4 Webにおける情報統合
——セマンティックWebとWebサービス

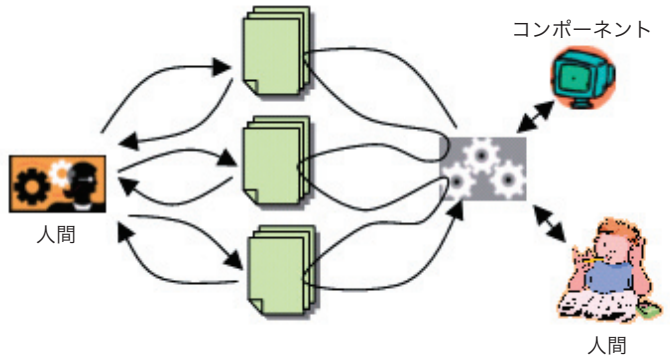
浦本　直彦　日本IBM(株)東京基礎研究所／国立情報学研究所
uramoto@jp.ibm.com

本稿では，古くから人工知能やデータベースの分野で知られている情報（データ）統合の立場から，大量で動的に変化する情報の集合であるWebにおける次世代の統合モデルであるセマンティックWebとWebサービスがどのように位置づけされ，どのような特色を持つかを議論する．情報統合とは，複数の情報源を，グローバルな表現に翻訳し，それらを組み合わせることで，単一の情報源を利用するだけでは解決できない問題を解くための仕組みである．本稿では，まず典型的な情報統合システムの構成を説明し，次にセマンティックWebとWebサービスについて最新動向とともに説明する．さらに両者が，どのように情報統合システムに対応しているかを説明しながら現状の課題について考察する．

はじめに

情報統合（information integrationあるいはdata integration）は，複数の異なる情報源に対し，統一的なインタフェースを提供することで，単一の情報源にアクセスするだけでは解決できない問題を処理するための技術である（**図-1**（b）を参照）☆1．古くから，人工知能やデータベースのコミュニティで研究が進められてきた[1), 4)]．

Webの普及によって，地球規模での情報統合のための基盤が急速に整備されつつある．Webでは，Uniform Resource Identifier（URI）を使って，さまざまな情報源を識別しアクセスすることができる．しかし，情報へのアクセスはURI単位であり，複数の情報源をたどってそれらを組み合わせながら，単一の情報源だけでは解決できない問題を解くのは人間の仕事である（**図-1**（a）を参



（a）従来の検索　　（b）情報統合

図-1　情報統合の例

☆1 情報統合はもっと広い概念で使われることも多い[4)]が，本稿では狭義に用いる．

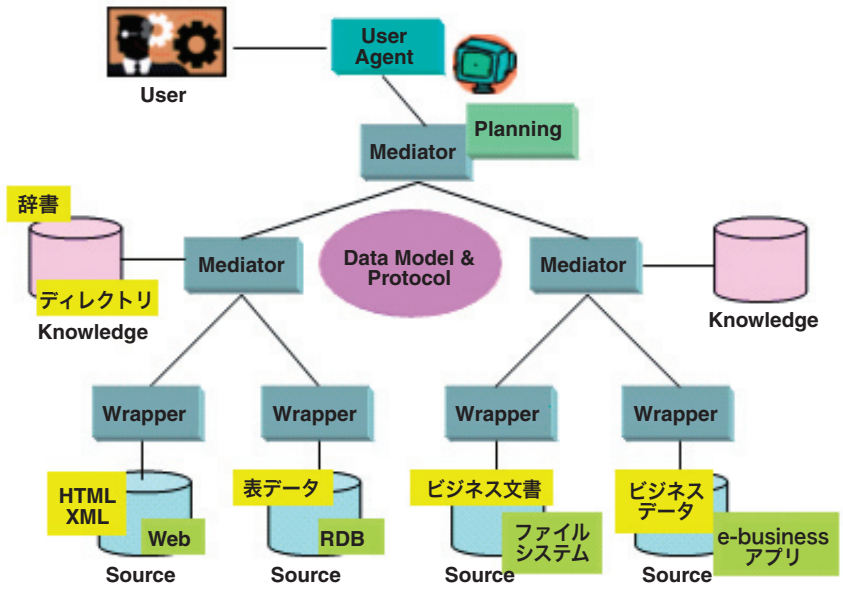


図-2　情報統合の構成要素

照）．Webという環境においては，従来の情報統合システムではあまり考慮されなかった以下の問題が，情報統合を容易でないものにしている．

- 情報量が膨大で，必要な情報にたどり着くのが困難である．
- 情報が動的である．Webページは分散環境で独立に更新され，ある時点では意味があったリンク関係が，別の時点では成り立たない場合がある．
- 情報の質が多様であり格差が大きい．信頼できる組織，個人の情報もあるが，チャットや掲示板のように匿名性が高く正確でない情報も多い．
- 専門家だけでなく，さまざまな レベルのユーザが対象となる．
- 統一的な構文（HTML）が用いられているが，内容に関する制約が存在しない．また，使用されている言語もさまざまである．

このような問題点を踏まえながら，進化し続けるWeb環境をより高度で役に立つものにするために，さまざまなモデルや手法が提案されている．本稿では，その中から代表的なアプローチである，セマンティックWeb（Semantic Web）とWebサービス（Web Services）を取り上げ，両者がどのように情報統合を実現しようとしているかを議論する．

情報統合の構成要素

情報統合（データ統合）は，古くから人工知能やデータベースの分野で研究が進められてきた[4]．Ullmanは，情報統合システムの共通アーキテクチャとして，Source, Wrapper, Mediatorによるモデルを紹介している[2]．これに関連する構成要素を加えたモデルを**図-2**に示す．

情報源（Source）は，それぞれ固有の（ローカルな）スキーマで記述された情報源を示す．XML文書におけるDocument Type Definition（DTD）やXMLスキーマ，関係データベースにおけるデータベース（表）スキーマ，アプリケーションが用いるデータの型やスキーマなどがローカルスキーマの例である．HTML文書のように意味的なローカルスキーマが定義されていない場合もある（たとえば，HTMLでは表（<table>要素）の各カラムの意味を定義していない）．

ラッパ（Wrapper）は，個々のSourceを定義するローカルスキーマを共通のグローバルスキーマへ「翻訳」するために用いられる．グローバルスキーマは，各Sourceの違いを吸収し，Source使用者に対して，共通のビューを提供するものである．セマンティックWebの世界ではRDFスキーマを基本とした個々のメタデータスキーマが，Webサービスでは，サービス記述であるWSDL文書がグローバルスキーマに相当する．

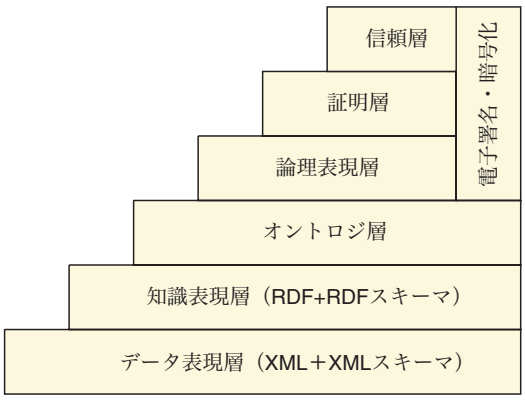


図-3 セマンティックWebの階層

メディエータ（Mediator, Facilitatorとも呼ばれる）は，グローバルスキーマへ翻訳されたSourceを統合し，使用者からの要求に必要な情報を提供する[3]．Wiederholdによれば，Mediatorは，（1）データベースのビュー定義やテンプレートを使ったデータベースの変形（2）複数データへのアクセスとマージ（3）（複数）データの抽象化と生成（4）知的ディレクトリ（4）不完全で矛盾したSourceの扱い，など広範囲な機能を受け持つ．また，Mediatorは，別のMediatorと連携することができる．また，Mediatorを統合し複雑な処理を行うための戦略を受け持つ推論エンジン（Planning engine）がMediatorとは別に用意されることがある．

WrapperやMediatorが参照するグローバルスキーマに関する情報を供給するのが辞書，オントロジ，ディレクトリからなる知識源である．辞書はグローバルスキーマで用いられる単語あるいは概念の定義，オントロジは概念間の関係や制約（概念定義がオントロジに含まれる場合もある），ディレクトリは，Mediatorにアクセスするための情報（たとえば所在情報やアクセス形式）を提供するために用いられる．

それでは，次章からセマンティックWebとWebサービスを簡単に説明しながら，本章で説明した情報統合の構成要素にどのように対応づけられるかを見ていく．

セマンティックWeb

セマンティックWebは，Webの生みの親であり，Web技術の標準化団体であるWorld Wide Web Consortium（W3C）の代表であるTim Berners-Leeが提唱するWebの新しいモデルである（詳細については，W3CのWebページhttp://www.w3.org/2001/sw/や，情報処理学会誌2002年7月号「セマンティックWeb」特集などを参照のこと）．HTML文書を中心に構成されるWebの空間は，従来，人間が読み理解するものであったのに対し，セマンティックWebでは，計算機に処理可能なXMLで記述された知識（メタデータ）の空間を構築する点に特徴がある．Web空間を巨大な知識システムだと考え，人間と計算機間の高度な協調を可能にするのがセマンティックWebのゴールである．

セマンティックWebは**図-3**に示す階層構造によって構築される．セマンティックWebにおけるSourceは，URIによってアドレスすることができる任意のWeb資源である．Web上の多くのSourceはHTMLで記述されているが，HTMLは文書の体裁を記述するために設計されており，知識表現言語としての記述力に欠ける．そこで提案されたのがW3Cによって標準化されたExtensible Markup Language（XML）とXMLスキーマである．XMLを用いることで，HTML文書ではうまく記述できなかった構造を持った情報を記述することが可能になる．さらに，XMLスキーマ言語を用いることで，整数や日付といったデータ型を記述することができる．

XMLおよびXMLスキーマは，XML文書の構造と基本的なデータ型を表現するものであるが，知識表現に必要な概念や概念間の関係を記述するためには不十分である．単なる構造データではなく，そこで，機械処理に向いた知識を表現するために提案されているのが，Resource Description Framework（RDF）である（http://www.w3.org/RDF/）．RDFは，Web上のリソースに対する「メタデータ」を記述するためにW3Cが標準化した仕様である．RDFは，メタデータおよび後述のオントロジ層でも用いられるグローバルスキーマとして位置づけられる．

RDF仕様は主に，モデルと構文を規定したRDFモデ

ルおよび構文と，スキーマを定義したRDFスキーマの2つからなる．RDFでは，有向グラフによるデータモデルを用いて，メタデータを表現する．RDFの基本構造は，
(1)（リソース，プロパティ，リテラル値）あるいは
(2)（リソース，プロパティ，リソース）
という3つ組で表現され，ステートメント（statement）と呼ばれる．例(1)は，あるリソース（メタデータの付与対象となる文書や概念）に対して，属性名―属性値を与えるものであると見ることができる（プロパティが属性名，リテラル値が属性値に対応する）．また，例(2)のように属性値がリソースとして与えられ，さらにプロパティを持つものとして表現することもできる．これらを組み合わせることでグラフ表現を構成する．さらに，ステートメントに対するステートメントを記述することも可能である．RDFスキーマは，概念やプロパティの意味的な制約や関係（たとえば継承関係）を定義するために用いられる．

さらに，メタデータで記述された個々の概念を関連付ける体系がオントロジ層によって構築される．オントロジを記述するためにRDFを拡張した表現言語OWLがW3C Web-Ontologyワーキンググループで開発されている．OWLは，米メリーランド大学のJames Hendlerらが提案したDARPA Agent Mark-up Language（DAML, http://www.daml.org）と，ヨーロッパを中心に開発が進められていたOntology Inference Layer（OIL）がベースになっている．

OWLは，RDF構文およびRDFスキーマをベースにオントロジ記述のための機能を追加したものである．たとえば，“Human”クラスが“Man”クラスと“Woman”クラスの和によって表現されることを，OWLでは以下のように定義する．

```
<owl:Class rdf:ID="Human">
  <owl:unionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Man" />
    <owl:Class rdf:about="#Woman" />
    </owl:unionOf>
</owl:Class>
```

ほかにも，disjointWith, disjointUnionOf, sameClassAs, equivalentTo, intersectionOf, unionOf, complementOfなど概念定義のためのプロパティが用意されている．

情報統合の立場からは，メタデータ・オントロジ層の上に位置する論理層が重要である．RDFで書かれたメタデータ，OWLで記述されたオントロジを組み合わせて推論や検索処理を行うMediatorが論理層に相当するからである．RDFデータモデルは，第一階述語論理における述語に対応づけることができる．また，OWLが提供するクラス間の階層関係を論理の含意だと考え，いくつかの基本的な推論規則を導入すれば，論理式を使った推論機構を導入することが可能である．ただし，いくつかの問題がある．前述したように，Web上の情報の質はさまざまであり，従来の情報統合システムのように，知識が正しいことを仮定できない．なぜなら，分散された情報源に対する分散されたメタデータの中には，矛盾する知識が存在する可能性があり，そのため，推論機構の前提となる公理系を組み立てるのは，困難であるからである．DAMLおよびOWLでは，軽量で直観的な記述が可能な記述論理（Description Logic[5]）による形式化が試みられている．記述論理は，概念定義や概念間の制約を自然に表現することができる．形式的な表現力は，述語論理を制限したものに相当し，述語論理を用いた推論システムよりも高速に動作することが期待されている．セマンティックWebプロジェクトでは，論理層に関する議論は緒についたばかり，今後の展開が期待される．DAMLプロジェクトでは，DAML-L（DAML-Logic）と呼ばれるサブプロジェクトが進行しているが，こちらもまだ具体的な仕様は公開されていない．記述論理や，論理規則をXMLで表現するRuleML（http://www.dfki.uni-kl.de/ruleml）の使用が検討されている段階である．

Webサービス

Webサービスは，Web環境におけるコンポーネントおよびアプリケーションの統合のための基盤技術である．広義には，HTML, Java Script, JSP, サーブレットなどの従来技術に基づくものも含むが，特に，Simple Object Accessing Protocol（SOAP），Web Services Description Langugage（WSDL），Universal Description, Discovery and Integration of Web Services（UDDI）に代表される，XMLに基づく標準技術を用いて構築されたものを指す．

Webサービスはビジネスアプリケーションやコンポーネントの統合が大きなゴールであり，情報源となるのは，データベースに格納されているデータやJSP, サーブレ

ットなどを介して提供されるアプリケーションデータである．セマンティックWebでは主にHTML文書のような静的な情報源に対してメタデータを付与するのに対し，Webサービスでは，動的なデータをターゲットにしている．

Webサービスでは，コンポーネントは特定のプラットフォームや実装に依存しない「サービス」として記述される．サービス記述を行うためのXML構文（メタデータ）がWSDLである．

WSDL文書には以下に示す情報が記述されている．

- メッセージ（message）メッセージの型情報
- オペレーション（operation）メッセージを入出力として持つオペレーション（メソッド）
- ポート型（port type）オペレーションの集合を提供するポート
- バインディング（binding）ポートにアクセスするためのトランスポート
- サービス（service）サービスの接続情報

WSDLは，Webサービスを記述するのに最低限必要な情報を記述するための構文であり，RDFと比較すると非常に限定されたメタデータしか記述できない．オントロジに相当する知識もなく，XMLスキーマを用いて，サービスの入出力のデータ型を記述するだけである．逆にいうと，Webサービスは，より少ない知識でサービスを結合し動作させることを目的としているといえる．

Webサービスで用いられるメッセージを運ぶのが，抽象的なメッセージトランスポート層とメッセージの構文を規定するSOAPである．SOAPは，XMLメッセージを運ぶための「封筒」の役目をする．実際，SOAPメッセージは，全体をくるむEnvelope要素（タグ），ヘッダ情報を格納するHeader要素，メッセージ本体を格納するBody要素からなるXML文書である．このメッセージング層の存在は，セマンティックWebでは定義されていないが，プラットフォームやプログラミング言語の非依存性など，現実的な情報統合システムを組むためには重要な概念である．

Webサービスは，WSDLで記述されたサービス記述とともに公開される．UDDIは，Web サイトの登録および検索機能を提供するビジネスディレクトリである．Webサービス提供者は，Webサービスを提供するビジネスエンティティ情報（例：企業名や業種）や提供できるサービス記述，インタフェース仕様のID（tModelキーと呼ばれる），接続情報を登録する．Webサービスの使用者は，UDDIに対し業種やtModelキーを使ってサービスを検索することができる．

WSDL，SOAP，UDDIに基づくWebサービスアーキテクチャは，サービスの提供者（プロバイダ），ブローカ，使用者（リクエスタ）間での，公開，検索，結合のモデルを基にしている．プログラミング言語やトランスポートの詳細をサービス記述（WSDL）とメッセージング（SOAP）を用いて隠蔽することで，柔軟な分散環境を実現する．また，UDDIを用いて，自動的にサービスを検索したり結合することができる．たとえば，電子商取引において，動的に変わる条件に合致した取引先を動的に発見し結合することで，あらかじめ取引先が固定される従来の分散アプリケーションでは困難であった新しいビジネス形態をサポートすることができる．このようなサービス発見と結合の動的かつ遅延結合がWebサービスアーキテクチャの特徴である．

どのようにサービス間の結合を行うかを決定することが，情報統合では，MediatorによるSourceの結合に相当する．ユーザ（アプリケーション）の要求に応じて，必要なサービスをディレクトリであるUDDIから検索し，複数のサービスを動的に結合することで，非常に柔軟で自由度の高いシステム（統合サービス）を構築することができる．

Webサービスを結合（あるいは合成）するための仕様がいくつか提案されている[6]．たとえば，XLANGは，Microsoftを中心に提案されたWebサービス結合の枠組みであり，ソフトウェアエージェント間の通信と動的な挙動変化を記述するために使われるπ計算に基づいている．IBMは，従来のワークフロー言語を基本にしたWeb Services Flow Language（WSFL）を提案しており，2002年に両者を統合したBusiness Process Execution Language for Web Services仕様がIBM, Microsoft, BEA, Seibel, SAPにより公開された．一方，SUN, SAP, BEA, Intalioは，Webサービス結合の仕様として，Web Service Choreography Interface（WSCI）を提案している．これらの仕様は，W3Cに新設されたWeb Services Choreography Working Groupで議論され，標準化される予定であるが，本稿執筆時は，複数の仕様が並立している状態である．

議論―情報統合の立場から―

前章までで，セマンティックWebとWebサービスを簡単に説明した．情報統合の立場から，両者の技術が，どのように位置づけられるかを**表-1**に示す．

どちらも，XMLに代表される標準技術をベースに，

構成要素	セマンティックWeb	Webサービス
Source	主に静的なWeb資源（HTML文書）	主に動的なWeb資源（cgi, JSP, ASP, データベース，業務アプリケーション）
Wrapper	人手でWeb資源に対するメタデータを作成する．あるいは半自動的にメタデータを構築する．	（人手あるいはツールを使って自動的に）既存のコンポーネントに対するWebサービスのインタフェースを構築する．
Mediator/ Planning engine	論理に基づく推論（第一階述語論理，記述論理，etc.）	ワークフローに基づく結合
オントロジ	必要とする（OWL）	必要としない．サービスの型が厳密に一致しないと結合できない．
ディレクトリ	未知のメタデータを検索する場合，サーチエンジンを用いる．	未知のサービスを検索する場合，UDDIを用いる．
データモデル	RDF, OWL	WSDL
プロトコル	HTTP	SOAP

表-1　情報統合の観点から見たセマンティックWebとWebサービスの比較

データモデルを定義しており，各種ツールも整備されつつある．ただし，以下に示すいくつかの点を考慮する必要がある．

まず第1に，MediatorおよびPlanning engineにあたる機能が，現時点では，両者共に不明確である．セマンティックWebでは，論理層がまだあまり議論されていないこともあるが，どのように複数のSourceからのメタデータを結合し，処理を行うかが大きな問題である．Webサービスではメタデータの記述を限定することで，サービス間の動的な結合を可能にしている．しかし，結合されるサービスのインタフェースは正確に一致しなければならない．

第2に，Wrapperの機能，つまり，ローカルスキーマとグローバルスキーマの翻訳を行う機能および必要な知識を開発するには，大きなコストがかかる．最近，既存のプログラムからWSDL文書を生成したり，WSDL文書を読み込み，定義されているサービスの雛形となるプログラムを自動生成するツールが提案されているが，いかに簡単にWebサービスを構築するかが大きな問題である．一方，セマンティックWebの場合，Wrapperを構成するということは，ローカルスキーマの一種であるHTML文書から，グローバルスキーマであるRDFで記述されたメタデータを付与するという作業に相当する．Web上の膨大なHTML文書に1つ1つメタデータを付与するのは現実的でなく，メタデータのオーサリング・自動生成ツールやメタデータの再利用のための仕組みが必要である．

おわりに

人工知能やデータベースの分野で古くから知られている情報統合の立場から，セマンティックWebとWebサービスがどのように位置づけられるかについて議論した．世界最大の分散環境であり，情報が日々変化し続けるWeb環境において，セマンティックWebとWebサービスは，有望な次世代Webのモデルとして注目を集めている．従来の情報統合システムでは考えられなかった情報の動的な変化や多様性，質の格差など，Web環境特有の条件をうまく克服することで，現在のWebでは実現できない高度な環境が生まれることを期待したい．

参考文献

1）Chawathe, S., Garcia-Mokina, H., Hammer, J., Ireland, K., Papakonstantinou, Y., Ullman, J. D. and Widom, J.: The TSIMMIS Project: Integration of Heterogeneous Information Sources, In 16th Meeting of the Information Processing Society of Japan, pp.7-18, Tokyo, Japan (1994).

2）Ullman, J. D.: Information Integration Using Logical Views, Theoretical Computer Science, Vol.239, No.2, pp.189-210 (2000).

3）Wiederhold, G.: Mediators in the Architecture of Future Information Systems, In Michael N. Huhns and Munindar P. Singh, editors, Readings in Agents, pp.185-196, Morgan Kaufmann, San Francisco, CA, USA (1997).

4）武田英明：ネットワークを利用した知的情報統合，人工知能学会誌，Vol.11, No.5, pp.680-688 (1996).

5）兼岩　憲，佐藤　健：D1: Description Logics, 人工知能学会誌，Vol.18, No.1, pp.73-82 (2003).

6）浦本直彦，佐藤直人，中村祐一：Webサービス合成のための振る舞い記述の仕組み，人工知能学会誌，Vol.17, No.4, pp.408-415 (2002).

（平成15年4月1日受付）