

DTD を用いた XPath 式生成ツールの開発

清水 優太

近年、様々な目的に応じて汎用的かつ柔軟にデータを記述可能なマークアップ言語として、XML(Extensible Markup Language)が普及している。XML を用いる事で、ユーザは自らタグを定義して構造的な文書を作成可能である。XML では、自らの利用したいタグなどを予め DTD(Document Type Definition)等のスキーマ言語を用いて定義しておくことができる。大量の XML データを管理する場合などは、DTD を予め定義しておいた状態で、それに対し妥当な XML データを作成するのが一般的である。また、XML 文書に対する問い合わせが可能な XPath(XML Path Language)という言語が存在する。この XPath は比較的容易に記述できる言語であるため、現在広く利用されている。

本研究では、XPath に関する研究において必要となる XPath 式生成ツールの開発を行う。ここで、予め定義された DTD とそれに妥当である XML 文書が存在している状況で、XPath を手作業で記述することを考える。XPath 式を記述するには、それ以前に存在する XML 文書の内容や DTD の構造を理解していなければならない。DTD は定義の内容などが複雑に記されているため、知識を持たない人間が DTD を参照して XPath を記述することは困難である。また、XML 文書を参照するとしても、その XML 文書全体のサイズが大きいかつ構造が複雑である場合、ユーザが問い合わせ用の XPath を作成することは非常に手間がかかる。XPath 式に関する研究では通常、XPath 式を複数必要とするので、手作業による作成は容易でない。また、XML 文書は管理者により変更が加えられることも多いため、XML 文書から XPath 式を作成した場合、その XPath 式が妥当ではなくなってしまう可能性がある。さらに、ある XML 文書には含まれる要素が別の XML 文書では含まれないことがあるため、XML 文書から作成した XPath 式は、他の XML 文書に対する問合せとしては適切でない場合がある。以上から、DTD に基づいて複数の XPath 式を自動的に生成するシステムが有用であると考えられる。

そこで本研究では、DTD の構造とユーザによる条件指定に基づいて XPath 式を自動で生成するアルゴリズムの考案、およびシステムの実装を行い、それらの評価を行った。本アルゴリズムは、まず child 軸のみを用いた複数の XPath 式を生成し、軸の変換などを行なって多様な XPath 式を生成することができる。評価実験の結果、本システムは、従来のシステムより高速に適切な XPath 式を生成可能であることが分かった。

(指導教員 鈴木伸崇)