

Chapter 17

Integration of Relational and Native Approaches to XML Query Processing

Huayu Wu

National University of Singapore, Singapore

Tok Wang Ling

National University of Singapore, Singapore

ABSTRACT

Existing XML twig pattern query processing algorithms fall into two classes: the relational approach and the native approach. Both kinds of approaches have their advantages and limitations. Particularly, the relational approach can search for data values (content search) efficiently using tables, but it is not efficient to match query structure to documents (structural search). The native approach processes structural search efficiently, but it has problem dealing with values. In this chapter, a hybrid approach for XML query processing is introduced. In this approach, the content search and the structural search in a twig pattern query are performed separately using the data structures in the relational approach and the native approach, i.e. relational tables and inverted lists. The authors show that this hybrid style technique can process both structural search and content search efficiently, and then improve the query processing performance comparing to the existing approaches. Furthermore, when more semantic information on object class and relationship between objects in the XML document is known, the relational tables used can be optimized according to such semantic information to achieve a better performance. Finally after performing twig pattern matching, value results can be extracted easily using relational tables, rather than navigating the document again in many other approaches.

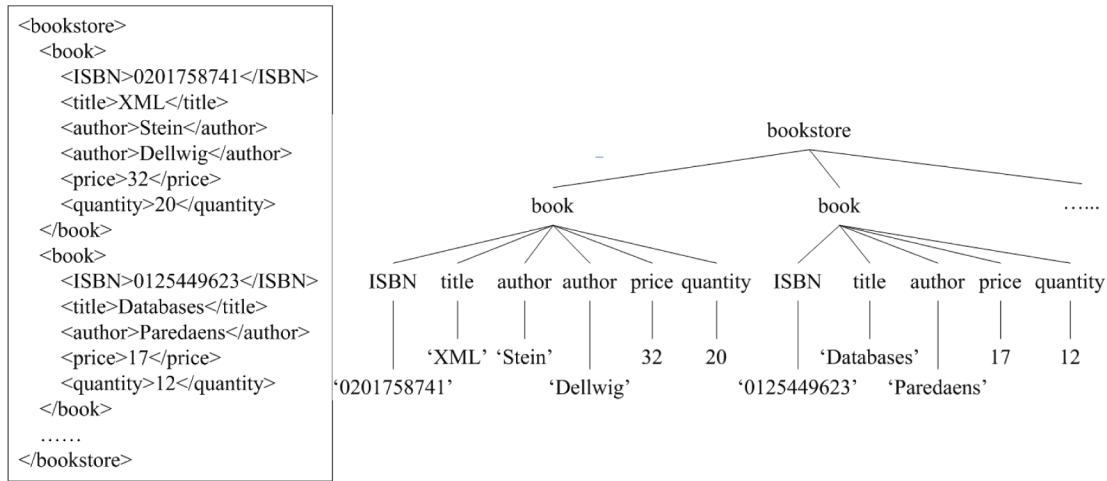
INTRODUCTION

XML is emerging as a standard format for data representation and exchange over the Internet. As a result, there is a compelling need of developing

efficient algorithms to query XML data. An XML document is normally modeled as a tree, without considering ID references. In an XML tree, the nodes represent the elements, attributes and values in the document and the edges reflect the nested relationships between elements, attributes and values. Figure 1 shows a small XML document with

DOI: 10.4018/978-1-61520-727-5.ch017

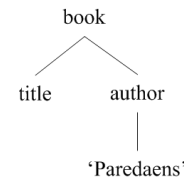
Figure 1. An XML document with tree model representation.



its tree model representation. In most standard XML query languages, e.g. XPath and XQuery recommended by W3C, the core pattern of a query is also a tree structure (or a set of tree structures). We call such a tree-structured query pattern a *twig pattern*. Finding all occurrences of a twig pattern query in an XML document is considered as an essential operation for XML query processing. Consider a query to find the titles of all books written by Paredaens in the bookstore document shown in Figure 1, the corresponding XPath expression is `//book[author="Paredaens"]/title`, and the twig pattern representation of this XPath query is shown in Figure 2.

Normally, an XML query consists of structural search and content search. The structural search in a query aims to find all matches that satisfy the structural constraints between query nodes in the document. Whereas the content search filters the query result based on the value comparison in query predicates. For the above XPath query, `//book[author]/title` is a structural search and `author="Paredaens"` is a content search. How to efficiently process structural search and content search are equivalently important to achieve a good query processing performance.

Figure 2. Example twig pattern query.



The XML query processing algorithms are categorized into two classes: the relational approach and the native approach. In the early stage, many research works focus on using relational databases to store and query semi-structured XML data. The main idea of the relational approach is to shred XML documents into relational tables, and transform XML queries into SQL queries to query the database. During query transformation, an XML query structure is normally expressed by multiple selections and joins in the corresponding SQL statement, thus the overhead on these table operations, especially join, may seriously affect the efficiency of the structural search. Later how to process XML queries natively becomes a hot research topic, because the native approach does not use relational database and avoids the overhead on relational operations. The structural join based approach is an important native approach

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/integration-relational-native-approaches-xml/41513

Related Content

Supplementing UML with Concepts from ORM

Terry Halpin (2001). *Unified Modeling Language: Systems Analysis, Design and Development Issues* (pp. 168-185).

www.irma-international.org/chapter/supplementing-uml-concepts-orm/30578

Extension of the Unified Modeling Language for Mobile Agents

Cornel Klein, Andreas Rausch, Marc Sihling and Zhaojun Wen (2001). *Unified Modeling Language: Systems Analysis, Design and Development Issues* (pp. 117-129).

www.irma-international.org/chapter/extension-unified-modeling-language-mobile/30575

A JSON-Based Fast and Expressive Access Control Policy Framework

Hao Jiang and Ahmed Bouabdallah (2019). *Emerging Technologies and Applications in Data Processing and Management* (pp. 70-91).

www.irma-international.org/chapter/a-json-based-fast-and-expressive-access-control-policy-framework/230684

Using Rule-Based Concepts as Foundation for Higher-Level Agent Architectures

Lars Braubach, Alexander Pokahrand and Adrian Paschke (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches* (pp. 493-524).

www.irma-international.org/chapter/using-rule-based-concepts-foundation/35872

Temporal OCL: Meeting Specification Demands for Business Components

Stefan Conrad and Klaus Turowski (2001). *Unified Modeling Language: Systems Analysis, Design and Development Issues* (pp. 152-167).

www.irma-international.org/chapter/temporal-ocl-meeting-specification-demands/30577