

WWWにおける文書記述言語としてのXMLについて

eXtensible Markup Language on WWW

濱崎 雅弘, 渡邊 真也 (知的システムデザイン研究室)

Masahiro HAMASAKI, Shin WATANABE (Intelligent Systems Design Laboratory)

Abstract In this paper, I introduce the language for world wide web. That language called XML(eXtensible Markup Language). Every networker expect XML to transform World Wide Web from only document expediter into infrastructure. Why XML is prospective language on WWW so much. I give an explanation with making a comparison between HTML(now defact standard language on WWW) and XML.

1 はじめに

一時期の Internet ブームのおかげで, WWW は世間一般に広く普及した。いまや WWW の無い Computer 環境は想像できない物となっている。WWW は Homepage を見るメディアとして一定の地位を築いたと言っても間違いはないが, WWW はメディアとして次の段階へと移ろうとしている。具体的に言うとデータ配送手段としての WWW である。しかし, WWW における標準文書記述言語である HTML は見出しや強調などの文書表現としての構造しかなく, WWW に期待されるデータの流通経路としての役割を HTML は実現する事ができない。なぜならば HTML は HyperText を実現するためのものであり, 文書内の意味的構造, すなわちデータとしての文書を表現する事ができない。これが HTML の WWW における限界である。

そういった環境で現れたのが XML である。XML は HTML と違って, SGML という強力な文書記述言語のサブセットという存在であるため, HTML 以上に文書構造を正確に伝達する事が出来る。また, 設計段階から Internet を意識しており, SGML よりも WWW に対応した能力も備えている。本発表では, XML が現在の標準言語である HTML と比べて, 具体的にどの様な点で次世代の WWW に適しているかを検証する。

2 HTML について

HTML(Hyper Text Markup Language) は HyperText を実現するマーク付け言語である。HyperText とは, 文書間を自由自在に越える環境を持った文書を実現する文書記述言語である。すなわち HTML と

はリンク機能を持った文書を記述する言語である。前出の通り, HTML は現在の WWW 上のほぼすべての文書を記述している。対応しているブラウザやエディタも多く, ほぼ全てのマシンにおいて扱うことが出来る。しかし, Internet に対する要求が単に閲覧する以上のモノとなっている昨今, HTML は WWW 上の文書をすべて記述する言語として限界が来ていることが指摘されている。主な問題点を以下に示す。

HTML の問題点 HTML は元々, SGML 宣言と文書型定義 (DTD) の組み合わせによって定義されたマーク付け言語である。HTML の文書型定義は「異なるプラットフォーム間で受け渡し可能な HyperText を作成するのに使われる、簡素な表記法」であることを目的としている¹。なぜならば, WWW がまだ普及していなかった頃はネットワークもそれに接続したコンピュータも脆弱であったために, 簡易的な形式にせざるを得なかったことに起因する。

しかし, 現在はネットワークやコンピュータは以前と比べて飛躍的に高性能化し, それに伴い WWW も大規模・複雑化した。それに応じて HTML も W3C の主導のもとに拡張し続けている。しかし, WWW が文書以外のテキストを扱うこととなった今, HTML では WWW 上の全ての文書を記述することは不可能になった。

このような問題点を解決するために XML が作られた。XML は技術の向上によって作られたものではなく, 時代の要求に合わせて作られたものである。それは決して HTML のプラスではなく, HTML が対応できなくなった WWW に合わせた, 特殊な言語である。

¹W3C の HTML3.2Reference Specification を参照

3 XMLについて

XML(eXtensible Markup Language)はHTMLとはともにSGMLから生まれたものである。SGMLは論理構造と文書型定義をもつ、強力な文書記述言語である。すでに歴史も長く、欧米での利用も多くあり、実績はHTMLやXMLの比ではない。しかしその複雑さが利用者の多いWWWに向いていないため、WWWに向けて設計されたSGMLができた。それがXMLである。対して、SGMLの文書型定義によってHyperTextを実現したMarkup言語、それがHTMLである。両者の関係は密接であり、共にマーク付け言語ではあるが、二つは全く別のものであり、直接的にはXMLとHTMLは関係はない²。それ故に、多くのXML関係者はXMLが出てHTMLが無くなるわけではない、と述べているのである。

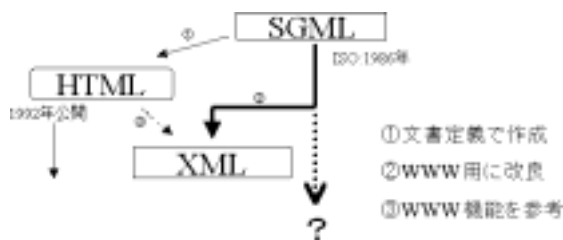


図 1: SGML-HTML/XML

XMLは論理構造と文書型定義によって構成された構造化文書記述言語である。HTMLとの決定的違いは、文書型定義を独自で作ることが出来る(つまり独自にタグが作れる)ところにある。この機能によって、XMLはHTMLには無い多くの特徴を持つことが出来る。主な特徴を以下に示す。

検索の容易性 文書型定義によって作られた独自タグによって、HTMLの様に画面表示を目的としたタグではなく、情報の意味表示を目的としたタグを作ることができる。これによりWWW上の電子文書は「自分を知っている」文書になることが出来る。

電子文書が自らの情報を表現できる事になれば、既存のサーチエンジンの様に、ページ紹介にそのページの頭何文字かを取るような必要もなくなり、適切な紹介と検索結果を出すことが出来る。この検索の容易化の実現は、大規模・複雑化するWWWにおいて非常に重要な位置を占める。

表現の多様性 XMLは文書型定義によって、様々な書式を生み出すことが出来る。数式を表すための

²ただしHTMLのリンク等のWWW技術はXMLに反映されている

XMLであるMathMLは以前から公表されている。さらに、先日IBMからSpeechMLという、音声を表現するマーク付け言語が発表された。もはやXMLはテキストという領域さえ越える表現を記述している。この柔軟な表現力こそがXMLの最大の強みであり、WWWにとどまらない記述言語として期待される所以である。

データの記述 XMLは独自タグを使うことによって、XMLは業務システムなどに必要な情報を体系的に表すことが出来る。WWWにおいてXMLが使われるようになると、WWWは完全に文書だけでなくデータを送ることが可能になる。現時点ではXMLは文書記述言語としてではなく、むしろデータ記述言語として利用されている。(なぜならエディタやブラウザの普及が必要な文書記述言語としてのXMLよりも、アプリケーションとセットで扱われるデータ記述言語の方が利用が簡単だからである)

具体例を挙げると、ポイントキャストのチャンネルやIE4の自動更新などのデータはすべてXMLによって記述されている。さらにBSデジタルのデータ放送仕様はXMLが中心となると決まった。また、現在期待されているものとして電子商取引におけるデータ記述があげられる。

4 まとめ

HTMLの問題点を挙げた際にも述べたが、HTMLが不可能になったのは、「WWW上の文書全てを記述すること」である。あくまでHTMLは簡単な文書表現を目的とした記述言語であり、XMLは多彩な文書表現を実現することを目的とした記述言語である。XMLがWWWの標準記述言語になったとしても、HTMLの役目が終わるわけでもなく、HTMLは続いていくものと思われる。

WWWの検索の容易化、文書の多様化、そして高度なデータ記述を可能にするXMLは、WWWを間違いなく新しいメディアへと変える力を持っている新言語である。

参考文献

[1] 村田真『XML入門』(日本経済新聞社, 1998)
[2] リチャード・ライト『XMLを知る』(プレンティスホール出版, 1998)