A Distributed Model based on Ontologies to query Relational Databases

F. Edgar Castillo-Barrera, Sandra E. Nava-Muñoz and Claudia A. Mendez-Hernandez
Facultad de Ingeniería
Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico,
Av. Dr. Manuel Nava No.8
Zona Universitaria, 78290
E-mail: {ecastillo,senavam,claudia.mendez}@uaslp.mx

R. Carolina Medina-Ramirez
Department of Electrical Engineering
Universidad Autónoma Metropolitana
Distrito Federal, Mexico
E-mail: cmed@xanum.uam.mx

Abstract—The use of relational databases is still very popular to this days. The bases of its success are characterized by almost error-free performance, reliability, and migratability. With the emergence of the Internet and proposed standards to achieve a semantic web, we must have mechanisms that allow the integration of data from relational databases to the semantic web. In this paper, we propose a distributed model based on ontology technology and relational databases which combines both models. A table with ontologies is created from the tables in the relational database and stored by means of XML serialization.

I. INTRODUCTION

The History of Database Management Systems[16] began in the 40’s where it was allowed through programming languages such as Cobol and Fortran to generate flat files with basic features of reading and writing. The first commercial DBMS, IDS Integrated Data Store, was created under the concept of Network Data Model[4]; then IMS was developed: Information Management System, on the concept of Hierarchical Data Model. In the 80’s some relational DBMS arose (RDBMS) such as Oracle, Informix, Ingres and DB2. At that time need arose to manipulate complex objects and structures such as those used in CAD and CASE systems, among others. This beginning gave way to two major trends: the ORDBMS (Object Relational Database Management System) which project as an extension to the RDBMS to the OO paradigm, and OODBMS (Object Oriented Database Management System) which would be available to store and manipulate classes, objects, and the association between them and their methods. Today, the most popular relational database managements (Oracle[19], DB2-Express C[10], etc) allow us to store XML documents in a native form.

A. XML and its success on the Web

XML (Extensible Markup Language)[17] is a markup language created in 1998 by W3C (World Wide Web Consortium); it is a free and open international standard. Markup languages are computer languages that use markings or labels to define the structure, presentation and text format. The fundamental goal of XML is to interchange data among different systems, mainly through Internet data. An XML field attribute in a relational database could be the cornerstone of its success on the Semantic Web proposal because a XML document can be mapped to OWL ontology[2].

Fig. 1. Ontology-Relational Database Schema (ORDS)

II. AN ONTOLOGY FOR DESCRIBING A RELATIONAL DATABASE

An Ontology was defined by Gruber[24] as a specification of a conceptualization. An Ontology consists of terms used to describe and represent an area of knowledge, also it is the model (set of concepts) for the meaning of those terms, thus defines the vocabulary and the meaning of that vocabulary which are used by people or machines to share information.

An Ontology is a formal representation of knowledge which allows us to obtain information by means of the checking of its consistency, using Reasoners such as Pellet[3] and queries in SPARQL[23].

An Ontology-Relational Database Schema(ORDS) was created to describe the information stored in a Relational Database.
This ontology consisted of 4 classes and 3 Object Properties. The Ontology was written using RDFS and OWL-DL.
languages[8]. The Protégé editor[22], was used to visualise the ontology and its RDF graph. OWL-DL and Protégé editor are W3C [21] recommendations. The Ontology was validated by means of checking its consistency, using the Pellet Reasoner[3].

III. COMMUNICATION BASED ON ONTOLOGIES: A DISTRIBUTED MODEL

In ORDS model the communication among agents for searching information on the web is based on ontology languages. These ontologies describe the schema and structure of relational databases stored in the web server. The distributed model is shown in figure 2.

IV. CREATING AN ONTOLOGY FROM RELATIONAL DATABASE SCHEMA

The University database described in this paper was created using IBM DB2 Express-C. University relational database consists of 5 tables. Only professor and student tables have an xml field (kardex and resume). The source code is shown below.

```sql
create table student (  
id int primary key not null,  
name varchar(30),  
srp decimal(7,2),  
kardex xml
);
create table professor (  
id int primary key not null,  
name varchar(30),  
srp decimal(7,2),  
resume xml
);
create table departament (  
id_departament primary key not null,  
departament_name varchar(30)
);
```

V. CASE STUDY: ENTERPRISES SEEKING QUALIFIED GRADUATES TO HIRE

It is very common for companies looking for graduates or students in universities with certain characteristics to recruit. The Autonomous University of San Luis Potosi frequently receives requests to recruit students or graduates with some skills. This searching can be done by the companies themselves on a web page and using queries in an Ontology serialized in a XML document. In our proposal, the student's table in the University Database has a field in XML with the structure necessary to save the resume of the student. In addition another attribute of the student's table in XML is used for saving the academic record (we called it Kardex). In this example according to the ontology of Figure 1, it is possible to know the structure of the "student's" table and know that the field "kardex" is a document in XML, which can be accessed through the XQuery language as shown in figure 4.

In the case of the professor's table shown in figure 1 we can find the field resume which is a document in XML. This field contains an XML document that allows the Enterprises to
make a queries searching professors with skills or experience of interest to them. The structure of the XML document about the professor is shown in figure 5.

A. Populating the Ontology with the information about the University Relational Database

The Ontology-Relational Database Schema has to be populated using the information about the structure or schema of the relational database. The table name in the relational database has to be an instance of table class, its attributes have to be an instance of attribute class. If some attribute is an Ontology (serialized in an XML document)[14] it has to be an instance of ontology class. From the names of the tables and their attributes in the relational database, you create instances of the ontology. The created instances are shown in figure 1. The source code of the ontology populated with instances and their relations is shown below.

UniversityRD has_table Student
UniversityRD has_table Professor

Student has_attribute id_student
Student has_attribute student_name
Student has_attribute kardex_ontology
kardex_ontology is_an ontology_XML_OWL

Professor has_attribute num_professor
Professor has_attribute name_professor
Professor has_attribute resume_ontology
kardex_ontology is_an ontology_XML_OWL

VI. DB2 EXPRESS-C A NATIVE XML STORAGE

In the literature, we have found that there are three popular database management systems of commercial relational databases that store XML documents natively, which are IBM DB2[10], Informix[11], Oracle[18] and SQL Server[13]. We have selected the DB2 Express-C version because it is free to develop, free to deploy, and free to distribute.

VII. QUERIES USING SQL OR XQUERY FOR XML

In our model, using the information obtained from the ontology relational database schema, it is possible to make queries using a SQL script (with authorization) or using XQuery only for XML fields. This process is shown in figure 3.

VIII. CONCLUSION

Using the DB2 Express-C Relational Database Management Systems (RDBMS) allows us to store ontologies in a XML document such as a field of a table. The XML document is storage in Native form and we can make queries using XQuery[6] or XPath[12]. We have chosen IBM DB2 Express-C because it is free, but in our future work, we want to explore the use of ORACLE DBMS for Ontologies [9]. Ontologies are the basic of the semantic web and a method to create an ontology called Ontology Relational Database Schema(ORDS) was described. In our proposal every relational database stored in the server has to have a ORD ontology in our model. This ontology is the first access point for searching agents in the world of semantic web.

ACKNOWLEDGMENT

The authors would like to thank Mr. Robert Bryce for reviewing this paper.
REFERENCES


