Journal of Network and Computer Applications 37 (2014) 348–364

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Active XML (AXML) research: Survey on the representation, system architecture, data exchange mechanism and query evaluation

Binh Viet Phan, Eric Pardede*

Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia

ARTICLE INFO

Article history:
Received 17 July 2012
Received in revised form
20 November 2012
Accepted 5 February 2013
Available online 27 February 2013

Keywords:
XML
Active XML
Intensional XML data
Web service
Peer-to-peer network

ABSTRACT

Active XML (AXML) is an extension of XML to exploit the powerful computation ability of peer-to-peer network and Web services technologies. AXML is considered a distributed XML DBMS which extends the capability of XML by embedding intensional XML data inside XML documents. The management of intensional XML and XML data together in XML documents raises issues such as representation for intensional XML data, AXML–XML data exchange and AXML data query processing. This paper will study these issues, comparing as well as discussing the current solutions to AXML systems.

Contents

1. Introduction

Traditional database systems are well-suited to manage and organize structured data. They are also used as one of the foundations for building and publishing web sites over centralized network architectures and the internet. The strength of

* Corresponding author. Tel.: +61 3 9479 3459; fax: +61 3 9479 3060.
E-mail address: E.Pardede@latrobe.edu.au (E. Pardede).

1084-8045/$ - see front matter
http://dx.doi.org/10.1016/j.jnca.2013.02.002

these database systems is derived from strict requirements of structured data as well as their own data formats.

However, there is a considerably high cost implication for data exchange and integration between traditional database systems and between applications and web sites built on top of these database systems. Moreover, along with the developments of the internet as well as diversified data formats, there is a high demand for exchanging and storing unstructured data which is not well-suited to be stored inside traditional database systems. These problems are solved by the emergence of eXtensible Markup Language (XML).

XML (Bray et al., 2008) is a flexible and scalable text-based language. In the last decade, it has been very popular for data representation and exchange over the internet. XML is applied as a means to communicate between applications over networks, regardless of different platforms. XML can be used as a repository for unstructured data as well as for data exchanges. In the early development stage of the internet, client-server architecture was a typical model for web applications. Nevertheless, centralized network models do not support the various properties needed in current web applications as well as trends of current technology such as exchangeability, heterogeneity, data scalability, autonomous systems and Service-Oriented trends. Therefore, peer-to-peer architectures are proposed as an alternative solution. The development of peer-to-peer architecture is also supported and facilitated by the emergence of web services technologies that help to remove the antagonism of diversified systems and platforms.

Web services are technologies that have been rapidly developed. This technology has been creating and contributing to fundamental infrastructures for Web applications. Web services are expected to be efficient sources of data, which are dynamic, distributed and interoperable. Hence, Web services can be considered as data objects and can be manipulated like data in DBMSs (Yu et al., 2008). It can be seen that the needs of infrastructures which rely on XML, apply to peer-to-peer architectures to manage and process Web services, such as database systems, are imperative.

Gemo, a research group on Integration of Data and Knowledge Distributed over the Web, proposed a declarative framework to integrate data and services that is provided by an extension of XML, namely Active XML (AXML) (Abiteboul et al., 2003; Abiteboul et al., 2004a, 2004b, 2004c, 2004d; Milo et al., 2003). This extension of XML with special characteristics offers advantages to web technologies, particularly in peer-to-peer (P2P) architectures. Two other groups aim to improve the original AXML system (Ferraz et al., 2010; Phan et al., 2011). The first group focuses on the improvement of platforms for storing and processing AXML (Ferraz et al., 2010). LTU (a research group from the La Trobe University), as the second group, proposed LTUAXML that focuses on the representation, AXML data exchanges, and query algorithms (Phan et al., 2011; Phan and Pardede, 2011).

2. AXML as intensional XML data

AXML embeds dynamic XML data, which is provided by the means of Web services, into XML data. The dynamic data is described by the representative XML elements, which contain information on Web services and other parameters. These XML elements are called intensional (XML) elements. Collections of intensional XML elements are called intensional XML data. AXML data are XML documents which contain both XML data and intensional XML data.

AXML data is employed in AXML systems to exploit the power of Web service technology, the computation capability of P2P architectures and the wide use of XML.

An AXML system contains an AXML database that maintains a collection of AXML data, which can be located in different peers and under the control of different and separate DBMSs running on independent computer systems. These peers are independent and autonomous but can participate in the execution of a query with other peers. These queries can require data from some other peers by means of Web services.

AXML can be employed in P2P architecture, centralized architecture and a hybrid of the two. AXML can take full advantage of P2P architecture because of its capability to share computing power between peers in the networks.

For end-users, AXML perform like normal centralized XML DBMSs. Additional computations such as invocation of Web services, materialization of AXML data and querying AXML data are executed by additional modules along with XML DBMSs, which is transparent to the end users.

We provide several scenarios to demonstrate how AXML systems can be used in our daily life.

Scenario 1: *Newspaper Web Content*. A homepage of a newspaper usually comprises: (1) explicit data such as the newspaper's name, the current date and time etc.; and (2) implicit data such as the weather forecast and local events. It is easy to see that the information about weather and local events should not be given as explicit and static data because this information only makes sense at a specific time. This information is usually retrieved by Web service calls.

Scenario 2: *Collaborative Management for a Real Estate Agency*. It is common for a real estate agency to design a database that can manage, facilitate and provide collaborative and mobile workspace for their staff as well as information to customers. A staff member can access their related documents, create new information and insert these into documents from their portable devices via the internet.

The most suitable solution is using intensional data on P2P architecture because it reduces the load on the central server and each peer can have their own initiative and independency. As described in (Abiteboul et al., 2003; Goldman and Widom, 1997), the agency can divide the database into three separate documents; namely, properties, requests and status (see Table 1). These documents contain comprehensive information about the properties managed by the real estate company. This information will be updated on the central repository. Staff can also examine this information on the central database. All information will be exchanged and organized between peers and the central repository by intensional information via Web services.

Scenario 3: *Electronic Patient Record Management*. Electronic Patient Record (EPR) is a document under the control of a number of peers such as hospitals, patients, insurance companies, and the

Table 1
AXML Documents and Web services for a Real Estate Agency.

AXML Documents	Contain	Web services
Properties.xml	A list of properties	getProperties(assignedTo, type, location, price); getPicture(propertyID); addClient(propertyID, clientDescr,by)
Requests.xml	Requests	getRequest(requestDescr); putRequest(requestDescr, handleBy)
allStatus.xml	Status of properties	getStatus(propertyID); updateStatus(propertyID)