意味情報を用いたXQuery問合せ作成支援システムの開発

古川 夏子[†] 森嶋 厚行[††]
[†] 芝浦工業大学大学院 工学研究科 [††] 芝浦工業大学 工学部情報工学科

1. はじめに

近年，インターネットにおけるデータ交換のための標準形式としてXMLが普及しており，XMLデータの検索や操作の必要性が増大している．XMLを対象とした様々な操作系の中でも，XQueryはXML問合せ言語の標準としての地位を確立しつつある．しかし，XMLはRDBに比べ構造が複雑であるため，RDBの問合せ言語であるSQLに比べXQuery問合せの記述は一般に複雑・大規模なものになる傾向がある．特にこれは，XMLを用いたデータ交換のために，与えられたメタデータに従ったXMLを作成する必要がある場合に顕著である．このような，大規模・複雑なXQuery問合せを記述することは容易でなく，バグなども入りやすいため，ソフトウェア開発の生産性を下げる一因となりうる．

本稿では，大規模・複雑なXQuery問合せの作成を支援するシステムについて述べる．本システムは変換元と変換先のXMLの構造情報(DTDなど)および意味情報を入力とし，半自動的なXQuery問合せの生成を行う．ポイントは，今まではXQuery問合せの作成者が暗黙のうちに用いていたXMLの意味情報を，XMLのメタデータ設計者に明示的に用意させることによって，仕事をXMLメタデータ設計者にシフトすることである．メタデータ設計者はそのXMLが表す情報のドメインのプロであるため，構造情報だけでなく意味情報も用意させることはそれほど大きなコスト増にはならないと考えられる．本稿ではこのシステムの核となるアルゴリズムの概要について説明する．

2. システム概要

本システムの概要を図1に示す．変換元の構造情報及び意味情報をそれぞれ$d1$及び$s1$で表し，変換先の構造情報及び意味情報を$d2$及び$s2$で表す．構造情報はXML Schema[1)]やDTDなどを想定しているが，ここでは総称してXML structureと呼ぶ．意味情報とは，キー制約をはじめとする意味に関する制約などを記述したメタデータである．

本システムは$d1$及び$s1$に従ったXMLデータを入力とし，$d2$及び$s2$に従ったXMLデータを出力する問合せ作成の支援を行う(図1)．

3. XML Semantics

我々は，XMLデータの意味を記述するメタデータ表現として，XML Semanticsを導入する．XMLに関する意味制約はXML Schemaなどでも一部表現可能であるが，それらはキー制約などに限定されている．XML Semanticsの表現は，これらでは不十分であった，XML意味制約を記述することを可能とする．

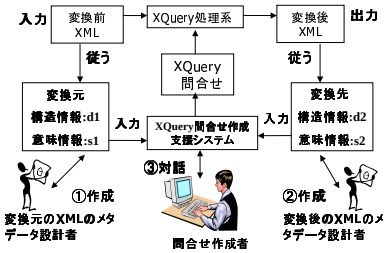

図1 XQuery問合せ作成システム概要図

XML Semanticsは，XMLがエンコードしている情報の意味を，ERダイアグラムを用いて表現したものである(図2)．具体的にはERダイアグラムの各構成要素毎に，その構成要素をXMLでどのように表現しているかを表すbinding expression集合を追加したものである．ここでbinding expression(以下be)とは，XQueryの`for`節におけるv in p(vは変数，pはXPath式)と同じものである．各be集合は，ERダイアグラムの構成要素の種類に応じて，次のように記述される．

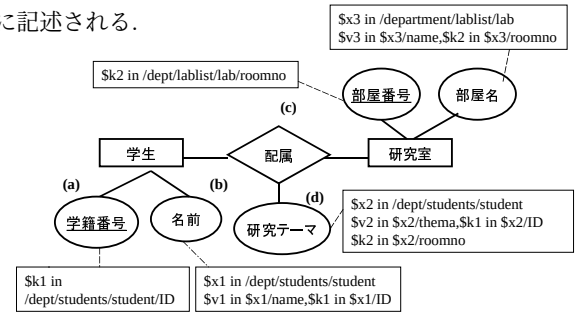

図2 XML Semantics $s1$

実体のキー属性: $\$k_i$ in pを持つ．ただし、pはキー属性を表すXMLの値(要素，属性，文字列)などへのXPath式である．例えば図2の実体集合「学生」のキー属性「学籍番号」(図2(a))はXMLでは/dept/students/student/IDの値として表現されている．“ k ”で始まる変数は，キーとなる値に束縛される変数(キー変数)であることを表す．

実体の非キー属性: $\$v_i$ in qを持つ．ただし，qは属性をあらわすXMLの値へのXPath式である．さらに，この属性が属する実体のキー属性との結びつきに関する制約を表すための be 群が追加される．例えば，図2の実体集合「学生」の属性「名前」(図2(b))の beは，$\$x$ in /dept/students/student, $\$v1$ in $x/name, $\$k1$ in $x/IDである．変数$\$k1$と$\$v1$は同じ`student`要素を共有するという制約が表されている．

関連の属性: 3種類の be 群を持つ．(1) 関連が接続されている全ての実体のキー属性の be 群．(2) 属性に対応するXML値を表すXPath式をもつ be．(3) それらの間の関連

A Rapid Query Development Tool for XQuery Queries Based on Semantic Infomation
[†]Graduate School of Eng., Shibaura Inst. of Tech.
[††]Dept. of Info. Sci. and Eng., Shibaura Inst. of Tech.

を表現した制約を表す be 群．例えば，図 2 の関連集合「配属」(図 2(c)) の属性「研究テーマ」(図 2(d)) の be 集合は，「配属」に接続されている実体「学生」「研究室」の変数$\$k1$, $\$k2$ と属性値を表す$\$v2$ を定義する be 群のおよびそれらの間に成立する制約を表す be 群を含む．

関連: 関連の属性から (2) を除いたもの (図 2 では省略)．

4. 問合せ作成アルゴリズム概要

本システムにおける問合せ生成アルゴリズムは Phase1 と Phase2 に分けられる．Phase1 は，XML structure $d2$ には従うが，XML Semantics $s2$ に書かれている制約は満たさないような XML データを出力する問合せを作成する．Phase2 は，XML Semantics $s2$ に記述されている XML データの制約を満たすようにその問合せを変更する．

4.1 Phase1

図 3 は，説明のための例として用いる XML structure $d1$ および $d2$ を木構造で表したものである．子要素の出現回数が 1 のとき，各辺に 1 をラベル付けする (1-エッジと呼ぶ)．例で用いる XML Semantics $s1$ は図 2 に，$s2$ は図 4 に示す．

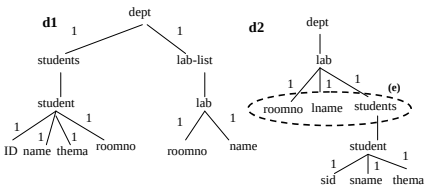

図 3 XML structure $d1$ と $d2$ の木構造

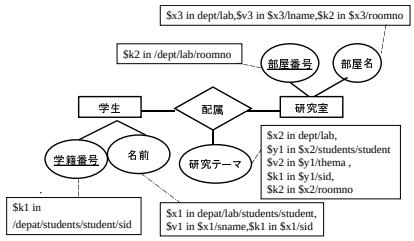

図 4 XML Semantics s2

Phase1 のアルゴリズム $phase1(e)$ の概要は次の通りである．e の初期値は $d2$ のルートノードである．例では，図 3 の `dept` 要素である．

(1) e と 1-エッジのみで到達可能な $d2$ のノード集合を $E1$ とする．

(2) e がルートでなければ `for` 節を出力する．具体的には，e に対応する $s2$ の構成要素 (属性など) を識別し，それに対応する $s1$ の構成要素に付随した be を `for` 節の式とする．例えば，e を `lab` とする．そのとき，$E1$ は `lab` の全ての子要素となる (図 3(e))．$s2$ で $E1$ 中のいずれかのノード（この例では `roomno`）に対応する構成要素が「部屋番号」であることを識別する (図 4)．さらに，$s1$ の「部屋番号」がそれに対応するので，その be を出力する ((図 5)f)．

(3)`return` 節を出力する．そこでは，$E1$ の中の要素のタグ及び `for` 節で出力した be の変数を出力する (図 5(g))．e がルートの場合には，(2) を省略して (3) が実行される (図 5(h))．

(4)$E1$ 中の要素の子のうち，1-エッジで親と接続されていないものの集合を $E2$ とする．

(5) 全ての $E2$ の要素 e' に対して $phase1(e')$ を実行する．

```
<dept>  (h)
for $k2 in/dept/lablist/lab/roomno, (f)
    $k2' in/dept/lablist/lab/roomno,
    $x3 in/dept/lablist/lab,$v3 in $x3/name,
                        ...
return <lab>
        <roomno> { $k2' } </roomno> (g)
        <lname> { $v3 } </lname>
                        ...
```

図 5 Phase1 の実行結果の一部

4.2 Phase2

Phase2 の概要は次のとおりである．

(1)$s2$ の全ての属性 a の be について次を行う．a の be のうち v を定義する XPath 式の末端要素を e とする．その e を出力する `flwr` 式の `where` 節に $k_i = k_i'$ を挿入する (図 6(k))．ここで，k_i および k_i' は問合せに存在する変数のうち，$s2$ において a のキー変数を定義する XPath 式が表す要素に一対一対応する値に束縛されるものである．

(2)$s2$ 中の全ての関連 r について次を行う．まず，r の be 中でキー変数を定義する XPath 式の中で，最も長いパスの末端要素 (複数ある場合がある) の一つを e とする．その e を出力する `flwr` 式の `for` 節に，$s1$ における r の be を追加する (図 6(i))．次に，その `flwr` 式の `where` 節に $k_i = k_i'$ を挿入する (図 6(j))．ここで k_i および k_i' は問合せに既に存在する変数のうち，$s2$ において r のキー変数 (群) を定義する XPath 式が表す要素に一対一対応する値に束縛される．

```
return <dept>
 for $k2 in/dept/lablist/lab/roomno,
    $k2' in/dept/lablist/lab/roomno,
    $x3 in/dept/lablist/lab,$v3 in $x3/name,
                ...
 where ...   $k2 = $k2'        (k)
 return <lab>      ...
         <lname> { $v3 } </lname>

                ...
        for
        $k1 in /dept/students/student/ID,
        $x2 in /development/students/student,$v2 in $x2/thema,
        $k1' in $x2/ID,$k2''' in $x2/roomno, ...(i)
        where $k1 = $k1' and  $k2 = $k2'''              (j)
        return <sid>{$k1}</sid>
...
```

図 6 Phase2 の実行結果の一部

5. おわりに

本稿では，意味情報を用いて XQuery 問合せの作成を支援するシステムについて提案し，そのキーとなるアルゴリズムの概要を説明した．実際には，変換元と変換先の ER ダイアグラムが完全に一致することはまずないため，それらの間のマッピングが重要となる．今後の課題は，そのようなマッピングの支援機構の開発，UML などへの拡張およびシステムの実装である．

参 考 文 献

1) W3C. XML Schema. http://www.w3.org/XML/Schema.