

入れ子区間モデルに基づく XML-RDB マッピングに対応した XPath 式変換手法

入江 直人

XML は標準的なデータの記述フォーマットとして広く普及しており、今後ますます XML で記述されたデータの増大が予想される。そのため、データベースを用いた効率的な管理・運用が注目されている。本論文では、XML データを関係データベースに格納するための手法である入れ子区間モデルに着目する。入れ子区間モデルは、他の手法に比べてデータの更新コストが小さいという利点がある。しかし、XML データを単一の関係表に格納しているため、関係表のサイズが増大し検索効率の低下を招くという問題がある。そこで本論文では、この問題点に対処し検索効率の向上を図った改良版モデルを提案する。さらに、この改良版モデルに基づいて、XPath 式から SQL 文への変換アルゴリズムを与える。

関係データベースに XML データを格納する方法には、XML データを文字列化して格納する方法と、木構造から関係表への写像を行い格納する方法の 2 つがある。さらに後者は、構造写像アプローチとモデル写像アプローチの 2 通りに大別される。構造写像アプローチは、DTD のような XML データの論理構造を利用して、関係データベースのスキーマを定義する。モデル写像アプローチは、XML データの論理構造を利用せず、新たに独立した関係データベースのスキーマを定義する。モデル写像アプローチに関する研究には、枝アプローチと経路アプローチ、そして入れ子区間モデルがある。枝アプローチと経路アプローチに基づくマッピングでは、ノードに対して唯一の識別子を与えることで木構造としてのトポロジーを保存し、検索の高速化を図る。しかし、更新性能の観点から見ると識別子を与えるという部分が制約になっており、頂点ノードに近いノードの更新ほど、性能が低下する問題がある。他方、入れ子区間モデルでは上記のような更新性能低下の問題を持たない。すなわち、他のノードに影響を与えることなく、追加・挿入・削除することが可能である。しかし、現在までの入れ子区間モデルに関する研究では、XPath 式から SQL 文への変換は手動で行うしかない状況である。

そこで本論文では、まず、入れ子区間モデルの写像スキームに対応させて XPath 式を SQL 文に自動変換するアルゴリズムを提案する。次に、入れ子区間モデルは XML データを単一の関係表に格納しており、XPath 式のサイズが大きくなるにつれて検索効率が著しく低下するという問題に対処する。具体的には、更新範囲を最小限に抑えて検索高速化を実現する入れ子区間モデルの改良版を提案する。更に、変換アルゴリズムの改良版モデルへの対応を行う。評価実験では、従来モデル上での変換アルゴリズムと改良版モデル上での変換アルゴリズムを用いて、変換結果(SQL 文)の実行時間を比較した。その結果、後者の方が実行時間を大幅に短縮できることが分かった。

(指導教員 鈴木伸崇)