

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 Ingenieria
Universidad Autonoma de San Luis Potosi,
San Luis Potosi, 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 Autonoma 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 the 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].

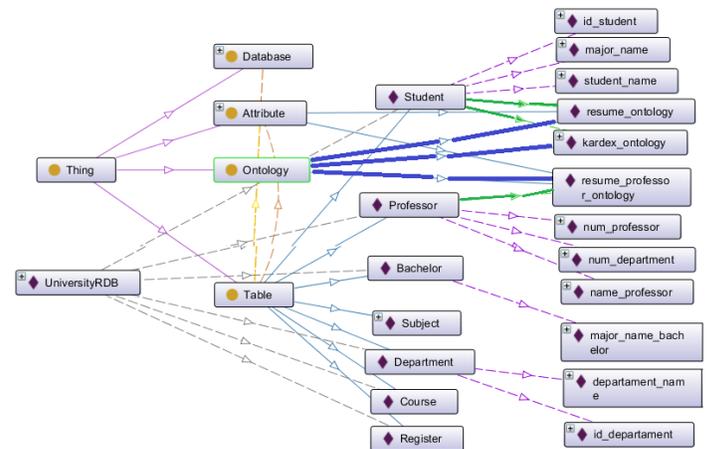


Fig. 1. Ontology-Relational Database Schema (ORDS)

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

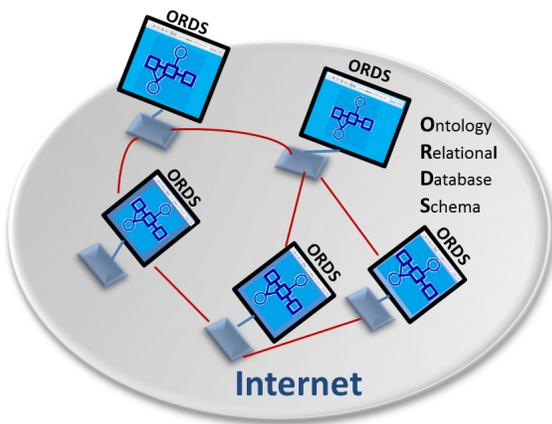


Fig. 2. Distributed Model using an Ontology Relational Database Schema

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.

```
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)
);
```

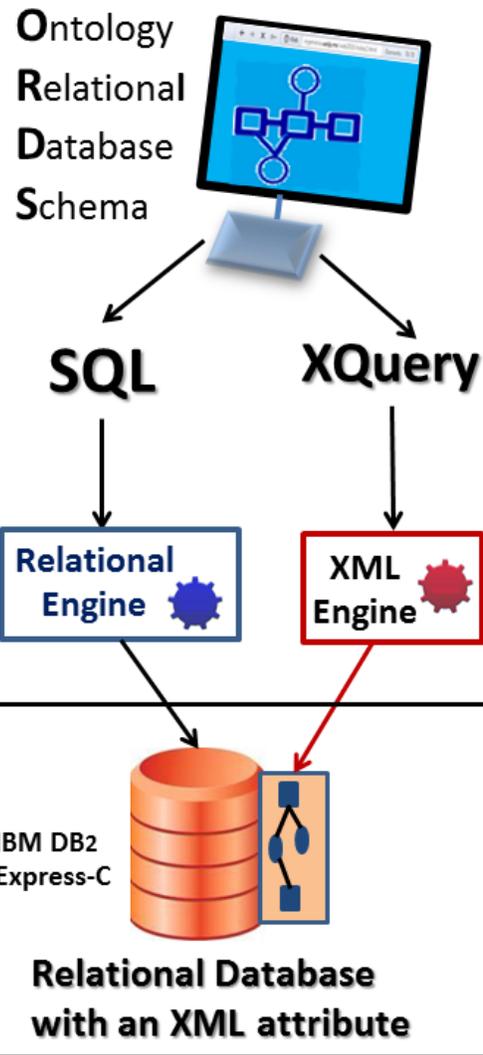


Fig. 3. Merging Relational with Ontologies Databases

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 and 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

```

<?xml version="1.0" ?>
<kardex_student
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
  <subject>
    <id> 540 </id>
    <name> Introduction to Databases </name>
    <date> 2014.7.8 </date>
    <grade> 85 </grade>
  </subject>
  <subject>
    <id> 520 </id>
    <name> Human Computer Interaction </name>
    <date> 2014.7.14 </date>
    <grade> 90 </grade>
  </subject>
  ...
</kardex_student>

```

Fig. 4. XML Structure on field student's table

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 figure1. 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.

```

<?xml version="1.0" ?>
<curriculum_professor
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <scholar_degree>
    <type> bachelor </type>
    <name> Computer Engineering</name>
    <date> 2014.7.1 </date>
    <university> UASLP </university>
  </scholar_degree>
  <scholar_degree>
    <type> master</type>
    <name> Computer Science</name>
    <date> 2016.8.1 </date>
    <university> Texas </university>
  </scholar_degree>
  ...
</curriculum_professor>

```

Fig. 5. XML Structure on field professor's table

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 choosen 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 basis 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

- [1] A. Madche et al. *An infrastructure for searching, reusing and evolving distributed ontologies*, en Proceedings of the 12th international conference on World Wide Web. ACM, 2003. p. 439-448.
- [2] B. Hannes and A. Sren. *Mapping XML to OWL Ontologies*. Leipziger Informatik-Tage, 2005, vol. 72, p. 147-156.
- [3] B. Parsia and E. Sirin, *Pellet: An owl dl reasoner*, En Third International Semantic Web Conference-Poster, 2004.
- [4] C. W. Bachman, *Data structure diagrams*, SIGMIS Database 1, 2 (July 1969), p. 4-10, <http://doi.acm.org/10.1145/1017466.1017467>
- [5] C. Nadine, G. Raji and Y. Kokou, *DB2OWL: A Tool for Automatic Database-to-Ontology Mapping*. En SEBD. 2007. p. 491-494.
- [6] DeveloperWorks, IBM, *Desarrollo con XQuery: un mejor lenguaje de programacin para el programador de bases de datos*, retrieved on 2014-10-03.
- [7] D. Dejing and L. Paea, *Ontology-based integration for relational databases*, en Proceedings of the 2006 ACM symposium on Applied computing. ACM, 2006. p. 461-466.
- [8] D. L. McGuinness et al., *OWL web ontology language overview*, W3C recommendation, 2004, vol. 10, no 10, p. 2004.
- [9] E. Cerrillo, A. Polo, J. M. Fernndez, L.J. Arvalo, M. Salas and J. C. Manzano. *Recuperacin de ontologas basada en contexto en Oracle 10g: Una aplicacin en la Arqueologa (ArqueOnto)*. CUORE XV, Congreso Nacional de Usuarios de Oracle. Palma de Mallorca, octubre de 2005.
- [10] IBM, *IBM express-c*, <http://www-01.ibm.com/software/data/db2/express-c/download.html>, retrieved on 2014-09-27.
- [11] Informix, *Informix and XML*, <http://www.informix.com/xml/>, retrieved on 2014-10-02.
- [12] J. Clark et al., *XML path language (XPath)*, W3C recommendation, 1999, vol. 16.
- [13] Microsoft, *Microsoft SQL Server XML support*, <http://msdn.microsoft.com/msdnmag/issues/0300/sql/sql.asp>, retrieved on 2014-10-03.
- [14] M. Erdmann and R. Studer, *How to structure and access XML documents with ontologies*. *Data & Knowledge Engineering*, 2001, vol. 36, no 3, p. 317-335
- [15] M. Erdmann and M. Studer, *Ontologies as conceptual models for XML documents*, En Proceedings of the 12th International Workshop on Knowledge Acquisition, Modelling and Mangement (KAW99), Banff, Canada, October. 1999.
- [16] R. Raghu and J. Gehrke, *Database management systems*, New York: McGraw-Hill, 2003.
- [17] T. Bray et al., *Extensible markup language (XML)*. World Wide Web Consortium Recommendation REC-xml-19980210, 1998, <http://www.w3.org/TR/1998/REC-xml-19980210>.
- [18] S. Muench, *Using XML and Relational Databases for Internet Applications*, Oracle Corporation (<http://technet.oracle.com/tech/xml/info/htdocs/relational/index.htm#ID795>), retrieved on 2014-10-03.
- [19] J. M. Fernndez, A. Polo, E. Cerrillo, L. J. Arvalo, M. Salas, J. C. Manzano. *Ontologas y BD relacionales en Oracle*. CUORE XV, Congreso Nacional de Usuarios de Oracle. Palma de Mallorca, octubre de 2005.
- [20] N. Guarino, *Formal Ontology in Information Systems*, Proceedings of FOIS'98, Trento, Italy, IOS-Press, 1998. p. 3-15.
- [21] *Semantic Web Activity: Semantic Web Activity State*, <http://www.w3.org/>, 2002, retrieved 2014-09-09.
- [22] Horridge, M., and Knublauch, H., Rector, A., Stevens, R. and Wroe, C., *Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0*, The University Of Manchester Printing, Vol. 27, August, 2004.
- [23] J. Pérez, M. Arenas and C. Gutierrez, *Semantics and Complexity of SPARQL*, En The Semantic Web-ISWC 2006, Springer Berlin Heidelberg, 2006. p. 30-43.
- [24] T. R. Gruber, *A translation approach to portable ontology specifications*. *Knowledge acquisition*, 1993, vol. 5, no 2, p. 199-220.