

XML、XSL-FO と「Antenna House XSL Formatter」

アンテナハウス株式会社

2000-11-16

Table of Contents

はじめに	3
これまでの XML の印刷方法.....	4
従来の XML ドキュメントの印刷方法	4
従来の XML 印刷方法の比較表	4
問題点.....	5
変換言語の標準化	6
XSLT の登場.....	6
XSLT の応用例：XML を Web ブラウザで表示する.....	6
XSL-FO 組版指定方法の標準化.....	7
XSL-FO とは？.....	7
XSL-FO ベースの組版プロセス	7
XSL-FO 組版エンジン.....	8
『Antenna House XSL Formatter』について.....	8
世界の XSL-FO 組版エンジン	8
実演.....	9
実演 1	9
実演 2	9
実演 3	9

はじめに

いままで、出版物を初めとするドキュメントは、紙に印刷して物流で配布すればそれで一応の役割を終えるものと考えられてきました。しかし、今後、ドキュメントは、印刷物としてのみでなく、WEB のホームページから配布されたり、CD-ROM、電子本などの様々な形態で配布されるようになります。

その場合、コンテンツをアプリケーションから独立したフォーマットで蓄積して管理することが大変重要になります。コンテンツとレイアウト指定がアプリケーションに依存した形式で蓄積されていますと、他のメディア用に編集する際に、まず、元のアプリケーション独自の制御情報を取り去るなどの前処理を行わねばならず、その負荷が非常に大きいからです。

このことは、もう 10 年以上前から指摘され、SGML が提案されてきました。しかし、SGML は普及しませんでした。SGML は、仕様が複雑で、習得が難しく、ツールの開発が大変、ツールが極めて高価で種類も少なかったからです。

SGML をベースに、1998 年 2 月に XML の仕様が決まりました。その後、まだ 2 年半ですが、短期間で様相が変わりつつあります。これからはドキュメントのコンテンツを XML で作成して、蓄積するというのが絵に書いた餅ではなく、現実に手の届くものになってきています。

今日は、ドキュメントを XML 形式で作成し蓄積する際、必ず大きな障害になっていた、**紙への印刷**の問題を解決する、新しい仕様「XSL Formatting Objects(XSL-FO)」と XSL-FO ベースの組版エンジンについて説明致します。

これまでの XML の印刷方法

最初に、これまで XML の印刷がどのような方法でなされていたかを簡単に説明します。

従来の XML ドキュメントの印刷方法

XML を RTF に変換

スクリプト(プログラム)を書いて、XML を RTF ファイルに変換します。RTF ファイルを Microsoft Word で読み込み、Word で最終的に体裁を整えて印刷します。Word を組版ソフトの代わりに使う方法です。

FrameMaker+SGML を使う

FrameMaker + SGML は、印刷業界において SGML を組版するためのソフトとして、一番人気があります。FrameMaker + SGML で、SGML 文書を読み込むには、独自の DTD 対応付け定義を作成する必要があります。これは、簡単なものは数時間でできるそうです。難しいものは専門の会社に依頼して作成してもらいます。問題は、FrameMaker に読み込んだ後、文書が FrameMaker 独自の形式になります。FrameMaker でレイアウトを変更・調整したりすると、元の XML 文書と FrameMaker 文書の 2 種類ができて管理が面倒です。

XML から TeX

XML を TeX に変換して TeX で組版する方法もあります。TeX に変換するスクリプトが必要です。TeX の組版ソフトは数式などに向いていますので、数式の印刷が必要な場合は良いかもしれませんが。

DSSSL

SGML の印刷のための標準仕様で DSSSL というものがあります。これを使えば XML も印刷できますが、DSSSL の組版ソフトは、非常に高価です。また、DSSSL 用のスタイルシートは、Scheme というプログラム言語で作成する必要があります、とても難しいので簡単にはできません。専門の会社に作成してもらいますと、実用的なスタイルシートはひとつで 150 万位するようです。

従来の XML 印刷方法の比較表

XML を印刷する方法について整理します。この表からすぐお分かりいただけますように、XML を印刷するには、前段の「変換」処理と後段の組版処理が必要です。これは XML の仕様上、大抵の場合必要になります。

方法	スタイル形式	指定方法
RTF にして印刷	アプリケーション独自	変換処理 → 対話処理 (WYSISYG)
FrameMaker+SGML にして印刷	アプリケーション独自	変換処理 → 対話処理 (WYSISYG)
TeX にして印刷	TeX ファイルになる	変換処理 + 自動組版 (バッチ組版)
DSSSL にして印刷	DSSSL 方式：標準だが普及していない	変換処理 + 自動組版 (バッチ組版)

問題点

では、いままでの方法でなにが問題だったのでしょうか？問題を大きくまとめますと、次の2つに集約されると思います。

変換処理のプログラミングがばらばら

XML の変換処理に使われますのは、Perl、Scheme、Tcl/Tk などのプログラム言語です。これを使って、XML の DTD 毎、場合によっては XML インスタンス毎に変換プログラムを組む必要があります。そして、プログラムが標準化されていないため、それぞれの方法毎に学習が必要になります。大きな印刷会社などですと、専門のセクションを置くことが可能で、技術者もいますが、普通の印刷会社、出版社ではなかなか専門家がいらないため手を出すのが難しくなります。

組版機能がアプリケーション依存を脱していない

組版機能が標準化されず、アプリケーション依存になっているということは危険性もあります。ご存知の方もいらっしゃると思いますが、Interleaf という構造化文書の編集では有名なソフトがあり、大手企業で SGML の編集に使われてきました。先年、Interleaf は買収されました。ところが、買収先が DTP 事業はしないということを決定し、最近、サポートが終了したそうです。このため Interleaf で大量のドキュメントを作成したところはパニックに陥っているようです。大きな機械の操作説明書のように数千ページのマニュアルを作成している場合、マニュアルの原稿形式をアプリケーション独自形式にするということは、そういう危険性が付きまといます。

変換言語の標準化

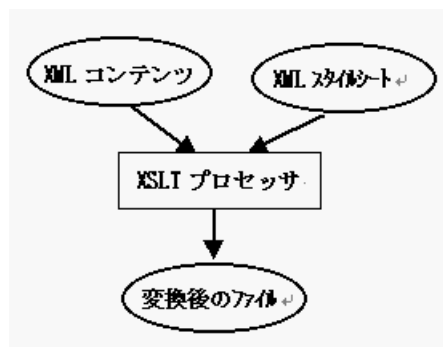
XML の印刷における最初の問題 **XML の変換** は、新しい変換言語の仕様「XSLT」の標準化で解決されつつあります。

XSLT の登場

XSLT の仕様は、当初 XSL-FO の仕様と一緒に検討されていましたが、一足先に 1999 年 11 月に勧告となりました。

- XSLT 「XSL Transformations Version 1.0」 1999 年 11 月 16 日 W3C 勧告
- XPath 「XML Path Language Version 1.0」 1999 年 11 月 16 日 W3C 勧告

XSLT では、XML 文書インスタンスとスタイルシートを「XSLT プロセッサ」に入力します。XSLT プロセッサは、2 つの入力ファイル进行处理して、新しいファイルを出力します。新しいファイルは、XML 文書でも良いですし、XML 文書以外の出力もできます。



大きなインパクトは、Microsoft が、上記の勧告仕様を実装した XSLT プロセッサを標準的なものとして配布を始めたということです。

Microsoft XML Parser V3.0 Release (MSXML3.DLL) 2000 年 11 月 1 日配布開始。 ⁽¹⁾

それから、Sun、Oracle、Apache など自分達の環境で使える XSLT プロセッサを開発し、装備し始めています。MSXML3.DLL は、Internet Explorer 6 には添付されるようですので、近い将来、標準仕様の XSLT はどこでも使えるようになります。XML を変換するには、ひとつだけ変換言語を学べば良くなります。そして、出来あがったスタイルシートはどんな環境でも使用できる、というようになります。

XSLT の応用例：XML を Web ブラウザで表示する

XML を「XSLT」と「スタイルシート」を使って変換するという方法は既にかなり使われ始めています。

Internet Explorer 5、5.5 には、Microsoft 独自仕様の XSLT プロセッサが付いています。この XSLT で XML ファイルを HTML + CSS に変換して HTML を表示する、という方法で、Internet Explorer を使って XML の表示と印刷ができます。現在、XML を HTML + CSS に変換するスタイルシートを作成する人はかなり増えています。書籍も何冊か出版されました。

この方法は、Web で見るには良いですが、ブラウザの表示機能に依存します。ブラウザはページ概念がありませんので、綺麗に組版したページを作成するのは難しいことになります。ブラウザの場合、HTML + CSS をコンピュータの画面にあわせて組版するという設計で、それがいろいろな環境で HTML のホームページを見ることができるという良さになりますが、A4 のような紙のサイズに最適な組版ということはできません。

⁽¹⁾ Internet Explorer バージョン 5、5.5 には MSXML.DLL が付いていますが、これは勧告以前の古い XSLT 仕様を実装しています。Internet Explorer バージョン 6 には、上記の MSXML3.DLL が標準で添付されるようです。

XSL-FO 組版指定方法の標準化

XML を印刷するにはもうひとつの課題があります。それは、従来で、Word や FrameMaker の組版機能、あるいは、DSSSL や TeX の組版機能を使っていた「組版」の標準化です。

XSL-FO とは？

XSL-FO は、Web の標準を定める団体である W3C で、「XML を綺麗にレイアウトして組版するための仕様」として、現在、策定が進んでいます。

XSL-FO 仕様は、今年の 3 月 27 日にラスト・コールになり、その後 10 月 18 日に少し改訂されています。そろそろ Candidate Recommendation になる見込みです。

XSL-FO で検討されている Formatting Objects とは、段落、リスト、表などの組版の対象になるものです。仕様では、ページマスターの記述方法、マージンや本文領域の設定方法、XSL-FO の種類、配置方法、パラメータなど設定できる属性値を標準化しようとしています。

XSL-FO ベースの組版プロセス

組版は、2 つのステップで行います。

XML ファイルとスタイルシートから XSL-FO に変換

これは、XSL-FO を作り出す処理です。通常は、先ほど説明しました「XSLT」を使って行います。XSL-FO に変換するスタイルシートの文法は、XML を HTML に変換する場合と同じです。違いは出力するタグが、HTML の時は、h1、h2、table、などの html のタグであるに対し、XSL-FO にする時は、XSL-FO で定義している Formatting Objects を指定するタグである、ということです。

XSL-FO を組版する処理

XSL-FO が出来あがりますと、次にそれを組版する処理になります。これは、XSL-FO 組版エンジンの仕事です。XSL-FO 組版エンジンに付きましては後で説明致します。

XSL-FO の標準化が完了すれば、XML ファイルを綺麗にレイアウトして紙に印刷するプロセスが標準化されることになります。標準的なプロセスで組版できるということは、非常に大きなインパクトであると思います。

XSL-FO 組版エンジン

仕様が標準化されても、プログラムによる処理系が実現しないと絵に描いた餅に過ぎません。次に、XSL-FO の組版エンジンには、どのようなものがあるかを説明致します。

『Antenna House XSL Formatter』について

弊社では、1999 年 12 月に最初の α 版 XSL-FO 組版エンジンを日本で初めて公開しました。その後、今年初めに α 2 版を公開。現在は、ホームページよりプレ・リリース版を配布中です。今月末に「製品版」を出荷開始するべく準備中です。

『Antenna House XSL Formatter』は、Windows 上で XML と XSL スタイルシートから XSL-FO を作成し、次いで XSL-FO を組版します。

- ・ 内部は組版エンジンとそれを制御する GUI から構成しています。
- ・ XSL-FO ベースの組版エンジンは、ActiveXControl として作成されています。
- ・ GUI では、いろいろな組版オプションを設定して組版エンジンを動かす機能も持ちます。
- ・ 組版エンジンは、ActiveXControl です。例えば、VisualBasic などからコントロールすることも可能です。
- ・ クライアント PC ではなく、サーバ上で動かす場合には、サーバ上でスクリプトから制御することになります。

実装レベルですが、XSL 仕様の最新ドラフトであります、2000 年 3 月 27 日、10 月 18 日の仕様に準拠しています。仕様で定めている Formatting Objects の中で基本部分は全て実装しており、拡張部分も一部実装しています。XSL-FO の仕様ではまだ高度なページレイアウトは困難ですが、操作説明書、報告書などのレベルであれば充分実用に耐えたと考えています。

世界の XSL-FO 組版エンジン

名称	特徴
Antenna House XSL Formatter	アンテナハウスの XSL-FO 組版エンジンとユーザ・インターフェイス。組版エンジンは Windows 版 ActiveXControl。
FOP	Web サーバの開発で有名な Apache が取り組んでいる XSL-FO 組版エンジン。FO から PDF を出力する。日本語は使えない。
XEP	米国の RenderX というベンチャー企業の XSL-FO 組版エンジン。やはり、FO から PDF を出力する。1999 年 4 月の XSL-FO 仕様に準拠。日本語は使えない。
PassiveTeX	FO を処理する TeX のマクロライブラリー（組版エンジンは TeX を使用）。
UFO	Unicorn 社の FO を TeX のマクロに変換する開発ツール（組版エンジンは TeX を使用）。

実演

実演 1

どの程度の組版ができるか？

- profile.xml、profile2c.xml + スタイルシート
- JapaX 形式 XML ファイル + スタイルシート
- hammer.fo (RenderX が公開している FO サンプルを最新仕様準拠に変換したもの)
- table-as-table.fo (RenderX が公開している FO サンプルを最新仕様準拠に変換したもの)

実演 2

Avenue.Quark は綺麗な XML ファイル書き出すことができます。Avenue.Quark で作成した、XML を Antenna House XSL Formatter で組版します。同じ XML にスタイルシートを切り替えて、横書き 2 段組、縦書き 2 段組する組版の実演。

実演 3

XSL チュートリアルの例

- XML ファイルから HTML ページを作成。
- 同じファイルを XSL-FO 経由で XSLFormatter で組版したページ。
- XSLFormatter から、Acrobat で PDF を作成。