

A GRAPH-BASED CONCURRENCY CONTROL PROTOCOL FOR XML METADATA KNOWLEDGE BASES

TSUI-PING CHANG¹ AND CHUN-HSIU YEH²

¹Department of Information Technology
Ling Tung University
No. 1, Lingtung Rd., Taichung 408, Taiwan
apple@teamail.ltu.edu.tw

²Department of Information Managements
Chung Chou Institute of Technology
No. 6, Lane 2, Sec. 3, Shanjiao Rd., Yuanlin Township, Changhua 510, Taiwan
chyeh@dragon.ccut.edu.tw

Received December 2009; revised March 2010

ABSTRACT. *Providing efficient access for XML APIs in XML metadata knowledge bases (XMKBs) is crucial, since XML is widely used to integrate data in XMKBs. In XMKBs, the types, contents and locations of heterogeneous data are illustrated in XML documents to provide a uniform interface for users to manipulate this data. As users update and query information about this heterogeneous data, the XML documents are manipulated by XML APIs (i.e., DOM APIs) to ensure consistency. Concurrency control protocols can ensure a better performance of the DOM APIs to manipulate XML documents. However, the existing protocols (i.e., traditional locking protocols and XML-based protocols) are not sufficiently adequate for DOM APIs in XMKBs. Traditional graph-based locking protocols cannot deal with new DOM operations, and the existing XML-based protocols treat XML documents as trees, with no regard for the ID/IDREF(s) used in them. This paper proposes a new XML graph-based concurrency control protocol, tailored for XMKBs, named XGP. The novel feature of XGP is to analyze the DOM APIs on XML document graphical structures to increase the transactional performance of XMKBs. Our simulated results show that our XGP outperforms other concurrency control protocols in terms of higher throughput.*

Keywords: Concurrency control, DOM, XML graph, XML metadata knowledge bases

1. Introduction. With the growth of the Internet, both the number and size of data sources available for public access is rapidly increasing. Thus, the need to combine data from different autonomous and heterogeneous data sources to provide public access to the Internet has become important [5,6,8,9,17,19]. However, most enterprises store data in relational database systems because of the reliability, scalability and performance associated with these systems. In addition, since many web-based systems publish their data using XML [4], a great deal of interesting and useful data can be found in valid and well-formed XML documents [4]. As a result, building an XML metadata knowledge base (namely XMKB) [5,6,8,9] which can provide unified access to diverse data sources is very desirable for linking data held in relational databases and XML documents.

Many XMKBs [5,6,8,9] are introduced to maintain the data source information (e.g., names, contents, types and locations) and meta-information about the relationship between paths of data sources. This is because XML has become the standard format to exchange information over the Internet, and it is suitable as metadata to integrate heterogeneous data. Thus, an XMKB contains multiple XML documents to describe the