

Semantic Web

～人と機械がコミュニケーションできる Web を目指して～

平井 聡, 森 隆史

Satoshi Hirai, Takashi Mori

1 はじめに

近年, インターネットが普及し, 生活は非常に便利になった. インターネットのおかげで様々な情報を容易に手にいれることができる. しかし, 現在のインターネットでは検索者が意図した情報とは異なる情報を多く検索してしまう. これは, コンピュータがインターネット上の情報を単なるビット列として認識しており, その意味を理解していないために発生する.

Semantic Web は, インターネット上の情報を意味のある文字として機械に認識させることを目的としている.

2 Semantic Web とは

Semantic Web とは, Web 創始者である Tim Berners Lee 氏が目標として掲げる次世代の Web のことである. これまでのように, 人間が理解すること (ヒューマンリーダブル) を目的とした Web ではなく, 機械が理解できること (マシンリーダブル) を目的としている.

2.1 従来の Web 体系

XML は, タグを内容に合わせて設計するコンテンツ処理が可能な言語である. XML の記述例を Fig. 1 と Fig. 2 に示す.

```
<Person>
  <Name>Satoshi Hirai</name>
  <Age>21</Age>
</Person>
```

Fig. 1 XML の記述例 1

```
<個人>
  <名前>Satoshi Hirai</名前>
  <年>21</年>
</個人>
```

Fig. 2 XML の記述例 2

Fig. 1 において, 人間は名前が "Satoshi Hirai" で年

齡が 21 歳であることが理解できる. しかし, 機械では Person や Name の意味を理解することができない. 例えば, Fig. 1 と Fig. 2 が同じ内容であることを機械は理解できない.

このように従来の Web 体系では, 機械が Web の構文の意味を理解することはできないのである.

2.2 Semantic Web の原理

Semantic Web の基本原理は, 全てのデータと情報にマシンリーダブルかつ構造化されたメタデータを付与させることである. それにより機械が Web データの意味を理解でき, 人の代替としてソフトウェアで自動処理することができる.

メタデータはデータを定義するデータのこと, つまり www 上に存在する情報に意味を与えることである. このメタデータの存在により Semantic Web の基本であるマシンリーダブルを実現することが可能である.

その記述の仕組みにはさまざまな階層が定められている. Fig. 3 にその概要を示す.

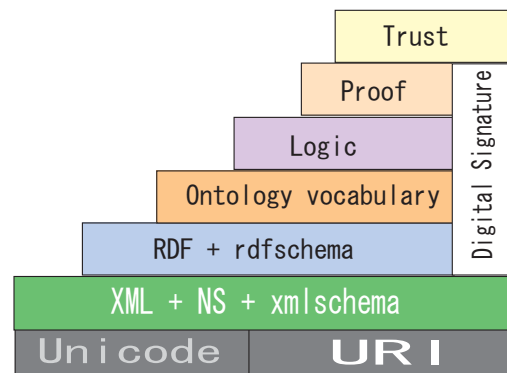


Fig. 3 Semantic Web の architecture

下位層には, 従来の Web にも存在した XML の使用方法や Unicode, URI といった階層が定義されている. XML の上位層に RDF¹, Ontology²といった Semantic Web を構成する上で不可欠な階層が定められている.

¹RDF (Resource Description Framework) とは XML によって文書に付加された構造に対して, その意味を付加するものである.

²Ontology とは「用語間の関係を定義している文書またはファイル」である. メタデータの表現方法が異なる場合, 意味が通じなくなるため, 複数の語彙体系で意味の共有化を図るものである.

2.3 Semantic Webの可能性

現在の Web では、情報者が意図しない情報まで検索される。例えば「今日開いている近藤歯科」で検索したいとき現在の Web では【歯医者 AND 近藤】というようなキーワードを考え検索を行う。そうすると歯医者に行ったという近藤さんの日記や近藤医院などの検索者が意図しない情報が大量に検索される。この中から必要な情報を見つけるのは面倒である。

しかし、Semantic Web ではメタデータの付与することにより検索結果は目的とする情報を多く含む「近藤歯科 = 歯医者という職業の近藤」というように理解でき意図しない情報は検索されない。また、現在では検索が困難な「この近くの」、「明日開いている」といった状況による検索も可能となる。

2.4 本研究室での Semantic Web の必要性

Semantic Web が必要となるのは本研究室でも同様である。研究テーマに関する専門知識を広げるためにインターネットを使うことが非常に多い。しかし、研究テーマについて必要なサイトを見つけるには時間がかかる。Semantic Web の実現により、Web の検索の時間が短縮されるため、効率よく知識を獲得できる。つまり、時間を効率よく使えるのである。

3 Semantic Web の課題と今後の展望

3.1 Semantic Web の課題

Semantic Web は有用な技術だが、実現には Ontology の確立やメタデータの付与といった課題がある。

Ontology に関しては、Semantic Web のおいて情報を記述する関係をきめるという意味で非常に重要な意味をもっている。しかし、この Ontology は世界で非常に多く存在するため、統一するのは困難である。

メタデータの付与に関しては、世界中の Web にメタデータを付与することは難しい。なぜなら、現在世界中には 20 億 ~ 40 億の大量の Web ページが存在しているためである。このような大量の Web ページにメタデータを付与する方法として以下の 3 つが考えられている。

- 機械による自動付与
- 利用者（とツール）による手動付与
- 機械と利用者による半自動付与

これらの方法で一番有効と見られている方法が、半自動付与である。半自動付与とは機械処理で半自動的にメタデータを付与し、その後専門家によりそのデータをチェックする方法である。

自動付与の場合は、確かに数がこなせるが、精度を上げることはできない。手動付与はより簡単に使えるユーザインターフェイスが研究が必要となる。

3.2 Semantic Web 実現における本研究室の課題

本研究室は Mikilab のホームページを外部公開用として持っており、基礎ゼミ資料や月例発表会の資料は、非常に多くの人々が閲覧している。Mikilab のページが Semantic Web の機能を持つと、多くの人々が効率よく研究ページを閲覧することができる。そのためには Mikilab のページにメタデータを付与しなければならない。つまりホームページに作成する際に、RDF、Ontology やメタデータの付与に関する知識が必要となる。Semantic Web の実現により効率的な検索はできるが、実現するためにはメタデータを付与する努力や知識が必要となる。

3.3 今後の動向

Semantic Web を実現するために、Semantic Web の認知度を高くしなければならない。しかし、現在の日本での Semantic Web の認知度はきわめて低い。これは、Semantic Web の開発に日本が出遅れているためである。

日本ではこの状況を打開すべく INTAP³が普及活動を今年から開催する予定である。Semantic Web は有用な技術であるため、認知度が高くなればそれだけ活動や研究が活発になると思われる。

つまり、認知度が高くなると、Semantic Web の研究が活性化し、さらに認知度が高くなる。このような好循環が発生することで Semantic Web は将来、一般的な Web 体系になると思われる。

4 おわりに

Semantic Web が実現すれば、Web 利用者にとって大きなメリットが与えられる。しかし、その実現のためには、いくつかの課題を克服し、Semantic Web が広く認知される必要がある。日本では Semantic Web の普及活動が始まったばかりであり、日本での実現にはまだ時間がかかってしまう。しかし、実現した世界はより快適で便利なものになっているであろう。

参考文献

- 1) 財団法人情報処理祖運用技術協会、セマンティック Web 技術の調査研究報告書
<http://www.net.intap.or.jp/INTAP/s-web/data/13-semanticweb-report.pdf>
- 2) 武田英明、日常世界のための Web、システム制御情報学会、システム制御情報学会セミナー 2002「人間と機械のコラボレーションを実現するための技術」、pp45-52、2002
- 3) 武田英明、Semantic Web: 人と機械のコミュニケーションの仕組みとしての Web
<http://www-kasm.nii.ac.jp/papers/takeda/01/SW-takeda-position.pdf>

³INTAP とは、Interoperability Technology Association for Information Processing, Japan の略で情報処理技術の進展を目的として、研究、調査、普及を行っている団体。正式名称は財団法人情報処理相互運用技術協会である。