

鬼沢 友和 安井 浩之 松山 実
武蔵工業大学

1 はじめに

コンピュータネットワーク上において、構造化された文章やデータを管理、交換、配布するためのマークアップ言語として XML^[1] (eXtensible Markup Language) が注目されている。従来の HTML に比べて文書要素の名前が自由に定義できるなど XML 文書を作成する際の自由度が高い反面、それゆえの問題点も多い。

例えば、既に存在する XML 文書に対して情報の追加、削除等を行う場合には XML インスタンスだけでなく文書型定義 (Document Type Definition) まで影響が及ぶ可能性が高く、さらに変更箇所が多岐に及ぶと DTD の修正が困難となる。(図 1)

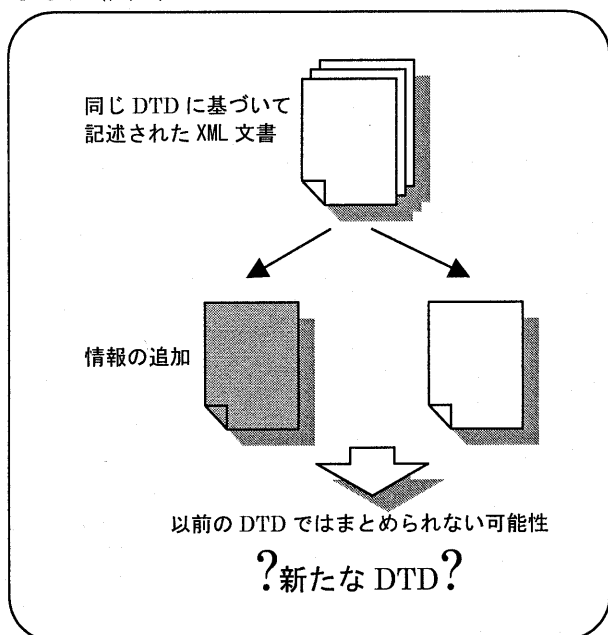


図 1 DTD の修正

前回の報告^[2]において手作業による XML 化時の問題と DTD の自動生成について論じたが、ここでは既存文書の変更に対する DTD の修正、統合を中心として、複数の XML 文書 (XML インスタンス) から統一された DTD を生成する手法を示す。

2 統一化プロセス

2.1 DTD の自動生成

複数の XML 文書进行处理する前段階として各 XML インスタンスから DTD を自動的に生成する。

●自動生成規則 (一部)

- 木構造の親と子のノードの関係を単純に DTD に反映させる
- 子ノードの出現順は走査順にする

この機能のみを使用すれば DTD を考慮しない (先に作らない) 状態で記述された XML インスタンスから逆に DTD を生成できる。

2.2 DTD の比較と統一

複数の XML 文書の DTD の比較は 1 対 1 で行い、双方に適合するような DTD を生成する。インスタンスが複数の場合にはさらに統一 DTD と残りの DTD を比較して新たな統一 DTD を生成する。

●統一規則 (一部)

- 一方の構造を基準にして他方の変更点を追加していく
- 基本的に子ノードの出現順は変更しない
- 同一要素名の出現回数条件を比較する場合には条件の緩い方に合わせる。
- 要素の追加、削除を複合的に行うと出現順序

が不明になるケースがあるが、“|”記号（いずれか1つが出現）を使用せず“?”記号で代用する

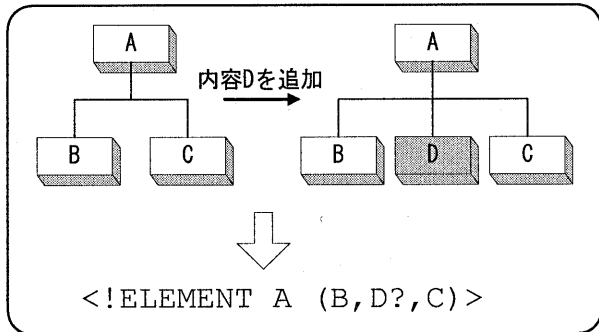


図2 統一 DTD の自動生成例-1

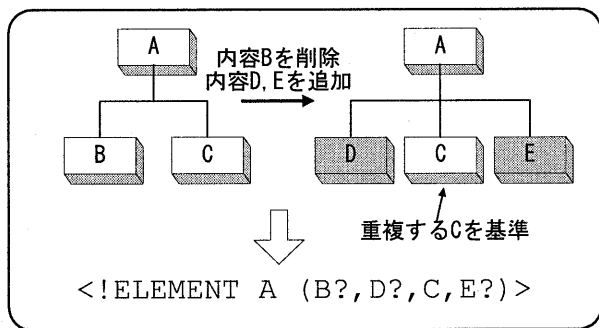


図3 統一 DTD の自動生成例-2

図2では単純な内容の追加の状態を、図3では情報の追加や削除を複合的に行った場合の例を挙げている。この手法を用いると図1のように情報の欠落がない場合には出現順を固定できるが、欠落があった場合には出現順があいまいになってしまう可能性がある。(図3ではBとDの順序があいまい)

3 実装と応用

3.1 実装環境

Java2 が動作する Windows98 上で DOM Level1, SAX version1 の各パーサ API を呼び出すことが可能な環境で実装テストを行った。

現段階ではコマンドラインで動作するアプリケーションであるが、外部から利用できるようなクラスとして設計した。

```
>java CreateUnifDtd sample1.xml
    sample2.xml sample3.xml ...
    Traverse "sample1.xml"
    >>Create "sample1.dtd"
    Traverse "sample2.xml"
    >>Create "sample2.dtd"
    ...
    ...
    Merge "sample1.dtd","sample2.dtd", ...
    >>Create "Unified.dtd"
```

図4 DTD 統一処理クラス (コマンドライン版) CreateUnifDtd.class の実行例

3.2 今後への応用

本手法の実装例として、XML エディタへの組み込みが考えられる。実際の構造化データから直感的に XML インスタンスを生成すれば、バックス記法を知らずとも DTD 付きの XML 文書が作成でき、修正を行うこともできる。ただし現状ではノードの出現順序がユーザの自由にはならないので、GUI による木構造の操作が必要になると考えられる。

4 おわりに

XML は今後多岐の分野で活用されるであろう。難解な DTD 部分を自動的に生成、解析することによって、容易に XML 文書を作成、編集できる手法を示した。今後さらに複雑な構造の文書解析能力や DTD 生成部の精度向上に取り組む予定である。

参考文献

- [1] W3C, “Extensible Markup Language (XML)1.0”,
<http://www.w3.org/TR/REC-xml/>, 1998
- [2] 鬼沢友和, 安井浩之, 松山実, “XML インスタンスからの統一的 DTD の作成”, 情報処理学会第 61 回全国大会, 4J-2, 2000.