Ontology in knowledge management

Simona Elena Varlan, Department of Computer Science, University "Vasile Alecsandri", Bacau, Romania, svarlan@ub.ro
Iulian Marius Furdu, Department of Computer Science, University "Vasile Alecsandri", Bacau, Romania, ifurdu@ub.ro

Abstract
The paper presents the theoretical study about the use of ontology in implementing knowledge management in virtual organizations and a proposed ontology for modelling the knowledge content in an e-commerce application within the Romanian virtual space. The paper represents one part of a larger research project that intends to implement an ontology-based knowledge management system for Romanian economical knowledge within virtual organizations using Semantic Web technologies.

Keywords: Ontology, Knowledge Management, Semantic Web, OWL, Inference Rules

Introduction

Knowledge management is a real important issue in economy because more than ever we need an efficient control of knowledge on the market used in day by day decisions. In a context where knowledge is vital for an organization, knowledge management becomes an imperative. Because knowledge management presume coding, collection, integration, and dissemination of knowledge our attention is focus on how to represent knowledge in order to obtain a successful knowledge management.

Unlike the intelligent activity of the human mind, computer cannot make alone acquisition and representation of knowledge, it needs a human to conduct knowledge storage in computer's memory and decide how to represent knowledge. So far have been developed some knowledge representation models each associated with a method for structuring and encoding of knowledge and a specific data structure, but we can say that no model is perfect or the best to represent any kind of knowledge regardless the specific domain (Dragan, 2006). According to Dragan et. al., a representation model must fulfil tree major characteristics:

- To be clear and concise without ambiguities,
- To be expressive enough for any human to understand it and,
- To be independent of context in order to make logical inferences.

From this, we can conclude that a representation model it has to be close enough to natural language but in the same time close to mathematical-logical formalism for logical processing. Each representation model and each data structure has deficiencies and cannot be adapted for any type of human knowledge.

Our *purpose* is to find a suitable model to represent the knowledge base within a Romanian e-commerce organization. We consider there are many ways to represent a knowledge base, of course according to the limits of the model, respectively the specific language representation of that model.

For the knowledge on the Web, there is the *semantic markup model*. This model implies representation in structured form and uniquely semantic markup of concepts and entities within Web documents. For this model, we use, naturally, Semantic Web representation languages.

Ontology it is a semantic markup representation model for knowledge on Web and the known representing languages used to describe ontologies are OWL (Web Ontology Language), OIL (Ontology Inference Layer) and DAML-OIL (DAML - DARPA Agent Markup Language) but the most popular is OWL. In recent years ontologies are used increasingly more, gaining popularity among web developers because of the simple but power enough way of representing knowledge as a set of concepts within a specific domain. The question formulated in the architectural phase was why

use Ontology. Better said, do *we need Ontology for* an *e-commerce* application and why not stick only to very well known and well-used RDBMS?

In order to answer to our questions first we will analyze what is Ontology. Then we will try to highlight the differences between a relational databases and ontology and the role of ontology in knowledge management in virtual organizations. We will end with a proposed ontology for a Romanian e-commerce system.

About Ontology

Among the first definitions used for defining the Ontology are that of (Gruber, 1993) "an ontology is an explicit specification of a conceptualization" and (Borst et. al, 1997) "Ontologies are defined as a formal specification of a shared conceptualization". Later, (Staab and Studer, 2004) explained what these terms mean "'conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared refers to the notion that an ontology captures consensual knowledge, that is, it is not primitive to some individual, but accepted by a group".

Better understand the Ontology and to be able to answer to our questions we conclude that Ontology, in a general sense, is a form of knowledge representation about the world and, in the narrow sense, is a formal representation of a set of concepts and relationships between these concepts.

Ontologies where created over the structure of RDF that only provides a way of describing Web resources by creating metadata. Ontologies provide further description of the structure of concepts and relationships between concepts, statements of classes and subclasses, types and cardinality for properties of concepts. That means ontologies are based on specific Description Logic (DL) concepts.

TBox and ABox, which incorporates elements of terminology, vocabulary domain modeling and conceptual assertions or individual axioms, represent ontology. Vocabulary consists of concepts and roles as binary relations between modeled concepts. DL is based on semantic description model therefore statements contained in the TBox and ABox can be identified by type formulas of first order predicate logic and inference rules (Baader and Nutt, 2002).

We propose a visual representation of the structure of Ontology on web in Figure 1 to capture the ideas subtitle above.

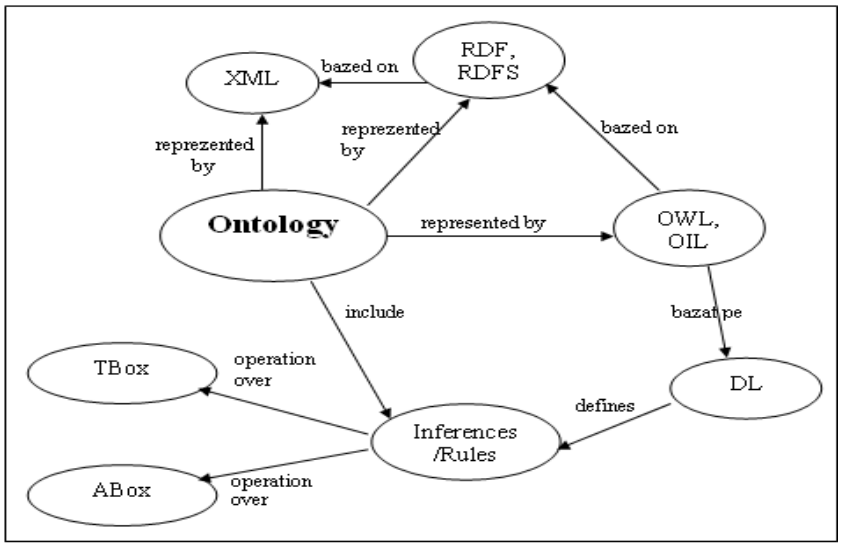

Fig 1: Structure of Ontology on Web

Way use Ontology – Ontology vs. Relational Databases

We saw what ontology is but we still need to answer the question why use ontology in steed of relational databases. To compare them we have to see what a database is. A database (BD) is a collection of data in interdependence, along with descriptions of the data and the relationships between them; the collection is automated, shared, rigorously defined (formalized) and controlled at a central level (Fotache, 1996). F. Codd introduced the relational database model, according to which, data can be organized as a collection of related tables.

If it is to compare the two definitions, of Ontology and database, we observe at a first glance, that ontology is our conceptualization (understanding) of the world, and databases is also a description of the world but with strict rules and constraints.

Related to e-commerce information (data and relations between them to describe), and if we want to create a basic and simple relational database with main information, will have three tables corresponding to Categories, Products and Producers, all three connected through table Products. The table Categories will contain a single level of categories, and if we want to create a hierarchy of categories then we will need to expand the database with other tables. The corresponding Ontology in stead would have:

- axiome of the form:

Class: Laptop
subClassOf: Computer

or

ObjectProperty: hasComponent
InverseOf: isComponentOf

- the facts of the form:

Individual: AB123 (the name is fictional to avoid publicity problems)
Type: Laptop
Facts: hasComponent HDD
Individual: Hard_Drive

The axioms are similar to database schema and the facts are similar to the data within the database. The differences are, however, important. (Horrocks, 200) highlighted these differences and we will test it on our example, to see if they apply also for e-commerce ontology.

- In a BD, the missing information is treated as false, when in ontology the missing information is treated as unknown. The database is closed world assumption (CWA) and Ontology is open world assumption (OWA). For example, if we query if Hard_Drive is component of Laptop, for database case we will receive a negative answer (if there is no relation between Laptop and Hard_Drive), and in Ontology case, the answer is "don't know" (OWA) because we did not say anything about that Hard_Drive not being a Laptop component.
- In a database, each instance has only one name, unique name assumption (UNA) and within Ontology, instances can have more than one name (no UNA). If we add to the individual AB123 the following fact: hasComponent Hard_Drive, and do the same thing for databases (connect product Hard_Drive to AB123), if we interrogate the DB how many computer components have AB123, the answer will be 2 (UNA). If we interrogate the Ontology the answer will be at least one, because HDD and Hard_Drive can be two names for the same product, which is kind of real in our world perception.
- Database schema imposes different constraints, Ontology axioms behave like implications (inference rules). If we have the following statements:

Individual: BC456
Individual: Hard_Drive
Type: LaptopComponents
Fact: isComponentOf BC456

Class:LaptopComponents
SubClassOf: isComponentOf only Laptop

for a database we are constraint to make relationships between BC456 instance and category Laptop, but for Ontology the fact infer that BC456 is a Laptop because only a Laptop has LaptopComponents. The conclusion is that for a database, to make queries, data must satisfy the corresponding constraints, and for Ontology, axiomos plays a very important role, because they can „hide" implicit facts that are descovered through implicit logical reasoning.

- Another difference between the two models regards query complexity. For databases, worst-case complexity is low (logspasce) that is exactly the size of databases. The complexity is limited because there is no missing data. This is the reason we can implement databases very efficiently. Ontology may have very high worst-case complexity (NP), the upper bound (ontology size) is an open problem because is considered missing information not yet declared. We have to recognize that a major problem in implementing Ontology is **scalability**.

There is, however, a significant difference between queries over a database and searches over knowledge base ontology type. We conducted a previous research (Varlan, 2011) about query complexity between databases and ontology. The study compared complexity of querying of an e-commerce web site that use an ontology to define concepts and their relations in the domain, to the complexity of querying of a conventional e-commerce web site that use the relational database to store the content information. For our research, we used the method used by Abrahams (2006) with measures for query expression complexity elaborated by Vardi (1982), formulas of first-order predicate formulated by Baader (2002) and terminology and studies published by Horrocks and Tessaris (2000) to translate conjunctive queries into ontology concepts. We succeeded to demonstrate that the query over an ontology has a lower expression complexity than the query over a database, and also a reduced number of queries and a better processing of knowledge through semantics.

For the purpose of knowledge sharing and reuse, we found in the literature five reasons to use Ontology. The reasons are, according to (Noy and McGuinness, 2003) the followings:

1. *Sharing common understanding* of the structure of information among people or software agents (Gruber, 1993) from different e-commerce web sites, then computer agents can extract and aggregate information from these sites.
2. *Enabling reuse of domain knowledge* means, simply reuse a specific ontology for this domain by others for their domains. To develop a large ontology we can integrate several existing ones or we ca reuse a general ontology and extended it to describe our domain of interest.
3. Making *explicit domain assumptions* are useful for new users who must learn what terms in the domain mean.
4. *Separating the domain knowledge from the operational knowledge* means the possibility to describe a task of configuring a product from its components according to a required specification and implement a program to do this configuration (McGuinness and Wright 1998). For example, we can develop an ontology of PC-Components and apply the algorithm to configure mate-to-order PCs.
5. *Analyzing domain knowledge* is possible once a declarative specification of the terms is available. Formal analysis of terms is extremely valuable when both attempting to reuse existing ontologies and extending them (McGuinness et al., 2000).

Another reason to use Ontology that we succeeded to identify is the use of RDFa tags to add semantics to e-commerce web sites content. Using these tags, we can assure higher visibility on the Internet, which is important if not vital for virtual organizations with an e-commerce activity. It is a simple, effective solution, since the main search engines used already recognizes these RDFa tags, and sustainable in time because RDFa standard is validated and recommended by W3C. To use these tags we need a corresponding ontology to describe concepts and their properties. We conducted a study about the effective use of RDFa tags inside existing e-commerce web sites and their utility and we find that the BestBuy Company started to use RDFa tags together with GoodRelations (Hepp, Martin, 2008) ontology in 2008. Conform to their declaration the number of web page access per

click increased with 30%, which in today's business concurrency is a major improvement. In the present, BesBuy, doesn't use anymore the GoodRelations ontology.

Analysing all the aspects presented in this section we arrived at the conclusion that ontologies represent a powerful tool to add semantics to web content, they are easy to develop, maintain, update, and integrate with other ontologies or semantic web vocabularies, easy to query with support for incomplete information and possibility of share and reuse of knowledge. The main disadvantage that the use of ontology would have is scalability problems.

At the and of these analyses we realised that the advantages are promising in the use of ontology so, in conclusion we decided to develop our Ontology dedicated to PC components in stead of using a relational databases.

The proposed ontology for an E-commerce Knowledge Management System

Our proposed ontology was created upon a survey on the most visited e-commerce Romanian sites, in order to conform to the information contained on their sites. Ontology was developed using the Protégé editor and then we applied some inference rules to verify our tests presented in the above section. All the tests where successful. We present the ontology and an inference rule, separately, on Figure 2 and respectively on Figure 3.

According to study we established the creation of six basic classes: Category (*Categorie*), Products (*Produse*), Manufacturers (*Producator*), Country of Production (*Tara Producatoare*), the Liquidation (*Lichidari Stoc*) and Promotions (Promotii). Class Category is one that contains other subclasses according to general classifications of IT sites. The words written in paranthesis represent the corresponding Romanian word for every concept. We create the ontology using Romanian language for better promote and a change to integration among Romanian e-commerce web sites.

Future work includes designing and implementing the KM in the e-commerce application. It will represent the final stage of the project.

Fig 2: The PC-Components Ontology

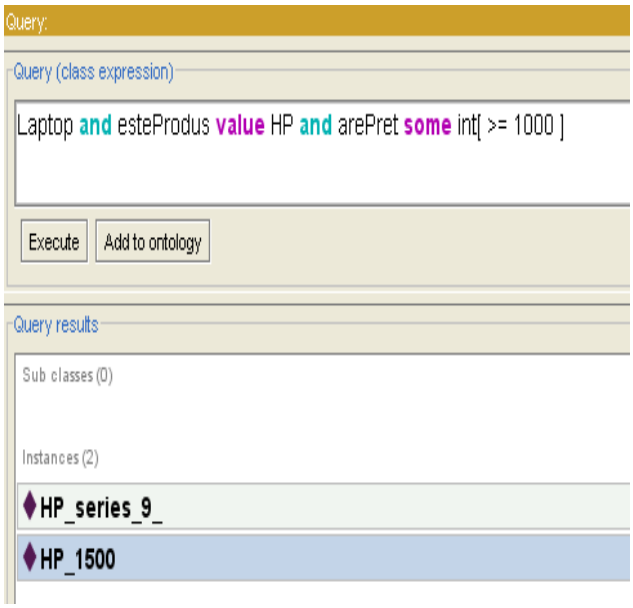

Fig 3: An inference rule over Ontology

Conclusions

During our analyses conducted in this paper some conclusions where drowned:

- ontologies are analogous to databases, axioms represents the schema of a database and facts the data within the database;
- there are some important differences in semantics between theme, a databases is UNA, CWA and must respect constraints, while ontology is OWA and axioms behaves like inference rules;
- ontology is NP – the size is indeterminable because of the missing information considered not yet declared (worst-case scenario);
- there can be scalability problems;
- in a concrete case, query expression complexity over ontology is lower than the one over databases;
- Internet visibility can grow for a web site through integration of ontologies and corresponding RDFa tags.

As a final conclusion, the question from the beginning of the paper: "why use Ontology?" was successfully transformed in "why NOT use Ontology?".

Acknowledgment

This work was supported by CNCSIS-UEFISCSU, project number PN II-RU code 188/2010.

References

[1] Abrahams, B. (2006) Tourism Information Systems Integration and Utilization Within The Semantic Web, PhD Thesis, Victoria University, School of Information Systems, Faculty of Business and Law, Australia

[2] Baader, F., Nutt, W., (2002), 'Basic Description Logics', [Online], [Retrieved 30 Octomber], http://www.inf.unibz.it/~franconi/dl/course/dlhb/dlhb-02.pdf

http://www.inf.unibz.it/~franconi/dl/course/dlhb/dlhb-02.pdf, p64

[3] Borst, P., Akkermans, H., and Top, J. (1997). Engineering ontologies. The International Journal of Human Computer Studies. In Special Issue: Using explicit ontologies in knowledge-based system development, Vol. 2/3 pp.365-406., 46(2-3):365–406.

[4] Dragan, G., Dragan, D., Vladan, D., (2006) „Model Driven Engineering and Ontology Development", Springer, 2006, p. 5

[5] Fotache, M. (1996), Baze de date relaţionale. Organizare şi interogare ISBN 973-37-0248-X, Junimea, Iasi, Romania

[6] Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5:199–220.

[7] McGuinness, D.L. , Wright, J. (1998). Conceptual Modeling forConfiguration: A Description Logic-based Approach.Artificial Intelligence for Engineering Design,Analysis, and Manufacturing - special issue on Configuration

[8] McGuinness, D.L., Fikes, R., Rice, J., Wilder, S. (2000). AnEnvironment for Merging and Testing Large Ontologies.*Principles of Knowledge Representation andReasoning: Proceedings of the Seventh International Conference (KR2000)*. A. G. Cohn,F. Giunchiglia and B. Selman, editors. San Francisco, CA, Morgan Kaufmann Publishers.

[9] Noy and McGuinness, (2003), Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University

[10] Hepp, Martin, The developer of GoodRealtions Ontology, (2008), [online] http://www.heppnetz.de/projects/goodrelations/

[11] Horrocks, I.,Ontologies and Databases, Information Systems Group, Oxford University Computing Laboratory, note de curs., p.2

[12] Horrocks, I., Tessaris, S., (2000), A conjunctive query language for description logic ABoxes, Conference on Artificial Intelligence AAAAI/IAAI, 399-404
[13] Staab, S. and Studer, R. (2004), Handbook on ontologies. International handbooks on information systems. Springer-Verlag.
[14] Vardi, M., (1982), 'The complexity of relational queries', *ACM SIGACT*, vol. 143(issue no.2), pp. 137–146
[15] Varlan, Simona-Elena, (2011), Query Evaluation of Ontology Data Model in E-commerce Organizations, 17th IBIMA Conference, Milan, Italy