

XML-OLAP における効率的な構造ベースの集約計算を行うアルゴリズム

キット チヤントラ⁺ 天笠 俊之^{+,++} 北川 博之^{+,++}
⁺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
⁺⁺筑波大学計算科学研究センター
〒305-8573 茨城県つくば市天王台 1-1-1
kchantola@kde.cs.tsukuba.ac.jp, {amagasa,kitagawa}@cs.tsukuba.ac.jp

Algorithms for Efficient Structure-based Grouping in XML-OLAP

Chantola KIT⁺, Toshiyuki AMAGASA^{+,++}, and Hiroyuki KITAGAWA^{+,++}

⁺Department of Computer Science,
Graduate School of Systems and Information Engineering

⁺⁺Center for Computational Sciences

University of Tsukuba

1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

kchantola@kde.cs.tsukuba.ac.jp, {amagasa,kitagawa}@cs.tsukuba.ac.jp

1. はじめに

Only in a short period has XML become a hugely popular format for marking up all kinds of data from web contents to data used by applications. XML has been used in a wide spectrum of application domains, such as web documents, business documents, and log data. Also, we have observed that the required data are often available as XML data.

Meanwhile XML has been broadly used; the more complex analysis on XML data is becoming extremely important in order to discover precious information from massive XML data. In fact, XML query languages such as XPath and XQuery have been used for querying XML data. However, the query languages do not support complex analytical functionalities such as OLAP.

Related to this issue, Bordawakar et al. proposed various concepts for XML data analysis in reference to OLAP for relational data. They also proposed new syntactical extensions to XQuery, such as “GROUP BY”, “ROLLUP”, and “TOPOLOGICAL ROLLUP”. The ROLLUP operations enable users to make data grouping by rolling up the multi-level of XML data, especially “TOPOLOGICAL ROLLUP” for structure-based hierarchy.

To contribute to the complex analysis of XML data, in our previous research, we discussed an approach to XML-OLAP system based on relational database systems. We discussed a logical data model of XML datacube and implementation using relational database technology.

Accordingly, in our recent research, we forwardly move our focus to the ROLLUP operation for XML's structure-based hierarchy that is “GROUP BY TOPOLOGICAL ROLLUP”. Specifically, we propose the following algorithms: Top-down (TOD) and Bottom-up (BUC) based on Structural Join algorithms. Structural Join enables us to find the primitive tree structured relationships, parent-child and ancestor-descendant, in XML database. By using Stack Tree Join Ancestor, one of Structural Join algorithms, TOD computes the grouping from the grand level to the detail level of XML hierarchical levels. In reverse, BUC computes grouping from the detail level to grand level, and reducing the processing cost by using the intermediate result from the lower level. However, we claim that both TOD and BUC are not so effective because they repeat Structural Join many times according to the number of XML's hierarchical levels.

Furthermore, we propose more effective algorithms, *Single Scan by preorder number (SSC-Pre) and by postorder number (SSC-Post)*, for TOPOLOGICAL ROLLUP grouping. SSC-Pre and SSC-Post will have only single Structural Join for scanning all XML nodes by manipulating “stack” and “list” respectively, for maintaining the intermediate results. We will also show the feasibility of our new algorithms by performing experimentation with large and complex structured XML data.