

# Toward Multi-Approach Model for Semi-Automating a Data Warehouse Design from an Ontology

**Morad Hajji, Mohammed Qbadou, Khalifa Mansouri**

*Laboratory: Signals, Distributed Systems and Artificial Intelligence (SSDIA)*

*ENSET Mohammedia,*

*University Hassan II of Casablanca, Morocco*

*morad.hajji@gmail.com, qbmedn7@gmail.com, khmansouri@hotmail.com*

## ABSTRACT

The proliferation of projects that are part of the semantic Web is truly impressive. In fact, ontologies become increasingly present in information systems, they constitute great data sources that arouse the interest of being analyzed. Ontologies are used for standardizing, structuring and formalizing the Web, Web Service, E-learning systems, and other fields. Regarding multidimensional approaches, researches in this field have focused on direct Data Warehouse conception from an ontology, which do not integrate the intervention of the expert in this process. In this case, the transformations are global and not very customizable; it can reproduce the inherent defects from the data sources into the resulting data warehouse.

In this paper, we propose a new multi-approach model based on the coupling of relational database design approaches from an ontology with Data Warehouse design approaches from a relational database. Our model is semi-automatic allowing Data Warehouse design from an ontology by giving the designer more ability to intervene in this process and closely control the transformations. To assess the usefulness of our approach, we evaluated it by applying it on an example case study. The results of the example show that our approach is more accurate in terms of useful data filtering and adaptation of the multidimensional model to the end-users business-needs.

**Keywords**-Semantic Web; Business Intelligence; Data Warehouse; Ontology; Database; Coupling; Automatic; Design.

## 1 Introduction

The primary goal of the Semantic Web is to provide a solution to the structural problem of contemporary Web content. The non-structuring of this content makes its exploitation difficult and inefficient. As advocated by TIM BERNERS-LEE, Semantic Web is intended to promote the formal structuring of Web content and the interoperability between machines themselves and between machines and humans. In fact, the Semantic Web reflects a perception of an environment design that promotes exchanges between software agents and users (humans). The Semantic Web is fundamentally based on the notion of ontology. Ontology reference definition is stated by Gruber [6] and refined by Studer and al. [7]: an ontology is a formal and explicit specification of a shared conceptualization.

Technologies revolving around the Semantic Web have multiplied in a breathtaking way. Indeed, since its polemical birth, the Semantic Web continue to promote the interest of researchers and standards organisms. In particular, W3C (World Wide Web Consortium), which is the source of multiple recommendations such as Ontology Web Language (OWL), Resource Description Framework Schema (RDFS), Resource Description Framework (RDF) and SPARQL (SPARQL Protocol and RDF Query Language). In addition, the notable emergence of numerous systems for the management of ontologies such as KAON [4] and a variety of tools for manipulating ontologies such as the software 'Protégé' [8].

The modern era is experiencing an impressive proliferation of projects that are part of the Semantic Web and thus integrate the notion of ontology. Ontologies are used for the standardization, structuring and formalization of the Web, Web Service, e-learning systems, etc.

Any Semantic Web implementation requires the manipulation of ontologies that are considered as the main pillar of this discipline. In addition, the integration of the Semantic Web into several information systems relating to several domains is growing continually and multiplying day after day. As a result, ontologies have become widely used in many fields and their number and size are extensively increasing. While some ontologies are standardized and normalized, others remain local and specific to the needs of the communities that create them. Ontologies are knowledge bases that store semantically structured information. Thus, ontologies may be interrogated and analyzed in several ways and according to conventional approaches that are based on query languages or non-conventional approaches using multidimensional models.

Concerning multidimensional approaches that deal with the analysis of ontologies, many approaches have been proposed by researchers for the design of a Data Warehouse from an ontology, as [28], [29], [30] and [31]. However, most of them ignore the end-user business-needs and do not integrate the verification of the results by an expert. Thus, these approaches may not satisfy the specific needs of the end user. On the other hand, they promote the propagation of anomalies from the data source (ontology) to the data warehouse. Therefore, they lead to an inaccurate or inadequate analysis. Moreover, they only deal with ontologies expressed in a specific language such as OWL and RDF. In this case, the transformation is global and not customizable. In fact, it can reproduce the inherent defects in the data sources into the resulting data warehouse. There is a partial satisfaction for the analysis of ontologies. These approaches are not suitable, because loading data from an ontology into a multidimensional scheme can lead to facts related to non-specified or multiplied dimensions/measures.

The aim of our paper is to propose a multi-approach model for the semi-automatic and controlled design of a Data Warehouse from an ontology. Relational database approaches (RDB) design from an ontology are integrated with approaches of Data Warehouse (DW) design from a relational database. Each integration is a new approach for designing a Data Warehouse from an ontology.

This article is part of our research that aims to develop models in the fields of Semantic Web and Business Intelligence, in particular new models favoring models coupling from these two disciplines. In the course of this work, we have proposed in a previous paper a model that reflects our systemic and integrated perception of the analyzing process of an ontology generated from a documentary corpus [1]. In this article, our main aim is to complete this approach by proposing a model dealing with the ETL (Extraction-Transformation-Loading) and Data Warehousing, which is part of the model proposed in our first article [1].

## 2 Background

Since the beginning of our research, we have worked on the link between the Semantic Web and Business Intelligence, when our article [1] was developed; our approach has adopted a roadmap to establish a model of a decision-making system adopting ontological sources among the data sources. The new model proposed for this coupling clearly identifies the "Extraction-Transformation-Loading (ETL)" and "Data Warehousing" phase, ensuring the transposition of the data from an ontological source to the multidimensional model.

On the other hand, the approaches of a Data Warehouse design from an ontology differ fundamentally according to the level of representation considered of this ontology. An ontology can be considered as a database consisting of its structure (T-Box) and the data it contains (A-Box). Therefore, the design of the Data Warehouse is based on the T-Box and the loading is based on the A-Box [1].

### 2.1 Designing a RDB from an ontology

The relational model was initiated by its founder Edgar Frank Codd in 1969. The strength of this model stems from its robust mathematical basis. This model, advocated by large IT organizations as Oracle and Microsoft, is currently present in almost all information systems. At the coupling level between the relational model and the ontologies, we identify mainly at least three orientations:

- Ontology construction from a relational database.
- Relational database construction from an ontology.
- Starting from a relational database and an ontology to design links between their components. It is about the mapping between a database and an ontology.

In the context of our contribution, we are interested in the second type of coupling between an ontology and a database, namely the design of a relational database from an ontology. There are many research studies dealing with this coupling. The work of Gali et al. [11], Vysniauskas and al. [12], Astrova and al. [13], Fankam and al. [14], J. Trinkunas and O. Vasilecas [15], Astrova and al. [16], E. Vysniauskas and L. Nemuraite [17], M. Mahfoudh and W. Jaziri [18] to cite but a few.

### 2.2 Designing the DW from a RDB

According to founding father W. H. Inmon, a Data Warehouse is a collection of thematic, integrated, non-volatile, historized and organized data for decision-making [19]. A Data Warehouse (data mart) is a critical component for designing a decision support system. Kimball et al. [23] have introduced the first methodology for designing the logic scheme of a data warehouse. However, the process of this design requires expertise and is too tedious and costly in terms of effort, resources and time. To overcome the complexity and heaviness of this design, several methodologies for automating this design process have been proposed in the literature such as the work of Phipps and Davis [20], Song and al. [21], Mior and al [22] and J. Feki and al. [24].

Approaches to multidimensional modeling, according to Winter et al. [2], can be classified into three classes: demand-driven, supply-driven and hybrid approaches of the two previous classes. The approaches that form part of the 'demand-driven' class begin with the determination of the specific needs of the end-users of the multidimensional model. Contrariwise, the approaches of the supply-driven class begin with the detailed analysis of the data sources in order to determine the concepts of the multidimensional

model. However, hybrid begin with the determination of multidimensional concepts while confronting them with the specific needs of end users. In this work, the hybrid class is adopted.

### 3 Architectural Model

Our work proposes a multi-approach model for semi-automating the design of a Data Warehouse from an ontology. Each integration of a database design method from an ontology to a design approach of a Data Warehouse from a database constitutes a new approach for designing a Data Warehouse from an ontology. Fig.1 presents the overall scheme of the proposed model. Our proposal sets up a model that is an intermediate infrastructure between an ontology (data source) and the design of a corresponding Data Warehouse.

The architectural model of our approach brings together a multitude of phases, each of which is dedicated to a specific task of the design process. Indeed, as illustrated in figure 2, this infrastructure can be divided mainly into the following phases: A) acquisition of the ontology B) design of the corresponding relational databases to the ontology C) implementation of the resulting database within a RDBMS, D) design of the corresponding Data Warehouse and E) implementation of this Data Warehouse within a DBMS. In the multidimensional domain, there are several implementation approaches such as ROLAP, MOLAP, HOLAP, Web-based OLAP, DOLAP, and Real-Time OLAP. For our model, we will use the ROLAP approach by implementing the multidimensional schema within a relational environment. In accordance with this approach, facts and dimensions are materialized by relational tables.

#### 3.1 Ontology Acquisition

It is a preliminary phase whose aim is to make the ontology, which is considered as a source of data and subject of analysis, available to the model. Knowing that an ontology is composed of two parts: Terminological-Box (T-Box) and Assertion-Box (A-Box) [25]. T-Box corresponds to the conceptual model of an ontology representing the concepts linked together by semantic relations. To each concept are attached specific properties, which uniquely identifies it in a domain [5]. A-Box regroups all the instances of the concepts in addition to the instances of the relations as well as the values of the properties [25].

#### 3.2 Relational Database Design

The interest of going through the relational model is to give the designer the hand to closely control the transformation and intervene in the process. Indeed, the model provides the designer with the ability:

- To analyze the data,
- To verify the validity of the data,
- To filter the data,
- To select the useful data,
- To delete the useless data,
- To ensure the coherence of the data,
- To ensure the integrity of the data,

In addition, through the relational model, the designer is able to act on the model itself. Consequently, it can add or remove entities, attributes, and relationships.

In fact, the relational model is characterized by its proven capabilities to cope with large amounts of data.

Finally, the passage through relational model allows us to get rid of the inherent complexity of ontologies.

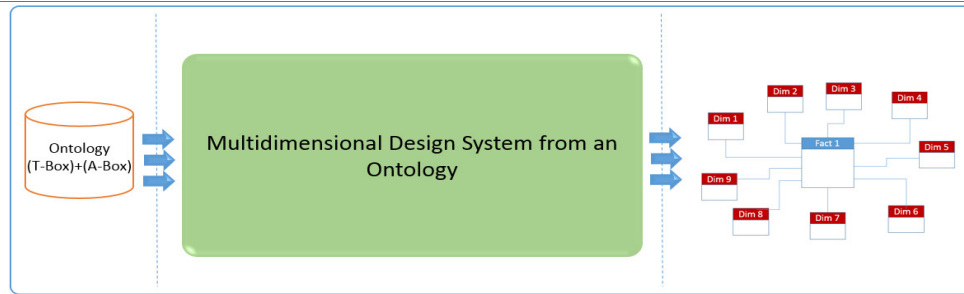


Figure 1. General architecture of a model

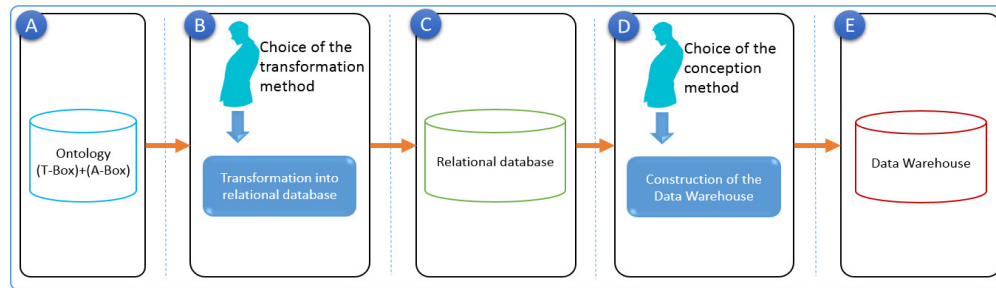


Figure 2. Proposed model architecture

The use of the relational model has substantial advantages, namely maturity, performance, robustness, scalability, reliability and availability [9]. This phase aims at designing a relational database from an ontology taking into account the specific approach chosen by the designer among those to implement in the context of our model.

Several implementations of several approaches to build a relational database from an ontology will be therefore made available to the designer. The implementation of each approach requires the development of a corresponding algorithm based on the transformation rules dictated by this approach. Among these rules is the transformation of a class into a table (Rule1 in [13]), a 'Data type' property into a column (Rule2 in [13]), and so on. Generally, any class is transformed into an entity at the conceptual level and a table at the logical level. However, the transformation of the attributes of the classes is more complex requiring two types of transformations: the transformation of attributes of type 'Data type' and the transformation of attributes of type 'Object'. Overall, the transformation of a 'Data type' attribute consists of associating it with an attribute of the corresponding entity (conceptual model) or a column of the corresponding table (logical model). On the other hand, the transformation of an attribute of type 'Object' class is much more complex. It consists in matching to this attribute an association at the conceptual level, an intermediate table, or an integrity constraint of foreign key type to the Logical level.

The intervention of the designer allows the parameterization and control of this phase in order to design the relational database corresponding to the basic ontology. This database constitutes the data source (input) for the next phase of the design process.

Each approach for designing a relational database implemented made available to the designer is an alternative for the design process of this database. Therefore, it constitutes an alternative for the design of the Data Warehouse corresponding to the basic ontology. Each choice made corresponds to a particular

result (relational database). The figure 3 shows details regarding the relational database design from an ontology.

### 3.3 RDB Implementation within a DBMS

The implementation phase of the relational database within a RDBMS is an intermediate phase.

There are two types of ontology-relational coupling methods depending on the type of generated schemas. The first kind of methods aims to design the Conceptual Data Model (CDM) from an ontology as mentioned by Y. Lv and C. Xie in [26], while the second kind of methods aims to design the Logical Data Model (LDM) from an ontology as stated by I. Astrova, N. Korda and A. Kalj in [13]. In the first case, the generated CDM will be transformed into a LDM. The logic model is then translated into a physical model (PDM: Physical Data Model) according to the chosen RDBMS. The physical model is used to generate the SQL DDL (Data Definition Language) script whose execution produces the physical structure of the relational database within the RDBMS.

This phase promotes independence from the underlying RDBMS since it is based on the CDM or LDM of the generated relational database depending on the method kind chosen by the designer.

### 3.4 Designing of the Data Warehouse

It is at this stage that the design of the Data Warehouse takes place through an approach chosen by the designer from among the approaches to be implemented. The implementation of a specific approach concretizes the vision of automatically designing a Data Warehouse from a relational database. This automation takes place in several steps. Among them are the search for candidate measures as well as candidate facts, the search for candidate dimensions, the search for hierarchies of dimensions, etc. There are methodologies that are based on a data-driven process by focusing on an in-depth analysis of data sources to derive the data warehouse schema in a reengineering process [3]. However, in order to meet users' needs and expectations, the methodologies integrate an analysis phase to avoid users' frustration by allowing them to analyze the data. These approaches are data-driven and requirement-driven as proposed by J. Feki and .al [24].

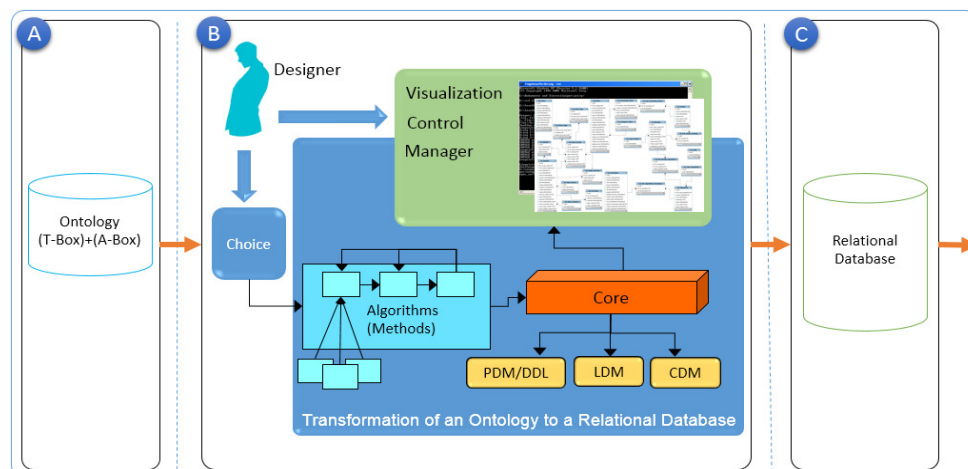
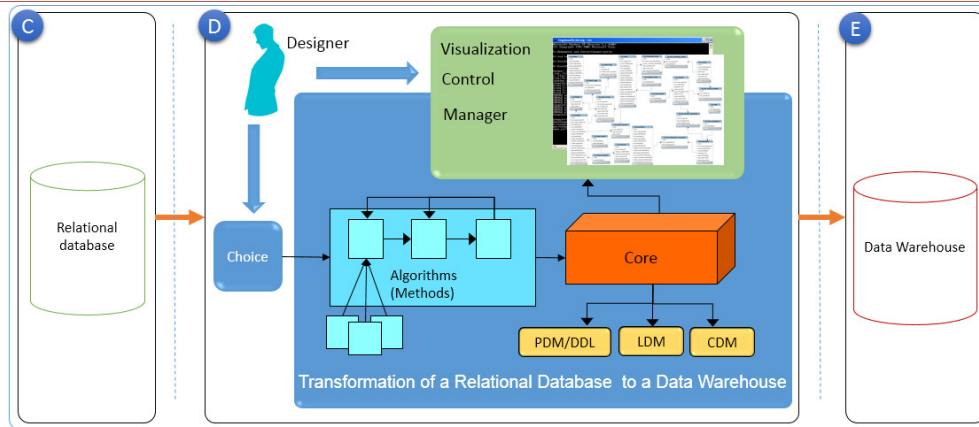


Figure 3. Designing a relational database from an ontolog





**Figure 4. Design of Data Warehouse from a relational database**

Figure 4 shows details regarding the Data Warehouse design from a relational database.

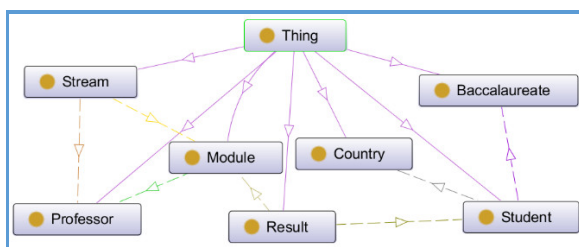
By implementing  $N$  design approaches of a relational database from an ontology and  $M$  design approaches of a DW from a RDB, our model integrates  $N$  times  $M$  ( $N * M$ ) Approaches of designing a DW from an ontology thus constituting an extensible multi-approach model (Fig. 8).

#### 4 Example Case Study

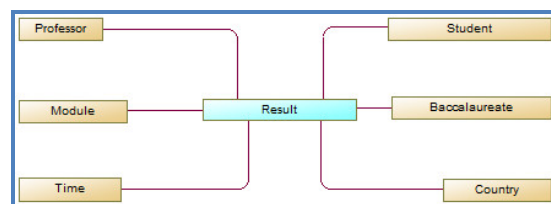
As an illustration of our approach in the design of a DW from an ontology, we put forward an example of an application to a domain ontology spelt out by the transitions of the different stages of the design of a corresponding DW. This ontology is relative to the pedagogical domain that we conceived based on a corpus of pedagogical documents, namely the descriptive sheets of the university streams, the results of the students they have obtained in a university stream module, and so on.

Figure. 5 shows a part of the representative graph of this ontology that represents the students and their performances, the university streams and their modular compositions, and so on.

Figure. 6 shows the Entity-Relationship diagram (ERD) of a relational database obtained by applying of the design/transformation rules as dictated by Y. Lv and C. Xie in [26] on the domain ontology. This resulting conceptual schema comprises 7 entities and 7 relationships.



**Figure 6. Entity-Relationship Diagram of RDB**



**Figure 7. Scheme of Data Warehouse**

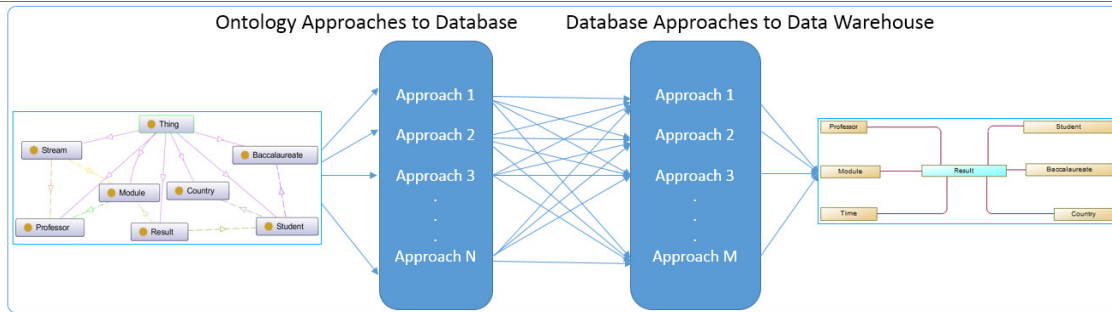


Figure 8. Combination of approaches

By applying the SAMSTAR approach [21] to design a DW from a relational database on the database whose ERD is shown in Figure 7, we identify a candidate scheme with one candidate fact, namely 'Result'. For specific purposes, we keep a scheme centered on the fact 'Result' (because it has highConnection Topology Value) as illustrated in Figure 7. This diagram includes in addition several dimensions namely 'Module', 'Student', 'Time', Etc. We add the 'Time' dimension in accordance with SAMSTAR Algorithm rule 8(iv): Add 'time' dimension to this list.

## 5 Conclusion

We presented a new multi-approach model to semi-automatize the Data Warehouse design from an ontology. The main objective of the proposed model is to maximize the adequacy of the Data Warehouse designed from an ontology to the end-users business-needs. We have shown that this model provides more ability to adjust the Data Warehouse design from an ontology to the user specific needs. By giving designer the capability to control the process, this model allows to overcome the complexity inherent in ontologies, thus minimizing the gap between data inconsistencies contained in the ontologies and data integrity in the Data Warehouse. In further work, the model should include the expert system in order to further support the designer.

## REFERENCES

- [1] M. HAJJI, M. QBADOU, K. MANSOURI, "Proposal for a new Systemic Approach of Analytical Processing of Specific Ontology to Documentary resources: Case of Educational Documents", *Journal of Theoretical and Applied Information Technology*, July 2016, Vol.89, No.2, pp. 481-51.
- [2] R. Winter, B. Strauch, "A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects", *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, pp. 231.1, 2003.
- [3] O. Romero and A. Abello, "A Survey of Multidimensional Modeling Methodologies," *International Journal of Data Warehousing and Mining* 5 (2), April 2009, pp. 1-23.
- [4] A. Maedche and S. Staab. Kaon: The karlsruhe ontology and semantic web meta project. *Künstliche Intelligenz, Special Issue on Semantic Web*, Mar 2003.
- [5] CHARLET J., BACHIMONT B. & TRONCY R., *Ontologies pour le Web Sémantique*, in *Le Web sémantique*, CHARLET J., LAUBLET P. & REYNAUD C. (Ed.), *Hors série de la Revue Information - Interaction - Intelligence (I3)*, 4(1), Cepadués, Toulouse, 2004, pp. 69-100.
- [6] T.R. GRUBER, "A translation approach to portable ontology specifications," in *Knowledge Acquisition Journal*, 5(2), Academic Press, 1993, pp. 199-220.



- [7] R. STUDER, V.R. BENJAMINS and D. FENSEL, "Knowledge engineering: principles and methods," in IEEE Transactions on Data and Knowledge Engineering, 25(1&2), 1998, pp.161-197.
- [8] W.E. Grosso, H. Eriksson, R.W. Fergerson, J.H. Gennari, S.W. Tu, and M.A. Musen, "Knowledge Modeling at the Millennium (the Design and Evolution of Protege-2000)," Proc. 12th Workshop Knowledge Acquisition, Modeling and Management (KAW '99), 1999.
- [9] E. F. Codd. 1990. The Relational Model for Database Management: Version 2. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- [10] Dan Brickley and R.V. Guha, "Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation", 2000.
- [11] A. Gali, C. Chen, K. Claypool and R. Uceda-Sosa, "From ontology to relational databases," in Conceptual Modeling for Advanced Application Domains, LNCS, vol. 3289, 2005, pp. 278–289
- [12] E. Vysniauskas, and L. Nemuraite, "Transforming ontology representation from OWL to relational database," Information Technology and Control, vol. 35A, no. 3, 2006, pp. 333-343.
- [13] I. Astrova, N. Korda and A. Kalja, Storing OWL Ontologies in SQL Relational Databases, International Journal of Electrical, Computer and System Engineering, 2007.
- [14] C. Fankam, L. Bellatreche, H. Dehainsala, Y. Ait-Ameur and G. Pierra, "SISRO : Conception de bases de données à partir d'ontologies de domaine", TSI Volume 28, pages 1-29, 2009.
- [15] J. Trinkunas and O. Vasilecas, "A Graph Oriented Model For Ontology Transformation Into Conceptual Data Model", Information Technology and Control, 2007, Vol. 36(1A), pp. 126–132.
- [16] I. Astrova, N. Korda and A. Kalja, "Storing OWL ontologies in SQL Relational Databases," Engineering and Technology, 2007, Vol. 23, pp. 167–172.
- [17] E. Vysniauskas and L. Nemuraite, "Mapping of OWL ontology concepts to RDB schemas," in Information Technologies' 2009, Proceedings of the 15th International Conference on Information and Software Technologies; 2009. pp. 317-327.
- [18] M. Mahfoudh and W. Jaziri, "Approche de couplage de BD et d'ontologie pour l'aide à la décision sémantique : contribution pour la satisfaction des requêtes SQL et SPARQL," Techniques et Sciences Informatiques, volume 32 of Hermes, pages 863-889, 2013.
- [19] W.H. Inmon, "Building the Data Warehouse", Second Edition, New York: John Wiley & Sons, 1996.
- [20] C. Phipps, K.C. Davis, "Automating data warehouse conceptual schema design and evaluation", Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW'2002), vol. 58, pp. 23-32.
- [21] Y. Song, R. Khare, B. Dai, "Samstar: a semi-automated lexical method for generating star schemas from an entity-relationship diagram", in Proceedings of the ACM tenth international workshop on Data warehousing and OLAP, pp. 9-16, 2007.
- [22] M. N. M. Nazri, S. A. M. Noah and Z. Hamid, "Automatic data warehouse conceptual design," 2008 International Symposium on Information Technology, Kuala Lumpur, Malaysia, 2008, pp. 1-7.

- [23] R. Kimball, L. Reeves, W. Thornthwaite, M. Ross, "The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing" in , 1998, John Wiley & Sons, Inc.
- [24] J. Feki, A. Nabli, , H. Ben-Abdallah and F. Gargouri, "An Automatic Data Warehouse Conceptual Design Approach". Encyclopedia of Data Warehousing and Mining, John Wang Edition, 2008.
- [25] I. Horrocks, "Ontologies and the semantic web", Commun. ACM, vol. 51, no. 12, pp. 58-67, 2008.
- [26] Y. Lv and C. Xie, "An ontology-based approach to build conceptual data model," 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, Sichuan, 2012, pp. 807-810.
- [27] T. Podsiady-Marczykowska, T. Gambin, and R. Zawiślak, "Rule-Based Algorithm Transforming OWL Ontology Into Relational Database," in Beyond Databases, Architectures, and Structures: 10th International Conference, BDAS 2014, Ustron, Poland, May 27-30, 2014. Proceedings, S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, and D. Kostrzewa, Eds. Cham: Springer International Publishing, 2014, pp. 148–159.
- [28] V. Nebot and R. Berlanga, "Building data warehouses with semantic web data," Decis. Support Syst., vol. 52, no. 4, pp. 853–868, 2012.
- [29] D. Colazzo, I. Manolescu, A. Roatis, and A. Roati, "Warehousing RDF Graphs To cite this version : Warehousing RDF Graphs \*," Hal, 2013.
- [30] D. Colazzo, I. Manolescu, A. Roatis, and A. Roati, "Warehousing RDF Graphs To cite this version : Warehousing RDF Graphs \*," Hal, 2013.A.
- [31] B. B and A. Bagnall, "A Value-Added Approach to Design BI Applications," vol. 9263, pp. 257–269, 2015.