Chapter 10

Content–Based Publish/ Subscribe for XML Data

Yuan Ni
IBM China Research Laboratory, China

Chee-Yong Chan
National University of Singapore, Singapore

ABSTRACT

Content-based publish/subscribe system is an effective means to deliver relevant data to interested data consumers. As the emergence of XML, it quickly becomes the de facto for data exchange on the Internet. Therefore, to use XML format in the content-based publish/subscribe system attracts increasing interests. This chapter firstly introduces the content-based publish/subscribe system for XML data; followed by the discussion of issues in this system and introduction of solutions to these issues. Finally, this chapter points out some possible research directions for the content-based publish/subscribe system for XML data.

INTRODUCTION

The Internet has considerably increased the scale of distributed information systems, where information is published on the Internet anywhere at any time by anybody. To avoid overwhelming users with such huge amount of information, the publish/subscribe systems have emerged, where users subscribe a set of queries to the system to express the kinds of information they are interested in and the system will automatically deliver newly published information to proper users. With the emergence of XML, it quickly becomes the standard for data exchange

on the Internet. There is a new trend to publish the data contents in XML format and to provide users with a more expressive subscription language such as XPath to address both the content and structure of the data, which makes the publish/subscribe system for XML data becomes increasingly important.

The publish/subscribe system could be divided into two categories: (1) topic-based publish/subscribe system which has been implemented by many industrial solutions, such as VITRIA (Skeen, 1999), TIB/Rendezvous (TIBCO, 1999), JEDI (Cugola, et al., 2001). Publishers associate some keywords with each message to indicate the topic the message belongs to; subscribers express their interests using keywords. Then all messages belonging to a

topic will be delivered to the users who subscribe to this topic. (2) content-based publish/subscribe system which improves the expressiveness by allowing the subscribers to use some subscription language to address the content of the information in which they are interested. In topic-based publish/subscribe, the information is delivered towards a group of users; while in content-based publish/subscribe, the information is delivered towards each individual user. The content-based system guarantees the users to receive accurate information they are interested in, which makes it more attractive than the topic-based system. A variety of content-based publish/subscribe systems are implemented by academy or industry, such as Gryphon (Banavar et al. 1999), Siena (Carzaniga et al., 2000), Elvin (Segall et al., 2000) and ONYX.(Diao et al, 2004). Compared with topic-based publish/subscribe, content-based publish/subscribe provides a fine grain way to address the content of the documents. Thus this chapter focuses on the content-based publish/subscribe system for XML data.

This chapter firstly gives an introduction about the content-based publish/subscribe system for XML data and three components, i.e. publisher, subscriber and XML routing network, in such system. The content-based publish/subscribe system for XML data has two important issues, i.e. the efficiency of the system and the functionalities of the system. Therefore, secondly this chapter categories and introduces the existing approaches that are proposed to improve the above two aspects of the XML publish/subscribe system. Finally, this chapter will point out some possible future work in the content-based publish/subscribe system for XML data domain.

BACKGROUND

In the content-based publish/subscribe system for XML data, the information is published as XML documents and the subscriptions are expressed using some XML query language such as XPath (1999) or XQuery (2006). Figure 1 illustrates the architecture for the content-based publish/subscribe system. There are three components in the system:

- **Publishers.** The left part in Figure 1 shows the data publishers, which are also called the data producers for the system. They generate the information and encode it as XML documents, and send these documents to the system. Published XML documents could come from various resources, such as newspapers, databases, libraries, mobile sensors, etc. Various publishers generate the XML documents independently, thus XML documents for the same domain by different publishers may conform to different schemas.

- **Subscribers.** The right part in Figure 1 gives the subscribers which are also called the data consumers, who receive the information from the data publishers. The subscribers register their interests to the system by subscribing their profiles to the system. In the system for XML data, their profiles are rewritten using some XML query language such as XPath (1999) or XQuery (2006). The subscribers would receive all and exactly the information that matches their subscriptions. When the subscribers do not want the information anymore, they need to unsubscribe their queries.

- **XML routing network**. The central part in Figure 1 illustrates the XML routing network, which contains a set of XML routers that are inter-connected. Each XML router receives the subscriptions from end-users or other XML routers; and receives the XML documents from the publishers or other XML routers. A routing table is stored at each router to store the set of queries subscribed to the router, and the routing table also maintains the information about the

Related Content

Integration of Relational and Native Approaches to XML Query Processing

Huayu Wuand Tok Wang Ling (2010). *Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies (pp. 385-405).*

www.irma-international.org/chapter/integration-relational-native-approaches-xml/41513

Towards the Integration of a Formal Object-Oriented Method and Relational Unified Process

Jing Liu, Zhiming Lui, Xiaoshan Li, He Jifendand Yifeng Chen (2005). *Software Evolution with UML and XML (pp. 101-133).*

www.irma-international.org/chapter/towards-integration-formal-object-oriented/29611

A Logic Programming Perspective on Rules

Leon Sterlingand Kuldar Taveter (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches (pp. 195-213).*

www.irma-international.org/chapter/logic-programming-perspective-rules/35860

A Disciplined Approach to Temporal Evolution and Versioning Support in JSON Data Stores

Safa Brahmia, Zouhaier Brahmia, Fabio Grandiand Rafik Bouaziz (2019). *Emerging Technologies and Applications in Data Processing and Management (pp. 114-133).*

www.irma-international.org/chapter/a-disciplined-approach-to-temporal-evolution-and-versioning-support-in-json-data-stores/230686

Modeling and Programming by Commitment Rules in Agent Factory

Rem Collierand Gregory M.P. O'Hare (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches (pp. 393-421).*

www.irma-international.org/chapter/modeling-programming-commitment-rules-agent/35868