

XML および XHTML の現状と今後の動向

the actual situation and the future of both XML and XHTML

佐野 正樹, 川崎 高志
Masaki SANO, Takashi KAWASAKI

Abstract: This is introduction of XML(eXtensible markup language) and XHTML(eXtensible HTML). XML was settled on as the new standard of markup language by W3C. And, W3C also decided on XHTML, which succeeded to HTML and is based upon XML. In this paper, We try to give consideration to the actual situation and the future of both XML and XHTML.

1 はじめに

本発表では, XML と XHTML の現状を説明し, 将来の動向について考察する.

XML(eXtensible Markup Language) は Web 上での使用を想定した, 文章の構造を定義する言語であり, HTML に代わるものとして W3C¹によって策定された.

既に民間企業では, 業務文書やデータベース (以下 DB) において XML の採用が行われている^{1, 2)}. しかし, Web ページ記述を直ちに XML に移行することは不可能なので, HTML²に XML の仕様を盛り込んだ XHTML(eXtensible HTML) が考案された (Fig. 1).

以降では, この 2 つの構造化文書記述言語の特徴を踏まえ, 今後の可能性について考察する.

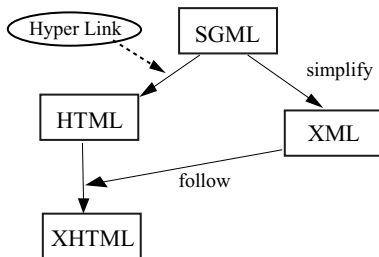


Fig. 1 markup languages

2 XML の背景と現状

2.1 HTML の肥大化と XML の出現

XML は HTML の反省をもとに作られたものである. HTML は本来, 文書構造を記述するためのものであった. しかし, 文書のレイアウトを定義したり, 特定のブラウザに依存した機能など, 仕様の肥大化が進んだ. ま

¹World Wide Web Consortium. プロトコルの相互運用性を高めること, World Wide Web の進化促進を図ることを目的として設立した, 米国マサチューセッツ工科大学コンピュータ科学研究所 (LCS), フランス国立計算機科学・制御研究所 (INRIA), 日本慶應義塾大学湘南藤沢キャンパスが共同運営するコンソーシアムである.

²Hyper Text Markup Language. SGML を拡張し, ページ間のリンクの指定やデータ入力フィールドを指定できる.

た, WWW で受信したデータに対し, 閲覧や印刷だけでなく, プログラムを用いた多様な処理 (解釈や変換など) を行いたいという要求も生まれた. 標準的で汎用的な構造化文書記述言語としては SGML³があるが, これは多機能で複雑である. そこで W3C は, SGML を簡略化した言語である XML を策定した.

2.2 HTML との比較

HTML と比較した XML の特徴としてまず挙げられるのは, DTD(Document Type Definition) との分離である. DTD とは, 文書の型を定義するものである. HTML ではこれが言語の仕様に組み込まれていた⁴. XML では, DTD やタグ⁵を独自に定義できるため, その構文は柔軟である. 文書の送信側と受信側で DTD を統一すれば, 文書の構造に関する情報の損失を抑えることができる. また, スタイルを記述する XSL などをそろえることにより, レイアウトのばらつきを防ぐことができる.

さらに, 文法が厳格であり, プログラム⁶による処理に適している. 例えば, 文書の検索・検証や, 目次などの派生文書の生成を自動化できる. これにより, Web 上での文書 DB の連携や, WWW エージェントによる文書情報の選択などが容易になるのである (Fig. 2).

2.3 現在の実用化例

現在のところ XML は, Web ページ記述よりも, 企業間での電子化文書の交換が主な用途である.

2000 年 5 月から, NTT データオフィススマートや, プラスから分社する「Biznet」が, XML を導入した企業向け文房具販売サービスを相次いで開始する¹⁾. 2.2 節で述べたように, XML では独自の構造を定義でき, レイアウトの統一も容易である. 企業同士の取引伝票の

³電子文書交換のために ISO が標準化した汎用の文書定義言語. マークアップ方式によって文章の論理構造を記述する.

⁴HTML3.0 などが, それぞれ一つの DTD を指す.

⁵SGML やそれに準拠した言語では, 文章をエレメントの集まり (構造) とみなす. これらの言語では, エレメントを区切るためにタグと呼ばれる文字列を用いる.

⁶DOM や SAX などの標準 API や, Java, Javascript, VBscript など

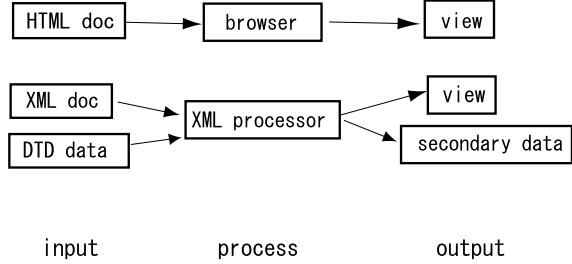


Fig. 2 The comparison of HTML with XML

内容を XML のデータ構造で置き換えることで、小売・卸売・メーカーの間で、同じ内容の伝票が人手で重複起票されるのを防ぎ、注文から伝票印刷までの工程を自動化できる。

また、保険見積もりサイト「bang!」を運営するウェブグループは、8月から保険会社とのデータ交換に XML を採用する。これは、ユーザに対する保険の見積もりの表示を、リアルタイムに表示することが目的である¹⁾。

このサイトでは、まず顧客の入力データを保険会社のサーバーに送る。サーバではデータを処理できる形式に変換して見積もりを算出し、サイトに返す。このデータ交換に XML を用いることにより、サーバでのデータ形式変換を自動化し、見積もりがユーザに届く時間を短縮できるのである。

こうした XML の導入では、運用面の問題が大きい。データ交換を行う企業間で、データの形式を揃えなければならないからである。業界全体に適用するには、さらに大掛かりな標準化が必要となる。この点が、XML 導入の最も大きな障壁であろう。

3 XHTML とは

3.1 HTML の後継

XML は、HTML に代わる Web ページ記述言語として策定されたものである。しかし、すぐに XML に移行することは困難である。なぜなら、XML では文書構造を自分で設定できる仕様になっており、既存の HTML 対応ブラウザに表示できないからである。作者が独自に設定した構造に基づく XML 文書を既存のブラウザに表示させるためには、XML 文書を HTML 文書に変換するソフトが必要となる。

そこで、HTML との下位互換性を維持しつつ XML に準拠する仕様として、XHTML が策定された。XHTML は、HTML4.01 の後継に位置づけられ、2000 年 1 月 26 日に W3C の勧告により事実上の標準となった³⁾。

3.2 Web ページへの適用における利点

XHTML は、HTML の文法を XML に沿って厳格化したものである。よって、XML の良い面のいくつかを

継承している。Web ページ作者の立場でのメリットは、ブラウザ依存性がないことである。今までの HTML のように、閲覧環境に応じて記述の仕方を変える必要が無い。

また、XML 対応ソフトでも処理が可能である。XML と同様に、文書の検索や整形形式かどうかの検証などを行うことができる。

4 今後の展望

Web ページ記述の XML への移行は、WWW を単なる娯楽でなく、情報交換の場とするために不可欠である。その足がかりとして、XHTML の普及がある。

しかし、実際に普及させるには、大きく分けて 2 つの課題がある。一つは、今まで蓄積された大量の HTML 文書をどう扱うかであり、もう一つはこれから発生する文書をどうやって XHTML で書かせるかである。

前者については、現時点では保留であろう。XHTML が、HTML との下位互換性を備えていることが、それを示唆している。

後者の課題を解決するためには「書く」環境と「見る」環境の整備が必要である。現在は、HTML の記述法を知らなくても Web ページを作成できるツールが多数存在し、またそれだからこそ HTML が普及したと言える。マークアップ方式による記述は、一般人にとって馴染み難いものだからである。XHTML に関しても「XHTML を知らなくても XHTML が書ける」ようなエディタを広めることが必要である。しかし、既存のページ作成環境が直ちに消滅するわけではなく、それによって出力された文書が、今後も Web 上に発生し続けるであろう。「見る」に関しては、HTML との下位互換性が完全なら問題ない。

以上より、Web 上では今後しばらくの間、HTML と XHTML が混在し、XML や XHTML の利用は特定のグループ間に限られるような状況が続くと予想される。

参考文献

- 1) 国米敏弘編『『XML を使え』が合言葉 企業間 EC で利用が急加速』『日経ネットビジネス』第 58 号、(日経 BP 社、2000)
- 2) 吉沢美行「XML を取り込み始めたデータベース」『日経コンピュータ』第 493 号、(日経 BP 社、2000)
- 3) "HTML4 の後継、XHTML1.0 登場。HTML と記述法などが一部変わる", <http://findx.nikkeibp.co.jp/ws/sp00xhtml1.html>