

RSSは新たな時代の入口にすぎない セマンティック Web 入門

ウェブ上に、人間だけでなくコンピュータにも理解できる「知識」を構築することで、現在のウェブをより進化させるための標準化作業が、ウェブ技術の国際標準化組織 W3C(ワールドワイドウェブコンソーシアム)で現在進められている。

「セマンティック Web」と呼ばれるその技術が果たしてどのようなもので、どのようなメリットが生じるのか、また実現に際してどのような技術が必要なのかについて、それぞれ紹介する。

平川 泰之

World Wide Web Consortium(W3C)Asian Communications Officer
慶應義塾大学 大学院 政策・メディア研究科 助手

ウェブの誕生と ポジティブスパイラル

1989年にCERN(ヨーロッパ素粒子物理学研究所・ジュネーヴ)に勤務していた若き日のティム・バーナーズ・リー(Timothy John Berners-Lee)によってウェブは発明された。分断されていない1つのウェブを全てのユーザーに保証すべく、開発したプログラムのソースコードをロイヤリティフリーに基づいて無償提供したことで、1990年当時、CERNにたった1台しか設置されていなかったウェブサーバーは、僅か15年で今日の数え切れない程多くのウェブサーバーへと普及し、ウェブは人類にとって不可欠な情報インフラへと成長している。

ウェブは、XHTMLやHTMLといったウェブページ記述言語、記述されたページのいわば「住所」にあたるURI、そして実際にページを取得するための通信プロトコルHTTPという、3つの簡潔な技術によって構築されたインターネット上の情報空間である。その簡潔さゆえ、ある時期を境に急速にユーザーの裾野を広げ、ウェブ上の情報量も爆発的に増大し

た。それに伴ってウェブ自体の有用性も増し、さらにユーザーが増えるというポジティブスパイラルをもたらした。

今日、ウェブ上では膨大な量の情報が日々刻々と公開、更新、蓄積されており、それらが互いに「リンク」されることで、時には情報を公開したユーザーの予想をはるかに超えるような様々な付加価値をもたらすことさえある。

一方、情報を閲覧する側のユーザーは、リンクをたどったり、キーワード検索をすることで、膨大な量の情報の中から必要とする情報を見つけることができる。しかしいつも必要とする情報にたどり着けるとは限らず、必要とする情報が見つからなくて悶々とするこもあれば、何となくたどっていった先で思わぬ掘り出し情報に出くわすなど、ウェブの奥は深い。

近年では、携帯電話でウェブにアクセスし、時刻表や乗換情報を調べたり、最寄の流行のレストランを検索して、そのメニューや混雑状況を確認したり、さらにはオンラインフォームから予約を入れたり、といったようなことも十分実現可能となっている。

人にしか分からないウェブ

このように生活の隅々にまで浸透したウェブではあるが、実際の利用ではまだまだ不便なところも多い。

例えば、GPSなどの位置情報検索には頼らずに、最寄のレストランを検索する場合、店名や地名などを知っていれば、最近の検索技術を持ってすれば、かなりの確率で検索が可能だ。もちろんそのレストランがウェブ上に情報を公開していることが前提だが、仮にレストラン側で公開していない場合でも、一般のユーザーが感想などを交えて情報を公開している場合などもあり、有名店などであれば検索できる可能性は高い。しかし、実際に足を運んでみたら定休日だった、なんてことになったら何とも興ざめである。

もちろん、レストランのウェブページや、一般ユーザーの感想ページに、定休日の情報がきちんと書かれていたかもしれないし、書かれていなかったかもしれない。いずれの場合も、ユーザーが該当するウェブページを実際に閲覧し、確認しない限り分からないことには違いがない。また定休日の情報を検索したいがために検索ワードに「定休日」などと追加してしまうと、かえって必要とするページが検索されなくなってしまう可能性もあるなど、実際にはなかなか難しい。

コンピュータにも分かるウェブへ

このような問題は、現在のウェブを構成する情報の多くがXHTMLやHTMLで記述された、人間が実際に見たり読んだりして利用することを前提としたものであることに起因する。人が情報を利用するのだから当たり前、と考えるのは早

図1 現在のウェブとセマンティック Webの違い

計である。前述した通り、ウェブ上には日々膨大な量の情報が蓄積されている。そのような情報空間から必要な情報を探すのは、人よりもむしろコンピュータが得意とする分野である。

そこで、人が見るための情報だけでなく、コンピュータにも分かる情報を用意することで、ウェブ上の膨大な情報をコンピュータ自身が解釈し、一定レベルの処理の自動化を実現しようとする技術こそ、セマンティック Web である。

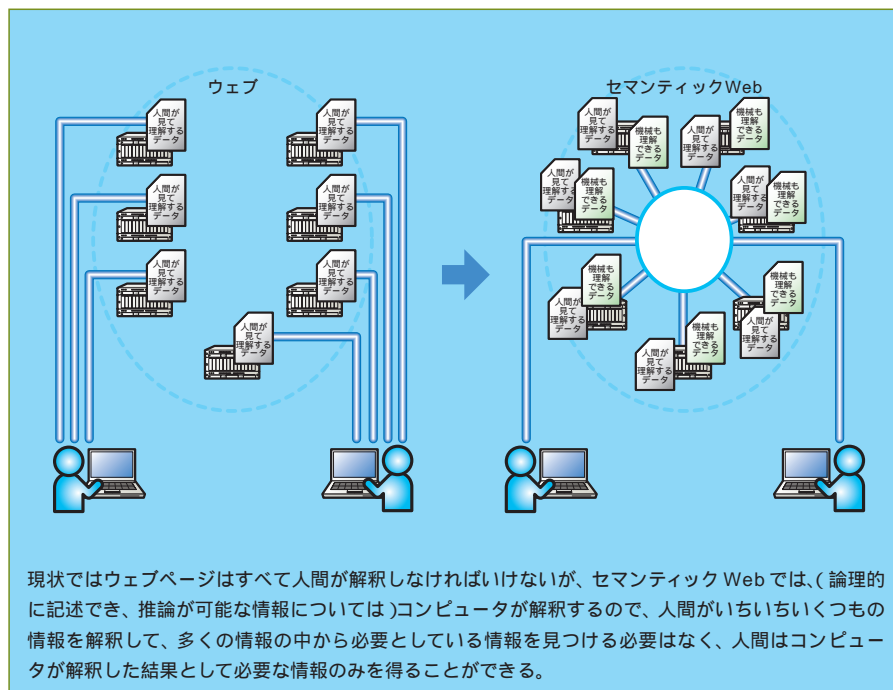
レストラン検索の例に戻れば、XHTML や HTML を用いて記述されたコンテンツに対し、そのコンテンツを利用する上で有用と思われる情報、例えば、店名、住所、開店時間、定休日、メニューなどを、メタデータとして用意し、コンピュータでも解釈できるようにしておく。検索エンジンもメタデータに対応し、指定した属性に対して検索を行うことで、検索率を格段に高めることができるようになる。

このように人が閲覧することだけを前提とした文書で構成されたウェブ(Web of Documents)から、コンピュータにも解釈可能なデータも含むウェブ(Web of Data)を実現し、その上で様々なデータ処理を実現することが、セマンティック Web の目標である(図1)。

セマンティック Web は現在のウェブの置き換えではない

セマンティック Web の実現に向け、すでに様々な技術が標準化されている。しかし、セマンティック Web が現在のウェブを置き換えるものではない、という点については誤解の無いようにしなければならない。

図2からも分かるとおり、セマンティック Web は様々な技術で構成されるウェブの1側面に過ぎない。従って、既存の XHTML や HTML で記述されたウェブページが利用できなくなってしまう、とい



ようなことではない。様々なウェブ技術は、より利用し易いウェブの実現に向けて目的に応じて有機的に連携し、ユーザーの利便性向上に寄与するであろう。

一方コンテンツ制作者にとっては、適切にメタデータを用意することによって、自身が制作したコンテンツがより適切に利用されることが期待できる。もちろん、メタデータをページごとに記述するのはページ作成の際に追加的な作業が発生する可能性を示唆しているが、コンテンツ制作ツールがメタデータの記述支援機能を備えることによって、十分コストに見合った作業に収まるであろう。またウェブアプリケーションプログラマにとっては、用意されたメタデータを適切に利用することで、より効率的で高機能な様々なウェブアプリケーションの実現が期待できるであろう。

当初から検討されていたセマンティック Web

そもそもセマンティック Web に通ずる、コンピュータによる自動処理が可能なウェブ上の情報についての考え方は、

ティム・バーナーズ・リーの1989年当時のプロポーザルの中ですでに言及されていた。

1994年にW3Cが設立されると、業界を挙げての技術の標準化が開始される。この中で、メタデータを取り扱う最初の技術として技術として、ウェブ上のコンテンツをフィルタリングするためのPICSという技術が1996～7年にかけて標準化された。当時既に問題となっていたウェブ上の暴力や性表現を含むコンテンツに対してそれを示すメタデータを付与することで、フィルタリングを実現する技術である。

PICSは当時ある程度の成功を収めたものの、実際にはPICSだけでは記述できないことも多く、汎用的なメタデータ記述の必要性が明確になった。そこで、策定されたのが1999年版のRDF(Resource Description Framework)である。同時期に、ティム・バーナーズ・リー自身によって実際に「セマンティック Web」という表現が初めて用いられ、この頃から現在検討されているようなセマンティック Web の姿が明確化されるようになった。

図2 W3C技術階層図

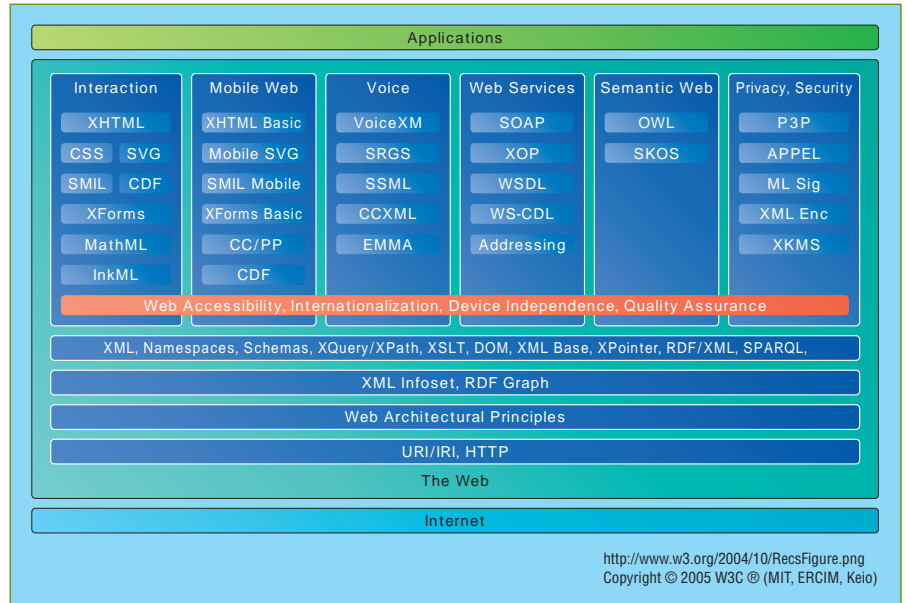
RDFの第2版とOWLの策定

その後RDFを中心とした活動は、その応用が中心となり、ウェブ上でのプライバシー方針を記述するP3Pや、携帯機器向けの機器特性及びユーザー嗜好の記述を行うCC/PPなどに応用される。並行してRDF自体も他のW3Cの最新仕様に合わせて更新され、OWL(Web Ontology Language)とともに2004年に再度勧告として策定しなおされる。

RDFはXメタデータを記述するための汎用言語で、リソース、プロパティ、値の3つの組でメタデータを記述する。リソースはURIで記述するため、理論上あらゆるものに対してメタデータが記述できる。さらに値もURIを記述することによって、リソース間の関係などを記述することもでき、いわば、任意のソース間に「リンク」を張ることができる(図4)。

一方、OWLはRDFを応用した技術で、「オントロジ」と呼ばれる、メタデータの記述に使われる語彙間の関係を記述することができる。例えば、2つのレストランがそれぞれ店名をメタデータとして記述する際に、一方は「名前」というプロパティを、もう一方は「name」というプロパティを用いてそれぞれ記述した場合、どちらも店名を示しているにもかかわらず、プロパティ名が異なるために同一内容のプロパティとして取り扱うことができない。これでは、せっかく店名がメタデータで記述されても、どちらか一方のメタデータにしか対応していない場合などは、メタデータが適切に利用できない。そこで、「名前」と「name」が同じ意味を持っていることを記述できるようにし、メタデータ間の関係を記述できるようにするのがOWLである。OWLにより、いうなれば、語彙間に「リンク」を張ることができる。

この2つの技術を用いることで、セマンティックWebの基盤となるコンピュー



タが解釈可能なデータをウェブ上に掲載することができる。

現在策定中のセマンティックWeb技術

2004年にRDFとOWLが勧告となり、他の基盤技術と合わせて、セマンティックWeb構成図の中程まで標準化が完了し、データを記述する下地は整えられた。それを受けて、現在はそれらデータへのアクセス及び問合せを実現するSPARQL:Protocol and RDF Query Languageの策定が進められている(図3)。

セマンティックWebとAIの遠くて近い関係

ところで、セマンティックWebはAI(人工知能)と比較されることがある。しかしセマンティックWebとAIとでは目標に対する立場が異なる、というのがセマンティックWeb技術の標準化を推進するW3C関係者間での共通認識である。

どちらも「コンピュータで知識を処理する」には違いないが、AIでは人間並みの推論の実現という目標を設定することが多いが、セマンティックWebでは、ウェブ

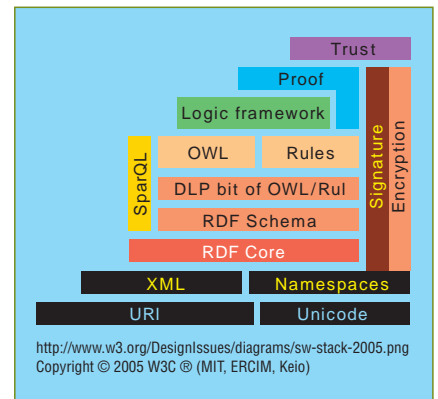


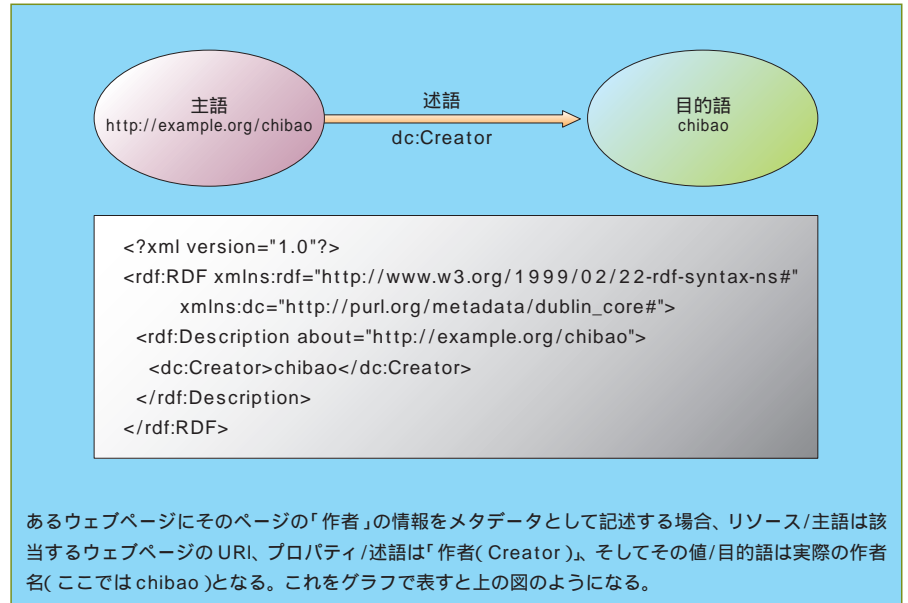
図3 セマンティックWebを構成する技術

上で知識をコンピュータ処理するための共通基盤を構築し、それに基づいてさまざまな知識を相互運用できるようにすることに重点が置かれている。

では、セマンティックWebとAIが全く相容れないものかといえ、そういうわけでもない。実際2005年4月末には、セマンティックWeb構成図において次の標準化対象とする統一なルール記述言語の策定に向けたW3Cワークショップも開催され、これまで個別に進められてきた様々なルール記述言語に関する業界活動や研究成果を受けて、ウェブ上での相互運用に向けた標準化の可能性が検討されている。

セマンティックWebはいわば、これま

図4 RDFとメタデータ記述の例



でのAI分野における膨大な研究開発資産をウェブというオープンな情報空間に応用する試みともいえる。AIの研究成果が全て応用可能ということではないだろうが、これまでのAI研究の知的資産の活用も十分に考えられる。

応用も進む セマンティック Web 技術

今後は、人が閲覧する情報に対するメタデータだけでなく、元々データとして存在する情報をウェブで公開し、広域的に共有することも視野に入れている。これに関連して、昨年10月には生命科学分野へのセマンティック Web の応用に関するW3Cワークショップも開催されている。

生命科学分野では、遺伝子の構造や、たんぱく質の分子構造など、様々な膨大な生命科学データが研究機関ごと蓄積されている。これらをウェブ上に公開し、相互に融通することで、それらデータを用いた仮想空間上でのたんぱく質の合成実験などを可能にし、研究開発スピードの飛躍的な向上を目指している。W3Cでは、元々データ量が膨大で、研究機関ごとに異なるデータ形式を用いていたとしても、オントロジーなどのセマンティック Web 技術を用いてそれらを融通し、相互運用性を確保しようとしている点に高い関心を寄せている。W3Cではそれらの組織と連携し、技術的な支援を行う一方で、実際の運用におけるフィードバックに期待している。

この他、W3C内で並行して標準化が進められているウェブサービスとセマンティック Web の相互運用性なども検討されている。特にウェブサービスにおけるあらゆる側面の記述にセマンティック Web が応用できるのではないかと考えられており、2005年6月にはウェブサービス分野におけるセマンティクスフレームワークに関するW3Cワークショップも開催されている。

信頼できるウェブの実現に向けた 第2のポジティブスパイラルに期待

セマンティック Web が最後に行き着くのは、セマンティック Web 構成図にもある通り、「Trust」の実現である。まだ技術的な開発には着手されていないため、具体的などのような技術になるかは分からないが、より信頼できるウェブを実現する基盤技術として、セマンティック Web が果たす役割が大きいことは確かである。そしてそのセマンティック Web 普及の鍵は、ウェブの爆発的な普及と同様、ひとえにメタデータに対するポジティブスパイラルが生まれるかどうかにかかっている。

RDF と RSS 編集部

RDFの応用の1つがRSS(RDF Site Summary)で、XMLフォーマットでドキュメントの見出しや要約を提供する。「ウェブサイトのサマリーを示すRDF(メタデータ)」という名のとおり、サイトの更新情報を通知する際のデータフォーマットとして利用されている。最初のRSSは、1999年3月にネットスケープが開発/公開したものだが、その後いくつかの派生規格が登場し、中にはRDFに沿わないものも存在する。ウェブサイトの情報以外にも様々なメタデータの公開に応用ができるため、新しい情報伝播/共有の手段として注目されている(44ページの本誌特集を参照)。

参照URI一覧

- ワールドワイドウェブコンソーシアム(W3C)
<http://www.w3.org/>
- ティム バーナーズ-リー
<http://www.w3.org/People/Berners-Lee/>
- Design Issues for the World Wide Web
<http://www.w3.org/DesignIssues/>
- W3C 技術勧告仕様一覧
<http://www.w3.org/TR>
- W3C セマンティック Web アクティビティ
<http://www.w3.org/2001/sw/>

- 生命科学分野へのセマンティック Web の応用に関するW3Cワークショップ
<http://www.w3.org/2004/07/swls-ws.html>
<http://www.w3.org/2004/10/swls-pressreleases>
- ルール記述言語に関するW3Cワークショップ
<http://www.w3.org/2004/12/rules-ws/>
<http://www.w3.org/2005/04/swrules-pressrelease>
- ウェブサービス分野におけるセマンティクスフレームワークに関するW3Cワークショップ
<http://www.w3.org/2005/04/FSWS/program.html>
<http://www.w3.org/2005/06/sws-pressrelease>



[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp