

携帯電話機向け XML コンテンツ復号・解析用ソフトの自動生成

松本 一則

小林 亜令

井ノ上 直己

KDDI 研究所

1. はじめに

伝送速度が比較的低い携帯網では、冗長性の高いテキスト形式の XML 文書をそのまま伝送するのは得策でなく、圧縮伝送することが望まれている。しかし、処理能力の低い携帯電話上に汎用デコードソフトを実装するアプローチは現実的でなく、使用する符号化規則にのみ対応可能なコンパクトで高速な専用デコードをアプリ毎に実装する必要がある。ただし、アプリ毎に専用デコードソフトを開発することは明らかに生産性の点で問題がある。そこで、本稿では符号化規則から専用デコードソフトを自動生成するアプローチについて述べる。

2. XML コンテンツの圧縮

2.1 各種圧縮手法

データ圧縮の手法方法として、gzip 等の汎用ツール以外に、XML-conscious な手法が種々提案されている。XMILL^[1]は文書を (1) 要素名や属性名のシンボルテーブル、(2) 属性値等の文字列、(3) 木構造情報、に分解し、テキスト圧縮手法で圧縮する。xmlppm^[2]は Prediction by Partial Match と呼ばれる手法で木構造を圧縮する。また、対象文書のスキーマを使用する手法には、Millau^[3]や XEUS^[6]等がある。このほか、特定種類の XML 文書を圧縮する BXML^[4]も存在する。本稿では筆者らが提案している XEUS 方式による圧縮手法を扱う。

2.2 XEUS 符号化方式の概要

XEUS¹は、圧縮対象文書の論理構造、データ型、符号化テーブルを記述した「XEUS シート」に従って符号化を行う汎用符号化方式である。XEUS は単に要素や属性のタグや値をシートの記述通りに符号化するだけでなく、以下の符号化戦略の選択が可能である。

- 符号シンボルをバイト整列する / しない²
- 要素に含まれる属性値を、属性値の有無を表すビット列と属性値列で表す / 属性識別子とその値の組の繰り返しで表す³
- 要素値や属性値を算術符号化する / しない⁴

¹“Generation of XML-Aware Decoder for Mobile Phones”, Kazunori MATSUMOTO, Asei KOBAYASHI and Naomi INOUE: KDDI R&D Laboratories Inc.

²「XML document Encoding with Universal Sheet」の略

³バイト整列すると圧縮効率は悪くなるが、復号処理は軽くなる

⁴属性の種類が多く、属性値が存在する確率が小さい場合、識別子とその値の組で表現した方が圧縮率が高い

値の出現頻度に偏りがある場合、算術符号化で処理負担が大きくなるが圧縮率は向上する

表 1 に、代表的な圧縮ツールと XEUS 符号化の圧縮性能を示す。使用した文書は、地図を記述した SVG ファイルと XML 圧縮のベンチマークによく使用される Xmark^[5]である。文献^[2]では、xmlppm が XML-conscious な圧縮ツールの中で最も圧縮率が高いと述べられているが、算術符号化を行った XEUS の方が圧縮率は高い。XEUS の場合、シートを読み込む処理が必要になるが、同一種類の XML 文書を連続処理する場合、1 回あたりの圧縮時間を減らすことは可能である。

表 1: 各手法の圧縮性能

文書とサイズ	gzip	xmll	xmlppm	XEUS	XEUS +算術 符号化
SVG 地図 100m 四方 6.0KB	32.7 0.01	32.0 0.02	31.5 0.06	32.5 0.32	30.8 0.32
SVG 地図 200m 四方 9.6KB	30.0 0.01	28.6 0.02	27.2 0.08	32.1 0.34	27.1 0.35
SVG 地図 500m 四方 30.1KB	25.6 0.02	23.3 0.03	21.7 0.21	31.4 0.46	22.0 0.50
SVG 地図 1Km 四方 95.0KB	23.5 0.05	21.0 0.07	19.4 0.62	29.4 0.88	18.8 0.97
SVG 地図 2Km 四方 354.2 KB	22.7 0.20	20.3 0.22	19.0 2.50	26.4 2.37	16.5 2.71
XMARK 116.5 MB	32.7 73.0	実行 不能	16.2 2,272.6	23.9 411.6	13.0 477.4

上段:圧縮後のサイズ [%] / 下段:符号化時間 [秒]

3. XEUS 用デコーダ

3.1 汎用デコーダと専用デコーダ

XEUS 方式で符号化したデータは使用された符号化戦略や符号化規則が分からないと復号できない。よって、デコーダに以下の 2 つのアプローチが存在する。

- (1) XML 文書の種類に応じて XEUS シート相当の情報を取り込む汎用の XEUS デコーダを用意する。
- (2) 符号化戦略・符号化規則ごと (XEUS シートごと) に専用デコーダを用意する。

処理能力が低い携帯電話の場合、汎用デコーダに比べ、専用デコーダは以下の点で有利である。

- シートの読み込みが不要で、符号戦略や符号化規則を保持するメモリが最低限で済ませれる。
- 汎用デコーダでは、バイト整列する時としない時のように各戦略で使用するルーチンを全て用意しておく必要がある。専用デコーダの場合、余分な処理ルーチンを用意する必要がなく、デコーダの

オブジェクトサイズが小さくなる⁵。

- 文書中の出現位置によっては要素名や属性名が自明な場合がある。このため、XEUS では、要素や属性の識別子を符号化結果に含むとは限らない。また、デコーダは元テキストを復元するのではなく、パース結果を直接アプリに渡す。このため、要素名や属性名をアプリが識別するための手段が必要である。汎用デコーダの場合、識別子テーブルをメモリに確保し、アクセス手段を提供することになる。専用デコーダの場合、ヘッダファイルとしてアプリに渡すことができ、実行時のメモリおよびプログラムサイズの軽量化が可能である。

3.2 XEUS シートからのデコーダ生成

現状の携帯電話の実装状況から当面は専用デコーダを用意すべきと判断した著者らは、XEUS シートから専用デコーダを生成するためのソフトを試作した。

3.2.1 デコーダのソフトウェア・アーキテクチャ

以下のデコーダ API の違いは、デコーダ内部の処理ルーチンに大きな影響を与える。

- 符号化結果をストリームで受け取る、もしくは、符号結果を保存したバッファで受け取る。
- イベント駆動型の API、もしくは、ツリー構築型の API を持つ。
- (イベント駆動の場合) 汎用コールバックの呼び出し後にアプリで要素種別を判断する、もしくは、特定タグで専用コールバックを呼び出す。

デコーダ生成ソフトでは、起動引数で生成するデコーダのタイプを決めた。同時使用しない処理ルーチンを含めないことがデコーダの軽量化に役立った。

3.2.2 生成したデコーダの特徴

デコーダはイベント駆動型の SAX のように使用したり、ツリー構築型の XML DOM のように使用することが可能である。また先の専用デコーダの利点で述べた軽量化の工夫がなされている。そして携帯電話端末で動作するため、下記の特徴を備えている。

- 内部資源へのアクセスが容易で、native 同様に高速に動作する BREW⁶がデコーダの実行環境である。デコーダは BREW のクラスライブラリとしてダウンロードされる。デコーダは、適合条件を満たすアプリから自由に使用することができる。
- 携帯電話では、資源を占有し続けられないなどのお行儀の良さがアプリに必要なになる。そこで、生成され

たデコーダはタイムアウト機能を持ち、復号処理の中断やアプリからの再開指示に対応可能である。

3.3 オブジェクトサイズによる評価

表 1 の XEUS 符号化した SVG 地図を復号するイベント駆動型のデコーダと、ツリー構築型のデコーダを生成し、BREW 上でのオブジェクトサイズを測定した。表 2(a) に、測定結果と zlib(gzip のライブラリ) の BREW 移植時のサイズを示す。同表 (b) は復号用ソフトのバイナリサイズである。xdemill, xmlunppm の BREW 上でのサイズは不明だが、(a),(b) の比較から著者らが生成したデコーダは十分に小さいことがわかる。

表 2: プログラムのサイズ

(a) BREW クラスライブラリの大きさ

zlib	SVG 用 XEUS デコーダ	
	イベント駆動	ツリー構築
46.1KB	21.0KB	25.0KB

(b) Solaris 9 上でのバイナリサイズ

gzip	xdemill	xmlunppm
63.8KB	213.5KB	53.8KB

4. おわりに

スキーマを使用して XML 文書を圧縮する XEUS 符号化方式の符号化結果を携帯電話上で復号するため、専用デコーダの使用を提案した。デコーダ生成ソフトを試作し、BREW 上で動作するデコーダのオブジェクトサイズを測定した結果、現実的なサイズで動作する見込みを得た。今後は動作速度等の評価を進める予定である。

参考文献

- [1] H. Liefke and D. Suciu. XMill: an efficient compressor for XML data. In Proceedings of the 2000 ACM SIGMOD, pp. 153-164, 2000.
- [2] James Cheney, Compressing XML with Multiplexed Hierarchical Models, in Proc. of the 2001 IEEE Data Compression Conference, pp. 163-172
- [3] Girardot. M, Sundaresan. N, Millau: an Encoding Format for Efficient Representation and Exchange of XML Documents over the WWW, 9th International World Wide Web Conference, 1999
- [4] Wireless Application Protocol Forum, Binary XML Content Format Specification Ver. 1.3, 2001
- [5] A. Schmidt, F. Wass, M. Kersten, D. Florescu, M. Carey, I. Manolescu, R. Busse, Why And How To Benchmark XML Database SIGMOD record 30(3): 27-32(2001)
- [6] 小林亜令 他, “XML 文書汎用符号化方式「XEUS」, 信学技報 DE2001-9

⁵動的ライブラリのロードが可能ならば、汎用デコーダでもオブジェクトを小さくできるが、携帯電話端末には負担が大きい

⁶「Binary Run-time Environment for Wireless」の略。C 言語のプログラムを携帯電話にダウンロードし、実行できる。