



同社はこのミドルウェア・ソリューション技術をベースとして、フロリダ州裁判所やジョージア州検事委員会（Prosecuting Attorney's Council of Georgia）をはじめとする司法関連機関にソリューションを提供している。たとえば、フロリダ州では、法執行機関、裁判所、刑務所などに従事する約 3,500 のユーザが、1 ヶ月間に 300 万件もの犯罪歴調査を行うために、同社の製品を利用している。特に、2005 年 5 月に発効したフロリダ州対性犯罪者規制法「Jessica Lunsford Act（ジェシカ・ランスフォード法）」を効果的に施行していくため、フロリダ州は同社の技術を導入して、関連する複数の司法機関に個別にあるシステムの情報を共有し、包括的な素性調査の実施が可能となった上、容疑者の逮捕手続きを行う国家システムの自動化を図ることもできるようになった。

司法関連機関に加え、同社の製品は Regions Financial Corporation 社などの大手金融機関でも利用されている。また、ライフサイエンス、製造業も同社のターゲット業界とされており、業界別のニーズに合わせたデータ統合ソリューションを提供している。同社の製品は多様な業界で利用されている。

Metatomix 社の提供するソリューションの種類と効果

対象業界	ソリューションによる効果	同社のソリューションを利用している主な機関
司法	同社のソリューションをミドルウェアとして導入することで、州や地域レベルの法執行機関、裁判所、刑務所の間で効果的に情報の共有ができる。異なる機関の情報を同社のセマンティック・モデルによって照合させることで、一貫性のある見解が導き出せる。また、包括的な素性調査を自動化させることで、素性調査にかかる時間の大幅短縮およびコスト削減につながる。	フロリダ州裁判所、ジョージア州検事委員会など
金融サービス	金融サービス機関は、同社のソリューションを利用して、異なるシステム間におけるエンドツーエンドの電信送金の自動化を促進するほか、企業の合併吸収によるシステムの移行を迅速に処理することができる。	Regions Financial Corporation、Merrill Lynch、Wells Fargo など
ライフ・サイエンス	ライフ・サイエンスでは、何百もの人間の遺伝子から酵素の相互作用などまで、ライフ・サイエンスに関わるあらゆる物事を説明するためのオントロジーを作成している。同社の提供するソリューションによって、異なる研究・学術機関が所有するデジタル化された膨大な量の知識の共有を可能とし、研究の重複を防ぐことができる。	利用組織の情報は非公開。
製造業	製品設計、製造、流通、サプライチェーン、デマンドチェーンなど、製造業者の所有する膨大な量のデジタル知識は、異なるシステムに保存されている。同社のソリューションを利用して、異なるシステムの情報をひとつのセマンティック・モデルに統合することで、コンピュータが様々な情報の相互関係を把握し、異なる分野における要因を考慮した結論を導き出すことができる。	利用組織の情報は非公開。

(2) TopQuadrant 社

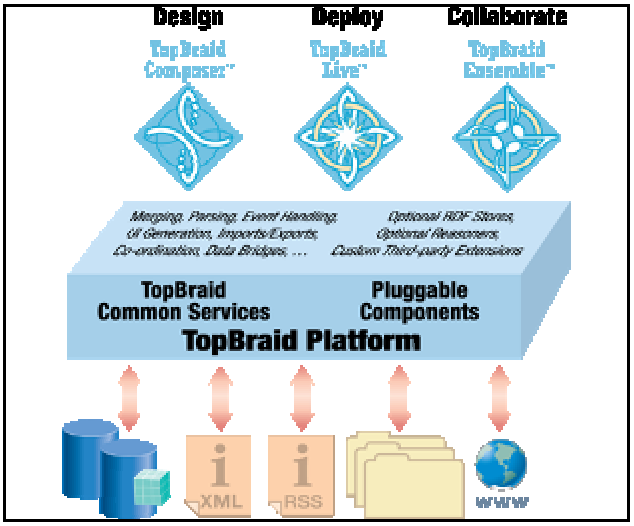
2001 年に創設された TopQuadrant 社は、バージニア州アレクサンドリアを拠点とするセマンティック・ウェブのソリューション会社である。同社は、航空宇宙局（NASA）、一般調達局（US General Service Administration）、連邦航空局（Federal Aviation Administration）などの政府機関のみならず、産業界においては、半導体メーカーの Intel 社、製薬メーカーの Pfizer 社、米航空宇宙・防衛産業大手

の Lockheed Martin 社、大手自動車会社の General Motors 社など様々な分野における顧客を抱えている。また、後述するテキサス大学健康科学センター・ヒューストン校（University of Texas Health Science Center at Houston: UTHSC）も TopQuadrant 社のソリューションを導入している。

同社では、主に社内における情報管理、情報検索・ナビゲーション、会社方針管理、エンタープライズ・アーキテクチャ、情報データ統合などに関連したエンタープライズ・ソリューションを提供している。また、同社では、セマンティック・ウェブ、エンタープライズ・アーキテクチャ、エンタープライズ・システムのモデルベース開発・設計などの分野における専門家を揃え、セマンティック･ウェブ・ソリューションに関する計画・設計から導入およびトレーニングまでの包括的なコンサルティング・サービスも提供している。

様々な分野におけるエンタープライズ・ソリューションをラインナップしている同社は、セマンティック関連では「TopBraid Suite」と呼ばれるセマンティック製品群を揃えている。「TopBraid Suite」には、「TopBraid Composer」、「TopBraid Live」、「TopBraid Ensemble」の 3 つの製品が含まれており、セマンティック・アプリケーション環境の開発から導入までのすべてのライフサイクルをカバーしている。同社の「TopBraid Suite」の製品を利用することで、企業はセマンティック・アプリケーションを全社的な規模で迅速に展開していくことができるとしている。

TopQuadrant 社の提供する「TopBraid Suite」



TopQuadrant 社の提供するセマンティック製品の概要

製品名	概要
TopBraid Composer	セマンティック・ウェブのオントロジーおよびセマンティック・アプリケーションを開発・構築していくための企業規模のプラットフォーム。ワールドワイドウェブ・コンソーシアム（W3C）の標準に完全準拠し、知識モデルやインスタンス知識ベースの設定における開発、管理、検査を行うための包括的なサポートを提供する。
TopBraid Live	データ、コンテンツ、ユーザ・インタラクションの統合や集約を目的としたマッシュアップ・システム、セマンティック・ウィキやソーシャル・ネットワーキング・アプリケーションなどの導入のために最適化されたサーバ・プラットフォーム。TopBraid Composer と統合させることで、オントロジー・モデルおよびセマンティック・アプリケーションを一度に展開することを可能とする。
TopBraid Ensemble	情報管理を共同で行うためのマルチ・ユーザ用セマンティック・ウェブ・アプリケーション。RDF（リソース・ディスクリプション・フレームワーク）にエンコードされたデータの作成・閲覧できる機能を提供する。セマンティック・ウィキなどのように柔軟性のあるコンテンツ管理システムとしても利用できる。

「TopBraid Live」および「TopBraid Ensemble」は、2007 年 5 月から提供開始となった新製品である。これまで、企業におけるセマンティック・アプリケーションの開発は複雑で、時間のかかる作業であるとされていたが、両製品によってそのような企業の悩みは解決され、包括的なセマンティック・アプリケーション環境を提供されることになった。

同社の共同創立者である Ralph Hodgson 氏は、「同社はセマンティック・アプリケーション開発におけるバリアを取り除いた。これまで企業は、セマンティック・アプリケーション開発のためのインフラを自ら構築する必要があったが、TopBraid Live および TopBraid Ensemble を導入することで、セマンティック・アプリケーションの開発のみに専念することができるのである」と述べている。

3. ユーザ機関・企業によるセマンティック技術の利用動向

次に、既にセマンティック技術を利用しているユーザ機関・企業の導入事例を紹介する。ここでは、2007 年 5 月 20 日から 24 日にかけてカリフォルニア州サンノゼで開催された「2007 Semantic Technology Conference」において発表された事例として、航空宇宙局（NASA）、Yahoo!社、CNET Networks 社、テキサス大学健康科学センター・ヒューストン校を取り上げる。

サンノゼで開催されたこのイベントでは、まだ無名のスタートアップ企業からの参加者が半数を占めたが、その他は、NASA、国防総省(DOD)、US Air Force、スタンフォード大学、ロッキード・マーチン社、ボーイング社、フォード・モータース社、マイクロソフト社、IBM社、オラクル社、サン・マイクロシステムズ社、グーグル社、ウォールマート社といった有名機関が多く名を連ねていた。

ここで取り上げる４つの機関で導入されているセマンティック技術は、現在の市場動向を反映するように、情報集約型ナレッジワークの自動化やセマンティック・インフラに含まれるセマンテック・ウェッブ関連のソリューション導入事例となっている。

(1) 複雑なデータの統合：航空宇宙局（NASA）

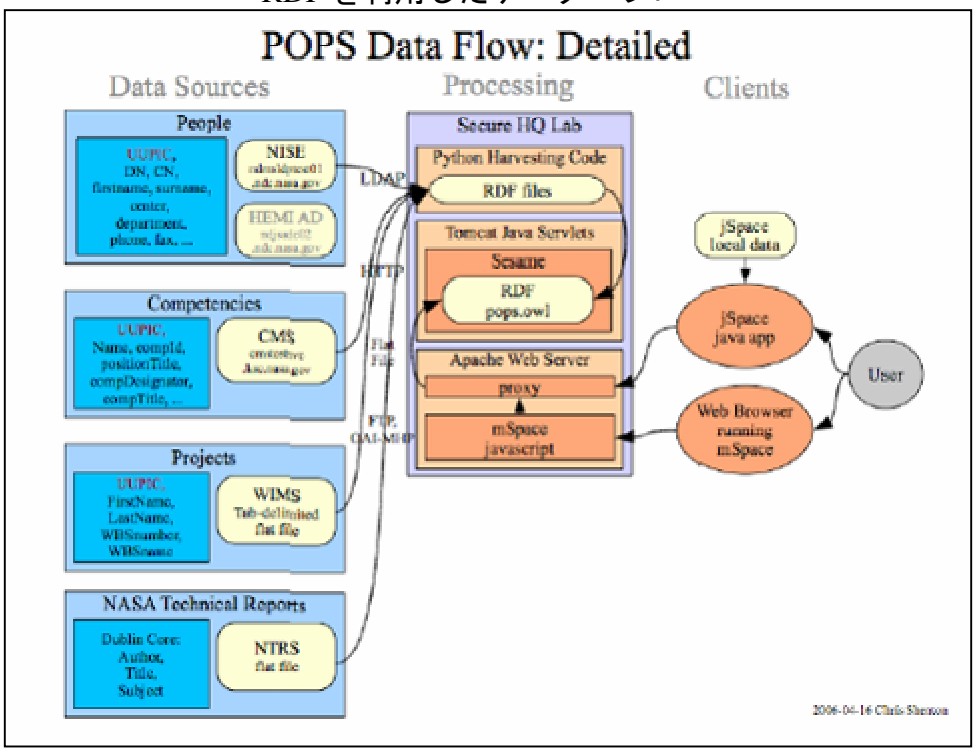
全米11ヶ所に宇宙センターおよび研究機関などを抱えるNASAでは、毎日膨大な量のデータが生成されている。しかし、同局では、これら11機関によって生成されるデータをひとつに集合させるといった中央集中型のデータ構造を採用しておらず、データ統合が非常に複雑となっている。また、同局のデータは、異なるデータベースに保管されており、データ･フォーマットが統一されていないため、データの検索が困難であり、見つかりにくいデータなどは何度も作成されるなど、データ重複の原因となっていたという。

こうした状況を改善するために、同局は現在、既存データ・ソースを利用して効率的なデータ管理を行っていくために、同局内のグループやプロジェクトに対して、セマンティック・ウェブ技術のRDF（Resource Description Framework：情報に関するメタデータをXMLで記述してWeb上で処理するための仕様。RDFで記述することで、コンピュータが扱う情報の分類や検索などの自動化・効率化を図ることが可能となる。）やOWL（Web Ontology Language：RDFによって定義された語彙を関連付けていくためのルールを定義するオントロジー言語。）の利用を推進し、NASA全体におけるデータの統合を進めている。

同局では、地球科学分野における情報の発見、利用、共有を促進するために大規模なオントロジー「SWEET（Semantic Web for Earth and Environmental Terminology）」が開発され、既に複数のプロジェクトによって使用されているほか、JAVAを使ったセマンティック・ブラウザ・アプリケーションである「mSpace」や「jSpace」などのユーザ・インターフェースも開発している。
NASAの最高技術責任者（CTO）であるAndrew Schain氏によると、同局のチーフエンジニア室（Office of Chief Engineer）では、セマンティック・ウェブの研究開発に取り組んでいるClark & Parsia社と協力し、4つの異なるデータベースの情報

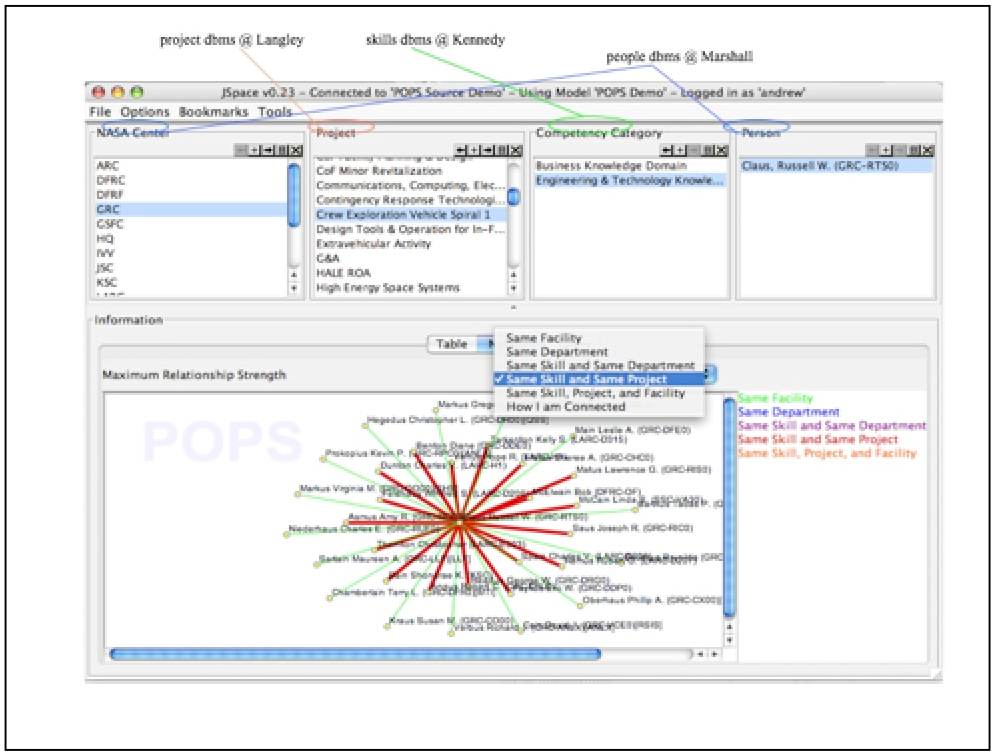
をRDFでエンコードし、ブラウズした情報を表示するユーザ・インターフェース「jSpace」を構築したとしている。　ユーザは、同ユーザ・インターフェースを利用することによって、4つの異なるデータベースにある情報を自由にブラウズすることが可能となった。

RDF を利用したデータ・フロー



Clark & Parsia 社との協力で開発された「jSpace」ブラウザは、異なるデータベースの情報を包括的に検索し、その結果を人物（People）、機関（Organizations）、プロジェクト（Projects）、スキル（Skills）の 4 つのフィールド（通称：POPS）にあわせて表示できるように、それらに適合した情報を導き出すことができるとしている。以下は、「jSpace」ブラウザのイメージ写真であり、ラングレイ研究センター、ケネディ宇宙センター、マーシャル宇宙飛行センターの 3 つの異なるデータ・ソースから、「NASA Centers」（機関）、「Project」（プロジェクト）、「Competency Category」（スキル）、「Person」（人物）の順に、各フィールドで選択した項目に当てはまる人物を最終コラムである「Person」の欄に表示している。

「jSpace」ブラウザ



(2) 複数のサイト間でコンテンツ共有：Yahoo!社

大手検索エンジンの Yahoo!社では、ニュース、金融、スポーツなどといった各サイトの担当部門がそれぞれ独立してウェブサイトを構築しているため、ウェブサイトに使用されている技術やコンテンツなどに関する膨大な量の知識・情報が分散して蓄積されており、異なる部門のデータベースにアクセスすることが非常に困難であった。このような問題を抱えていた同社は、2005 年 10 月に RDF およびセマンティック･ウェブの開発者の Dave Beckett 氏を Yahoo! メディア・グループに迎え、RDF などのセマンティック技術を使ったセマンティック・ウェブサイトの構築に力を入れている。

Yahoo!メディア・グループでは、セマンティック・ウェブ技術を使ったシステムを開発し、異なるサイト間でコンテンツの共有や再利用を可能にした。Beckett 氏によると、異なるサイトのメディアコンテンツをつなげるプラットフォームには、コンポーネント・ベースのアーキテクチャを採用し、共通の API（Application Program Interface）を使ってコンポーネントを接続しているとのことである。また、同社が採用しているコンテンツ・プラットフォームは、セマンティック・ウェブ技術を基に構築されており、同社は、RDF および OWL のメタデータを中心としたコンテンツ戦略を展開している。

RDF は、分散された情報の関係を説明するのに適しており、オープン・スキーマを採用したフォーマットであるため、変更に柔軟であることが特徴であるほか、タギングなどの機能も追加しやすく、REST（REpresentational State Transfer）をベースとしたウェブサービスとの相性もいいという。データベースを含めた情報システムは、柔軟性に欠けているため、社内における変化や新しい情報に対応しきれず、18 ヶ月で時代遅れになってしまうというが、セマンティック・ウェブを利用することで、既存のデータベースをキープしながら、新しい情報を柔軟に取り入れることが可能となったという。

このような特徴を持つ RDF を中心としたセマンティック・ウェブ技術を取り入れた同社では、異なるサイト間においてコンテンツの共有・再利用を可能にし、ユーザが分野を横断した関連コンテンツを検索することが可能となり、全体的なユーザ・エクスペリエンスが改善されることになると期待されている。

(3) 関連コンテンツ検索の強化：CNET　Networks 社

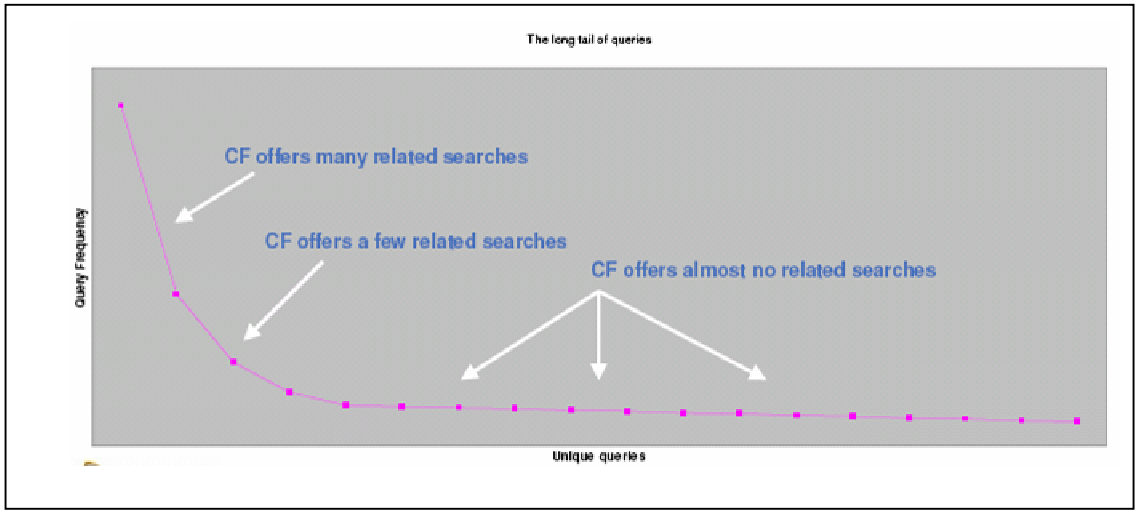
大手オンライン・ニュース会社の CNET Networks 社は、一般的な IT 関連ニュースを提供している CNet News、パソコンやデジカメなどの情報家電製品のレビューを紹介する CNet Review、情報家電製品のオンライン・ショッピング・サイト CNet Shopper、IT 専門家を対象に IT 関連ニュースやブログサービスを提供している ZDNet などをはじめ、数多くのウェブサイトを運営している。

同社の運営しているウェブサイトには、コラボレティブ・フィルタリング（Colaborative Filtering: CF）を使った関連検索機能が搭載されており、ウェブサイトユーザによる過去の検索記録を集めた全てのクエリーを基に、入力したキーワードおよびそのキーワードに関連した情報も検索できる。コラボレティブ・フィルタリング（CF）とは、ユーザの嗜好を過去の行動という形で記録し、そのユーザと似たような行動を取っているユーザ嗜好情報をもとに、ユーザの嗜好を推測

するシステムであり、リコメンデーションサービスを提供する際に使用される代表的な手法である。

しかし、CF をベースとした関連検索機能の問題点は、頻繁に検索されるキーワードに関しては多くの関連情報が検索されるが、検索頻度が少ないキーワードに関しては検索される関連情報の数は少なく、ロングテール現象が起こっている。

CF ベースの関連検索によるロングテール化



このような CF ベースの関連検索のロングテール化を解決するために、CNET Networks 社は、ウェブサイトの関連検索に TextDigger 社の開発した多義性解消技術（Word Sense Disambiguation: WSD）を採用している。WSD とは、自然言語処理技術の一つで、多義語の意味を文脈に応じて特定するものであり、キーワード検索に多い検索質問の曖昧性を解消する。

TextDigger 社の創業者兼 CEO である Tim Musgrove 氏によると、同社の WSD 技術によって、ユーザが入力したキーワードに当てはまる情報がない場合でも、そのキーワードに似た意味を持つ情報を取り上げた検索結果が表示されるため、キーワード検索のロングテール化を防ぐことができるとしている。

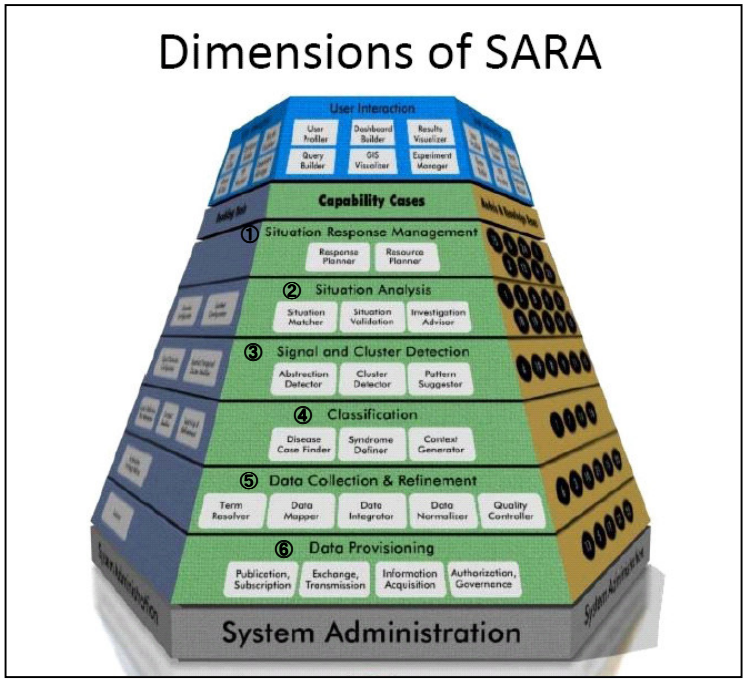
また、TextDigger 社の WSD 技術は、キーワード検索のロングテール化を防ぐだけでなく、キーワード広告の収益向上にもつながるという。現在のキーワード広告プレースメントでは、キーワードに関連した広告を網羅していないため、400%もの関連広告がユーザの目に触れる機会を逃している。WSD 技術を利用することで、キーワードに関連した広告をより多く掘り出すことができるほか、関連広告の検索結果ゼロということがなくなるため、広告のクリック率も通常の CF ベースのキーワード検索に比べて高くなる。

WSD 技術を使ったキーワード関連広告の表示

(4) 情報交換の促進：テキサス大学健康科学センター・ヒューストン校

テキサス大学健康科学センター・ヒューストン校（University of Texas Health Science Center at Houston: UTHSC）では、公衆衛生面における準備および意思決定を向上させることを目的として、公衆衛生準備システム（Public Health Preparedness System）を設計・評価するためのフレームワーク「Situation Awareness Reference Architecture（SARA）」を開発した。SARA は、①状況対応管理（Situation Response Management）、②状況分析（Situation Analysis）、③信号・クラスタ検出（Signal and Cluster Detection）、④分類（Classification）、⑤データ収集・詳細化（Data Collection & Refinement）、⑥データ・プロビジョニング（Data Provisioning）の 6 つの機能ケース（Capability Cases）によって構成されている。

SARA フレームワークにおける機能ケースの構成図

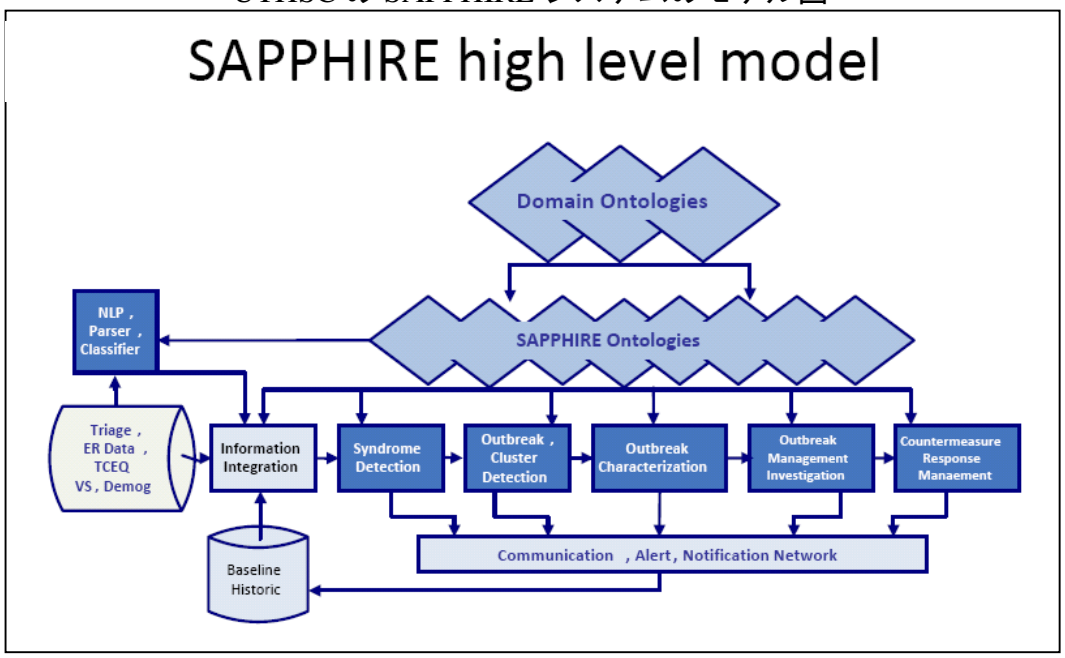


UTHSC は、SARA フレームワークの実現に向けて、Oracle 社および同社のパートナー企業である前述の TopQuadrant 社が提供するセマンティック・ウェブ技術を導入し、Situational Awareness and Preparedness for Public Health Incidences Using Reasoning Engines （SAPHIRE）システムを構築した。

同システムは、地域のヘルスケア・プロバイダ、病院、薬局などから健康や疫学に関する幅広いデータを統合している。ソースの異なる様々なデータに対してソーシャル・ネットワーク分析を適用することで、ヘルスケア・プロバイダは、ひとつの統合されたメタデータ・モデルを使って、公衆衛生に関する問題を探知・分析し、適切な対応を実施していくことができるとしている。

UTHSC の SAPPHIRE システムは、Oracle 社の RDF（Resource Description Framework）データモデルを採用し、Oracle 社データベースの基本機能のひとつである空間データベース「Oracle Spatial 10g」に TopQuadrant 社が提供するセマンティック・ウェブのオントロジーを開発・構築していくためのプラットフォーム「TopBraid Composer」を統合して構築されている。Oracle 社の RDF データモデルは、オープンで、拡張性があり、安全で信頼できる RDF 管理プラットフォームを提供している。

UTHSC の SAPPHIRE システムのモデル図



テキサス大学健康情報科学ヒューストン校バイオセキュリティおよび公衆衛生情報科学研究所（The Center for Biosecurity and Public Health Informatics Research at University of Texas School of Health and Information Sciences at Houston）の Parsa Mirhaji 所長は、「これまでの公衆衛生提供システムは、症状の探知を中心に行ってきたが、Oracle 社と TopQuadrant 社のセマンティック・ウェブ技術を採用することによって、（公衆衛生に関する）傾向監視や症状の探知・対応など、幅広い分野をカバーした公衆衛生監視が可能となった。セマンティック・ウェブ技術によってソースの異なる多種多様のデータが実用的な情報へと変換され、より効果的な意思決定行うのに役立っている」と述べた。

2005 年秋、パイロット・プロジェクト中であった SAPPHIRE システムは、テキサス州ヒューストンのアストロ･ドーム（Astrodome）、リライアント・パーク（Reliant Park）、ジョージ・R・ブラウン・コンベンション・センター（George R. Brown Convention Center）に避難していたハリケーン・カトリーナの被災者の健康に関する情報収集・分析に利用されたという。SAPPHIRE システムの PDA を使い、300 人を超えるボランティアによって、約 9,000 件におよぶ患者に関する情報が収集された。

Oracle 社健康産業部門の副部長を務める Mychelle Mowry 女史によると、テキサス大学がハリケーン・カトリーナ採用した公衆衛生情報の収集・分析フレームワークは、全米における公衆衛生情報に関連した取り組みのモデルになるとしてい

る。同女史は、「Oracle 社と TopQuadrant 社のセマンティック・技術を統合させたシステムを導入することで、ソースの異なる健康データへのアクセスを容易にし、ヘルスケア・プロバイダによる情報交換を促進することにつながる。また、バイオ監視に関する取り組みを向上させると共に、より改善された公衆衛生サービスを提供していくことができる」と述べている。

（参考資料）

Mills Davis, 2006: ”Semantic Wave: Executive Guide to Billion Dollar Market”
http://www.project10x.com/
http://www.metatomix.com
http://www.metatomix.com/industry/index.html
http://www.topquadrant.com/
http://www.uthouston.edu/
http://www.w3.org/
http://www.topbraidcomposer.com/tqWebsite/documents/topbraidLiveAndEnsembleFinal.doc
http://sweet.jpl.nasa.gov/
http://www.mspace.fm/
http://clarkparsia.com/projects/code/jspace/
http://xtech06.usefulinc.com/schedule/paper/147
http://www.businessweek.com/technology/content/apr2007/tc20070409_248062.htm
Dave Beckett, 2007 Semantic Technology Conference: “A little semantics goes a long way at Yahoo!”
Tim Musgrove, 2007 Semantic Technology Conference: “TextDigger-CNET Case Study on Related Search using Semantics”
Narendra Kunaparenddy&Zhe Wu, 2007 Semantic Technology Conference: “A Semantic Web Approach to Integrative Biosurveillance”
http://www.textdigger.com/
http://www.uthouston.edu/
http://www.oracle.com/corporate/press/2007_feb/UT%20Houston-TopQ.html

このレポートに対するご質問、ご意見、ご要望がありましたら、
hiroyoshi_watanabe@jetro.go.jp までお願いします。

なお、本レポートは、上記の参考資料を利用して作成しているものであり、本レポートの内容に関しては、その有用性、正確性、知的財産権の不侵害等の一切について、執筆者及び執筆者が所属する組織が如何なる保証をするものでもありません。また、本レポートの読者が、本レポート内の情報の利用によって損害を被った場合も、執筆者及び執筆者が所属する組織が如何なる責任を負うものでもありません。