Chapter VII
Recent Advances and Challenges in XML Document Routing

Mirella M. Moro
Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil

Zografoula Vagena
Microsoft Research, UK

Vassilis J. Tsotras
University of California, Riverside, USA

ABSTRACT

Content-based routing is a form of data delivery whereby the flow of messages is driven by their content rather than the IP address of their destination. With the recognition of XML as the standard for data exchange, specialized XML routing services become necessary. In this chapter, the authors first demonstrate the relevance of such systems by presenting different world application scenarios where XML routing systems are needed and/or employed. Then, they present a survey of the current state of the art. Lastly, they attempt to identify issues and problems that have yet to be investigated. Their discussion will help identify open problems and issues and suggest directions for further research in the context of such systems.

INTRODUCTION

Content-based routing is a form of data delivery whereby the flow of messages is driven by their content rather than the IP address of their destination. Specifically, in XML Routing, there is a continuous stream of XML messages (usually, one message has one XML document) from
Recent Advances and Challenges in XML Document Routing

data producers to consumers, without any of the parties having knowledge of the other (Snoeren, 2001). Message transmission is performed by a sophisticated overlay network of application-level, content-based routers (called message brokers or XML routers) that match data messages against registered client subscriptions, and forward those messages (based on such matching) to output links, i.e. other routers or clients. The task of matching incoming messages to the set of client subscriptions is called message filtering.

This form of communication is widely employed by content-based information dissemination services, which are usually instantiated as publish/subscribe systems (pub/sub for short). For example, pub/sub systems have created opportunities for new applications such as a plethora of alert and notification services that notify interested users of new products in the market, stock price changes, currency variation, better offer deals and so on. Furthermore, with the expansion of Web services, new pub/sub systems are released every week. For instance, online travel agencies such as Priceline.com and Hotwire.com inform their clients of price changes and hot deals that take into consideration the subscriber’s interests. Likewise, Ticketmaster.com sends to its users email alerts about upcoming events and pre-sale information according to the user’s signed up artists and locations.

With the recognition of XML as the standard for data exchange, specialized XML-aware information dissemination services become necessary (Diao, 2004). These services can be implemented as publish/subscribe systems in which the information to be routed is encoded

![Figure 1. General architecture of XML routing system](image)

Producers inject XML messages into the system. Consumers subscribe to the system with XML query statements. A router processes the subscriptions, usually by storing and indexing them into a routing table. Then, it evaluates the set of subscriptions over the incoming messages (such processing may include parsing operations and indexing). It then filters the messages and forwards the results to the consumers. Finally, each router also interacts with the overlay network, by forwarding (receiving) queries and results to (from) other routers.