

RDBを利用したXML Storage環境におけるXPathの実装

Development of a XPath processing system using RDBMS

油井 誠¹⁾
Makoto Yui

1) 株式会社 NEC 情報システムズ (〒213-0212 神奈川県川崎市高津区坂戸 3-2-1
かながわサイエンスパーク R&D ビル E-mail: m-yui@pi.jp.nec.com)

ABSTRACT. XML is becoming the realistic technology for the various users as a common data format. The database management system which stores a lot of XML documents permanently, and enables operation of reference, updating, etc. is developed actively. The purpose of our project is providing various functions XML database environment, and is contributing to various research development and commercial use in the future. In this project, we developed the XPath processing system built towards the XML Storage environment using PostgreSQL, an open-source RDBMS, and XMLPGSQL, an XML extension. This result is opened freely in GBorg, a PostgreSQL-related project portal.
<http://gborg.postgresql.org/project/xpsql>

1 背景

XML は計算機環境におけるデータ及び文書の表現形式の一標準として、先進ユーザや研究者による興味の対象から、より幅広い利用者による現実的な技術へととなりつつある。

XML は一般的にそれ自体で“データベース”としての側面を持つとされ、データベース管理システムが備えるいくつかの機能を有している。スキーマ (DTD, RELAX, XML schema languages), 問合せ言語 (XPath[1], XQuery, XQL, etc.), プログラミングインタフェース (SAX, DOM, JDOM) といったものが、その機能として挙げられる。しかし、マイナス面から見るとブレインテキストの XML は、データベースが備えるべき多くの機能を欠いており、データ管理に必要な機能に乏しいという問題がある。記事 [2] では、XML をデータベースと捕えた時のマイナス面として、効率的なデータ格納、インデックス、セキュリティ、トランザクション、データの整合性、マルチユーザアクセス、トリガ、複数の文書に跨った問合せといった点を指摘している。

こうした問題を受け、大量のデータを永続的に格納し、検索、更新などの操作を可能とする為の XML データベースシステムは、研究、商用目的共に盛んに開発が進められている。このような競争の中で、本プロジェクトが目指す立場、目的としては、多機能な XML データベース環境を提供することであり、今後の各種研究開発や商業利用に貢献することである。

近年のソフトウェア開発において、オープンソースソフトウェアが果たしてきた役割は大きい。データベースの分野では PostgreSQL[3], MySQL[4] などが代表的なものである。また、Free Software Foundation[5] や Apache Software Foundation[6] などが提供してきた各種ツールは特に多くの利用者に利用されている。

本プロジェクトは、IPA による未踏ユース事業で支援を受け、PostgreSQL と XMLPGSQL[7] を利用した XML Storage 環境をベースに、XPath 処理系を構築した。成果

物は PostgreSQL 関連のオープンソースポータルである GBorg[8] に無償公開される。

2 目的

本環境開発の原初の動機は、オープンソース RDBMS である PostgreSQL に本格的な XML 格納・検索機能を追加したいということであった。特に日本において、PostgreSQL は最もメジャーなオープンソース RDBMS であり、実務に利用されている例も多数存在する [9]。しかし、現在 PostgreSQL は商業データベースがサポートを始めているような XML を効率的に扱う機能は存在しない。少なくともデファクトとして利用されているようなツールは存在しない。H12 未踏プロジェクトで開発された XMLPGSQL は、DOM ライクの API を有しているが、XPath や XQuery のような高度な XML 問合せ機構が容易されていないこと、大容量の XML データに関する扱いに問題があり、国内に数十の利用例がある程度である。試用された形跡があったが、XPath サポートが無いこと、パフォーマンスを理由に採用が否決された例も数例存在した。

本プロジェクトでは、XPath 処理系の構築と並行して、XMLPGSQL のデータベーススキーマも互換性を極力保つ範囲で、設計の見直しを図った。

今後の展開としては、GBorg でプロジェクトを運営しつつ、ドキュメントを充実させ、またコミュニティに対して積極的にアピールも行なっていく事で、PostgreSQL での XML 利用ケースでデファクトスタンダードとなる事を目的としたい。また、国内外での先端事例に採用される事で、新たな価値の創出の助けとなる事ができたら幸いである。

3 本文部分

to appear

4 参加企業及び機関

プロジェクト管理組織 株式会社メディアフロント

参考文献

- [1] W3C. XML Path Language (XPath) Version 1.0.
<http://www.w3.org/TR/xpath>.
- [2] Ronald Bourret. XML and Databases, 1999-2003.
- [3] PostgreSQL Homepage.
<http://www.postgresql.org/>.
- [4] MySQL Homepage.
<http://www.mysql.com/>.
- [5] GNU's Not Unix! - the GNU Project and the Free Software Foundation. <http://www.gnu.org/>.
- [6] The Apache Software Foundation.
<http://www.apache.org/>.
- [7] 小松誠. 進化する XMLPGSQL. Software Design, pp. 19-26, 2002 年 4 月.
- [8] GBorg - PostgreSQL related projects
<http://gborg.postgresql.org/>
- [9] 日本 PostgreSQL ユーザ会. PostgreSQL 使用サイトリンク集.
<http://www.postgresql.jp/projects/links/ippan.html>