Semantic Technologies for Business and Information Systems Engineering:

Concepts and Applications

Stefan Smolnik
EBS Business School, Germany

Frank Teuteberg
University of Osnabrueck, Germany

Oliver Thomas
University of Osnabrueck, Germany

BUSINESS SCIENCE Reference

Managing Director: Lindsay Johnston
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Michael Killian
Acquisitions Editor: Erika Carter
Typesetters: Jennifer Romanchak, Adrienne Freeland
Print Coordinator: Jamie Snavely
Cover Design: Nick Newcomer

Published in the United States of America by
Business Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Semantic technologies for business and information systems engineering: concepts and applications / Stefan Smolnik, Frank Teuteberg and Oliver Thomas, editors.
p. cm.
Includes bibliographical references and index.
Summary: "This book investigates the application of semantic technologies to business and information systems engineering"-- Provided by publisher.
ISBN 978-1-60960-126-3 (hardcover) -- ISBN 978-1-60960-128-7 (ebook) 1. Semantic computing. 2. Semantic networks (Information theory) 3. Management information systems. I. Smolnik, Stefan, 1970- II. Teuteberg, Frank, 1970- III. Thomas, Oliver, 1971- IV. Title.
QA76.5913.S47 2012
005.7--dc23
2011031983

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 1
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: 10.4018/978-1-60960-126-3.ch001

(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 the categorization and classification of concepts, the definition of the relationships between these concepts and the treatment of the terminology employed in the concepts and structure relations.

Soergel (1999) highlights the lack on the communication among the many fields which work with conceptual structures, like ontologies. Silva (2008) corroborated this vision showing how many ontology engineering methodologies are available whilst no true standard is accepted. This may explain why, according to the words of Gómez-Pérez et al (2004), methodologies for ontology engineering are never used in practice

Literature Analysis

The first methodological step in the research was the identification and selection of the available literature on methodologies for ontology building and engineering. We have chosen to search in the knowledge bases that were available, as Citeseer[1], ACM Portal[2] and Google Scholar[3], being all of them references in the Information Science and Computer Science Fields. We have sought also the main references from these articles, in order to enhance the results.

After selecting the knowledge bases, we have performed the following steps: i) selection of articles with the keywords "ontology" or "ontologies" associated to "building", "methodology" and/or "engineering"; ii) selection of the most commonly cited sources from the previously found articles, adding those to the selected documents; iii) identification of the methodologies for building ontologies in the selected documents; iv) frequency analysis of the citations to the methodologies, to choose the most preeminent; and v) definition of the set of methodologies to be used in the research.

The process of analysis and interpretation of retrieved documents took place through the usage of bibliometrics and content analysis (Bardin, 1977) techniques, which enabled the selection of the relevant documents. Content analysis includes a set of communication analysis techniques which aim at obtaining indicators (quantitative or not) that allow the inference of knowledge present in the messages.

According to the criteria, we have obtained 25 relevant documents from the survey. When we applied citation analysis, we found that the TOVE was the most cited methodology (68%), followed by the Enterprise Ontology (60%), Methontology (56%), Cyc (36%), Kactus (24%), Sensus and On-To-Knowledge (20% each), method 101 (12%) and KBSI IDEF5 (4%). Then we have analysed which of these methodologies aimed to build domain ontologies, discarding the On-To-Knowledge and KBSI IDEF5.

The methodologies and methods used for the construction of ontologies chosen for the comparative analysis were: a) Gruninger and Fox's TOVE methodology; b) Uschold and King's Enterprise Methodology; c) the Methontology methodology; d) the Cyc method; e) the Kactus method; f) the Sensus method; and g) the 101 method. It was assumed that, by analyzing the most discussed methodologies in the literature, it would be possible to achieve a reasonable result regarding a comparative study of methodologies used in ontology engineering.

Throughout the activity to identify methodologies for the building of controlled vocabulary, the existence of standards and manuals created by nationally and internationally acknowledged institutions was also verified. As a main reference, this research used one standard, which is justified by the following criteria: a) it is the most current standard (from 2005), accepted and consolidated in the Information Science community; and b) it is the standard which presents an interdisciplinary approach of the theories derived from Information Science and from Terminology – The Theory of Faceted Classification; the Concept Theory; and Terminology Theory – in the building of controlled vocabulary. This standard, named ANSI/NISO Z39.19-2005 (ANSI, 2005), was built based on several American and international standards concerning the creation of thesauri, including the ISO 2788. Besides the standard, the manual available in the Librarianship, Information and Information Technology site - BITI (Campos, Gomes & Motta, 2004), about thesauri elaboration, was used. The manual, even though focusing on a specific type of controlled vocabulary, proved to be coherent as regards the recommendations given in standard ANSI/NISO Z39.19-2005.

During the preliminary content analysis carried out in the materials regarding the ontology building and controlled vocabulary development, we noticed a similarity between some stages of development of these instruments and other similarities derived from the development of software process. Some similarities were identified especially throughout domain analysis activities and in the technical approaches devised for the creation of conceptual patterns. Therefore it was decided we would use, under this study perspective, the internationally accepted standard for the development of software (IEEE1074-1997), to define the categories of analysis. This choice is justified by the fact that this standard describes a structured and methodical way for product development (Pressman, 2004), and because it derives from Software Engineering, a discipline considered mature in the sense it possesses widely accepted methodologies (Fernández et al., 1999, p.1). As ontologies are considered components of software products (Fernández et al., 1999), the utilization of the standard as instrument for data qualitative analysis was deemed pertinent.

THE CATEGORIES FOR ANALYSIS

The content analysis categories of the empirical material were defined from principles elucidated by Bardin (1977), who advocates the use of categories for procedures of qualitative analysis. According to the author, the choice of categories may involve several criteria: i) semantic; ii) syntactic (verbs, adjectives, pronouns); iii) lexical (group according to the meaning of the words, group synonyms and antonyms); and iv) expressions and phrases. In this study, the chosen criterion for the categories was the semantic one, that is, it was carried out according to the standard IEEE1074 (1997) and to the literature from the ontologies field. The categories were then adapted to the processes extracted from the standard and from the characteristics particular to ontologies (formalization and integration). They are: i) project management; ii) pre-development; iii) requirements specification; iv) conceptual modelling; v) formalization; vi) implementation; vii) maintenance; viii) integration; ix) evaluation; and x) documentation. Afterwards, each category is established according to standard IEEE-1074 (1997) and methodological principles designed for the building of ontologies (Fernández, Gómez-Pérez, & Juristo, 1997; Uschold, & Gruninger, 1996).

- **Project management:** activities related to the initial stage of a project, such as the software creation and life-cycle; to the monitoring and control of the software project in its entire life-cycle.
- **Pre-development:** consists of analyzing ideas or concepts of a system and, due to problems observed in the environment, allocate the system requirements before the beginning of the software development. This stage includes feasibility study activities and the analysis of the system requirements
- **Requirements specifications:** encompasses restrictions or standards that the software must abide by according to the definitions of the requestor needs. The requirements must serve as an initial document for the realization of modelling and prototyping tasks, and the process is normally interactive.
- **Conceptual modelling:** aims at developing a well organized and coherent repre-

sentation of the system, able to comply with the software requirements specified in the requirements activities.

- **Formalization:** consists of transforming the ontology conceptual model (or conceptualization) into a formal model in order to accurately define its meaning. The professional involved in the ontology building focus on the computer modelling of the problem, using, for example, first-order logic and its extensions (representation systems based on frames, semantic webs, descriptive logic, etc). The techniques used in this stage derive from the Artificial Intelligence area.
- **Implementation:** results in the transformation of the software's engineering project version into a programming language. In the case of ontologies, the implementation consists of mapping the formal model into a language able to respond to the demand, such as the Web Ontology Language – OWL (Van Harmelen et al., 2003).
- **Maintenance:** regarded as a post-development stage consisting of identifying problems and improvements in the products. It can result in new versions of the same products.
- **Integration:** an alternative to facilitate the building of ontologies is to integrate the ontology being created to existing ontologies. This stage considers the reutilization of existing concepts in other ontologies. The proposal is to examine the conceptualization of meta-ontologies (called high level ontologies) and to select (partially or completely) the ones that better fit into the model being constructed. In the integration process the activities can be carried out during the conceptual modelling and ontologies implementing stages. As a result, it is regarded as an integral process.
- **Evaluation:** the activities are carried out at the same time as the activities of processes turned to the software development, such as: revision and auditing of the processes, development of testing procedures, testing and result assessment.
- **Documentation:** development and distribution of documentation for developers and users involved in the processes, in order to provide, in due time, information on the software.

The next step was the development of a comparative chart depicting the development stages presented in the methodologies and methods used for the building of ontologies and controlled vocabularies. This scope was devised through a matrix structure representing the objects investigated in its columns and each phase of its life cycle on the lines. Based on the treatment and organization of the content in the structure, it was possible to comparatively analyze every methodology, method and standard, making theoretical and empirical conclusions about the ontologies and controlled vocabulary building processes.

METHODOLOGIES ASSESSED

This section presents the methodologies and methods used in the building of ontologies as well as the methodology and standard used for the building of controlled vocabulary. Their analysis is described in detail in Silva (2008, p.132). Hence, the section *The Cyc Method* introduces the Cyc Ontology and the method used for its development. The section *Gruninger and Fox's Methodology* presents a few considerations on the Tove Project and the procedures of Gruninger and Fox's methodology. The section *Uschold and King Method* introduces the Enterprise Ontology Project, regarding the procedures of Uschold and King's method. The section *Kactus Method* presents this method. The section *Methontology Methodology* introduces the methods and techniques of Methontology. The section *Sensus*

Method presents the SENSUS ontology and the method that was based on it, the Sensus Method. The section *101 Method* presents this method. Finally, the section *Methodology and Standard for the construction of controlled vocabulary* ponders about the standard used in the building of controlled vocabulary and demonstrates the methodological procedures in the BITI manual which are used for the building of thesauri.

The Cyc Method

In the 1980's, the company Microelectronics and Computer Technology – MCC began creating Cyc, a broad knowledge base which considers the consensual knowledge on the world, including standards and heuristics for deductions about quotidian objects and events (Reed, & Lenat, 2002). Cyc's language of representation is the CycL, which is considered hybrid because it merges frames to predicate calculus. The language has got an inference machine which allows multiple heritages, automatic classification, maintenance of inverse links, restrictions verification, ordinate search, detection of contradictions and resolution module. The Cyc knowledge base was developed in 1990 by Douglas Lenat and Ramanathan Guha (Fernández, Gómez-Pérez, & Corcho, 2004), and three processes were considered: i) extraction of knowledge from the common sense; ii) extraction helped by computers; and iii) extraction managed by computers. In the first process, the knowledge required for the building of the ontology was manually obtained from different sources, such as articles, books and newspapers. The second process was carried out automatically, with the use of computational tools for the processing of natural language and machine learning capable of using common sense knowledge sufficient to investigate and discover new knowledge. Finally, the third process was carried out by a larger number of tools in order to manage the knowledge extraction from the common sense (parts which are deemed difficult to be interpreted in the knowledge sources involved) in the Cyc base.

Gruninger and Fox's Methodology

The methodology was proposed by Michael Gruninger and Mark Fox in 1995 (Gruninger, & Fox, 1995), having as groundwork for its development the experience gained in the Toronto Virtual Enterprise project, known as the Tove project (Fox, 1992), whose theoretical and methodological principles are found in the field of Artificial Intelligence. Project Tove's goal is to create a common sense model about the company, namely a shared knowledge regarding the business, able to lead to the deduction of answers to the questions concerning the domain (Fox, 1992). In order to accomplish that, ontologies are created to specify models for public and private organizations, taking under consideration the following characteristics: a) the capacity to provide a shared terminology for organizations, which can be understood and used by each application, that is, by every kind of business; b) the definition of the semantics of every term through a logical theory; c) the implementation of the semantics into a set of axioms which allows the ontology to automatically deduce the answers to the questions common in the organizations scope; and d) the definition of a simbology used to graphically represent terms or concepts (Gruninger, & Fox, 1996). Gruninger and Fox's methodology was used in the University of Toronto Enterprise Integration Laboratory for the project and assessment of integrated ontologies, including propositions for the building of new ontologies and extensions of existing ontologies. The following procedures were proposed in the methodology: i) the devising of motivational frameworks which aim at identifying problems in the current environment; ii) the specification of informal questions, which aim at specifying, in natural language, the requirements that the ontology should be able to meet; iii) the devising of the formal terminology in which, through first-order logic statements, the concepts and their properties are organized in a taxonomy; iv) the specification of formal questions, in which problems are consistently defined before the axioms in the ontology; v) the specification of formal axioms which restrict the interpretation of the terms involved in formal questions; and vi) the verification of complete theorems, which determine the conditions under which the solutions to the questions are completed.

Uschold and King Method

The method was initially proposed by Mike Uschold and Martin King in 1995 (Uschold, & King, 1995) and expanded in 1996 by Mike Uschold and Michael Gruninger (Uschold, & Gruninger, 1996) in the development experience of the *Enterprise Ontology*. Such ontology was developed as part of the project *Enterprise* by the Artificial Intelligence Applications Institute of the University of Edinburgh and partners such as IBM, Unilever, among others.

Uschold and King (1995) consider the following stages necessary for a comprehensive methodology: i) identification of the ontology objective, aiming at identifying the need of construction, the degree of formality (from the informal, using natural language, to the strictly formal, employing logic declarations) and the classes of users of the ontology, including developers, maintainers and users of applications; ii) the ontology construction, which is divided in: a) capture or conception of the conceptualization of the ontology; b) codification or implementation through a language of representation of ontologies and c) integration with existing ontologies; iii) ontology evaluation through specified requirements; and iv) documentation concerning the ontology intentions and the primitives used to express definitions in ontology.

Kactus Method

The emphasis of the European project, *Esprit Kactus,* is in the organization of knowledge basis which can be shared and reused in different knowledge based systems. In order to do that, domain ontologies are used to organize knowledge independently from the software application which will be built.

Based on the Kactus Project, Amaya Bernaras and colleagues (Bernaras, Laresgoiti, & Corera, 1996) investigated the viability of reusing knowledge systems of technical complexity, such as the domain of electric networks, and the role of ontologies as a support for such systems. This investigation resulted in an ontology construction method in which the processes involved would be conditioned to the development of the application, i.e., every time an application is built, the ontology that represents the knowledge required for that application would be refined. These processes would involve the: i) development of a list or requirements or prerequisites which have to be fulfilled by the application; ii) identification of relevant terms for the application domain based on such requirements, building, therefore, a preliminary model, iii) refinement and structuring of the ontology in order to achieve a definite model; iv) search for ontologies already developed by other applications so as to reuse them. The reused ontologies would demand refinement and extension to be employed in the new application.

Methontology Methodology

The Methontology was developed in the laboratory of Artificial Intelligence of the Polytechnic University of Madrid between 1996 and 1997 by the group of researchers (Fernández, Gómez-Pérez, & Juristo, 1997; Gómez-Pérez, Fernández, & Vicente, 1996).

Methontology involves a set of development stages (specification, conceptualization, formalization, integration, implementation and maintenance), a life cycle based on prototypes evolution (Pressman, 2004) and techniques to carry out planning, development and support activities. The planning activity includes task echeloning and control, aiming at reaching the desired quality. The support activities involve knowledge acquisition, documentation and evaluation, and they take place throughout the ontology's life cycle. The initial development stages (specification and conceptualization) demand a great effort within the support activities, such as knowledge acquisition and evaluation. This is due to many reasons: a) most of the knowledge is acquired in the beginning of the ontology construction process; and b) the conceptual model should be correctly evaluated to avoid future errors in the life cycle of the ontology. Finally, the detailed documentation should be produced at the end of each stage anticipated in the life cycle.

Sensus Method

The SENSUS ontology was developed by the group of natural language *Information Sciences Institute* – ISI aiming to be used in the processing of natural language. The SENSUS ontology comprehends approximately 70.000 concepts, organized hierarchically according to their abstraction level, which ranges from medium to high. However, its structure does not contemplate domain-specific terms (Swartout et al., 1996); to do so, the domain-specific terms are linked to the comprehensive SENSUS ontology, in order to build ontologies for private domains.

The Sensus method, based on the SENSUS ontology, presents some procedures to establish links among the specific terms and the high-level ontology terms (Swartout et al., 1996). The result of such process is a structure of a new ontology, which is automatically generalized through the OntoSaurus tool (Fernández, Gómez-Pérez, & Corcho, 2004; Swartout et al., 1996). According to the method, the procedures involved in the building of specific domain ontology would be: i) to identify the relevant concepts in the domain; ii) to link manually the concepts to the SENSUS ontology; iii) to add paths to the concept of SENSUS top hierarchy; iv) to add new domain concepts; and v) to complete the hierarchies.

101 Method

The 101 method (Noy, & McGuinness, 2001) was conceived based on the experience in the development of wine and food ontology, using the ontology-editing environment *Protégé-2000* (Horridge et al., 2004).

The 101 method proposes basically four activities for the development of an ontology: i) define classes in the ontology; ii) organize the classes in a taxonomy; iii) define slots (or properties) for the classes and describe their allowed values (called facets); and iv) fill the slots values for the instances. Such activities imply in modelling decisions, which the method tries to emphasize, besides being an iterative process of an ontology life cycle.

Methodology and Standard for the Construction of Controlled Vocabulary

The standard proposed by the *National Information Standards Organization* (NISO, 2005) sets the general guidelines for the construction, formatting and maintenance of controlled monolingual vocabulary. Concerning the construction standards, the ANSI/NISO Z39.19-2005 standard allows the construction of various kinds of controlled vocabulary, including thesauri, taxonomies, lists and synonyms rings, in a known and structured order, aiming at making clear the equivalence, associative and hierarchical relationships, when applicable to each type (see appendix B of the standard, page 135). This flexibility is important, as it allows a better adaptation of the instrument to the needs of the informative environment, as, for example, the Web.

An innovation of the ANSI/NISO Z39.19-2005 in relation to the previous standards is the inclusion of the faceted analysis, whose basis is concentrated in the work of Ranganathan (Ranganathan, 1967) and in the refinements made by the *Classification Research Group* – CRG (Campos, Gomes & Motta, 2004), in England, for the elaboration of thesaurus in specific areas of knowledge. Another important question is the orientation regarding the interoperability among controlled vocabularies.

Finally, the ANSI/NISO Z39.19-2005 standard recommends special treatment in relation to the term with regards to activities involved in the construction process. Yet, the BITI manual (Campos, Gomes & Motta, 2004) on the elaboration of a type of controlled vocabulary, the thesauri, describes in details and in sequence the necessary procedures to its construction. In this way, the BITI manual is more efficient than the ANSI/NISO Z39.19-2005 standard as regards methodological principles explicitly defined for the building of controlled vocabularies. The procedures proposed in the manual can be summarized in: i) planning, which consists in delimitating the subject area to be covered by the specialized vocabulary, define the target public to whom the instrument is destined to and elaborate the planning for future maintenances which become necessary; ii) vocabulary survey, which consists in selecting the terms representing the subject and define them according to the nature of the subject; iii) concept organization, consisting in grouping the terms of the same nature in categories or facets, aiming at allowing a better understanding of the concept and organization of the relationship between the concepts; iv) final presentation, including the kinds of exhibition (simple alphabetic lists or graph visualizations) and the format which can be printed or electronic; and v) evaluation criteria, which determines if the controlled vocabulary is giving satisfactory search results, implying in a good relationship between precision and revocation. After presenting the investigated objects, the content is classified within each category, denominated, from this point on, process building phases. The phases are shown in Table 1, below, summarizing the analysis of the methodologies, methods and norms. The phases which were not proposed in the investigated objects had their cells filled in with "Absent", when contemplated, the cell was filled in with the adopted methodological principles.

COMPARATIVE ANALYSIS

In order to carry out the comparative analysis, it was necessary to understand the methodological procedures of the investigated objects, presented in the previous section. Such understanding was possible through content analysis conducted in empirical materials by means of the categories of analysis explained in particular section.

Based on the information displayed in the Table 1, some considerations were drawn about the methodologies and methods for the building of ontologies and about the methodology and standard for the construction of controlled vocabularies which were analyzed here. Such considerations are listed below:

- There are several strategies for the development of ontologies, proving the hypothesis that different groups present diverse approaches and characteristics, for different aims and applications (Fernández et al., 1999).
- In the ontologies context, some approaches follow a life cycle model, whereas others do not. In this aspect, the one that stands out is the Methontology as it is almost complete regarding a development cycle, since it does not propose only a pre-development phase. Such remark can be confirmed in Table 1.
- Regarding details of the activities and procedures to carry them out, some methodologies and methods prove to be superficial in the elucidation of the steps to be followed in construction of ontologies. This is the case of the Cyc, Kactus and Sensus

Table 1. Synoptic table of the methodologies, methods and standards with respect to the categories of pre-defined analysis

Lifecycle phases		**Cyc**	**Gruninger e Fox**		**Uschold and King**	**Kactus**	**Methontology**	**Sensus**	**101 Method**	**Z39.19-2005**	**BITI Manual**
Project Management		Absent	Absent		Absent	Absent	Staggering of activities; task control and quality guarantee.	Absent	Absent	Recommends characteristics necessary to management systems	Project planning activities.
Guided Projects	Pre-development	Absent	Motivation Scenarios		Absent	Absent	Absent	Absent	Absent	Recommends a construction feasibility study	Absent
	Development	Requirements Specification	Manual knowledge extraction	Informal competency questions.	Determine ontological aims.	Requirements list.	Definition of ontological scope	Absent	Competency questions definition	Absent	i) Area delimitation; and ii) target public definition.
		Conceptual Modelling	Absent	Formal terminology conception.	Consensual vocabulary construction.	Identification of relevant terms.	Activities involving the construction of conceptualization of the ontology.	Identification of the domain key-terms	Definition of classes, *slots*, facets (restrains) and instances.	Recommends the facet analysis in the organization of knowledge; organization of relationships in hierarchical, associative and equivalent.	Terms selection; definition of concepts and relationships (logic and ontological); facets classification.
		Formalization	Absent	Formalization of competency questions. Axioms specifications.	Absent	Absent	Formalize the conceptual model in a formal language as descriptive logic.	Through the semantic network.	Representation language based on *frames*.	Absent	Absent
		Implementation	Knowledge codification and extraction with computer-supported tool	Axioms specifications are implemented in Prolog.	Requires a language representation such as Prolog and Ontolingua.	Absent	Recommends developing environments that fits to meta-ontologies and ontologies selected in the integration phase.	Uses the OntoSaurus tool for development.	Interactive use of the language through the *Protégé* tool.	Absent	Absent

continued on following page

Table 1. Continued

Lifecycle phases		Cyc	Gruninger e Fox		Uschold and King	Kactus	Methontology	Sensus	101 Method	Z39.19-2005	BITI Manual
	Post-development	Maintenance	Absent	Absent	Absent	Absent	Orientation concerning the necessary adjustments in the conceptualization activities.	Absent	Absent	Recommends procedures related to the addition, alteration and exclusion of terms.	The inclusion of new terms should be done in a specific spreadsheet.
Integrative processes	Integration	Integration of computer systems to Cyc micro theories.	Integration of common core ontologies.		Integration with existing ontologies.	Search for already existing ontologies.	Integration document with meta-ontologies.	Manual integration to SENSUS ontology.	Considers the reuse of ontologies.	Absent	Absent
	Evaluation	Absent	Through complete theorems.		Can be done through competency questions.	Absent	Divided in ontology verification and validation.	Absent	Absent	Recommends the use of usability testing.	Criteria adopted for the adoption or not of a thesaurus.
	Documentation	Absent	Absent		Claims and primitives.	Absent	Recommended in each phase.	Absent	Via *Protégé*	Recommendations on the documentation content.	Data records in spreadsheets.

methods, which seem to assume that the ontologist already dominates the subject of ontology construction and does not need details about the activities and procedures involved. Yet, the Methontology stands out in supplying, most of the times, details concerning how to proceed in conducting a given activity.

- Some approaches emphasize the development activities, specially the ontology implementation (Cyc method and 101 method), disregarding important aspects related to project management, feasibility studies, and maintenance and evaluation of ontologies.
- The BITI manual presented a virtually complete life cycle according to the IEEE-1074 (1997) standard, as shown in Table 1. The formalization, implementation and integration phases were not taken into consideration as they did not belong to the controlled vocabularies purpose. In this way, the maturity of the methodology in relation to a life cycle model can be checked.
- Finally, it is worth saying that the thesaurus construction methodology stood out in relation to the other methodologies and methods for building ontologies in the theoretical and methodological aspect, for the identification, definition and organization of concepts. Although these indicate methods to identify concepts, present resources to define concepts and organize them in taxonomy, they do not explain clearly the theoretical principles which govern the classification theory (Ranganathan, 1967) neither the concept theory (Dahlberg, 1978) in the specification of their elements. Such principles would be relevant in methodologies for ontology constructions, since both tools, ontologies and thesaurus, present semantic and conceptual relationships.

FUTURE RESEARCH DIRECTIONS

The ontology development, in spite of many works being presented in the last decade, is still a relatively unexplored field, in which much work is still needed. The authors are working at the moment in compiling the gold standard pieces of each methodology, along with some novelty research, to propose an integrated framework for ontology building. For the moment, it is known that it would be less likely that a general framework would work for any ontology type, in any domain, but the different approaches must be integrated in order to an adequate process to take place. We also acknowledge the fact that we have not analyzed the potential influences that each software tool and formal language, chosen to build ontologies, can have. They potentially interferes with the final steps of the process, given the extent in which those tools and languages can represent entities, relationships, instances, constraints, etc.

CONCLUSION

This chapter proposed an assessment of the most representative methodologies and methods for ontology construction in literature, and also pointed out similarities between standards of construction of software (IEEE-1074) and methodological principles used in the elaboration of ontologies and controlled vocabularies. Such similarity was evident in the analysis of methodologies, methods and norms investigated, presented in section *METHODOLOGIES ASSESSED*. The methodology presented in the BITI manual for the construction of thesaurus is mature as a life cycle model, since the construction processes fit most categories of analysis based on IEEE-1074 standard (1997). Although the recommendations of the ANSI/NISO Z39.19-2005 standard were classified in the categories of analysis, such norm does not aim at proposing an activity cycle for the construction of controlled vocabularies, but to recommend a series of treatments concerning the term in activities that fit the building process. Regarding the methodologies for ontology construction, the Methontology was prominent in the maturity aspect according to the IEEE-1074 (1997) standard, as only the pre-development category was not taking into consideration in its development process.

The chapter also confirmed some problems related to the lack of a standard for ontology construction and the lack of systematic explanations of how, where and under which limits the theoretical approaches can be used within the elaboration process. This fact was verified in the methodology analysis and the ontology building methods investigated in the research, which, in most cases, were not efficient in explaining clearly the construction procedures. Thus, the solution for these problems would be centred on a methodological proposal based on theoretical and methodological principles which would give scientific support to the ontology building process. Finally the presentation of a comparative analysis as a preliminary step can be used to define methodological standards for ontology building.

REFERENCES

Almeida, M. B., & Barbosa, R. R. (2009). Ontologies in knowledge management support - A case study. *Journal of the American Society for Information Science and Technology*, *60*(10), 2032–2047. doi:10.1002/asi.21120

Bardin, L. (1977). *L'analyse de contenu*. Paris, France: Presses Universitaires de France.

Bernaras, A., Laresgoiti, I., & Corera, J. (1996). Building and reusing ontologies for electrical network applications. In W. Wahlster, (Ed.), *The European Conference on Artificial Intelligence* (pp. 298-302). Chichester, UK: John Wiley & Sons.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43. doi:10.1038/scientificamerican0501-34

Campos, M. L. A., Gomes, H. E., & Motta, D. F. (2004). Tutorial de Tesauro. UFF. Retrieved February 8, 2009, from http://www.conexaorio.com/biti/tesauro

Corazzon, R. (2008). *What is ontology? Definitions by leading philosophers. Ontology: A resource guide for philosophers*. Retrieved November 9, 2008, from http://www.formalontology.it/section_4.htm

Dahlberg, I. (1978). A referent-oriented, analytical concept theory for interconcept. *International Classification*, *5*(3), 142–151.

Fernández, M., Gómez-Pérez, A., & Corcho, O. (2004). Methodologies and methods for building ontologies . In Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (Eds.), *Ontological engineering* (pp. 107–153). London, UK: Springer.

Fernández, M., Gómez-Pérez, A., & Juristo, H. (1997). *Methontology: From ontological art towards ontological engineering*. Retrieved April 16, 2009, from http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-005.pdf

Fernández, M., Gómez-Pérez, A., Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. *Intelligent Systems*, *14*(1), 37–46. doi:10.1109/5254.747904

Fox, M. S. (1992).*The TOVE Project: Towards a common-sense model of the enterprise*. Retrieved December 20, 2005, from http://www.eil.utoronto.ca/enterprise-modelling/papers/fox-tove-uofttr92.pdf

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies - An etymological note. *The Journal of Documentation*, *59*(1), 7–18. doi:10.1108/00220410310457984

Gómez-Pérez, A., Fernández-Lopez, M., & Corcho, O. (2004). *Ontological engineering: With examples from the areas of knowledge management, e-commerce and the Semantic Web*. Springer.

Gómez-Pérez, A., Fernández-Lopez, M., & Vicente, A. J. (1996).*Towards a method to conceptualize domain ontologies*. Retrieved March 24, 2005, from http:// webode. dia. fi. upm. es/ Asun/SSS97.ps

Gruber, T. (1993). *What is an ontology?* Retrieved April 16, 2009, from http://www-ksl. stanford. edu/kst/what-is-an-ontology.html

Gruninger, M., & Fox, M. S. (1995). *Methodology for the design and evaluation of ontologies*. Retrieved January 9, 2007, from http:// citeseerx. ist. psu. edu/ viewdoc/ summary? doi=10.1.1.44.8723

Gruninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In D. Skuce (Ed.), *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI 1995)* (pp. 1–10). Montreal, Canada: AAAI Press.

Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). *Pratical guide to building OWL ontologies using the Protégé-OWL plugin and CO-ODE tools*. Retrieved June 30, 2006, from http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf

Institute of Electrical and Electronics Engineers. (1997). *IEEE 1074 standard for developing software life cycle processes*. Retrieved February 10, 2006, from http:// ieeexplore. ieee.org/ Xplore/ login.jsp?url=/ iel4/5984/16018/00741936. pdf?temp=x

Lancaster, F. W. (1986). *Vocabulary control for information retrieval*. Richmond, VA: IRP.

NISO. (2005). *ANSI/NISO Z 39.19: 2005 - Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Retrieved April 20, 2008, from http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-19-2005.pdf

Noy, F. N., & Guinness, D. L. (2001). *Ontology development 101: A guide to create your first ontology*. Stanford University, 2001. Retrieved February 15, 2003, from http:// protege. stanford. edu/ publications/ ontology_development/ontology101.pdf

Pressman, R. S. (2004). *Software engineering: A practitioner's approach*. New York, NY: McGraw-Hill.

Ranganathan, S. R. (1967). *Prolegomena to library classification*. Bombay, India: Asia Publishing House.

Reed, S. L., & Lenat, D. B. (2002). *Mapping ontologies into Cyc*. Retrieved February 12, 2005, from http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf

Silva, D. L. (2008). *A proposal for a methodology of building ontologies: Interdisciplinary perspective between the Information Science and the Computer Science*. Master Thesis, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, BR, School of Information Science.

Soergel, D. (1997). *Functions of a thesaurus / classification / ontological knowledge base*. Retrieved August 2, 2005, from http://dl.lib.brown.edu/intranet/ws/userx/docs/Soergel--Functions%20of%20a%20thesaurus%20%20classification%20%20ontological%20knowledge%20base.pdf

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science American Society for Information Science*, *50*(12), 1119–1120. doi:10.1002/(SICI)1097-4571(1999)50:12<1119::AID-ASI12>3.0.CO;2-I

Swartout, B., Patil, R., Knight, K., & Russ, T. (1996). *Toward distributed use of large-scale ontologies*. Retrieved February 2, 2004, from http:// ksi. cpsc. ucalgary. ca/ KAW/ KAW96/ swartout/Banff_96_final_2.html

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, *11*(2). doi:10.1017/S0269888900007797

Uschold, M., & King, M. (1995). *Towards a methodology for building ontologies*. Retrieved July 20, 2002, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5357

Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. M., Patel-Schneider, P. F., & Stein, L. A. (2003). *OWL Web ontology language 1.0 reference*. Retrieved January 21, 2004 from http://www.w3.org/TR/2003/WD-owl-ref-20030221/

Vickery, B. C. (1997). Ontologies. *Journal of Information Science*, *23*(4), 277–286. doi:10.1177/016555159702300402

ENDNOTES

[1] http://citeseer.ist.psu.edu

[2] http://portal.acm.org/portal.cfm

[3] http://scholar.google.com/