

## スキーマ進化に伴う XML 変換の推測手法

人見 千尋

XML 文書の構造や要素の出現規則を定義するために、スキーマが用いられている。スキーマ定義は時間の経過と共に必要に応じて更新され、これをスキーマ進化と呼ぶ。スキーマ進化に伴いスキーマが更新された場合、更新前のスキーマに対して妥当であった XML 文書が、更新後のスキーマに対して妥当でなくなる場合がある。この場合、更新前のスキーマに対して妥当になるように XML 文書を修正する必要がある。本研究では、こうした場合における XML 文書の変換手法について考察する。

XML においては、スキーマで定義された要素が XML 文書中では複数回出現したり、一つのスキーマの下に複数の XML 文書が存在することが一般的である。そのため、スキーマの更新に対する XML 文書の修正量が膨大なものとなることが多い。したがって、それらを手作業で修正するのは非常に手間がかかる。その手間を軽減するため、本研究ではスキーマの更新情報を用いて自動で XML 文書の変換を行う手法を提案する。

本研究では、スキーマとして DTD を対象とする。DTD に対する更新操作として、要素の追加、要素の削除、要素名の変更、オペレータ('\*' や '+' など) の追加、オペレータの削除を用いる。DTD の個々の内容モデルを木構造として表し、要素やオペレータを木構造のノードとみなす。内容モデルに対する更新操作を木構造に対する更新として定義する。内容モデルやそれに対する更新操作に応じて変換テンプレートを出力し、それらを組み合わせることで 1 つの XSLT スタイルシートを生成する。生成された XSLT スタイルシートを用いて変換を行うと、XML 文書は更新後のスキーマに妥当な形式に変換される。

生成される XSLT スタイルシートはヘッダ、基本出力テンプレート、更新操作テンプレート、フッタの 4 つの部位で構成される。基本出力テンプレートにて、更新のない要素をそのまま出力する処理を行い、更に、更新操作テンプレートにて、更新のある要素に対する変換を上書きする形式で処理を行う。ヘッダとフッタは XSLT スタイルシートを生成する上で必要な記述を出力する。

更新内容によっては、変換結果が一意に定まらない場合や XML 文書の妥当性に影響を与えない場合がある。変換結果が一意に定まらない場合、必要最小限の修正を行って元の XML 文書を妥当なものに変換する。妥当性に影響を与えない場合、XML 文書の変換は行わない。

以上の手法を Ruby を用いて実装し、評価実験を行った。その結果、スキーマ進化に伴う XML 文書修正作業の手間の軽減という観点で、本手法の有効性が検証された。

(指導教員 鈴木伸崇)