

Chapter 2.10

Mediating RDF/S Queries to Relational and XML Sources

I. Koffina

Institute of Computer Science & University of Crete, Greece

G. Serfiotis

Institute of Computer Science & University of Crete, Greece

V. Christophides

Institute of Computer Science & University of Crete, Greece

V. Tannen

University of Pennsylvania, USA

ABSTRACT

Semantic Web (SW) technology aims to facilitate the integration of legacy data sources spread worldwide. Despite the plethora of SW languages (e.g., RDF/S, OWL) recently proposed for supporting large-scale information interoperation, the vast majority of legacy sources still rely on relational databases (RDB) published on the Web or corporate intranets as virtual XML. In this article, we advocate a first-order logic framework for mediating high-level queries to relational and/or XML sources using community ontologies expressed in a SW language such as RDF/S. We describe the architecture and reasoning services

of our SW integration middleware, termed SWIM, and we present the main design choices and techniques for supporting powerful mappings between different data models, as well as reformulation and optimization of queries expressed against mediator ontologies and views.

INTRODUCTION

A cornerstone issue in the realization of the semantic Web (SW) vision is the achievement of large-scale interoperability among multiple information sources spread worldwide. In order to capture source semantics in a machine process-

able way, various ontology-based formalisms have been recently proposed (e.g., RDF/S (Manola & Miller, 2004), OWL (Harmelen & McGuinness, 2004)). However, the vast majority of existing legacy data are not yet in RDF/S or any other SW language (Lee & Goodwin, 2005). As a matter of fact, most of the data are physically stored in relational databases (RDB) and they are actually published on the Web as virtual XML.

SW applications, however, require to view such data as valid instances of a virtual RDF/S ontology for a specific domain of discourse and to be able to manipulate them with high-level query and view languages (e.g., RQL (Karvounarakis, Alexaki, Christophides, Plexousakis, & Scholl, 2002) or RVL (Magkanaraki, Tannen, Christophides, & Plexousakis, 2003)). Such an ontology offers an intuitive conceptual interface to express queries, which can be easily reused and refined (compared to the relational or XML data models). Therefore, we need an expressive and flexible SW middleware to either publish RDB data directly as RDF, or republish XML as RDF, or—even better—be capable of doing both. A decade of experience with information integration architectures based on mediators (Cluet, Delobel, Simeon, & Smaga, 1998; Levy, Rajaraman, & Ordille, 1996; Papakonstantinou, Garcia-Molina, & Widom, 1995) suggests that it is highly beneficial to avoid costly intermediate data migrations, as well as to (semi)automatically generate such systems from succinct *formal specifications*, mapping directly local data to mediator ontologies (or schemas). This greatly enhances the maintainability and reliability of the publishing systems in an environment of often revised and shifting requirements.

MOTIVATING EXAMPLE

Let's assume two sources providing fine-art information as seen in Figure 1: the lower left handside represents a relational database (*art.db*), while the

lower right handside represents an XML document (*art.xml*) whose content is described by a DTD or an XML schema. In particular, *art.db* consists of three tables¹: Artist(Name, Year of Birth, Nationality) that stores information about artists, Museum(Name, Address, Country) that stores information about museums, and Artifact(Title, Creator, Kind, Museum) that stores information about artifacts, their creator, and exhibition place. Moreover, document *art.xml* contains information about museums and their collections (of either paintings or sculptures). Each collection comprises artifacts along with information about their creator (either under the form of artist elements or references when they are introduced elsewhere in the document). On the top of Figure 1, we consider a publicly available RDF/S ontology in the cultural domain. This ontology is represented as a directed graph whose nodes denote the classes (e.g., Artist, Artifact) and edges denote the properties (e.g., Creates). The domain of a property is always a class (drawn as an oval), while the range may be a class or a literal (drawn as a rectangle). Subsumption relationships may be defined for both classes (e.g., Painter is a subclass of Artist) and properties (e.g., Paints is a subproperty of Creates).

In this article, we propose a *semantic Web integration middleware* (SWIM) for *virtually* integrating relational and XML sources using RDF/S ontologies as depicted in Figure 1. In order to specify our data publishing services and reformulate on-the-fly mediator queries expressed against the RDF/S ontologies according to the syntax, structure, and semantics of local sources (see Table 1 for the translation of a query returning artifacts that are exhibited in the museum of Louvre), we rely on a uniform and expressive logic framework termed *semantic Web logic framework* (SWLF). More precisely, the main contributions of SWIM are:

- **Logic framework:** We advocate a first-order logic (FOL) representation of heterogeneous data models such as RDF/S, XML, and

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/mediating-rdf-queries-relational-xml/7933

Related Content

Optimal Information Acquisition and Sharing Decisions: Joint Reviews on Crowdsourcing Product Design

Jizi Li, Xiaodie Wang, Justin Z. Zhang and Longyu Li (2024). *Journal of Database Management* (pp. 1-34).

www.irma-international.org/article/optimal-information-acquisition-and-sharing-decisions/337971

Federated Process Framework for Transparent Process Monitoring in Business Process Outsourcing

Kyoung-Il Bae and Soon-Young Huh (2004). *Advanced Topics in Database Research, Volume 3* (pp. 272-293).

www.irma-international.org/chapter/federated-process-framework-transparent-process/4364

A Unified Approach to the Design and Generation of Complex Database Schemata

J. Peckham, B. Mackellar and J. Vorbach (1997). *Journal of Database Management* (pp. 16-26).

www.irma-international.org/article/unified-approach-design-generation-complex/51181

Blockchain-Based Banking: Theory and Applications

Fasel Qadir, Gulnawaz Gani and Zubair Jeelani (2022). *Applications, Challenges, and Opportunities of Blockchain Technology in Banking and Insurance* (pp. 1-18).

www.irma-international.org/chapter/blockchain-based-banking/306452

Optimization of Multidimensional Aggregates in Data Warehouses

Russel Pears and Bryan Houliston (2007). *Journal of Database Management* (pp. 69-93).

www.irma-international.org/article/optimization-multidimensional-aggregates-data-warehouses/3367