

1 はじめに

近年、各種文書データベースの構築やその利用の高度化が急速に進みつつある。特に、インターネットの著しい普及に伴い、XML をはじめとする構造化文書データベースの構築とその利用の高度化の要求が増大している。XML データベースに対する基本的な処理記述言語として、これまでに各種の問合せ言語が提案されている。

XML データベースに対する処理記述では、タグの階層構造で与えられる文書構造に対する処理と文字列テキストで与えられる記述内容に対する処理の両者を扱う必要がある。前述の問合せ言語の多くは高水準な宣言的記述による高い問合せ記述能力を持っており、特に XML-QL や半構造化データ問合せ言語のいくつかは文書構造に対する柔軟な再構成機能を持つ。しかし一方で、これらは記述内容処理に関しては極めて限定された機能しか提供していないため、より詳細かつ具体的な記述能力が要求される。

我々はこのような問題に対し、XML 文書に対する問合せ言語を持つべき文書構造操作機能に加え、ユーザ定義の外部関数による機能拡張性を備えた拡張可能 XML 問合せ言語 X²QL を提案し、開発を進めている [1]。

本稿では、大量の XML 文書を格納管理する XML サーバと連携した X²QL 処理系の設計とその実装について報告する。本 X²QL 処理系では XQL 問合せを処理可能な商用 XML サーバをそのコンポーネントとして利用する。X²QL 問合せが与えられると、XML サーバから必要な XML 文書を抽出するために、XQL に基づく絞込み問合せを発行する。また、その XML 文書の絞込みの結果、最初にユーザの与えた X²QL 問合せを簡略化することが可能な場合も考えられる。本稿ではこれらの手法についても言及する。

2 拡張可能 XML 問合せ言語 X²QL

X²QL 問合せは、XML-QL をベースとした構文によって記述される。また、問合せ中には外部プログラムとして与えられるユーザ定義の外部関数が利用可能であり、これによって高い機能拡張性が要求される処理に対応することも可能である。

X²QL 問合せの基本構文を次に示す。

```
where パターン式および変数束縛 in 対象
      [, パターン式および変数束縛 in 対象]* [, 述語]*
[rank-by 順位付け基準式 top 選択件数]
[order-by 出力順序基準式 [descending]]
construct 出力生成式
```

where 節は問合せ条件として適合すべき文書構造やそれが満たすべき条件を指定し、必要に応じて変数の束縛を指定する節である。変数はその記述位置に応じて、文書要素、文書要素内容、文書要素名、属性名、属性値のいずれかに束縛され、対応する述語によってその値の満たすべき条件が記述される。construct 節は where 節のパターン式中で束縛された変数のそれぞれの組合せに対して生成される出力の構成を指定する節である。rank-by、order-by 節について本稿では説明を省略する。X²QL の構文は基本的に XML-QL の構文に従うが、外部関数を適宜利用可能である。

Design and Implementation of an X²QL Query Processing System
Jun Kawada[†], Norihide Shinagawa[‡], Hiroyuki Kitagawa^{‡‡‡},
Yoshiharu Ishikawa^{‡‡‡}

[†] Graduate School of Sys. and Info. Eng., Univ. of Tsukuba

[‡] Doctoral Program in Eng., Univ. of Tsukuba

^{‡‡‡} Institute of Info. Sci. and Elec., Univ. of Tsukuba

3 X²QL 処理系の概要

現在、論文 [2] に述べたアプローチに基づいて X²QL 処理系の構築を行っている。本システムの概略図を図 1 に示す。

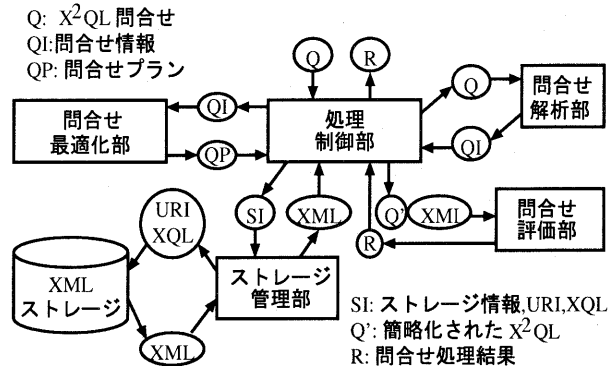


図 1: システム概略図

システムは、処理制御部、問合せ解析部、問合せ最適化部、ストレージ管理部、XML ストレージ、問合せ評価部から構成される。

この中の XML ストレージとは XML サーバ、ファイルシステム、Web 等であり、特に本研究で用いる XML サーバは大量の XML 文書を格納管理し、XQL に基づく XML 文書データの問合せ機能を提供する。

本 X²QL 処理系では、この XML サーバの機能を用いて、与えられた X²QL 問合せが対象とする XML 文書の効率的な絞込みを実現する。

以下に、それぞれの構成の各部分について説明する。

3.1 問合せ解析部

与えられた X²QL 問合せをパースし、内部表現である問合せ情報にして返す。

3.2 問合せ最適化部

与えられた問合せ情報を元に以下の処理を行う。

- 問合せ処理対象となる XML 文書を提供する XML ストレージの種別を判別
 - XML ストレージが XML サーバであるとき、XML 文書の絞込み問合せを生成 (4 節で詳述)
 - XML 文書の絞込みによって可能となる、ユーザの与えた X²QL 問合せの簡略化 (5 節で詳述)
 - 各 XML ストレージに問合せを行う手順の決定
- そして、これらの情報を表す問合せプランを返す。

3.3 ストレージ管理部

XML サーバ、ファイルシステム、Web 等のストレージ種別情報に応じ、問合せ最適化部から与えられた XML 文書のパス情報や絞込み問合せを用いて XML ストレージを操作し、XML 文書群を取得する。

3.4 問合せ評価部

X²QL 問合せを XSLT[†] 記述に変換し、既存の XSLT 処理系を利用して問合せの処理を行う。問合せ評価部については文献 [1] に詳細を示している。

[†]XSLT は W3C で策定中の XSL (XML スタイルシート言語) で用いられるルールベースの XML 変換言語である。

3.5 処理制御部

ユーザから与えられた X²QL 問合せ等の要求を受け取り、その要求に応じて必要な処理を呼び出し、適切な手順で処理を行う。

3.6 問合せ処理手順

ユーザが X²QL 処理系に X²QL 問合せを与えたとき、処理制御部によって次のような処理が行われる。

1. 問合せ解析部を用いて、与えられた X²QL 問合せに関する問合せ情報を取得する。
2. 問合せ情報を問合せ最適化部に送り、以降の問合せ処理に必要な問合せプランを取得する。
3. 問合せプランに基づいて、適切な手順でストレージ管理部にストレージ種別情報とストレージの種別に応じた XML 文書へのパス情報や絞り込み問合せを送る。ストレージ管理部は、XML ストレージから取得した XML 文書群を返す。
4. ストレージ管理部から取得した XML 文書群と問合せプランから得られる簡略化された X²QL 問合せを問合せ評価部によって処理し、その結果をユーザに返す。

4 絞り込み問合せ

今回、XML サーバとして Tamino を用いて、実装を進めている。この Tamino は問合せ言語として XQL のサブセットにいくつか機能を追加した Tamino XQL を用いる。従って、本研究では XML サーバへの絞り込み問合せとして Tamino XQL を利用している。

Tamino XQL では、問合せはパス条件によって記述され、それに適合する文書要素などのノード集合が得られる。X²QL 問合せの適用対象となる文書を取得するために絞り込み問合せを生成するが、その問合せによって得られる結果は元の文書構造を保存していなければならない。そのため、常にルート文書要素を取得するような絞り込み問合せが生成される。X²QL 問合せの条件は、取得されるノード集合、即ちルート文書要素に対する選択条件として絞り込み問合せに記述される。ここで、Tamino XQL では、選択されるノード集合に関する選択条件はそれを起点とした分岐のないパスでのみ記述可能であるため、X²QL の木構造の条件を直接表現できないことに注意する必要がある。そこで、この場合には問合せの条件を緩め、X²QL 問合せ処理対象外となる XML 文書も取得するような絞り込み問合せを生成する。

この絞り込み問合せ生成の簡単な例を、図 3 の中に示す。

絞り込み問合せの生成手順等、詳細については論文 [2] に示した。

5 X²QL 問合せの簡略化

絞り込み問合せによって、常に満たされるようになる条件が含まれる可能性があり、X²QL 問合せ処理に先立ってこれを簡略化することが可能である。しかし、4 節で述べたように、Tamino XQL では選択されるノード集合に関する選択条件は、それを起点とした分岐のないパスでのみ記述可能であるため、分岐無く表現可能な範囲、即ちルート文書要素の属性と直接の子要素に関する条件を簡略化の対象としている。

簡略化可能であるための条件と、そのとき可能な簡略化の内容の判定方法を図 2 に示し、この条件に従って簡略化

を行う具体例を図 3 に示す。

```

if(ルート要素が属性を持つ and その属性名が変数でない) {
  if(属性値がリテラルで与えられている)
    該当するルート要素の属性を削除可能 --- (1)
  else if(属性値で束縛される変数が
    Tamino XQL で記述可能な述語条件を持つ)
    該当する変数に関する述語を削除可能 --- (2)
}
if(ルート要素がある子要素を持つ) {
  if(そのルート要素の子要素のいずれの部分も変数に束縛されない)
    該当するルート要素の子要素を削除可能 --- (3)
  if(そのルート要素の子要素が変数に束縛されている
    and その束縛される変数が
    Tamino XQL で記述可能な述語条件を持つ)
    該当する変数の述語を削除可能 --- (4)
}
  
```

図 2: 簡略化対象の条件

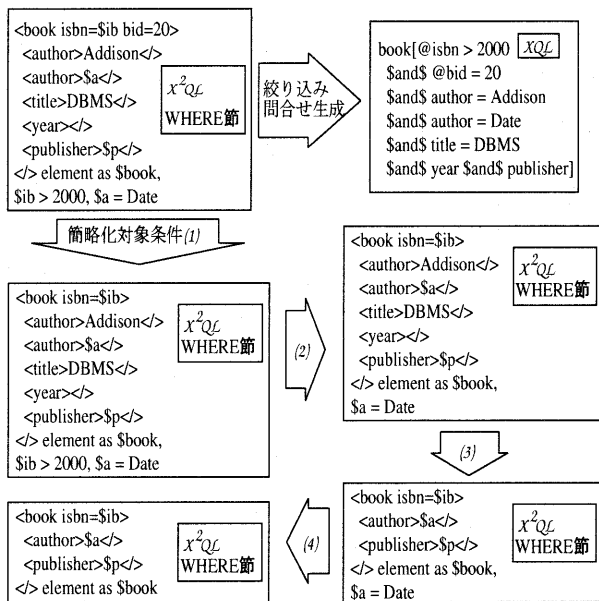


図 3: 簡略化の具体例

6 まとめ

本稿では、処理制御部、問合せ解析部、問合せ最適化部、ストレージ管理部、XML ストレージ、問合せ評価部から構成される X²QL 処理系の概要について報告した。また、絞り込み問合せの結果可能となる、X²QL 問合せの簡略化についても報告した。

今後、XML ストレージとして他の XML サーバへの対応についても検討する予定である。

参考文献

- [1] Shinagawa, Kitagawa, and Ishikawa, X2QL: An eXtensible XML Query Language Supporting User-defined Foreign Functions, *Proc. 2000 ADBIS-DASFAA Symposium on Advances in Databases and Information Systems*, Prague, Czech, September 2000.
- [2] 川田, 品川, 北川, 石川, 拡張可能 XML 問合せ言語 X²QL 処理系の開発, 情報処理学会第 61 回 (平成 12 年後期) 全国大会 2000 年 10 月 愛媛県松山市愛媛大学.