Chapter 1.4
Ontologies and Controlled Vocabulary:
Comparison of Building Methodologies

Daniela Lucas da Silva
Universidade Federal do Espírito Santo, Brazil

Renato Rocha Souza
Fundação Getúlio Vargas, Brazil

Maurício Barcellos Almeida
Universidade Federal de Minas Gerais, Brazil

ABSTRACT

This chapter presents an analytical study about methodology and methods to build ontologies and controlled vocabularies, compiled by the analysis of a literature about methodologies for building ontologies and controlled vocabularies and the international standards for software engineering. Through theoretical and empirical research it was possible to build a comparative overview which can help as a support in the defining of methodological patterns for building ontologies, using theories from the computer science and information science.

INTRODUCTION

The organization of information has increasingly became a crucial process as the volume of information available has exponentially increased, sometimes resulting in the chaotic information collections. In this sense, a lot of research has been made (Lancaster, 1986; Gruber, 1993; Berners-Lee, Hendler & Lassila, 2001) aiming at the construction of mechanisms for the organization of information with the sole objective of improving the efficacy of the information retrieval systems.

This fact contributes to the attention paid to the ontologies, which are originated in the theoretical field of Philosophy (Corazzon, 2008) and are researched and developed as a tool for the representation of knowledge in Computer and Information Sciences. For the Information Science, the ontologies are of interest because of their potential to organize and represent information

DOI: 10.4018/978-1-61350-456-7.ch1.4

(Vickery, 1997). According to Almeida & Barbosa (2009), the ontologies can improve the information retrieval processes as they organize the content of the data sources in a specific domain.

Gruber (1993) presents a definition which is widely accepted by the ontology community: "an explicit specification of a conceptualization" (Gruber, 1993, p. 2), where "explicit specification" would be related to concepts, properties and explicitly defined axioms; and "conceptualization" regards an abstract pattern of any real world phenomenon. As components of ontology (Gómez-Pérez, Fernández, & Vicente, 1996; Gruber, 1993), there are: a) conceptual classes which organize the concepts of a domain in a taxonomy; b) class attributes, which are relevant properties of the concept; c) instances, which are used to represent objects specific to a context; d) attributes of instances, which are relevant properties used to describe the instances of a concept; e) relationships between classes, which represent the type of interaction between the concepts of a domain; f) invariants, which always have the same values and are generally used in standards or formulations to infer knowledge in ontology; g) terms, which design the concepts of a domain; h) formal axioms, which limit the interpretation and usage of the concepts involved in the ontology; and i) standards, which determine conditions to the domain besides inferring values for attributes.

This chapter proposes an analytical study on methodologies and methods used for ontology building more commonly found in the literature and methodologies and standards designed to build controlled vocabulary, in order to delineate a comparative overview about the construction of such instruments. Such panorama can contribute to the definition of methodological standards for the construction of ontologies through the integration of theoretical and methodological principles from the Information and Computer Sciences as well as from contributions of known methodologies and methods employed to build ontologies and controlled vocabularies.

In order to accomplish the task proposed, the methodological steps taken in the research were the following: i) the identification and selection of documents referring to the subject methodologies for ontology building; ii) the identification and selection of methodologies for ontology building discussed in them; iii) the identification and selection of standards for the construction of controlled vocabulary; iv) the definition of content analysis categories in order to collect data relevant to the research; and v) the comparative analysis of the methodologies, methods and standards.

Background

Within the domain of ontologies development, the approaches for their building are, invariably, specific and limited. One problem, from the methodological point of view, is that there are neither patterns nor wide accepted methodologies for its building (Fernández et al., 1999; Uschold, & Gruninger, 1996). Despite the fact that great quantities of ontologies have already been developed by different communities – chemistry (Gómez-Pérez, Fernández & Vicente, 1996) and in business process modelling (Gruninger & Fox, 1995), just to give a few examples – under different approaches and using different methods and techniques, there is no consensus about a "gold standard" for the development process (Fernández, Gómez-Pérez & Juristo, 1997). The consequence is the absence of rigorous standardized techniques. Besides that, it is verified the lack of a systematic explanation on how and where the theoretical approaches will be used within their elaboration process.

Information Science researchers (Vickery, 1997; Soergel, 1997; Soergel, 1999; Gilchrist, 2003) often present similarities in their ideas about controlled vocabulary used in Library Science, like the thesauri, taxonomies and tools used in Artificial Intelligence, such as ontologies. The similarities lay especially in the way the structures of these tools are devised, which demands the organization of concepts into processes that include