

## XML とその周辺の標準化動向の概要

津田 宏 \*1 浦本 直彦 \*2 上田 隆也 \*3 佐藤 研治 \*4 野村 浩郷 \*5

- \*1 (株) 富士通研究所 htsuda@flab.fujitsu.co.jp
- \*2 日本 IBM(株) 東京基礎研究所 uramoto@trl.ibm.co.jp
- \*3 キヤノン(株) 情報メディア研究所 takaya@cis.canon.co.jp
- \*4 NEC C & Cメディア研究所 satoh@hum.cl.nec.co.jp
- \*5 九州工業大学 情報工学部 nomura@ai.kyutech.ac.jp

近年 W3C (World Wide Web Consortium) を中心に、XML (eXtensible Markup Language) や WWW の文書処理の標準化の動きが盛んになってきている。XML 自体は、1998 年 2 月に W3C により正式に構文が認可され、リンク機構やスタイルシートなどの周辺の標準化が進められつつあり、フリーの XML プロセッサも出回ってきている。ただし、XML は単にコンテンツを記述するためのものではなく、文書のメタ情報や、ユーザプロファイル、コンテンツのアクセス制御といった情報の記述枠組みとしても重要になってきている。本稿では W3C への提案を中心に、最近のネットワーク文書標準化動向を XML を軸に紹介し、今後の展開について考察する。なお、本稿は筆者らの所属する電子協ネットワークアクセス専門委員会の調査内容に加筆したものである。

## An Overview of XML and Its Related Standardizations

Hiroshi Tsuda \*1 Naohiko Uramoto \*2 Takaya Ueda \*3 Kenji Satoh \*4 Hirosato Nomura \*5

- \*1 Fujitsu Laboratories Ltd. (htsuda@flab.fujitsu.co.jp)
- \*2 IBM Research, Tokyo Research Laboratory (uramoto@trl.ibm.co.jp)
- \*3 Canon Inc. Media Technology Laboratory (takaya@cis.canon.co.jp)
- \*4 C&C Media Research Laboratories NEC Corporation (satoh@hum.cl.nec.co.jp)
- \*5 Kyushu Institute of Technology (nomura@ai.kyutech.ac.jp)

Recently, XML (eXtensible Markup Language) and its related standardizations have been intensively proposed to W3C (World Wide Web Consortium). XML is expected as not only a next WWW document format following HTML, but a framework to represent various kind of meta-level information of network documents. In this paper, we outline XML1.0, which was recommended in this February, and explain several proposals based on XML for network-based document processing. This survey is a revised version of the report from JEIDA network access working group.

## 1 はじめに

<sup>1</sup>コンピュータ一般に、また特にネットワークアクセスに関しては、

技術 → 標準化 → 技術 → 標準化 → …

という流れによりブレイクスルーが実現されてきている。TCP/IP, HTTP, HTML といった技術の標準化で爆発的に広まってきた WWW も、

- 膨大なフォーマットフリー情報によるコンテンツ検索の困難
- 膨大なゴミ、有害情報
- プッシュ型情報提示の非効率な実現
- ユーザに合わせた情報処理の不備
- エレクトロニックコマースの諸問題 (電子カタログの検索など)

といった点で、問題点が目立ち始めてきている。全文検索やフィルタリング、マルチキャスト、ユーザプロフィールといった技術により、これらのうち部分的に解決されるものもある。しかし、本質的にこのような弊害を解決するには情報コンテンツの標準化技術は必須と言える。また、そうした標準化のもとで、新たな技術開発のニーズが生まれ、ネットワークアクセスのさらなる進歩がもたらされると考えられる。

計算機関係の世界的な標準化については、ISO (International Organization for Standardization) が国際標準の制定機関としては有名である。文字コード (Unicode は ISO10646 の関連規約) や SGML (ISO8879) などが規定されている。また、インターネット全般、特にプロトコルに関しては、IETF (Internet Engineering Task Force) のワーキンググループで種々の規格が決められている (<http://www.ietf.org/>)。RFC (Request For Comments) は IETF でまとめられたもので、例えば電子メールの RFC822 が有名である。

しかし、最近 WWW 全般や XML に関しては、W3C (World Wide Web Consortium) による規格が重要なことが多い。W3C は企業や大学の協力で 1994 年 10 月に作られた、世界的で中立的な業界団体である。W3C の代表は、WWW の発明者として有名な Tim Berners-Lee である。W3C に提案/認可された規格は、<http://www.w3.org> から誰でもアク

セスすることができる。単なる提案は Note、現在仕様検討中のものは Working Draft、正式に認可された (勧告) されたものは Recommendation として公開されている。

W3C は中立団体とは言え、参加しているのは Microsoft, Netscape, Sun, 日本のコンピュータメーカ各社、数々のベンチャー企業など、WWW の世界で現在盛んにビジネスを展開している企業である。毎月のように W3C に提案される規格には、標準化のイニシアティブを取ることで、少しでも自社のビジネスを有利にしたいという企業のエゴが見え隠れしている。規格制定のワーキンググループも、非常に活発な議論が毎週繰り返されているということである。W3C に提案される規格群は、現在のインターネットビジネスの縮図と言って良いかもしれない。これらを調べていくと、インターネットが将来どのような方向へ進もうとしているのか、臆げながら掴むことができる。

本稿では、ネットワーク上の文書処理に関する最近の標準化動向を、XML (eXtensible Markup Language) を軸に紹介し、将来の技術的展望について考察する。XML は HTML のような文書コンテンツの記述言語としてだけではなく、文書の様々なメタ情報、さらにはネットワーク上の知識表現言語の枠組みとしても利用が望まれている。

以下、XML とその周辺規格、XML を利用した関連規格の順に現状を紹介する。なお、本稿で紹介する規格の状況や URL は、1998 年 4 月時点のものである。

## 2 XML (Extensible Markup Language) の動向

### 2.1 XML の生まれた背景

現在 WWW では、ほとんどの文書が HTML (HyperText Markup Language) で記述されているのは言うまでもない。構文が単純で、マルチメディア文書を容易に作ることができ、ブラウザを通じてレイアウトもすぐに確認できる、といった親しみやすさにより HTML は爆発的に広がっている。今後も、ブラウザの進歩とともに仕様が拡張されて使われ続けるであろうことは想像に難くない。

しかし、HTML は人間がブラウザで見ることを第一に考えているため、これを計算機で処理する文書/データの対象として見ると様々な問題が出てくる。そ

<sup>1</sup>本稿は筆者らの所属する電子協ネットワークアクセス専門委員会の 97 年度報告書 [JEIDA 98] が元になっている。

の大きな原因は、HTMLはタグが固定され、タグがブラウザ上のレイアウトと密接に結びついていることにある。

このHTMLの欠点を補うものとして、最近XML(eXtensible Markup Language)が脚光を浴びつつある。XMLは現在の所ホームページのようなデータの記述に使われるケースは少ないものの、XMLをベースとした文書群のメタデータの規格が次々とW3Cに提案されてきている。既に、Internet ExplorerやNetscapeのCommunicator 5(この4月にソース公開)といったブラウザもXMLに対応してきている。

もともとは、HTMLもXMLもSGML(Standard Generalized Markup Language)から派生した文書記述形式である。歴史をひもとくと、SGMLは文書記述のための国際規格として1986年にISO8879に制定された。HTMLはWWW時代の文書記述の枠組みとして1992年に公開。現在用いられているのはHTML3.2で、次バージョンのHTML4.0が1997年12月にW3Cにより正式に認可されている。XMLはSGMLを簡略にしてWWW時代に合わせたもので、1996年11月にW3Cに最初のドラフトが公開され、1998年2月にW3CによりXML1.0が正式認可された[XML98]。

XML1.0では、構文やDTDの部分だけが決められているだけである。実際のXMLの利用には、以下で紹介するようなXMLのリンク機構、XSL、XML-DataといったXMLを拡張したり表示したりする規格も必須である。そのようなXMLの周辺規格もW3Cで標準化が進められつつある。

こうしたXML自身や周辺規格もさることながら、XMLをベースとしたアプリケーションの規格も続々と提案されていることは注目すべきである。例えば、WWW文書群のメタデータのフォーマットとして、CDF(1997年3月)、MCF(1997年6月)、RDF(1997年10月)、ネットワークを通じて配信するソフトウェアの記述であるOSD(1997年8月)、数式記述のMathML(1998年3月)、アプリケーション統合の枠組みとしてのWIDL(1997年10月)などをW3Cの提案に見ることができる。

無論、これら全てが正式な規格として認められるわけではない。しかし、ここで言えることは、XMLは今や単なる文書フォーマットではなく、ネットワーク上の様々な知識を記述する言語として多方面から囑望されているということである。図1は、XMLを中心とした主な標準化の相関図である。

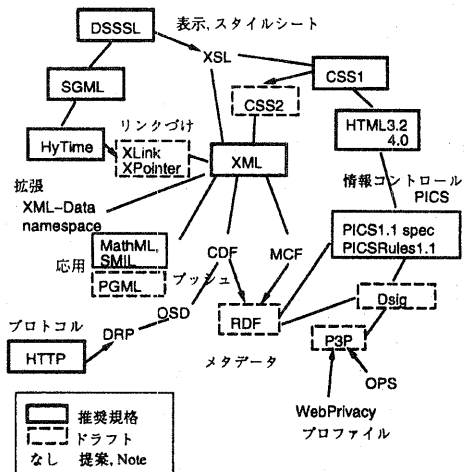


図1: XMLを中心とした標準化の関係

電子商取引や電子カタログにおいても、XMLが脚光を浴びている。XML/EDI(Electronic Data Interchange)グループでは、ビジネス上のデータ交換の protocols としてXMLを使うことを提案している(<http://www.geocities.com/WallStreet/Floor/5815/>)。また、スタンフォード大等で行われている電子カタログプロジェクトのCommerceNet(<http://www.commerce.net/>)も、XMLを採用する方針である。

## 2.2 XMLの特徴

XMLの構文の詳細等の仕様は、W3CのXMLページ(<http://www.w3.org/XML/>)、XML1.0仕様書[XML 98]、[富士通 97]、[村田 98]、富士ゼロックスSGMLカフェ(<http://www.fxix.co.jp/DMS/sgml/>)などを参照されたい。

### 2.2.1 XMLの構文

XML1.0 [XML 98]では、XMLの構文およびDTDについて定義されている。XMLでは、文字コードなど一部を除いてcase sensitive(大小文字を区別する)である。

XML文書には二段階の正しさがある。一つは、適正(well-formed)なXMLである。これは、構文が合っているXML文書のことである。もう一つは正当(valid)なXMLである。これは、適正であってかつ

宣言された DTD によるタグ構造を満たす XML 文書のことである。

## 2.2.2 DTD (Document Type Definition)

DTD により XML の文書構造が規定される。XML 文書の中でそれ自身の DTD を定義する場合と、外部で定義された DTD を参照する場合とがある。XML の DTD では以下の構造を規定することができる。(開始タグと終了タグで囲まれた部分をエレメントとも言うが、以下では特に区別しない)

- タグの構造 (element)。タグの入れ子関係や順序を規定する。
- タグの属性 (attlist)。
- 情報に名前付けする (entity)。エンティティとは、エイリアスやマクロのようなものである。
- グラフィック情報などの形式の記述 (notation)

## 2.2.3 リンク

HTML では、他文書や画像へのリンクは A (anchor) タグ、又は IMG タグによる画像の張り込みだけがサポートされている。XML のリンク機構は、当初は XML 仕様の一部に含まれていた (XLL) が、1998 年 3 月には独立して XML Linking Language (XLink) および XML Pointer Language (XPointer) として、ドラフト段階である [Xlink 98][XPointer 98]。これらの元となったのは、SGML のハイパーリンク、アドレッシング方法、時間・空間管理方法を記述する規約である HyTime (ISO10744 [HyTime]) である。

XLink は HyTime のサブセットになっており、次のような機能が実現される。

- 任意のタグに対して属性としてリンクを記述できる。
- リンク先の処理内容を記述できる。ソースに埋め込むか、文書を置きかえるか、新規ウィンドウを立ち上げて表示するか。また、こうした処理のタイミングも、自動かユーザが指定した時点かを選べる
- リンク先の指定で、文書中任意の要素をリソースとして指定できる。HTML では `<a name="xxx">` とある部分しか指定できなかった。
- 3つ以上のリソースをリンクで記述できる

- リンクとリソースの分離ができる。
- 双方向リンクが書ける。

## 2.2.4 表示系: スタイルシート

HTML ではタグとブラウザ上のレイアウトとの関係が密であった。しかし、XML ではタグとレイアウトに直接の関係はない。

XML 文書をブラウザ上に表示するための情報を与えるのがスタイルシートである。その一つが XSL (eXtensible Style Language) で、1997 年 8 月に Microsoft を中心に W3C で Note が公開されている [XSL 97]。

XSL の元となったのは、DSSSL と CSS である。DSSSL (Document Style Semantics and Specification Language) とは、SGML 向けスタイルシートであり、ISO10179 の国際規格になっている。Scheme 言語で記述する。DSSSL はやや難解でサポートするツールがないのが欠点である。CSS (Cascading Style Sheet) は、元々は HTML のスタイルシートである。CSS1 は 1996 年 12 月に W3C で正式認可されている。XSL は、DSSSL のサブセットであり CSS への互換性も有する。XML + XSL を HTML に変換という動きもあるようである。

なお、CSS1 を HTML4.0 や XML 向けに拡張した CSS2 (CSS level 2) も、1997 年 11 月に最初のドラフトが W3C に提出され、1998 年 3 月 24 日には最終仕様が表示され、正式認可の審査段階に入った [CSS2 97]。

## 2.2.5 意味論: XML-Data, 名前空間

DTD ではタグの名前やそれらの階層を定義することができる。しかし、それだけで情報の論理構造が全て定義されるわけではない。そこで、Microsoft を中心に、XML による広汎な情報 / 知識記述の枠組みとして XML-Data が提案されている。1998 年 1 月、W3C で Note が公開されている [XMLdata 98]。ただし、Microsoft は標準化と関係なく自社製品に XML-Data をサポートするという動きもある。

DTD では、タグの入れ子関係、タグの出現順や出現回数を定義することができる。XML-Data では、DTD ではなくスキーマ (Schema) により情報の枠組みを記述する。スキーマでは、DTD で記述できる内容に加えて、タグの取る値の型や制約、デフォルト値、タグのクラス階層と継承 (ただし単一継承のみ)、

オブジェクト識別子などがサポートされている。タグで囲まれたエレメントをオブジェクトと見ると、XML-Dataの記述は、ほとんどオブジェクト指向知識表現言語である。

このように、情報の統一的な記述の枠組みとしてのXMLの利用を考えると、必須となるのが名前空間(name space)である。つまり、タグや属性の名前を、ネットワーク上でどうユニークに、又は共通につけるかに関する取り決めである。XMLの名前空間については、1998年1月W3CでNoteが公開されている[XMLname 98]。あるURLを"E"という名前空間で指すためには、XMLの最初に次のような宣言をする。

```
<?xml:namespace name="
"http://ecommerce.org/schema/" as="E"?>
```

タグ名や属性名は、次のように名前空間:と共に記述する。

```
<E:ITEM E:PRICE="12.95">
```

また、名前空間やスキーマは HTML, HTTP, Metadata, XML, RDF とからめて、Web Architecture: Extensible Languages というW3C NOTEも1998年2月に公開されている[WebArchitecture 98]。

## 2.2.6 XML プロセッサ

XMLの重要性が認識されるにつれ、多くのXMLツールが開発されている。それらの多くは無償ソースコード付きで公開されており、XMLがインターネット文化から生まれたことを示す良い例となっている。XMLプロセッサは、XML文書を構文解析し、CやJavaのオブジェクトツリーに変換するパーザ、またその逆を行うジェネレータのことを指す(ソフトウェアなのにプロセッサというのはXML文書の解析と生成の両方を行うためらしい)。現在公開されているプロセッサには、MSXML (Microsoft), Lark (Tim Bray), DXP/NXP(DataChannel), XML for Java (IBM) などがある。

Microsoftは、XMLの普及を最も熱心に推進している企業の一つであり、同社のXMLに関するWWW Page (<http://www.microsoft.com/xml/>, このURLからもMicrosoftの意気込みが感じられる)から、MSXMLとよばれるXMLパーザがダウン

ロードできる。MSXMLにはC++版とJava版があり、またMSIE4.0にもすでに搭載されている。プログラムの更新も頻繁に行われている。

Larkは、W3CにおいてXMLの標準化に中心的な役割を果たしているTim Bray氏が開発したXMLプロセッサであり、他のXMLプロセッサに比べると高速であるという特徴を持つ。しかし、Larkはnon-validating、すなわちXML文書がvalidであるかを検証しないので、例えば電子商取引のような複雑かつ金銭の授受を伴うような場面で用いるのは危険である。<http://www.textuality.com/Lark/> からダウンロードできる。

XML for Javaは、IBM(東京基礎研究所)が開発したJavaで書かれたvalidatingなXMLプロセッサである。多言語コードサポートや、validating generationと呼ばれる、あるelementにどのようなsub-elementがどのような順番で付加されるかを知ることができる機能があることが特徴である。<http://www.alphaworks.ibm.com> からダウンロード可能である。

これら以外にも、多くのXMLプロセッサが公開されている。ユーザは、XML仕様への準拠の程度、速度性能、validatingかどうか、必要な言語コードをサポートしているか、などの観点から適切なプロセッサを選択することができる。

XSLに関しては、MicrosoftからXSLプロセッサが公開されている(URLは前述)。これとMSXMLを組み合わせることで、現時点でXML文書をMSIE4.0に表示することが可能となっている。具体的には、ActiveX controlとして提供されたプロセッサが、XMLファイルとXSLファイルからHTMLファイルを生成し、それをMSIEに表示する。

## 2.3 XMLと他フォーマットの比較

### 2.3.1 XMLとHTML

現在WWWの世界では、HTMLが事実上の文書形式の標準の位置付けにある。しかし、HTMLには次のような欠点がある。

- タグの定義や表示方法がブラウザの種類に依存する。
- タグと文書の論理構造との関係が希薄。
- タグが固定されている。

これらは特に、大量の HTML 文書を機械処理する場合に問題となる。例えば、[Kushmerick 97] では、HTML の中から表のような関係的な情報を取り出す Wrapper プログラムを学習を使って効率的に作成する試みである。HTML では、例えば <I> タグで囲まれた数値が何行かに渡ってあったとしても、それが何を意味するのかは不明である。

また、[Perkowitz 97] では、HTML の中にそのようなメタ情報を埋め込んだ A-HTML により、アダプティブな WWW サイトを実現するとある。

XML には HTML の上記の欠点がない。タグは DTD により自由に定義でき、表示の世界との関係はスタイルシートにより別に与えられる。そのため、情報の内容を表すタグ (<price> や <telephone> など) を自由に付けることができる。XML により文書情報が記述されれば、上記の論文で指摘されたような問題点は、最初から生じないことになる。また、メタデータや XML-Data のように情報記述の統一的な枠組みとして XML も用いられはじめてきている。XML による大量な文書が提供されていけば、より AI の本質的な研究も可能となるであろう。

また、現状では HTML は正しい構文でなくても、ブラウザでそれなりに表示されてしまうため、適正でない HTML ファイルが大量に出まわり、文書の機械処理を困難にしている。ところが、XML では少くとも適正でない文書は認められない。文法チェックは、XML でも必須と言える。XML は個人がエディタで書くよりは、オーサリングツールで自動生成する方が主流となるかもしれない。

### 2.3.2 XML と SGML

XML と SGML の関係は、ほぼ次のような式で表される。詳しくは、Comparison of SGML and XML (<http://www.w3.org/TR/NOTE-sgml-xml-971215>) を参照のこと。

XML = SGML - タグの省略 - SGML 宣言  
+ URL

省略可能なタグとは、例えば HTML の <p> タグのように、対応する終了タグ </p> を書かなくても良いタグのことである。XML については、そのようなタグは認められていない。ただし中身の無い (empty element) タグは、

```
<element type="title"/>
```

というような構文で記述する。

SGML 宣言とは、構文を定義したり、文書で使われる文字集合を定義したりするものである。XML においては、構文を固定し、文字集合は Unicode (UCS-2 または UTF-8) とした。URL は WWW 時代に合わせた仕様である。SGML 自体にはリソースの指定法はない。

## 3 その他の技術動向、今後の展望

### 3.1 その他の技術動向

ここでは、その他の標準化動向についても簡単に紹介する。メタデータ関係は別稿で述べる。

#### 3.1.1 Platform for Privacy Preferences Project (P3P)

P3P は、ユーザがネット情報をアクセスするソフトウェア (例えばウェブブラウザ) に設定した個人情報、どの程度までアクセスしたサーバへ公開するか、また、公開された情報を利用することでサーバ側がどの程度までコンテンツを利用してサービスを行えるかと言ったことを考慮したウェブ上のプライバシー情報伝達のためのプラットフォームである [P3P]。

P3P は W3C で行われている標準化のプロジェクトであるが、このプロジェクトは、Netscape 社を中心とした Open Profile Standard (OPS) [OPS 97] および Microsoft 社の Privacy and Profiling on the Web (Web Privacy) [PPW 97] の 2 つの W3C に対する提案が基となっている。この 2 つの W3C に対する提案は、別々に、どちらも 97 年 7 月 2 日に行われたが、その直後の 7 月 11 日に、Netscape 社および Microsoft 社は、OPS をウェブ上のプライバシーの標準として共同で制定して行くことに同意した。

#### 3.1.2 Platform for Internet Content Selection (PICS)

PICS とは、情報発信を制限することなく、情報に対するラベル付けを設定し、クライアント (受け手) 側で「見たくない、見せたくない」情報を排除できるフィルタリング方式のプラットフォームである。

PICS 自体はフィルタリングの機能は有してはならず、選別の機能が作動できるようなプラットフォームを提供するに過ぎない。情報の受け手は、PICS 準拠のソフトを利用して、利用者自身が主に格付け機関

を指定することによって、情報の選別が行われる。PICSの標準化を検討している人たちは、これにより情報の検閲なしに、適切な情報選別を行うことが可能としている [小倉 97, PICS]。

PICSは、W3CのPICSワーキンググループを中心に仕様がまとめられている。現在までに、以下の規格がW3Cで正式に認可されている。

- PICS Label Distribution Label Syntax and Communication Protocols
- PICS Rules 1.1

他にデジタル署名の規格などがドラフト段階である。

### 3.1.3 OSD/DRP

いわゆるソフトウェアの「プッシュ」を効率的に行う規格である。OSD (Open Software Description) では、ソフトウェアの種々のメタ情報 (OS、リソースなど) をXMLにより記述する [OSD 97]。言わば、IE4.0のアクティブチャネルの記述に使われているCDF (Channel Definition Format) のソフトウェア版のようなものである。

DRP (HTTP Distribution and Replication Protocol) は、ソフトウェアプッシュを行うためのプロトコルである。HTTPを拡張したものとなっている。いずれも1997年8月、W3CでNoteが公開されている [DRP 97]。

ちなみに、プッシュに対して、この1年間でインターネットの世界では評価が正反対になった。1年前は時代の寵児として、PointCastやMarimbaといった企業がもてはやされた。ところが、実際にはプッシュはあまりうまくビジネスになっておらず、最近では吸収されたり撤退する企業も数多い。ネットワーク負荷、過剰な広告、独自のインタフェース、標準がないことなどの複合的な要因が重なったためと言われている。 (<http://www.zdnet.co.jp/news/9802/09/push.html>)

### 3.1.4 MathML (Mathematical Markup Language)

数式記述のためのマークアップ言語で、XML1.0準拠である。現状のHTMLでは複雑な数式を表現する方法がなく、やむを得ず図の貼り付けなどで式を表しているWWWページが多い。MathMLでは、数式の表記用と意味を伝えるための2種類のタグを提

供し、表示だけでなく処理対象ともなるのが特徴である。1998年3月7日にW3Cで正式に勧告された [MathML 98]。

### 3.1.5 DOM (Document Object Model)

HTMLやXMLで記述されたコンテンツに、プログラムがダイナミックにアクセスするときの、インタフェースをまとめたものである。Core DOM, HTML DOM, XML DOM から成る。W3CのDOMワーキンググループにより仕様がまとめられて、1997年10月にドラフトが公開された [DOM 97]。1998年4月に公開されたNetscape Communicatorのソースも、DOMに基づきXML部分が記述されている。

### 3.1.6 SMIL (Synchronized Multimedia Integration Language)

「スマイル」と呼ぶ。W3CのSynchronized Multimediaワーキンググループにより提案された、マルチメディアコンテンツ作成の記述法である。音楽、動画、静止画、テキストといった異種のコンテンツを、同期させてTV番組のようなWWWコンテンツを簡単に作成することができる。SMILは、XML1.0に準拠したフォーマットで記述する。1998年4月10日にW3Cで正式に認可された [SMIL 98]。

### 3.1.7 HTML スレッド

1998年1月に、MicrosoftやロータスによるNoteがW3Cで公開されたのがHTML threadingである [HTMLthread 98]。HTML4.0を電子メールとして利用するための標準化である。従来の電子メールのヘッダ情報 (RFC822で定義) をXMLにより記述することで、ユーザーエージェントによる機械処理を可能にするものである。

### 3.1.8 Precision Graphics Markup Language (PGML)

ドロー系グラフィックのフォーマットとして、1998年4月10日にAdobe, IBM, Netscape, SunによりW3CのNoteとして提案された [PGML 98]。これもXMLベースである。

### 3.2 今後の展望

XMLのような文書フォーマットは、他の機械や処理手順の規格と異なり、それ自身が革新的という程の技術ではない。また、皆がXMLにより同じような文書/データを書いて初めて効果が出てくるものなので、導入して即座に種々のメリットを享受できるという種の性質を持ったものでもない。また、これまで溜まった膨大な多様なフォーマットの文書をどうすれば良いのか等々、XMLが本格的に文書記述の基盤となるには解決していかなければならない問題は多い。それを越えてなお、XMLを使いたいという気にさせるのに最も重要となるのは、やはりアプリケーションであろう。メタデータ、プッシュ型におけるメタ情報記述、電子商取引プロトコルとして、現在XMLには追い風が吹いている状況にある。それを、この1年ほどの間に生かせるかどうかが鍵である。

現時点でのXMLでは、以下の問題点があげられる。まず、XMLはあくまでも構文(データフォーマット)を定義したものであり、データの意味を規定することはできない。例えば、`<person>` タグで囲まれた部分には人名しか入れることができない、あるいは上位のタグで定義された属性は下位のエレメントに継承される、といった性質を記述することができない。また、XMLでは、DTDを定義することで自由にタグを定義できるが、このDTDを誰が作成するのか、あるいはどうやって共有すればいいか。また、上であげた問題点とも関連するが、DTDで記述できない性質を今後どのように付与していくかは大きな問題である。これらを解決するために、W3CではスキーマによるDTDの拡張仕様や、名前空間に関する議論が進められているが、これらの問題点がどのように克服されていくかは興味深い。

#### 【参考文献】

- [CSS2 97] Cascading Style Sheets, level 2, <http://www.w3.org/TR/WD-CSS2/>
- [DOM 97] Document Object Model Specification, <http://www.w3.org/TR/WD-DOM/>
- [DRP 97] The HTTP Distribution and Replication Protocol, <http://www.w3.org/TR/NOTE-drp/>
- [富士通 97] 富士通XMLチーム編, はじめてのXML, 日経BP, 1997.
- [HTMLthread 98] HTML Threading: Conventions for use of HTML in email,

<http://www.w3.org/TR/1998/NOTE-HTMLThreading-0105>

- [HyTime] HyTime User's Group Home Page, <http://www.hytime.org/>
- [JEIDA 98] 自然言語処理システムの動向に関する調査報告書, 日本電子工業振興協会, 1998
- [Kushmerick 97] N. Kushmerick, D. S. Weld, and R. Doorenbos, Wrapper Induction for Information Extraction, IJCAI97, pp.729-735, 1997.
- [MathML 98] Mathematical Markup Language, <http://www.w3.org/TR/REC-MathML>
- [村田 98] 村田 真 編著, XML入門, 日本経済新聞社, 1998.
- [小倉 97] 小倉利丸: PICSとは <http://www.jca.ax.apc.org/~toshi/NetCensor/WhatPICS.html>
- [OPS 97] Proposal for an Open Profiling Standard, <http://www.w3.org/TR/NOTE-OPS-FrameWork.html>
- [OSD 97] The Open Software Description Format, <http://www.w3.org/TR/NOTE-OSD>
- [P3P] Platform for Privacy Preferences Project, <http://www.w3.org/P3P/Overview.html>
- [Perkowitz 97] M. Perkowitz and O. Etzioni, Adaptive Web Sites: an AI Challenge, IJCAI97, pp. 16-21, 1997.
- [FGML98] Precision Graphics Markup Language, <http://www.w3c.org/TR/1998/NOTE-PGML>
- [PICS] Platform for Internet Content Selection, <http://www.w3.org/PICS/>
- [PPW 97] Privacy and Profiling on the Web, <http://www.w3.org/TR/NOTE-Web-privacy.html>
- [SMIL 98] Synchronized Multimedia Integration Language, <http://www.w3.org/TR/PR-smil/>
- [WebArchitecture 98] Web Architecture: Extensible Languages, <http://www.w3.org/TR/NOTE-webarch-extlang>
- [XLink 98] XML Linking Language <http://www.w3.org/TR/WD-xlink>
- [XML 98] Extensible Markup Language 1.0, <http://www.w3.org/TR/REC-xml>
- [XMLdata 98] XML-Data, <http://www.w3.org/TR/1998/NOTE-XML-data>
- [XMLname 98] Name Spaces in XML, <http://www.w3.org/TR/1998/NOTE-xml-names>
- [XPointer 98] XML Pointer Language, <http://www.w3.org/TR/WD-xptr>
- [XSL 97] A Proposal for XSL, <http://www.w3.org/TR/NOTE-XSL.html>