

3W-7 XML 複合文書の実現方式に関する一考察

○野村直之*1 川口浩司*2 坂川浩二郎*3 小林亜令*4 高木康夫*5 小林茂*5 水野雅裕*6

蔵田憲彦*7 松井善郎*1 横田寧子*1 藤岡慎弥*8

*1 ジャストシステム *2 セック *3 日立製作所 *4 KDD 研究所 *5 日本ユニシス

*6 IBM システム・エンジニアリング *7 CRC 総合研究所 *8 沖電気工業

*全員 Java コンソーシアム XML 部会兼務

1 はじめに

次世代 Web は、意味表現を備え、"Semantic Web" と呼ばれる世界知識庫になる、と規格標準化団体 W3C(World Wide Web Consortium)が公式に表明している[Lee2000]。ここでいう「意味」は、(1) 情報をその生来の構造に即して必要なだけ構造化して表現する、(2) リポジトリによって世界知識体系上で絶対的もしくは相対的に位置付ける、の2通りの仕組みで与えられる[野村 2000]。(1) を担うのが XML ファミリー言語の規格群であり、(2) を担うのが URI(Universal Resource Identifier)や RDF(Resource Description Format)などの、Web 上での書誌情報(メタ情報)記述の仕組みである。かように、ビジネスや生活の基盤となる知識インフラ、情報通信インフラの根幹が XML ファミリー規格に基づいたものになるようとしている中で、Word 文書や一太郎文書形式が交換、編集目的で今世紀中使われ続けるとは考え難い。特定のアプリ専用複雑・大規模化した XML ベースの専用言語も同様の運命を辿るかもしれない。

文書処理、特に文書交換、それも全地球規模で、何のアプリを使っているかもわからない相手と共同執筆したり知識の共同創成を日常的に行えるような、効率良い、且つ多彩な表現能力を備えた文書フォーマットはどうあるべきであろうか。本稿では、OLE Compound Document [OLE2000]等が備える、異なる表現メディア('オブジェクト')を枠の中に編集可能な形で埋め込めるといった機能性や、実際のビジネス現場での文書運用のニーズと対比しながら、XML ファミリー言語規格を組み合わせる XML 複合文書を構成するガイドライン構築の議論の一端緒を開いてみたい。

2 XML 複合文書の備えるべき要件と3つの実現法 ~XHTML, SVGによる複合文書の提案

以上の背景から本稿では、特定のアプリケーション(群)専用 XML ベースの言語を設計するのではなく、文書交換目的の汎用性、可搬性、可読性を維持しつつ、XML ファミリー言語のシンタックス、標準ボキャブラリを組み合わせる多重の埋め込みオブジェクトを2次元の文書イメージ上にレイアウトできるような手法を追求する。そこで、ベースとなる個別言語として、文書レイアウト論理記述言語の XHTML1.1 [XHTML2000]とベクトル図形記述言語の SVG [SVG2000] とを採用した。これらは共にモジュール化されていて汎用性・拡張性に富み、広く普及した標準ボキャブラリを備える。共にハイパーリンク・モジュールを備え、単独でも、他リソースとのインタフェースを備えた複合文書コンテンツ表現言語となり得る位の充実した仕様である。

この両者を組み合わせた複合文書として次の3種の実現方式を考えた：

- (1) XHTML1.1 の新モジュールとして、SVG(の一部)を'ドライバ'[XHTML2000]で定義
 - (2) SVG に一部の XHTML1.1 要素の定義を埋め込み DTD で記述するなどして個別拡張
 - (3) SVG と XHTML1.1 を名前空間でマージ；要素を互いに多重に入れ子で埋め込む
- 方法(1)は XHTML1.1 を主言語とし、それに従属するサブモジュールとして SVG を利用するものであり、逆に SVG を主言語としてその拡張部分に XHTML を埋め込むのが方法(2)である。いずれの場合も主従関係が存在し、モジュール定義の部分で独自拡張が必要となる。これに対して、方法(3)は、名前空間の語彙使用の構造パタン記述ガイドラインを設けるだけで、

独自拡張部分無しという利点がある。規格としては最も弱いものになるが、原則的に双方に主従関係無い点、および既存の処理系で早期に実用になる点から今回実験の対象として採用した。また、(2)についても、foreignObject という、SVG 拡張の作法に従って表示可能な処理系が存在したため実験の対象に加えた。

3 既存のブラウザ類による表示実験とその結果

XHTML と SVG とを、名前空間で混在させ、互いに多重に入れ子にしたソースを作成し [野村 2001b]、既存のブラウザ、プラグインによる表示実験を行った。実験 1 は方法(2)を、実験 2 で方法(3)を試した。それぞれの結果の概要を表 1、表 2 に示す。

表 1 実験 1 の結果

	Mozilla Milestone 18	W3C Amaya 4.1	Adobe SVG Viewer 2.0 beta 1	IBM SVGView 0.4a	Apache Batik SVG Toolkit 1.0β
外側の SVG の表示	○	○	○	○	△ fO 未対応
SVG の style 有効	×	○	○	○	○
ForeignObject	△ 無視	○ 認識	△ 無視	△ 無視	×エラー
内側の HTML の表示	○	×	×	×	×

表 2 実験 2 の結果

	Mozilla	Amaya	Adobe	IBM	Apache
外側の HTML の表示	○	○	—	—	—
内側の SVG の表示	○	○	—	—	—
SVG の style 有効	×	○	—	—	—
3 重以上の複合文書	○	○	—	—	—

4 おわりに

幅広い文書交換、共同執筆・編集用途を意識して、3種類の XML 複合文書の実現方式を提案し、その内 2 つについて、既存の最新のブラウザやプラグインにより、複合文書らしいレイアウトで表示可能であることを確認した。ソースの詳細、分析は別稿 [野村 2001b] に譲るが、特に方法(3)では、4重、5重と、XHTML、SVG が互いに交代する多重の入れ子についても正しい表示を確認できた。このことは、XMLらしい、比較的緩い規約による複合文書が軽い開発コストで早期に実現できることを確認したことを意味する。今後は、方法(1)や他の方法の考案と検証、さらに、OLE Compound Document [OLE2000]等にみられる、更新履歴記述スタックを含む書誌情報の埋め込み拡張や、内部リンク、構造検索のための拡張等の検討を踏まえて、さらなる得失の分析・比較や、複合文書規格の標準化を視野に入れていきたい。

参考文献：

- [Lee2000] Tim Berners Lee: "Semantic Web," presentation at XML, Sept. 2000
- [OLE2000] "OLE Documents," <http://msdn.microsoft.com/library/books/inole/s12a2.htm>, 2000
- [SVG2000] "Scalable Vector Graphics (SVG) 1.0 Specification," <http://www.w3.org/TR/SVG/>, 2000.11
- [XHTML2000] "XHTML™ 1.1 - Module-based XHTML," <http://www.w3.org/TR/xhtml11/>, 2000.1
- [野村 2000] 野村直之：「知の創出における XML の役割」, 第 2 回広島セミナー, <http://ph2.ed.hiroshima-u.ac.jp/xml/seminar/2-000925/index.html>, 2000.9
- [野村 2001a] 野村直之：「ナレッジマネジメントツールの配備、実践動向と次世代技術」, 人工知能学会誌 2001.1, pp. 39-41, 2001
- [野村 2001b] 野村直之, 川口浩司, 坂川浩二郎, 小林亜令, 高木康夫, 小林茂, 水野雅裕, 蔵田憲彦, 松井善郎, 横田寧子, 藤岡慎弥：「XML 複合文書の実現方式に関する一考察 ～XHTML, SVG のポキャブリ活用によるアプローチの有効性について」, 情報処理学会デジタル・ドキュメント研究会予稿集, 2001.3