

セマンティック・ウェブ・アウェアな 言語学習リソースのための基礎研究

木村 修平 (京都女子大学非常勤講師)

syuhei@papiko.com

1.はじめに

コンテンツとその利用者が主に閲覧という静的な関係性で結ばれていた従来の WWW に対して、近年、その関係性をより柔軟かつ動的なものに変える技術が次々と生み出されている。こうした次世代の WWW の方向性を探る試みを、WWW 技術の標準化団体である World Wide Web Consortium (W3C) ではセマンティック・ウェブ (Semantic Web) という名称のプロジェクトとして推進している。

本稿の目的は、セマンティック・ウェブを見据えた (セマンティック・ウェブ・アウェアな) コンテンツの記述方式について、主にメタ情報の付与という観点から、どのように言語の学習リソースとなるコンテンツに適用できるかを、特に英文法の学習コンテンツに焦点を当てて、いくつかの基礎的な適用形態を提言するものである。

2.セマンティック・ウェブの特徴とその背景

セマンティック・ウェブを意識して設計されたコンテンツの大きな特徴として、各種のメタ情報によってデータそのものが、人間に対してのみならずコンピュータにとっても意味を持つという点が挙げられる。これにより、データを持つ柔軟性は飛躍的に高まり、その結果、コンテンツ利用者の立場が、データ伝達されるだけの受動的なものから積極的にデータを活用できる能動的なものに変質することが期待される。

そもそもセマンティック・ウェブという発想の起源は WWW の誕生にまで遡る。WWW の生みの親と

言われる Berners-Lee[1]は、WWW を提唱した初期段階でその最終的な目標を、コンピュータによるデータ分析が人間の重要な知的生産ツールとなることと定義した。この意味で、いくつかの技術的革新によって可能となったセマンティック・ウェブの到来は WWW の根本的な思想への回帰と言えよう。

3.セマンティック・ウェブを可能にする諸技術

Berners-Lee[2]は、セマンティック・ウェブを9つの技術の階層上に成り立つものとして説明している (図 1)。これらの技術は W3C によって順次標準化がすすめられている。

XML (Extensible Markup Language) 層以下によって成り立つのが現在の WWW の中心的な構造である。この上に、データのメタ的な意味づけを行う RDF (Resource Description Framework) 層と、異種メタ情報間のスキーマ統合や調整を行うオントロジ (Ontology) 層を加えることで、セマンティック・ウェブの根幹部分が築かれることになる。

言うまでも無く、セマンティック・ウェブを見据えたコンテンツとは、XML 層以下での要件を満たしていることが前提となる。たとえば、セマンティック・ウェブを見据えたコンテンツを XHTML (Extensible HyperText Markup Language) で作成すると仮定すると、その大前提として、まず、Unicode によるマークアップの記述と URI (Uniform Resource Identifier) によるコンテンツの識別、さらに、コン

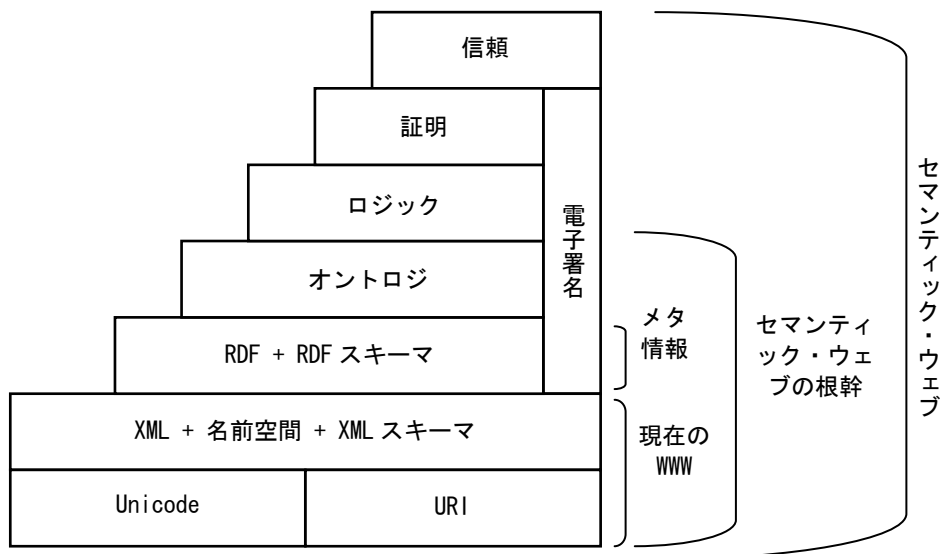


図 1: セマンティック・ウェブを可能にする諸技術

テンツ構造と名前空間の適切な記述と、主にユーザー・エージェント向けとなるメタ情報を `head` 要素内に記述することが求められる。

上記を鑑み、セマンティック・ウェブ・アウェアな言語学習向けのコンテンツを作成する上での最大公約数的な提言として、どのようなメタ情報をコンテンツに与えることができるかを次節で検討する。

4. セマンティックな言語学習コンテンツの作成

ここでは、セマンティック・ウェブの根幹を成すいくつかの基礎的な技術の、英文法学習用コンテンツへの適用方法を検討する。

4.1 HTML のメタ情報と DCMES による記述

WWW でのもっとも一般的なメタ情報の提供方法は、HTML の `head` 要素にメタ情報を記述することである。これらのメタ情報は、たとえば `link` 要素などのように主にユーザー・エージェントに対して意味を持つものであるが、たとえば `keywords` や `description` といったプロパティは各種の検索エンジンに対してドキュメントの情報を効率的に伝達する手段にもなりうるため、WWW の共有性を高める意味でも適切な記述が望ましい。

メタ情報の統一的な書式として昨今注目を集めているのが、メタ情報の記述の標準化を推進する Dublin Core Metadata Initiative が提唱する Dublin Core Metadata Element Set (DCMES) である[3]。DCMES では、コンテンツの作者、タイトル、作成日といった情報を記述するための汎用ボキャブラリを定めている。DCMES は、名前空間を宣言するだけで XHTML の `head` 要素や、後述する RDF のプロパティとして直接書き込める (図 2) ため、扱いが極めて簡便であることから、膨大になることが予想される英文法学習サイトのコンテンツの管理に大いに役立つと思われる。

```
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xml:lang="ja" lang="ja">
```

図 2: DCMES を XHTML の名前空間宣言に追加(下線部)

4.2 RDF によるリソース関係の記述

RDF は、リソースの持つデータ・モデルを記述する技術として W3C が勧告している枠組みである。HTML のタグ付けだけでは表現できないリソースの体系を、特定のアプリケーションに依存することなく表現するための手法として注目を集めている。

RDF を用いると、コンテンツに含まれる複数のドキュメントないしは名前空間をコンピュータにとって意味のある体系として記述できる。そのため、英語学習コンテンツの場合、各単元や項目ごとの繋がりが参照項目を RDF による記述で繋ぐなどの利用方法が考えられる。また、RDF の技術を活用した更 RSS (RDF Site Summary) を用いれば、学習コンテンツ提供者がコンテンツの更新情報を迅速に伝える

ことができる他、各種のオンライン端末にコンテンツを柔軟に配信できる可能性もある。

4.3 DTD を用いての英文構造の表現

セマンティック・ウェブでは XML によるコンテンツ記述が前提となるため、複数のスキーマ言語を混在させて多様なデータ表現を行える。

英文法の学習コンテンツには多数の英文が含まれることが予想される。スキーマ言語である DTD (Document Type Definition) に基づいてそれらの英文の構造を、たとえば品詞単位でタグ付けすることもできる。品詞の分類や識別には諸説あるため一元的な表現は難しい可能性があるが、英文に含まれる各単語に意味を持たせることにより、英文は単なる文字列の集合ではなくなる。結果として、全コンテンツの中から、特定の品詞組成構造に合致する英文のみを検索・抽出することも可能となる。

こうした試みの一例として、数式を記述するための言語として W3C が勧告している MathML (Mathematical Markup Language) があり、利用が進んでいる (図 3)。

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow><mrow>
    <msup><mi>x</mi><mn>2</mn></msup><mo>+</mo>
    <mrow><mn>4</mn><mi>x</mi>
  </mrow>
  <mo>=</mo><mn>4</mn></mrow><mo>=</mo><mn>0</mn>
</mrow>
</math>
```

図 3: MathML を用いた数式の記述例

5. おわりに

英文法学習コンテンツは、その製作に膨大な労力を必要とする種類のコンテンツである。それだけに、ドキュメント間の有機的な関連性とデータの適切な意味づけは、コンテンツの柔軟な利用を可能にする。

セマンティック・ウェブの技術は標準化の過程にあるものも少なく無いため導入には慎重を要するが、学習コンテンツの共有性を高めることができれば、教授者・学習者の双方にとってより広範な教育機会への利用可能性が広がることは間違いないだろう。

6. 参考文献一覧

- [1] Berners-Lee, T. 1996. *The World Wide Web: Past, Present and Future*. (online), available from <http://www.w3.org/People/Berners-Lee/1996/ppf.html>, (accessed 2006-06-10).
- [2] Berners-Lee, T. 2000. *Semantic Web on XML*. (online), available from <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html>, (accessed 2006-06-10).
- [3] Dublin Core Metadata Initiative. 2004. "Dublin Core Metadata Element Set". (online), available from <http://dublincore.org/documents/dces/>, (accessed 2006-06-10).