

XML ドキュメントデータの構造を利用した品質評価手法の提案 A Structured-Based Document Data Assessment Method for XML

金 京煥[†] 川口 真司[†] 谷口 真也[‡] 松下 誠[†] 井上 克郎[†]
Kim Kyonghwan Shinji Kawaguchi Shinya Taniguchi Makoto Matsushita Katsuro Inoue

[†]大阪大学大学院基礎工学研究科
Graduate School of Engineering, Osaka University

[‡]NTTソフトウェア(株)
NTT Software Corporation

1. まえがき

近年インターネットの普及に伴い、電子文書やデータ構造の記述の際に XML を用いることが多くなってきている。XML[1,2]で記述されたドキュメントデータにはさまざまな用途があるため[3]、その品質を評価する際には、用途に応じた評価基準を設定した上で、それに基づいて評価を行う必要がある。本研究では、XMLドキュメントデータを DTD に基づいて評価するための手法について提案する。本手法では、多様なドキュメントデータおよび評価基準に対応するため、DOMモデルより評価用のデータ構造記述を導出し、導出された構造に対する評価を行う。

2. XMLドキュメントデータの評価

XMLドキュメントデータを、DTD定義に基づいて以下の3つの手順により評価をする。

(1)DOMツリーの構築

XMLドキュメントデータと対応する DTD 定義を読み込み、構文解析して DOM ツリーを生成する。

(2)評価用 DOM ツリー作成

(1)で作成した DOM ツリーを、評価基準と連携した DOM 変換ルールを用いて評価用 DOM ツリーに変換する。

(3)評価用 DOM ツリーの構造評価

あらかじめ定めた評価目的に応じた基準に基づき、評価用 DOM ツリーの評価を行う。

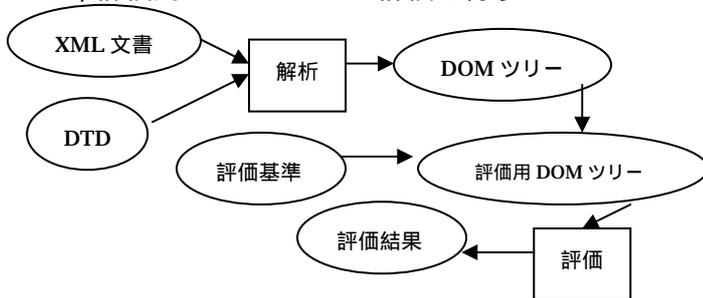


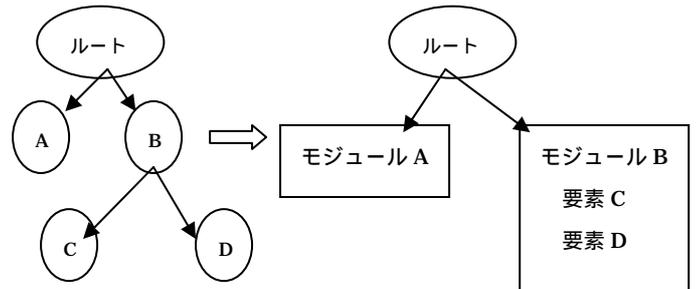
図 1.XML 文書の評価手順

3. 評価用 DOM ツリー

評価用 DOM ツリーに変換する際の変換ルールを記述するには次の2つの概念を用いる。

- ・モジュール:評価を行う際の基本単位であり、DOM ツリー上で1つ以上の要素の集合である。

- ・リンク:モジュール間の関係を表す。



DOM ツリー

評価用 DOM ツリー

例えば要素 A と B をモジュールとしてみなすと、モジュール A は子要素を持たないモジュールに、モジュール B は要素 C と要素 D の2つの子供を持つモジュールに変換される。このようにドキュメントデータを評価基準の目的に応じて評価用 DOM へ変換する。

4. 評価基準と評価

評価用 DOM ツリー上で、このドキュメントデータの評価目的に応じた評価基準を適用する。

例えば、上記の評価 DOM ツリーに対して、階層、モジュール数およびモジュール内の要素数といった評価基準を適用すると、階層:3, モジュール数:2, モジュール内の要素数は A モジュール:1, B モジュール:2 という結果値が出力される。これらの結果を用いてより抽象度の高い評価値の計算を行うことができる。

5. まとめ

本稿では、XMLドキュメントデータを DTD 定義に基づいて評価するための手法を提案した。本評価手法では、利用目的に応じて多様な XMLドキュメントデータを評価することができる。今後、本手法に基づく評価ツールを実装して、本手法の有効性について評価を行いたい。

参考文献

- [1] “ Extensible Markup Language(XML) ”, <http://www.w3c.org/XML> (2000) .
- [2]中山 幹敏, 奥井 康弘:改正版標準 XML 完全解説(上), 技術評論社(2001) .
- [3]日本ユニシス情報技術研究会編:WWW における XML の活用, 東京電気大学出版局(2000) .

