

SuperSQL による XML 生成の高速化

5W-1

慎 祥揆† 有澤 達也‡ 遠山 元道§

†慶應義塾大学大学院 理工学研究科 開放環境科学専攻 ‡慶應義塾大学大学院 理工学研究科 計算機科学専攻

§慶應義塾大学 理工学部 情報工学科 / JST さきがけ研究 2 1

1 はじめに

XMLはWWWでのビジネスデータ交換のフォーマットの標準になりつつある。しかし、現存の大部分のデータは相変わらず関係データベースシステムに貯蔵されていて、今後もこのような現象は持続すると思われる。そのために、関係データベースに貯蔵されているデータをXMLで出版するためのシステムが必要であり、多くの機能をそなえたSuperSQLはそのようなシステムうちの一つである。関係データベースに貯蔵されているデータをXMLに出版するMIDDLE-WAREシステムとしてのSuperSQLにおいて重要な問題は関係型の構造をもつデータから階層構造をもつXMLデータに変換する効卒の良い方法をさがすことにある。本研究ではSuperSQLを用いて関係データベースからXMLを出版するとき、どの時点でXMLデータを構成するか、反復データのグルーピングとJOINの実行をどういうタイミングで行うのかを考えて最も適切かつ高速なXMLの生成を目指すものである。

2 SuperSQL

SuperSQLの質問文は、SQLのSELECT節を拡張し‘GENERATE<メディア指定><TFE>’という構文を持つGENERATE節で置き換えたものである。<メディア指定>にはXML,HTML,LATEX等の出力したい媒体の指定ができ、<TFE>には出力

のレイアウトを記述する。TFEは、表の構造とレイアウトを規定する式の一つであり、オペランドと演算子から構成されている。オペランドはSQLのターゲットリストの要素であり、演算子は結合子及び反復子からなる。結合子は属性及びTFEをレイアウト上で並べるために用いられ、そのレイアウト方向を示す為に「,」(縦方向)、「!」(横方向)、「%」(深さ方向)、「#」(時間方向)によって二つの属性やTFEを区切ることによって示される。また、反復させた部分を「[]」で囲む反復子を用いて記述する。

GENERATE XML

```
[ a.value, [ b.value, [ c.value ] ! ] ! ] !  
FROM ex.a a, ex.b b, ex.c c  
WHERE a.id = b.ex.a and b.id = c.ex.b and  
c.id = 10000;
```

3 実行選択

関係データベースからXMLに出版するためには初めに関係データベースからの問い合わせのフラットな検索結果をXMLに合わせるための木構造に変換することが必要である。

SuperSQL処理系では、与えられた質問文に対して図1に示すように「構文解析部」において構造情報と装飾情報を抽出し、「データベースインタフェース部」において分解された情報をもとにSQLを関係データベースに発行する。そして、その検索結果を「データ構造化部」において構造情報に従ってグルーピングを行い、「レイアウト編集部」において装飾情報とメディア指定にあわせて、目的の応用データを生成する。

[4]ではデータ構造化部で特に着目し、SuperSQLの質問文処理に対して、ORDERBY節をSQLへ自動的に付加することにより、グルーピング時に

SHIN Sanggyu†, ARISAWA Tatsuya‡,
TOYAMA Motomichi§

†School of Science for OPEN and Environmental System,
Faculty of Science and Technology, Keio University.

‡Department of Computer Science, Faculty of Science and
Technology, Keio University.

§Department of Information and Computer Science, Faculty
of Science and Technology, Keio University. PRESTO, JST.

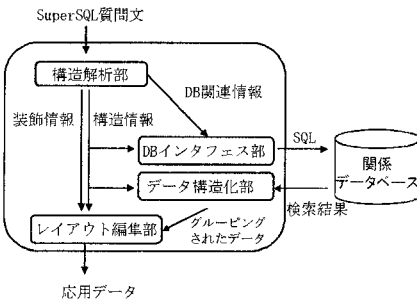


図 1: SuperSQL 処理系処理図

必要なソート処理を、データベースからの検索結果を得る時点で行うといった手法を提案した。

3.1 タグ化プラン

本論文で特に着目する部分は、データ構造化部とレイアウト編集部である。今までの SuperSQL での XML 出版はデータ構造化部とレイアウト編集部が分離されている(図 2)。つまり、関係データベースからの検索結果を構造化し、この構造を XML で出版している。本論文ではこの方法とデータを構造するときのグルーピング時点にタグする(図 3)ことを処理時間で比較する。すなわち、全ての構造が構成されたときではなく、反復されるデータに対してグルーピング時点で必要なタグをつける。

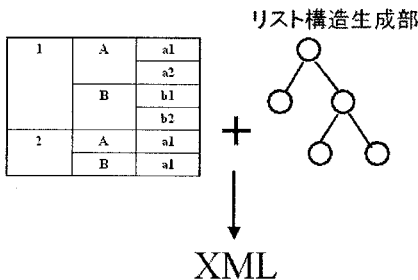


図 2: 全て構成の後にあとタグ化

3.2 問い合わせ処理プラン

また、もう一つの問題は分割された構造の ORDERBY によるグルーピングの問題である。さまざまなグルーピング木を想定した上で、結合処理上の

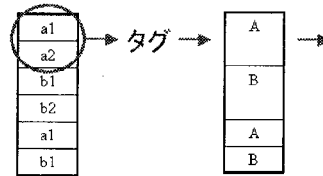


図 3: 反復子があるときのタグ化

クルーピングしたプランと結合処理上アウトユニオンする方法(図 4)に対して処理時間を比較する。

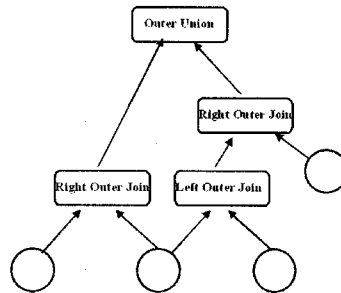


図 4: 問い合わせ処理プラン

4 まとめ

SuperSQL の XML 出版に対していろいろな方法の構造化とタグ時点に対して実験的に比較して効率的な方法を模索した。関係データベースから XML に効率的にデータ変換をするためには ORDERBY 節の付加、木構造の分化計画などの前処理とデータ構造とレイアウト時点などに関する色々な方法があるが、どのような方法が高速になるかを実験的に比較する必要がある。

参考文献

- [1] Motomichi Toyama : An Extended SQL for Database Publishing and Presentation, *Proc. of ACM SIGMOD '98*, pp. 584-586, 1998.
- [2] SuperSQL Home Page, <http://ssql.db.ics.keio.ac.jp/index.html>
- [3] Michael carey, et. al. : Efficiently Publishing Relational Data as XML Documents, *Proc. of the 26th VLDB Conference*, pp. 67-76, 2000.
- [4] 有澤 達也, 遠山 元道 : SuperSQL 処理系におけるグルーピング操作の効率的な実装, DEWS 2001