

XML スキーマの修正履歴から XML 文書形式の自動変換

光吉 和哉[†] 中富 竜一郎[†] 堀江 隆太[†] 杉本 典子[‡]崇城大学大学院工学研究科[†] 崇城大学情報学部[‡]

1. はじめに

XML (eXtensible Markup Language) は、さまざまなデータを表現できる柔軟性と拡張性をもつことから、時々刻々と変化する流動的なデータを扱うデータベースシステムの構築に適しているとされている。そのため、XML データベースを利用したシステム構築は増加しており、今後、さらに活用の場は広がっていくことが予想される。

しかし、未だ課題は残されている。その一つに、スキーマの変更への対応が挙げられる。現状では、XML データ形式を自動的に変換することが難しいため、スキーマを変更するためには、特殊な専門知識を有する技術者が不可欠となっている。

そこで、本研究では、XML データのスキーマ変換要求に従って、対応する XML データの形式を自動的に更新するための手法の提案を行う。本研究では、XDES(eXtensible Data Entry System) [1] と呼ばれる XML 用データエントリーシステムを活用する。XDES は、オンラインで Web ブラウザから XML データを入力するためのユーザインターフェースを提供するシステムである。具体的には、XML データスキーマの新規作成、入力データ項目の追加、削除、データフィールドの追加、削除、データ入力形式の変更および、Web 表示形式の変更等のスキーマ変更処理を行うことができる。

この XDES においても課題が残っている。一度、データの入力・蓄積が行われたデータ項目に対してスキーマを変更すると、すでに蓄積されている XML データの形式は、修正前のスキーマに従った形式となっているため、データの整合性がとれなくなる。これまでは、データを全て消去して再入力するか、または管理者が変換プログラムを作成して対応していた。

この問題を改善し、スキーマをデータ入力後であっても自在に変更できる機能を構築するため、XDES の機能として実装されている XML スキ

ーマの更新操作履歴を扱いやすい形式に変換して、XML 文書の形式変換プログラムを自動生成するための手法を提案する。さらに、XDES 上で実際に使用されているデータを対象としたデータ変換の実験を行い、本研究で提案する手法の妥当性と有用性を検証する。

2. XDES の機能拡張

従来の XDES では、図 1 にあるスキーマはスキーマの修正履歴を残すことしか行っていなかった。これを今回は、図 1 のようにスキーマからスキーマの修正履歴とそれのバックアップデータを用意するようにした。この理由として、万が一古い状態に戻したいなどの場合が生じても対応できるようにするためである。また、拡張機能としてスキーマの修正履歴から蓄積データの変換が行えるように自動変換プログラムを追加した。これによって、スキーマの変更が生じた場合でも変更内容に合わせて蓄積データの内容が修正されるようになる。

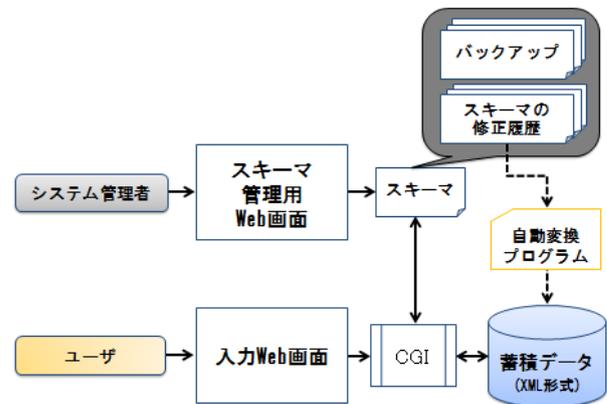


図. 1 XDES の拡張機能を追加した構成図

3. データ変換

データスキーマの変更要求に伴い、自動的に蓄積データの形式を更新するため、データ形式の自動変換を行う。本研究では、XDES での実装を行うため、スキーマの修正履歴から変換プログラムを自動生成する問題を考える。そのために、EFS(Elementary Formal System) と呼ばれる論理プログラムを用いて変換を行う。この手法

Automatic Translation on XML Documents from Modification History of Data Schema.

[†]K. Mitsuyoshi, R. Nakatom, R. Horie · Faculty of Engineering, Sojo University.

[‡]N. Sugimoto · Department of Computer and Information Science, Sojo University.

は、等価変換操作群[2]という手法で行われている変換や3つの編集操作を定義[3]している手法とは設定の異なる問題となる。本研究では、XDESの修正操作として実装されている以下のスキーマ変換について考察する：

- (1) データ項目の追加,
- (2) データ項目の修正,
- (3) データ項目の削除,
- (4) データ項目の挿入,
- (5) データ項目の前後の入れ替え.

上記(1)から(5)のスキーマ変更は、XDES上では、すべてWebブラウザ上でのボタン操作で実行できるものとなっている。上記(1)の「データ項目の追加」を例にあげて説明していく。図2は、蓄積データの一部である。この蓄積データに仮に郵便番号という項目を追加するとする。本来であれば郵便番号の中は空白のデータであるはずなのだが、図3のように追加した項目の前後がずれてしまうなどの問題が発生してしまう。

```

1 <アンケート>↓
2   <氏名>崇城花子</氏名>↓
3   <年齢>10代</年齢>↓
4   <住所1>福岡県</住所1>↓
5   <住所2>福岡市</住所2>↓
6 </アンケート>↓
7 <アンケート>↓
8   <氏名>崇城太郎</氏名>↓
9   <年齢>20代</年齢>↓
10  <住所1>熊本県</住所1>↓
11  <住所2>熊本市</住所2>↓
12 </アンケート>↓
    
```

図. 2 追加前

```

1 <アンケート>↓
2   <氏名>崇城花子</氏名>↓
3   <年齢>10代</年齢>↓
4   <郵便番号>福岡県</郵便番号>↓
5   <住所1>福岡市</住所1>↓
6   <住所2></住所2>↓
7 </アンケート>↓
8 <アンケート>↓
9   <氏名>崇城太郎</氏名>↓
10  <年齢>20代</年齢>↓
11  <郵便番号>熊本県</郵便番号>↓
12  <住所1>熊本市</住所1>↓
13  <住所2></住所2>↓
14 </アンケート>↓
    
```

図. 3 追加後

このような事態を回避するために、スキーマの変更要求が生じた場合でも瞬時に対応できる手法を考案した。具体的な流れとして、図4のようにスキーマの修正履歴から元のスキーマを

自動変換プログラムによって新たなスキーマを生成する手法を考案し、実装している。

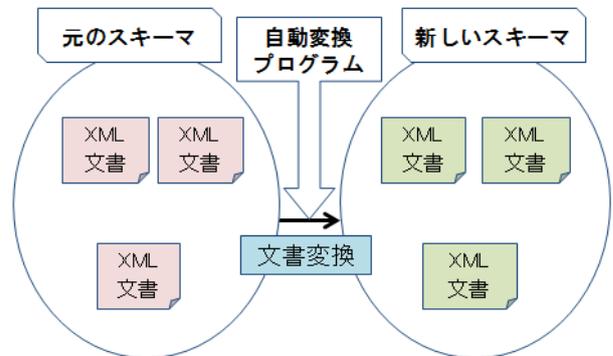


図. 4 自動変換の流れ

4. まとめと今後の課題

データスキーマの修正履歴から、その修正前のスキーマに従った形式のXML文書を新たな形式に自動変換する手法を提案した。本手法は、実際に九州大学の大学評価システム(教員データベース)のデータエントリシステムとして活用されていたXDESにおける実装を念頭におき、XDESでのスキーマ修正作業で実際に得られるスキーマ変更履歴を扱っている。このスキーマ変更履歴は、特定のXML文書のクラスにおいて、重要な意味をもつものであり、計算論的学習理論の分野で用いられる、背景知識を用いた学習の問題として、理論的にとらえることができると考えられる。今後は、この修正履歴情報をもつ本質的な意味を理論的に解明することを考えている。

参考文献

[1] 杉本典子, 金丸玲子, 関隆宏, 石野明, 竹田正幸, 廣川左千男: XDES-多様な構造と流動的変化に対応できるデータエントリシステムの構築, 第4回情報科学技術フォーラム(2005).

[2] 山田大, 橋本健二, 石原靖哲, 藤原融: XMLスキーマに対する等価変換操作群の提案とその評価, DEIM Forum(2010).

[3] Yuan Wang, David J. DeWitt, Jin-Yi Cai: X-Diff: An Effective Change Detection Algorithm for XML documents, 19th. ICDE(2003).