METHODOLOGIES FOR CREATING ONTOLOGIES

Marcus Rogério de Oliveira[1]

ABSTRACT

Recording, transmitting and consulting the knowledge in computers with application of inferences and reasoning made by computerized agents requires knowledge structured in formals constructions and standardized constraints called Ontologies. The data representing the knowledge formalized in Ontologies must be organized in data structures with some software requirements. Many researchers have proposed and designed concepts, applications and tools for Ontologies with different requirements. This paper discusses a historical approach on ontology, its philosophical aspects and principles and addresses their main methodologies for creation.

KEYWORDS: Ontologies. Philosophy. Knowledge. Methodology.

INTRODUCTION

The knowledge must to be recorded, transmitted and consulted. The application of inferences and reasoning over the knowledge is the base for the evolution of all the Sciences. Because of this, the knowledge must be structured in formals constructions with standardized constraints. The Ontologies have this goal.

Since Francis Bacon and Gottfried Leibniz, the intent of many philosophers and thinkers were to provide a language to disseminate ideas and to communicate without the natural ambiguities of the human language. As a consequent, the term Ontology was used for the first time in 1613, in an independent manner, by two Philosophers: Rudolf Göckel with his "Lexicon philosoficum" and Jacob Lorhard with his "Theatrum philosophicum" and, was definitively added to an English dictionary in 1721 (Bailey´s Dictionary) defined as "An account of being in the Abstract" (SMITH, *et al*., 2001).

Today, the big deal of many researches, besides recording, transmitting and consulting the knowledge with the use of computers, is to apply inferences and reasoning on it with computerized agents. So, the history comes back again with Ontologies and many Computer Science researches have obtained yours inspirations in the late years of the philosophy and have tried to organize the knowledge using Ontologies in computers.

Contributing to the convergence from Philosophy to Computer Science, Thomas Gruber, a research from Intelligence Artificial area, in 1993, defined ontology as "an explicit specification of a conceptualization" (GRUBER, 1993) and provided a vast explanation about it.

Parallel to Gruber, Willian J. Clansey, a Knowledge Engineering research, in 1993, said that the knowledge is more appropriately represented by models and the models are not the knowledge by itself (CLANSEY, 1993).

Ontologies are object of study of many areas. Nicola Guarino, in 1998, enumerated the following as areas with specific role for ontologies: Artificial Intelligence, Computational

[1]Professor Pleno I da Fatec-TQ (cursos de Tecnologia em Sistemas para Internet e Tecnologia em Desenvolvimento de Sistemas). Doutor em Biotecnologia pela UFSCar. Email: marcusrogerio@gmail.com

Linguistics, Database Theory and knowledge Engineering. As research field, the Ontologies, still according Guarino, are present in knowledge representation, qualitative modelling, language engineering, database design, information modelling, information integration, object-oriented analysis, information retrieval and extraction, knowledge management and organization and agent-based systems design. Concern to its use as applications, Ontologies have been applied in enterprise integration, natural language translation, medicine, mechanical, engineering, standardization of product knowledge, electronic commerce, geographic information systems, legal information systems, biological information systems (GUARINO, 1998).

For all areas and applications that use ontologies in a computational environment, there is a common and fundamental feature inherent to the persistence aspect of the knowledge based on ontology: the need for methodology for ontology construction that must to respond adequately to the requisites of semantic and knowledge representation.

1. ONTOLOGY'S PRINCIPLES

Present in many areas, as saw in the previous section, the term Ontology have different meanings. But its meanings tend to remain a common principle based on knowledge sharing. Guarino and Giaretta present the following interpretations for Ontology that are specific to its application area (GUARINO, *et al*., 1995):

a) Ontology as a philosophical discipline: is the most general interpretation for the term and its meaning is related to the Aristotle concept in with ontology is the Science of the being as such;
b) Ontology as an informal conceptual system: by this interpretation, Ontology is a conceptual system used to underly a specific knowledge base. Here, ontology has the intend of to be "not formal" in respect to the semantic level;
c) Ontology as a formal semantic account: here, ontology is assumed to be the inspiration to a knowledge base and is expressed in terms of structures with the intend of to be formal at the semantic level;
d) Ontology as a specification of a conceptualization: this interpretation was proposed recently by Gruber as shown in the previous section and will be discussed in detail later;
e) Ontology as a representation of a conceptual system via a logical theory: here, ontology is considered to be a collection of assertions about something. In Logic, this is called Theory. So, Ontology is nothing else than a Logical Theory;
f) Ontology as a vocabulary used by a Logical Theory: in this interpretation, the term Ontology is considered to be just a vocabulary used by Local Theory;
g) Ontology as a meta-level specification of a Logical Theory: under this interpretation, Ontology is considered a specification of the architectural components used in a specific domain theory.

The interpretation "d" is the most relevant to this paper. The other ones can be seen in details in (GUARINO, *et al*., 1995). Enumerated by Guarino and Giaretta, that interpretation is similar to the Gruber´s definition: Ontology is an explicit specification of a conceptualization (GUARINO, 1998). Fundamental to the validation of this definition, is the terms "explicit specification" and "conceptualization".

In a simplified way, a conceptualization is an abstract and simplified understanding of the world or of a domain of interest. According Genesereth and Nilsson (GENESERETH, *et al*.,

1987), a conceptualization is the objects, concepts and others entities that are assumed to exist in some area of interest and the relationships among them. The shared knowledge queried by some kind of agent must be committed to some conceptualization.

The term "explicit specification" represents the form in which the conceptualization becomes shared. To be explicit is to be precisely and clearly expressed or readily observable; leaving nothing obscure or implied. Specification is a detailed description expressed in some language or vocabulary.

The definition made by Gruber and enumerated by Guarino and Giaretta, is very compatible with the premise utilized in this paper: "Ontology describes, characterizes, distinguishes, identifies and categorizes the entities of the world (like physical objects, peoples, events, places, documents, cells,…), using concepts, properties, qualities, states, roles, constraints and establishes the relationships between entities, categories and instances; all expressed by a formal language". The computational purpose adopted in this paper is related to the persistent aspect of the Ontology and so, the definition given by Gruber is extended to attempt that purpose: "Ontology is an explicit specification of a conceptualization, persistent in computers and available for queries for some kind of agent". This extended definition is the one in which the Database Group at Federal University of São Carlos have considered in their works with ontologies.

In this aspect, Ontology can be considered as an application of the database system in with, not only the data is consulted, but the meaning of the data is considered during the consult. The schemas describe the data and the relations between the data.

2. METHODOLOGIES FOR ONTOLOGY CONSTRUCTION

There are several methodologies described in the literature for Design and Construction of ontologies. Some methodologies were built for some specific purpose, inside a specific project (like TOVE and CyC) and become widely considered in articles related to ontology construction. Others were specifically created for design purposes, but for some specific knowledge area, like Methondology that was created for the Chemical domain.

All the methodologies, project or area specific driven, are developed over a motivation derived from the definitions and standardizations considered as a life cycle. These definitions and standardizations include since requirement definitions until tests and maintenance of the finished product and obey techniques that drive their development.

2.1. DESIGN CRITERIA FOR ONTOLOGY CONSTRUCTION

One of the most important aspects related to ontologies is the design. The necessary phases to the creation of ontologies can be considered as an Ontology Life Cycle. As a life cycle, these phases should be outlined by some methodology. Uschold and King (USCHOLD, *et al.*, 1995) say that a comprehensive methodology to create or develop ontology includes the stages:

a) Purpose identification: is the definition of the ontology finality; is about why the ontology is necessary and what its users;
b) Ontology Construction: includes ontology capture, ontology coding and ontology integration. The ontology capture is the identification of the domain of interest or the scope definition; the coding is the representation of the conceptualization in some

formal language; the ontology integration is to determine the relationship between the ontology being created and the other already created;

c) Evaluation: is about the quality of the ontology, derived from a technical judgment of it, its software and its documentation;

In 2003, Corcho *et al* (CORCHO, *et al*., 2003) resumed the considerations about the several questions involving ontologies construction in three main aspects: the methods and methodologies to use; the tools that give support to the ontology development process and the language that supports adequately the representation of the conceptualization.

Gruber (GRUBER, 1993), in a more general way, enumerates a set of design criteria to guide the design decisions for building formal ontologies. Gruber want say that, before to think in the life cycle, the ontology designer must to have in mind the following characteristics desirable in ontology:

a) Clarity: all the defined terms should express their meaning in an unambiguous way;
b) Coherence: all the inferences should be in harmony and agreement with the definitions;
c) Extendibility: the designer must to consider that new facts may arise and the ontology must to be able to continue being applicable;
d) Minimal encoding bias: encoding bias is the phenomenon associated to the problem in witch different agents are implemented in different representation system. The conceptualization must be represented without depending on a particular symbol encoding;
e) Minimal ontological commitments: the ontology being designed must leave all its parts independent to extend, to specialize and instantiate. The terms defined by the ontology must be minimal, that is, they must be only the necessary to communicate the knowledge.

3. METHODOLOGIES FOR ONTOLOGY CONSTRUCTION

3.1. ENTERPRISE MODEL APPROACH

Uschold and King (USCHOLD, *et al*., 1995), already cited in these paper, proposed a methodology based in their experience of developing the Enterprise Ontology consisting of following four approach:

a) Purpose identification: is the definition of the level of formality in which the ontology should be described;
b) Scope identification: is the specification of the problems that the ontology should solve, that is, the motivating scenarios;
c) Formalization: is the use of some formal language to express the axioms and definition of the ontology;
d) Formal evaluation: is the application of tests and analyses with the intent of to verify the competency questions.

3.2. TOVE

The Tove methodology (GRUNINGER, *et al*., 1994) was created for specific purposes and based on experiences gained from the development of the Toronto Virtual Enterprise project. This methodology has the following approach:
Motivating scenarios: are the problems that provide motives for the ontology construction;

a) Ontology requirements: are the requisites that the ontology must meet;
b) Specification of the terminology: is the formal description of the objects, relations and attributes of the ontology;
c) Formal competency questions: the defined terminology is used to formalize the requirements of the ontology;
d) Axiom specification: constructed using first-order logic, the axioms must be necessary and sufficient to denote or express the competency questions;
e) Completeness theorem: is the definition of the conditions in which the competency questions are accomplished.

3.3. METHONDOLOGY

Created by Gomes-Perez and Vicente (GOMES-PEREZ, *et al*., 1996), the Methondology has its principle based in the following activities related to the construction of ontologies:

a) Specification: is the definition of the purpose and scope of the ontology expressed in natural language;
b) Knowledge acquisition: it is the taking of knowledge from experts, texts and other sources using any elicitation method;
c) Conceptualization: is the definition of the concepts, instances, relations and properties using a informal representation;
d) Integration: is the effort to provide uniformity with the existing ontologies;
e) Implementation: is the formal representation of the ontology expressed in some formal language;
f) Evaluation: is the application of analyses and tests to verify inconsistencies, incompleteness and redundancies;
g) Documentation: is the elaboration of documents for registering the actions and the results gained.

3.4. IDEF5

The objective of this methodology is to support not only the creation, but the modification and the maintenance of ontologies. The guidelines of the IDEF5 methodology are the following:

a) Organization and scope definition: is the identification of the purpose and context of the ontology including objectives and requirements;
b) Data collection: is the acquisition of the data needed for the development of the ontology. The data may be obtained by methods like interviews with experts and text analysis;
c) Data analysis: is the definition of which elements of the data collection is necessary to be present in the ontology;
d) Initial ontology development: is a prototype of the ontology in which the preliminary validations are made;
e) Refinement and validation: is the application of tests with real data, that is, the instantiated data.

CONCLUSION

This paper presented a historical approach on Ontologies and addressed its philosophical aspects. It described the principles and the main methodologies for creation and representation of knowledge. After that, in a conclusive way, it can be observed that the

creation of Ontologies is still an open problem. Although there are several methods of creating, there is no a single best solution neither a preferred approach. The choice for a methodology depends on the purpose of Ontology, the application in which the ontology will be used for and the real world aspects that the new ontology will be able to represent. Finally, the reader certainly had obtained a empirical evidence showing that the approach to representing and disseminating the knowledge is a effective role of the Ontologies.

RESUMO

A gravação, a transmissão e as consultas ao conhecimento armazenado em computadores com a aplicação de inferências e raciocínios feitos por agentes computadorizados requerem que o conhecimento esteja estruturado em construções formais chamadas Ontologias. Os dados que representam o conhecimento formalizado em Ontologias devem ser organizados em estruturas de dados que obedecem a um conjunto de requisitos. Muitos pesquisadores têm proposto e concebido conceitos, aplicações e ferramentas de ontologias com diferentes requisitos. Este artigo discute uma abordagem histórica sobre ontologias, seus princípios, seus aspectos filosóficos e ainda apresenta as principais metodologias para criação.

PALAVRAS-CHAVE*: Ontologias. Filosofia. Conhecimento. Metodologia*.

REFERENCES

CLANSEY, W. J. (1993). **The Knowledge Level Reinterpreted: Modeling Socio-Technical Systems**. Special issue of International Journal of Intelligent Systems .

CORCHO, O., FERNANDEZ-LOPEZ, M., & GOMES-PEREZ, A. (2003). **Methodologies, tools and languages for building ontologies: where is their meeting point?** Data & Knowledge Engineering .

GENESERETH, M. R., & NILSSON, N. J. (1987). **Logical Foundations of Artificial Intelligence**. San Mateo, CA: Morgan Kaufmann Publishers.

GOMES-PEREZ, A., FERNANDEZ, M., & VICENTE, A. (1996). **Towards a Method to Conceptualize Domain Ontologies**. ECAI-96 Workshop on Ontological Engineering .

GRUBER, T. R. (1993). **A translation approach to portable ontologies**. Knowledge Acquisition , pp. 5(2):199-220.

GRUBER, T. R. (1994). **Toward Principles for the Design of Ontologies Used for Knowledge Sharing**. International Workshop on Formal Ontology .

GRUNINGER, M., & FOX, M. (1994). **The Design and Evaluation of Ontologies for Enterprise Engineering**. Workshop on Implemented Ontologies, European Conference on Artificial Intelligence .

GUARINO, N. (1998). **Formal Ontology in Information Systems**. Proceedings of FOIS'98 , pp. pp. 3-15.

GUARINO, N., & GIARETTA, P. (1995). **Ontologies and knowledge bases: towards a terminological clarification**. Proceedings, Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing .

SMITH, B., & WELTY, C. (2001). Ontology: **Towards a new Synthesis**. Proceedings of the international conference on Formal Ontology in Information Systems .

USCHOLD, M., & KING, M. (1995). **Towards a Methodology for Building Ontologies**. IJCAI Workshop On Basic Ontological Issues In Knowledge Sharing, .