Contributions of Semantic Web Tools to Relational Models

Damien E. Zomahoun

***Abstract*—In the quest for models that could help to manage the data, relational model has been the most efficient data management solution and the data management bedrock of business information processing. Approaches have resorted to the integration of data analysis improvement knowledge and data interpretation using relational models. The availability of multiple heterogeneous, autonomous, distributed data sources containing related information has created a need for integrated access to these information systems. Although the relational models are recommended through their advantages, they have also some limits such as the data structure and relation and also the interoperability between databases. This paper studies the contributions of answering limits through relational model through the efforts of the Semantic Web Community (W3C).**

***Index Terms*—Relational model, database, semantic web, ontology, limitations.**

I. Introduction

The relational database model was developed using a branch of mathematics called set theory. In set theory a two dimensional collection of information is called a relation. A relational database management system allows users to query the tables to obtain information from one or more tables in a very flexible way. The relational database is attractive from a user's standpoint because end users often think of the data they need as a table. The capability of a relational database management is to handle complex queries is important [1].

The relational model is the primary data model for data processing applications. It consists of a collection of tables each of which is assigned a unique name and allow to store and retrieve data in a tabular form [2]. The relational world assumes that data naturally fits in tables; tables are easily and uniquely identified by a relational key, and that all views of the same tuple are consistent [3]. A row in a table represents a relationship among a set of values. The relational model represents the database as a collection of relations and each relation corresponds to a table of values or to some extent a flat file of records [4]. They are being used in a number of applications outside the domain of the traditional data processing. Relational databases are considered as the most popular storage solutions for all kinds of data and they have been recognized as a key factor in generating huge amounts of data [5].

The dramatic success of relational technology has propelled data modeling and management requirements beyond the modeling and processing capabilities of the relational technology [6]. As a result, most data modelers and data integrators work in a relational world.

Today, enterprise information systems of large companies typically store of data across multiple relational databases, each with hundreds or thousands of tables [7]. As more and more information becomes available to a growing multitude of people, the ways to manage and access data are rapidly evolving as they must take into consideration, on one front, the kind and volume of data available today and, on the other front, a new and larger population of prospective users.

This need on two opposite fronts has originated a steadily growing set of proposals for ways to manage and access data, which fundamentally rethink the concepts, techniques, and tools conceived and developed in the database field during the last years. Recently, these proposals have produced a new generation of data management systems, mostly proposed as effective solutions to the needs of an increasing number of large-scale applications for which traditional database technology is unsatisfactory [8].

Effective understanding of complex schemata is a crucial task for enterprises to support decision making and retain competitiveness on the market. Ontology-based data access (OBDA) [9] is an approach that has recently emerged to provide semantic access to complex structured (relational) data.

This paper discusses in detail two existing models used for data management, the relational model and Semantic Web model. Our goal, however, is to investigate the use of relational models and their limits in data management. Our work focuses on the relational model studies in important ways: the structure, the relationships and interoperability of data, and Semantic web contributions in these ways.

The paper is organized as follows: in Section II we present some related works. In Section III we present the relational model limitations. While in Section IV we introduce the Semantic Web tools contributions: advantages, limitations and proposal. The paper is finally concluded in Section VI.

II. Related Works

Despite the maturity of relational database products and the dramatic growth in computer power over the past decade, we still hear about projects that fail because the performance of the relational database used is just not good enough. Usually this is because of the way relational databases physically store data. For developers to assemble the data that they need, they often have to do multiple JOINs of one table to another. To retrieve the data, the database runs optimization routines to determine the best way to gather the data and then retrieves it. This process often takes a long time and can negatively impact performance [10].

Some approaches such as [1], [3] and [4] focus on access

Manuscript received September 4, 2013; revised January 6, 2014.

Damien E. Zomahoun is with the University of Burgundy, Dijon, France (e-mail: damien_zomahoun@etu.u-bourgogne.fr).

time. Authors confirm that a relational database requires much more computer memory and processing time. According to them, the slower search and the access time may result in processing efficiencies which lead to a lack of acceptance of the relational model. The databases based on relational model may be composed of many interrelated tables; the overall design may be complex and therefore have slower reach and access times in comparison to the hierarchical and network models [4]. The paper [3] suggests that the relational database management system needs comparatively powerful hardware as it hides the implementation complexities and the physical data storage details from the users. It also needs more powerful interoperability methods to run smoothly.

Other researchers have worked on the structure of the relational model. Among these works we can mention the papers [1], [3], [4], [10]-[12]. According to them, a relational database modeling gives very poor database. The ease of design would be a handicap. It can lead to the development and implementation of very poorly designed database management systems and the poorly designed database will slow the system down, and will result in performance degradation and data corruption [11]. The relational database systems are easy to use and implement, people or departments may create their own database and applications. This situation might hinder information integration that is necessary for the smooth and efficient functioning of the organization. Problems like data inconsistency, data duplication and data redundancy may also crop up [3]. According to the authors of the paper [4], some relational databases have limits in the field lengths. When a database is designed the amount of data must be specified which can fit into the field. Search queries or some names are shorter than the original and this can lead to loss of data.

Y. An *et al.* propose in [12] that although the design of relational databases is based on a conceptual model, which is defined beforehand and then transformed to the final relational model, the initial conceptual model is often not kept alongside the logical schema. According to them, the intention implied by the logical schema is missing and this represents a major obstacle in reusing it properly, e.g. in the case of legacy systems.

In [1], for example, the relational database allows only text and numerical information to be stored in the database. It did not allow the inclusion of complex object types such as graphics, video, audio, or geographic information. The desire to include these complex objects in databases led to the development of object oriented databases. Otherwise in [10], the storing and representing are considered. The fairly common data structures could be very difficult and relational databases only hold tables as unordered lists and can retrieve an ordered list only if a specially built index is added.

III. Relational Model Limits

Although the relational model offers several advantages, it is also limited in the effective data management and information processing. In summary, we distinguish three groups of relational model limits. These are: the structure, the relationship between the data and the interoperability of databases as we show in Fig. 1.

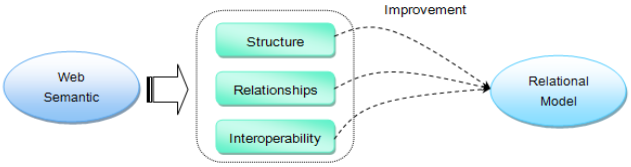

Fig. 1. Schema of relational model limits.

A. Structure

The structure of relational model defines the core of the data and the relationships involved. The model structure is described in terms of relations, attributes and domains. With the relational model, data is managed effectively with appropriate modeling. Due to their structures, data can be recorded, edited, displayed, delete, etc. Although the importance of structures roles, that play role in the relational model, we noticed some shortcomings among which we can mention the separation between the data structure and the data themselves (see Fig. 2).

Consider at first the following table (PRODUCT) for our experiment.

Id_prod	French_name	English_name	Qty	Threshold
1	Ordinateur	Computer	30	5
2	Imprimante	Printer	70	8

Fig. 2. Separation between structure and data.

It is necessary to extract the structure of the table to understand the meaning of each data. A datum is applied to a recording, as it is linked to a field and the relationship is induced by the structure of the table.

Id_delivery	Id_prod	Date	Store	Hour	Id_supplier	Qty	Cat_prod
1	1	12/10/2011	AC122	14h: 35	NULL	0	NULL
2	2	25/11/2011	B302	10h: 51	1	32	STIC

Id_supplier	French_name	English_name	Chinese_name	Date_birth	Address
1	Zhou Yang	Zhou Yang	周洋	12/02/1986	NULL
2	Gaspar Tina	Gaspar Tina	NULL	NULL	NULL

Fig. 3. Fictitious values added.

Id_delivery	Id_prod	Date	Store	Hour	Id_supplier	Id_supplier_Assi	Qty	Cat_prod
1	1	12/10/2011	AC122	14h: 35	NULL	NULL	0	NULL
2	2	25/11/2011	B302	10h: 51	1	2	32	STIC

Fig. 4a. Fields creation.

Id_supplier	Id_supplier_Assi
2	1
2	1

Fig. 4b. Table creation.

Fig. 4. Data duplicating in relational model

A datum is applied to a particular recording because it is associated to a field. The relationship is induced by the structure of the table. The data are not independent of each other. They are conceived in the context of the database, of a record and of a field. If a data is missing for a field in a record, a fictitious value "NULL" is added (see Fig. 3). Also if a datum is to duplicate in a record, it is necessary to create a field (see Fig. 4.a) or table (see Fig. 4.a).

Consider the following tables which records deliveries and suppliers (Table: DELIVERY and Table: SUPPLIER).

We note also that to manage multilingualism, it is important to create corresponding fields in each language, while the meaning of the field is exactly the same ... or create a specific table as we show in the Fig. 5.

Id_prod	French_name	English_name	Qty	Threshold
1	Ordinateur	Computer	30	5
2	Imprimante	Printer	70	8

Id_supplier	French_name	English_name	Chinese_name	Date_birth	Address
1	Zhou Yang	Zhou Yang	周洋	12/02/1986	NULL
2	Gaspar Tina	Gaspar Tina	NULL	NULL	NULL

Fig. 5. Multilingualism management.

B. Relationships

The relation defines how the data in the model will be accessed and manipulated which in turn provide the answer to some question posed by a user of the data. The manipulation is achieved though relational algebra or relational calculus [1]. The relationship between the data in two tables is induced by the use of common identifiers called foreign key and the nature of the relationship is not clearly expressed in either the structure or the data. The extraction of a database does not emphasize relationships. It is necessary to extract data from different tables to keep the relationship. We also note that the identifier of a record does not have a normalized form. It depends on the database (the table i.e. the structure), as we show in Fig. 6.

Id_prod	French_name	English_name	Qty	Threshold
1	Ordinateur	Computer	30	5
2	Imprimante	Printer	70	8

Id_delivery	Id_prod	Date	Store	Hour	Id_supplier	Id_supplier_Assi	Qty	Cat_prod
1	1	12/10/2011	AC122	14h: 35	NULL	NULL	0	NULL
2	2	25/11/2011	B302	10h: 51	1	2	32	STIC

Id_supplier	French_name	English_name	Chinese_name	Date_birth	Address
1	Zhou Yang	Zhou Yang	周洋	12/02/1986	NULL
2	Gaspar Tina	Gaspar Tina	NULL	NULL	NULL

Fig. 6. Relationships between tables.

C. Interoperability

Traditional information systems, as well known, are built for that purpose by using some data models and databases. That means to access information from other sources (or systems); an information system must transfer the data formats of these sources to hers. This process is time-consuming and not always easy [13]-[15].

In relational model, it is unable to identify two equivalent resources between two different databases. The identifiers of a record are a data as other, local and specific to a database (see Fig. 7). The same to the field names are specific to a database. The database structure is not based on any inheritance mechanism. There are no norms for naming properties and assign them to a normalization of this or that type of data. It's also impossible to relate a table to a generic model of local or external description which he can inherit characteristics. The identifier of a record does not have a standardized form. It's local and specific to a database i.e. it depends on the base, the table and the structure. Let the two following databases. It's impossible to identify two equivalent resources between two different databases. Impossible to relate a table to a generic model of local or external description which he can inherit characteristics [16].

Id_prod	French_name	English_name	Qty	Threshold
1	Ordinateur	Computer	30	5
2	Imprimante	Printer	70	8

Id_prod	French_name
ACB1	Ordinateur
WFG7	Imprimante

Id_delivery	Id_prod	Date	Store	Hour	Id_supplier	Qty	Cat_prod
1	1	12/10/2011	AC122	14h: 35	NULL	0	NULL
2	2	25/11/2011	B302	10h: 51	1	32	STIC

Id_supplier	French_name	English_name	Chinese_name	Date_birth	Address
1	Zhou Yang	Zhou Yang	周洋	12/02/1986	NULL
2	Gaspar Tina	Gaspar Tina	NULL	NULL	NULL

Fig. 7. Two equivalents resources from two different databases.

D. Summary

The relational model imposes a rigid structure, difficult to advance, independent of data themselves. It induces relationships and does not offer a clear identifier system. In relational model we have no standard representation for exchanging; relational model on a network, no standard for exchanging data in a database and merge with another basic syntax and no interoperability between different databases distributed over a network [13].

IV. The Semantic Web

The Semantic Web is a Web of data. The vision of the Semantic Web is to extend principles of the Web from documents to data. Data should be accessed using the general Web architecture using, e.g., URI-s; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data [17].

Ontologies, on the other hand, are one of the key concepts and main vehicle of knowledge in the Semantic Web research area [18]. It is playing a vital role in solving the existing web problems by producing semantic aware solutions. Ontology makes machines capable of understanding the semantic languages that humans use and understand by producing the abstract modeled representation of already defined finite sets of terms and concepts involved in intelligent information integration and knowledge management [19].

A. Web Semantic Contributions

With to Semantic Web, it is possible to be agreed to describe clearly resources and relationships between the resource described and a data with a common vocabulary. Each triplet is independent and the data structure is part of

the data.

With Semantic Web, the item that describes a resource must have a unique identifier, durable and universal and must be located on a network [16]. The identifier of a resource is not data. It is the entry point to the resource description (see Fig. 7). Semantic web requires that relationships must be clearly expressed between the resource described and a datum or another resource (see Fig. 8).

	French_name	...	Qty	Threshold
http://dbproduct.org/resources/product_info/ordinateur	Ordinateur	...	30	5
http://dbsupplier.org/resources/supplier_info/imprimante	Imprimante	...	70	8

Fig. 7. Resource described by a unique identifier.

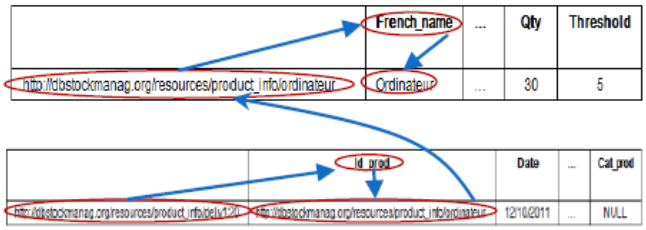

Fig. 8. Relationships between resources.

Remark: Each triplet is independent and the data structure is part of the data as we show in the next figure.

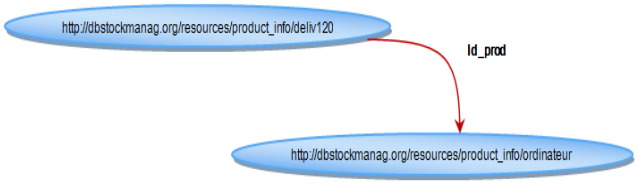

Fig. 9. Triplet RDF.

B. Ontology Limitations

The semantic web, with the new tools it offers, tries to solve the limitations of the relational model as we detail above. But it also presents some limitations and principal problems [20] among which we can mention that it is not possible to automatically handle the increase in size of ontology (due to the increase in number of classes and instances) and creating ontologies manually is a time consuming process which becomes very complex when there is a large amount of data to create large number of ontologies [9], [21]. Furthermore it is also quite impossible to perform automatic emergence of ontologies to create new ontologies and no multi user support is provided by any ontology supporting language [22], [23]. The process of manually developing of ontology is difficult, time-consuming and error-prone [24]. Existing natural language parsers used to parse the information to construct ontologies are limited because they can only work over a single statement at a time [25]. Currently available ontology validators are restricted and not capable of validating all kind of ontologies e.g. based on complex inheritance relationship [26]. Domain specific ontologies are highly dependent on the domain of the application and because of this dependency domain specific ontology's contained specific senses are not possible to find in general purpose ontology [27]. The process of semantic enrichment reengineering for the web development consists of relational metadata required to be developed at high speed and in low cost depending on proliferation of ontologies, which is currently also not possible.

V. Conclusion

Relational databases constitute significant structured sources of domain knowledge, allowing for, mainly automatic, ontology development methods. The fact that the design of relational databases is based on a conceptual model that is very much alike the ontology model, together with their frequent maintenance and timeliness of their data, especially in business environments [28], act as arguments in favor of using relational databases as knowledge sources for ontology development. Mapping databases to ontologies is a term often used to describe the above process [18]. Relational models present several advantages. Today, they are indispensable in the creation of databases. Although they are essential, their performance is limited since they are unable to meet all user requirements.

We have critically evaluated relational model and compared it with Semantic Web tools. Some important differences between the products are enumerated above. Three characteristics are studied in our analysis: structure, relationships and interoperability. Modern relational database systems also support data distribution with location transparency. The basic model of the distributed database, interconnected database instances with common transactional canals, lends itself well to a gateway model of interoperation. A real-world information system includes many types of database system, from different vendors, and capability. Gateways successfully connect these database systems so location-transparent requests may address heterogeneous data sources.

Acknowledgments

This work has been funded by Electronic, Computer Science and Image Laboratory (LE2I) and by Doctoral School SPIM, FRANCE.

References

[1] R. Swift, *Building Advanced Data Warehouse*, NCR Corporation, California, 1996.
[2] W. Inman, *Data Warehouse Building*, London: John Wiley & Sons, 1993.
[3] P. Suri and M. Sharma, "A comparative study between the performance of relational and object oriented database in data warehousing," *International Journal of Database Management Systems*, vol. 3, issue 2, pp. 116, May 2011.
[4] M. L. Brodie and J. T. Liu, "The power and limits of relational technology in the age of information ecosystems," in *Proc. Seminar on Digital Enterprise Research Institute*, NUI, Galway Ireland, April 29, 2011.
[5] J. J. Koh, "Relational database schema integration by overlay and redundancy elimination methods," in *Proc. International Forum on Strategic Technology*, Institute of Electrical and Electronic Engineers, IEEE Computer Society, 2007.
[6] C. J. Darter, *Introduction to Database System*, New York: Addison Wesley, 1995.
[7] SAP HANA. (2013). [Online]. Available: http://www.help.sap.com/hana /html/sql export.html.
[8] P. Atzeni, C. S. Jensen, G. Orsi, S. Ram, and L. T. R. Torlone, "The relational model is dead, SQL is dead, and I don't feel so good

myself," *ACM SIGMOD Record*, New York, NY, USA, vol. 42, issue 1, March 2013.
[9] D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. R. Muro, R. Rosati, M. Ruzzi, and D. F. Savo, "The mastro system for ontology-based data access," *Semantic Web Journal*, vol. 2, no. 1, 2011.
[10] B. Bloor, "The failure of relational database, the rise of object technology and the need for the hybrid database," © Baroudi Bloor International Inc., 2003-2004.
[11] M. Park, "Data warehouse designing on relational database systems," *Informix Co.*, Stanford, 1996.
[12] Y. An, A. Borgida, R. J. Miller, and J. Mylopoulos, "A semantic approach to discovering schema mapping expressions," in *Proc. IEEE 23rd International Conference on Data Engineering (ICDE'07)*, pp. 206–215, IEEE, 2007.
[13] T. D. T. Nguyen, "A DL-Based approach to integrate relational data sources into the semantic web," présenté le 5 mai, 2008.
[14] J. N. Gray, "Notes on database operating systems," in *Operating Systems: An Advanced Course*, R. Bayer, M. Graham, and G. Seegmueller, Eds. LNCS 60, Springer, Berlin, 1978, pp. 370-481.
[15] M. Stonebraker, P. Brown, and M. Herbach, "Informix Software, Inc., Interoperability, Distributed Applications and Distributed Databases: The Virtual Table Interface," *IEEE Data Engineering Bulletin*, 1998.
[16] A. Bertails and G. Poupeau, "Contribution of semantic web technologies for managing of structured data," *Slides, Atos Origin*, March 2009.
[17] (W3C) Semantic Web Activity Lead. (2009). [Online]. Available: http://www.w3.org/RDF/FAQ
[18] D. E. Spanos, P. Stavrou, and N. Mitrou, *Bringing Relational Databases into the Semantic Web: A survey*, IOS Press Amsterdam, The Netherlands, pp. 169-209, April 2012.
[19] F. Dieter, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Berlin Germany: Springer-Verlag, 2000.
[20] A. Zeeshan and G. Detlef., "Role of ontology in semantic web development," in *Proc. First International Workshop on Cultural Heritage on the Semantic Web in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference*, November 2007, Busan Korea.
[21] S. José and Q. Paulo, "A methodology to create legal ontologies in a logic programming information retrieval system," in *Law and the Semantic Web, 3369*, V. R. Benjamins *et al.*, Eds. Berlin Heidelberg Germany: Springer-Verlag, 2005, pp. 185–200.
[22] Antonkulaga, Limitations of Semantic Web; Ontology. (May 2, 2013). [Online]. Available: http://www.denigma.de/data/entry/ limitations -of-semantic-web-ontology
[23] E. Marc and Y. Sure, "Ontology mapping-an integrated approach," in *ESWS, LNCS 3053*, J. Davies *et al.*, Eds. Berlin Heidelberg Germany: Springer-Verlag, 2004, pp. 76–91.
[24] G. Perez, O. C. Garcia, and M. F. Lopez, "Ontological Engineering," New York, Inc., Secaucus, NJ, USA: Springer-Verlag, 2003.
[25] A. Galia, "Language technologies meet ontology acquisition," in *ICCS, LNAI 3596*, F. Dau, M. L. Mugnier, and G. Stumme, Eds. Berlin Heidelberg Germany: Springer-Verlag, 2005, pp. 367-380.
[26] G. Denker, S. Nguyen, and A. Ton, "OWL-S semantics of security web services: a case study," in *ESWS, LNCS 3053*, J. Davies *et al.*, Eds. Berlin Heidelberg Germany: Springer-Verlag, 2004, pp. 240–253.
[27] T. Amalia, R. Laurent, and B. Dalila, "Vulcain – an ontology-based information extraction system," in *NLDB, LNCS 2553*, B. Andersson, *et al.*, Eds. Berlin Heidelberg Germany: Springer-Verlag, 2002, pp. 64–75.
[28] S. Zhao and E. Chang, "From database to semantic web ontology: An overview," in *Proc. OTM 2007 Workshops on the Move to Meaningful Internet Systems*, LNCS 4806, Springer, 2007, pp. 1205–1214.

D. E. Zomahoun was born on April 13, 1983. He is a Ph.D. student at the University of Burgundy, Laboratory of Electronics, Informatics and Images (Le2i) in Dijon, France. D. E. Zomahoun develops the annotation applications of images. He is an author of the publication concerning Exploiting Semantic Indexing Images for Emergence Recommendation Semantics System.