

DTD の木構造表現を用いた問合せ式生成手法

片山 はなこ

XML の普及に伴い、その応用範囲が急速に拡大している。これに伴い、必要とされるデータの構造もより複雑化しており、XML のスキーマ定義が複雑化しその記述量も増加している。したがって、スキーマの内容を正確に把握するのは必ずしも容易ではなく、スキーマの規則を満たす正しい問合せ式を記述するのが困難となっている。このような場合に、スキーマの構造を視覚化し、GUI を用いて所望の問合せ式を容易に作成できるシステムが利用できれば有用であると考えられる。そこで本研究では、DTD の構造を GUI 上に視覚化し、必要な要素等の指定に基づいて問合せ式を自動生成する手法を提案する。

提案手法では、生成対象の問合せ言語として XQuery を用いる。XML の問合せ言語としては、XPath が比較的習得が容易で幅広く使用されている。しかし、XPath は言語の仕様上 1 つの要素を出力する問合せしか記述することができない。したがって、互いに関連する複数の要素を指定してそれらの関連を保持した検索結果を得たいとしても、XPath 単独では不可能である。ここで、要素は共通の祖先要素によって関連づけられていると考える。一般に共通の祖先は複数存在するが、そのうち最小の要素が指定された要素に最も近接しているので、最も強い関連を表すと考えられる。よって、本システムでは最小共通祖先ごとに指定された要素を纏めて出力する XQuery 式を生成する。

再帰を含まない DTD では、要素間の親子・兄弟関係は木構造として表せる。提案手法では、DTD を木構造として GUI 上に表示する。しかし、DTD は一般に再帰構造を含むので、DTD を単純に木構造として表示するだけでは再帰構造を表すことができない。そこで本システムでは、再帰の範囲内の要素のうち、最も祖先に位置する要素と最も子孫に位置する要素に再帰の範囲を識別するためのラベルを付与して表示し、再帰の範囲を表す。それぞれの再帰に対して、問合せ式において再帰を巡回する回数を指定できる。

提案手法では、まず、DTD の文書型宣言を解析して要素間の階層構造を把握し、GUI 上に DTD の概観を木構造として表示する。次に、表示された木構造に対してユーザが検索対象となる要素群や DTD が再帰を含む場合の再帰回数の指定を指定する。この指定に応じて XQuery 式を生成する。特に、複数の要素が指定された場合、それら要素の最小共通祖先を求め、それを子孫とする XML 文書中の要素をひと纏まりとして出力する XQuery 式を生成する。

提案手法を Ruby を用いて実装し、XQuery 式作成の手間に関する評価実験を行った。その結果、提案手法を用いた場合、用いない場合と比較してより少ない手間で所望の XQuery 式が得られることが分かった。

(指導教員 鈴木伸崇)