関係データベースに基づく XPath2.0 処理器の開発

-XPath 式の構文解析および中間表現への変換-

田中　勇也†　佐用　健†　天笠　俊之†　田中　二郎†
筑波大学大学院　システム情報工学研究科　コンピュータサイエンス専攻†

1. はじめに

XML は、柔軟なデータ構造を文字列で簡単に表現できることから、Web サービスでのシステム間のデータ交換、アプリケーションの設定ファイル、Web サイトの見出しや要約情報など様々な用途で利用されている。今後は XML 形式で記述された文書が増大し、データベースに蓄積されることが予想される。そのため、XML 文書のデータベースへの格納と、格納された XML 文書の検索処理に関する研究が盛んに行われている[1]。

ところが、このような研究の実験で使用されるシステムは、研究内容に応じて一から構築されることが多く、本来の研究に加えてシステム開発に時間を費やさなくてはならないという問題がある。オープンソースのデータベースシステム、先行研究を行っている研究者が作成したシステムを再利用するという方法も考えられるが、前者はシステムの規模が大きいため全体像を把握しづらく、後者は特定の研究用に作成されているため、汎用性に欠け、再利用することが難しい。

そこで、我々は研究活動の基盤となるシステム XPath2.0 処理器「XPathProcceingPlatform」の開発を行った。XPathProcceingPlatform は、関係データベースに XML 文書を格納するとともに、XML 問合せ言語である XPath2.0[2]のサブセットによる問合せを行い、問合せ結果を再構築する機能を持つシステムである。

研究者は、本システムを再利用することにより、短時間で実験に使用するシステムを開発することが可能となり、従来と比べて効率的に研究活動が行えると考えられる。

The development of XPath2.0 processor based on relational databases
† Yuuya Tanaka, Takeshi Sayo, Toshiyuki Amagasa, Jiro Tanaka
† Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

2. XPathProcceingPlatform の概要

XPathProcceingPlatform とは、我々が開発した XML データベースシステムであり、以下の機能を持つ。

・XML 文書を関係データベースへ格納する
・XPath 式による問合せを行い、その結果を XML 文書に再構築する

本システムの概要を図 1 に示す。

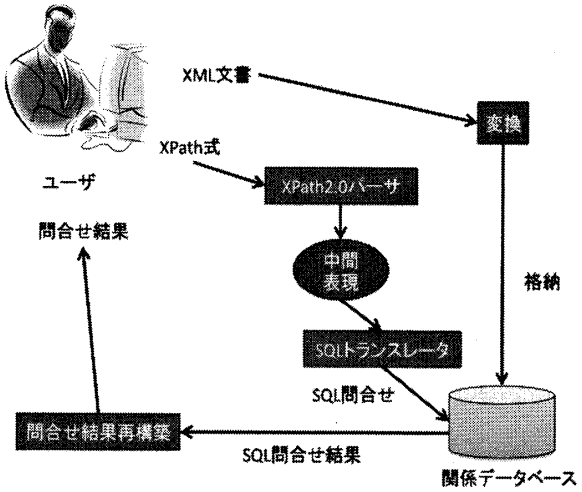

図 1　システム概要

ユーザは、CUI 画面上からコマンドを入力し、本システムがそのコマンドの内容を解釈して、関係データベースにデータを挿入するなどの操作を行う。

2.1 XML 文書格納コマンド put

put コマンドは、XML 文書を関係データベースに格納するコマンドである。

格納する XML 文書を指定し、put コマンドを入力すると、XML 文書が解析され、関係データベースに格納される(図 2)。

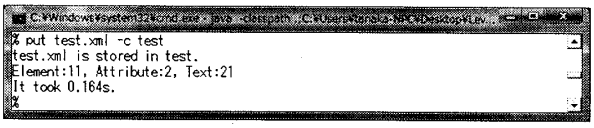

図 2　put コマンド実行結果

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.

XML 文書は木構造であり、そのままでは二次元の表である関係データベースへ格納できない。そこで、put コマンド実行時、XML 文書を関係データベースへ格納できる形に変換するとともに、XPath 式による問合せを行う際、必要な情報が付加される。

2.2 問合せコマンド find

find コマンドは、関係データベースに格納されている XML 文書に対して、XPath 式による問合せを行い、その結果を XML 文書に再構築し、画面に出力するコマンドである。

find コマンドによって入力された XPath 式は、まず、XPath2.0 パーサによって構文解析が行われ、中間表現へと変換される。その中間表現をもとに SQL トランスレータによって、XPath 式が指し示す部分を求める SQL が生成され、問合せが行われる。SQL 問合せ結果は、そのままでは XPath 式が指し示す部分の要素しか得られないので、問合せ結果再構築によって、その要素の子要素以下を取得し、格納した XML 文書の部分文書として出力される(図 3)。

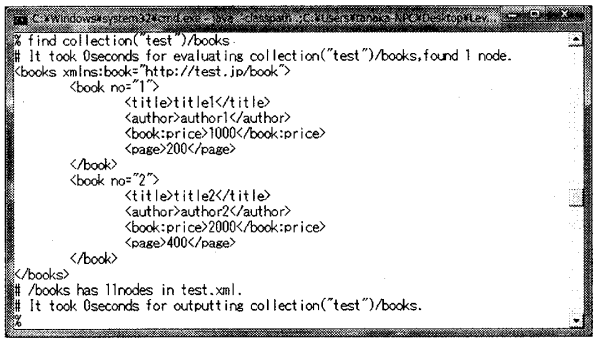

図 3　find コマンド実行結果

3. XPath 式の構文解析と中間表現への変換

XPath 式が入力されると XPath2.0 パーサによって構文解析が行われ、中間表現へ変換されるが、本システムでは LeviathanGraph という中間表現が用いられる。

LeviathanGraph は、四角形のノードとノード間を結ぶ矢印のエッジから構成される。以下の XPath 式が入力された場合、図 4 のような構文木を生成する。

/catalog/book[compare(title, “XML”) = 0]/author

XPath2.0 サブセットの文法に沿って、構文解析を行うと導出された記号に応じて、Step ノード、Func ノード、Bool ノード、Str ノード、Num ノードが生成される。それぞれのノードには、ステップ、関数、比較演算子、文字列、数字といった文法で定義された意味のある値が入る。また、それに伴い、次のステップを表す NextStep、述語を表す Predicate、関数、比較演算子の引数を表す Argument といったエッジがノード間の関係表すエッジとして結ばれる。

このように作成された LeviathanGraph をもとに SQL が生成される。

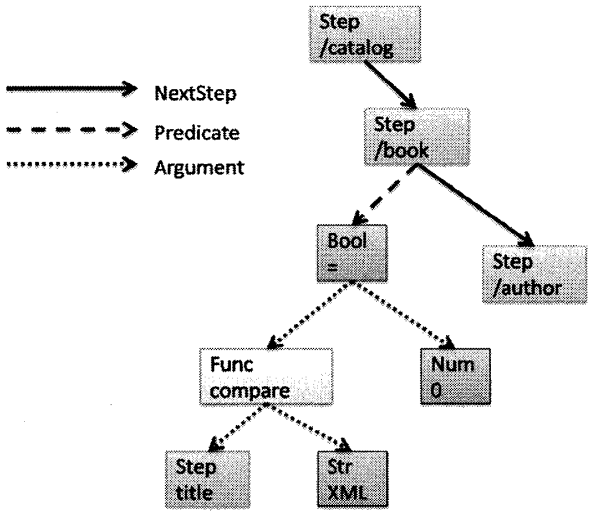

図 4　LeviathanGraph の構文木

4. まとめ

関係データベースに XML 文書を格納するとともに、XPath 式による問合せを行い、問合せ結果を再構築する機能を持つシステムを開発した。本システムを利用することで研究活動が促進されることを期待する。

謝辞

筆者とともに本システムの開発に携わったチームメンバーの上田保祐氏、羽鳥貴之氏に深く感謝いたします。

参考文献

[1] 清水 敏之, 鬼塚 真, 江田 毅晴, 吉川 正俊, “XML データの管理とストリーム処理に関する技術”, 電子情報通信学会論文誌：Vol.J90-D, No.2, pp. 159-184, 2007 年 2 月.

[2] W3C: ” XML Path Language (XPath) 2.0”, http://www.w3.org/TR/xpath20/, W3C Recommendation 23 January 2007

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.