

# Classifying XML Materialized Views for their Maintenance on Distributed Web Sources

Tuyêt-Trâm Dang-Ngoc, Virginie Sans, Dominique Laurent

LICP Laboratory, University of Cergy-Pontoise  
2, avenue Adolphe Chauvin  
95302 Cergy-Pontoise Cedex. France  
{*Prenom.Nom*}@dept-info.u-cergy.fr

**Résumé.** Ces dernières années ont mis en évidence la croissance et la grande diversité des informations électroniques accessibles sur le web. C'est ainsi que des systèmes d'intégration de données tels que des médiateurs ont été conçus pour intégrer ces données distribuées et hétérogènes dans une vue uniforme. Pour faciliter l'intégration des données à travers différents systèmes, XML a été adopté comme format standard pour échanger des informations. XQuery est un langage d'interrogation pour XML qui s'est imposé pour les systèmes basés sur XML. Ainsi XQuery est employé sur des systèmes de médiation pour concevoir des vues définies sur plusieurs sources. Pour optimiser l'évaluation de requêtes, les vues sont matérialisées. La difficulté est de maintenir incrémentalement des vues matérialisées lors de la mise à jour des sources, car dans le contexte de sources web, très peu d'informations sont fournies par les sources. Les méthodes habituellement proposées ne peuvent pas être appliquées. Cet article étudie comment mettre à jour des vues matérialisées XML sur des sources web, au sein d'une architecture de médiation.

## 1 Introduction

In industry as well as in research, data information systems were often built separately. This results in highly heterogeneous and distributed systems, which generates fragmented views of data and information. In 1992, Wiederhold proposed a mediation architecture (Wiederhold 1992) for combining information from multiple data sources. The goal of a mediation architecture is to integrate the sources in order to present a uniform view to the final user or application. Generally speaking, a mediation architecture is composed of mediators and wrappers. A wrapper is associated to each source with the roles of extracting and transmitting information from the sources to the integrator. Before transmission, the data are transformed into a predefined format used by the mediator. The mediator integrates information provided by the wrappers and presents a queryable uniform view to the user application.

To facilitate the integration of data across different systems, XML has been adopted as the standard format for information exchange and XQuery, a standard query language for querying XML, has become a major requirement for XML-based systems. Recently, mediation architectures based on XML and XQuery have been proposed and implemented (Dang-Ngoc and Gardarin 2003) (Draper *et al.* 2001). We also note