HTMLによるXML文書の入力方式について

5V－1

今村 誠　森口 修　鈴木 克志
三菱電機（株）情報技術総合研究所 音声・言語インタフェース技術部

1．はじめに

EC(Electronic Commerce)やEDI(Electronic Data Interchange)をWWW(World Wide Web)により実現するための文書記述言語として、XML(eXtensible Markup Language)が注目されている。XMLでは、文書中にタグで識別される論理構造しかもたないので、文書を表示するためには、文書の表示様式を定義するスタイルシートとそのスタイルシートを解釈して表示形式に変換するエンジンが必要になる。

現時点では、W3C(World Wide Web Consortium)で審議中のスタイルーシート規格案として、Microsoft 等が提案した XSL(eXtensible Style Language)[1]がある。XSL では、XML 文書の要素とその要素がもつ表示様式を結び付けるためるのルール記述を提供する。また、XSL 関連の既存 S/W として、Microsoft 社の XSL Processor [2]とArbortext社のXSL Styler[3]がある。XSL Processorは、XSLスタイルシートを解釈してXML文書をHTML文書に変換しWWWブラウザで表示する機能を提供する。XSL Stylerは、XMLの要素毎に表示様式を指定することにより、XSLスタイルシートを作成支援する機能を提供する。

我々は、HTML変換を用いたWWWブラウザによるXML文書入力方式を検討しているが[4]、[3]の方式では、XMLの論理構造を意識しながら入力用HTMLフォームをデザインするために、スタイルシートの作成に手間どるという問題に直面している。

本稿では、上記の問題を解決するために、既存のHTML作成支援ツールを用いて入力用HTMLフォームをまず作成し、そのHTMLフォームにXML文書との対応関係を指定するスタイルシート作成支援方式について検討する。

Input Methods of XML Documents by Using HTML
Makoto Imamura, Osamu Moriguchi, Katsushi Suzuki
Mitsubishi Electric Corporation, Information Technology R & D CENTER, Human Media Technology Dept., 5-1-1 Ofuna, Kamakura, Kanagawa 247, JAPAN

2．表示-論理対応方式によるスタイルシート作成支援

スタイルシートとは文書の論理構造と表示様式を結び付けるものである。XML/HTML変換によりXML文書を表示する場合には、スタイルシートの作成手順としては、XSL Stylerのように XML文書中の要素に表示様式を対応付ける方式(以下、論理-表示対応方式と呼ぶ)とは逆に、表示結果であるHTML文書中の要素(または属性)にXML文書の要素(または属性)を対応づける方式(以下、表示-論理対応方式と呼ぶ)が考えられる。

表示-論理対応方式によるスタイルシート作成支援S/Wでは、図1に示すように、HTML文書、XML文書、およびHTML文書中の要素(または属性)とXML文書中の要素(または属性)の対応づけを入力として、スタイルシートを出力する。

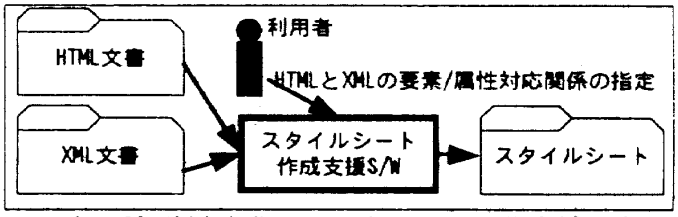

図1：表示-論理対応方式によるスタイルシート作成支援S/W

表示-論理対応方式には、以下の特徴がある。

(1) [利点] WYSIWYGでのスタイルシート作成

HTMLの要素とXMLの要素を対応付けるGUIを新たに開発し、既存のHTMLエディタを併用することにより、WYSIWYG(What you see is what you get)で簡単にスタイルシートが作成できる。

また、XML文書入力用のHTMLフォーム作成の場合には、XML文書中の普通の文字列(内容モデル#PCDATA)がすべてHTMLの入力用コントロール(input,select,textarea等のタグにより設定される)に対応するという制限を設けることにより、対応指定操作が単純になり、操作性を向上させることができる。

(2) [欠点]一般性の欠如

文書インスタンスを基にして、スタイルシートを

作成するので、繰り返し(+,*)や選択(|)を含んだ文書型定義で規定されるXML文書の表示スタイルを定義できない。

従って、表示-論理対応方式によるスタイルシート作成では、対象文書が帳票や伝票のような固定フォーマットに限定される。

3．表示-論理対応方式の実現方式

表示-論理対応方式の実現方式は、スタイルシートの形式の選択に応じて、HTML埋込み方式とXML要素毎指定方式とに分類される。2方式の実現方式とその特徴について述べる(表1)。

(1) HTML埋込み方式

HTML埋込み方式では、スタイルシートとして、HTML文書中にXML文書の要素を一意に指定できる識別子を挿入したものを出力する。HTML文書中に直接識別子を書き込むことにすれば、HTMLとXML間の要素/属性対応指定のための特別なユーザインタフェース(UI)は必要ない。また、スタイルシートを解釈するエンジンは、この識別子からXML文書の対応する部分を認識し、XML文書の表示/入力処理を行う。

本方式ではHTML文書の中にXML文書との対応を指定するので、XML文書中の要素の出現順序によらない文書表示が可能という利点がある。一方、XML文書中に繰り返し構造があっても、固定した繰り返し数にしか対応できないという欠点がある。

(2) XML要素毎指定方式

XML要素毎指定方式では、スタイルシートとして、XSLと同様の「XML文書の要素とその要素がもつ表示様式を結び付けたもの」を出力する。HTMLとXML間の要素/属性対応指定UIでは、HTMLのタグや属性毎にXML文書の要素を対応させる。このUIは、前述したように、XML文書中の内容モデルはすべてHTMLの入力用コントロールに対応するという制限を設けることにより、属性NAMEの値にXMLの要素名を対応させるといった簡単なものにすることができる。

スタイルシート出力処理では、HTML文書を解析して、対応指定されたHTMLのタグ毎に、「対応するXML文書要素の開始タグを処理する前に出力するHTML文書断片」と「対応するXML文書要素の終了タグを処理した後に出力するHTML文書断片」を同定する。

本方式では、上記処理のため、XML中での出現順序と同じ順序で文書表示するスタイルシートしか作成できないという欠点がある。但し、作成されたスタイルシートはXSL等の標準に従ったものになるので、既存のXSLエディタでの後編集が可能になる。

4．おわりに

最初に表示用のHTML文書を作成し、そのHTML文書にXML文書との対応関係を指定するスタイルシート作成支援方式を検討した。本方式は、WYSIWYGでのスタイルシート作成可という利点があるが、固定構造をもつXML文書用のスタイルシートしか作成できないという欠点がある。本方式は、帳票や伝票のような固定した入力フォーマット用、あるいは従来方式[3]の前編集用として利用できる。

参考文献

[1] W3C: A Proposal for XSL, http://www.w3.org/TR/NOTE-XSL-970910 (1987)
[2] Microsoft社: The Microsoft XSL Processor, http://www.microsoft.com/xml/xsl/msxsl.asp(1998)
[3] Arbortext 社: XSL-Styler 2.0, http://www.arbortext.com/xmlstyler (1988)
[4] 今村 誠 他: WWWブラウザによるSGML文書入力方式について、第56回情処学会全国大会 3-65～3-66(1998).

表1：スタイルシート作成支援方式の比較

	WYSISYG編集	一般性(繰返し/or構造対応)	その他の特徴
論理-表示対応方式	×	○	・既存方式(例：Arbortext の XSL-Styler)
HTML埋込み方式 (表示-論理対応方式)	○	× (繰り返し数は固定)	・帳票や伝票のような固定フォーマット用。 ・XML中での出現順序によらない表示が可。
XML要素毎指定方式 (表示-論理対応方式)	○	△(論理-表示対応方式での後編集可)	・XML中での出現順序に従った表示のみ可。