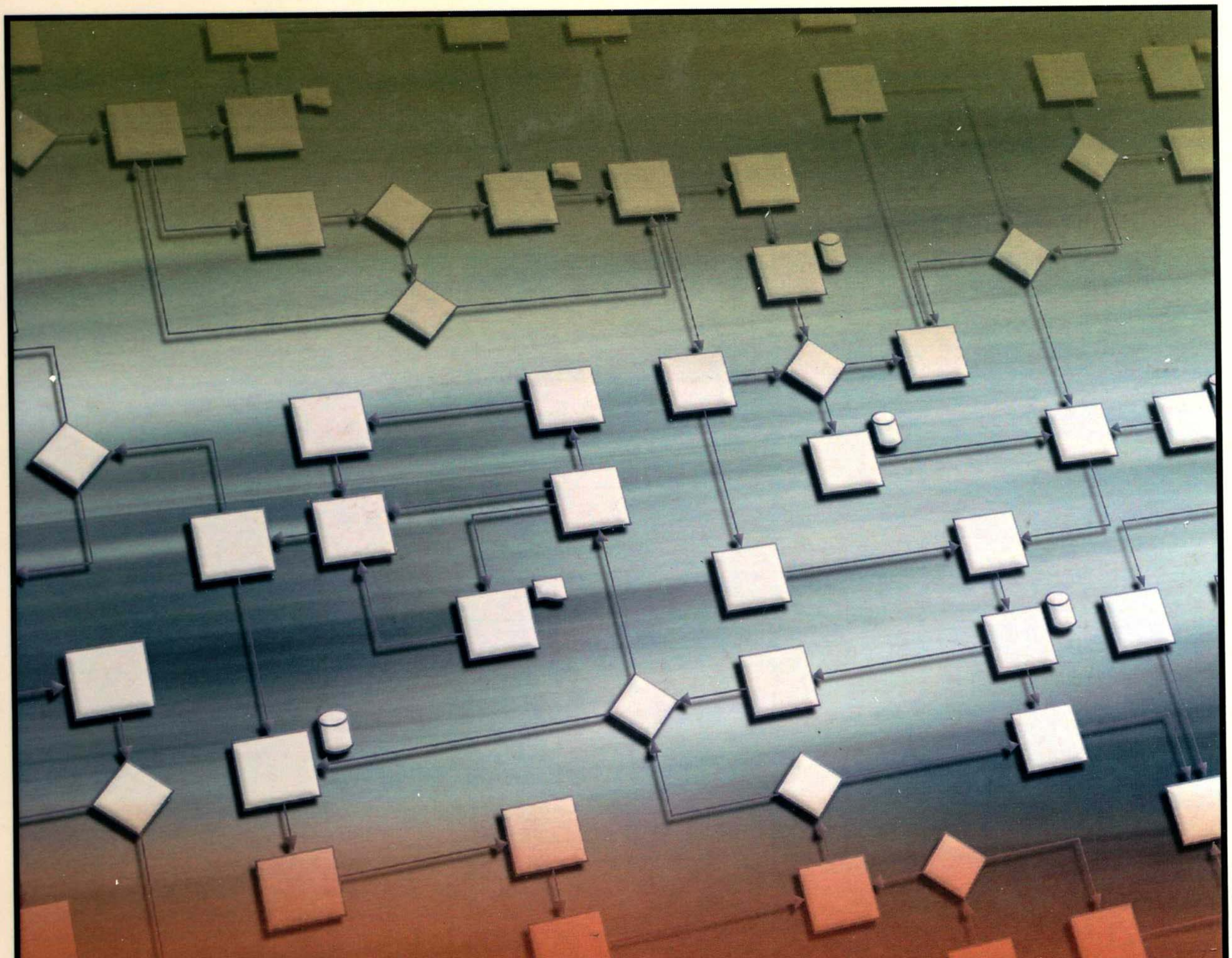
PREMIER REFERENCE SOURCE

SEMANTIC TECHNOLOGIES FOR BUSINESS AND INFORMATION SYSTEMS ENGINEERING

Concepts and Applications



Stefan Smolnik, Frank Teuteberg & Oliver Thomas

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.

Editorial Advisory Board

Preface

There are increasing opportunities to consider the application of semantic technologies for business information systems. Semantic technologies are expected to improve business processes and information systems, and lead to savings in cost and time as well as improved efficiency. However, the degree of automation in enacting business processes and monitoring information systems and value chains is still unsatisfying. Current problems are representational heterogeneities between the various modeling notations used, the various (subjective) perspectives of the modelers on the application domains, or the different stages in the life-cycles of business processes. Interoperability and integration of advanced business information systems is concerned with the application of semantic technologies. These technologies allow for annotating meaning to business concepts and processes, and allow for automatically monitoring business systems and processes without human interactions. The book at hand explores the potential of semantic technologies for business and information systems engineering and provides an overview of applying semantic technologies for business and information systems engineering.

The first section, *Models and Methods,* covers fundamental aspects of creating and applying knowledge structures. At first, methodologies are compared by Daniela Lucas da Silva, Renato Rocha Souza, and Maurício Barcellos Almeida. The authors conduct an analytical study compiled by the analysis of literature about methodologies for building ontologies and controlled vocabularies as well as by the analysis of international standards for software engineering. The application of an ontology-based method is then described by José González and Mathias Uslar. The authors develop an approach for constructing a domain specific reference model catalogue for the energy sector. In doing so, the authors consider both modeling requirements in the fields of ontology and reference model design. In the last chapter of this section, Peter Fettke and Peter Loos illustrate the use of ontology to evaluate Scheer's reference model for production planning and control systems. The authors use ontology as a tool or theory and thus complement the dominating view on ontologies as design artifacts by a view on ontology in line with the philosophical discipline.

The second section provides an overview of the plethora of applications of semantic technologies in the field of *Data and Knowledge Management*. A fundamental problem in this area is how to process data originating from heterogeneous data sources and in diverse data formats such as text, numerical data, multimedia and others. Farid Bourennani and Shahryar Rahnamayan propose a unified approach for representing and processing data of heterogeneous types which ultimately should augment the interpretation of qualitative and quantitative data with use cases in business and financial sectors. Further, Liane Haak presents new ways for the integration of structured and unstructured data focusing on the application to data warehousing and knowledge management. The chapter introduces a solution for generating an ontology from a data warehouse system and integrating it with a knowledge management systems ontology. The integrated ontologies are subsequently used for semantic navigation.

The last chapter of this section, by Alexey Alishevskikh and Tatiana Emshanova, demonstrates how personal knowledge management can be improved by semantic desktop technologies. It describes theoretical aspects as well as the implementation of a supportive technology and framework integrating several aspects such as content aggregation, search, natural language processing, metadata management, and tagging. The synthesis of these different techniques aims at augmenting the experience of knowledge workers when working with digital information assets.

The third section, *Semantic Technologies in Conceptual Modeling,* focuses on applying semantic technologies to support conceptual modeling. The improved construction and analysis of semi-formal models on a general level is targeted by an approach for conceptual model analysis contributed by Patrick Delfmann, Sebastian Herwig, Łukasz Lis, and Jörg Becker. The approach integrates semantic standardization and structural pattern matching, hence enabling an unambiguous analysis of the models' contents. It is intended for a number of purposes such as revealing syntactical errors, model comparison, model integration or the identification of business process improvement potentials.

Regarding the semantic verification of business process models, an ontology-based approach making use of background knowledge encoded in formal ontologies and rules is proposed by Michael Fellmann, Frank Hogrebe, and Oliver Thomas. The authors develop a model for the ontology-based representation of process models which is used in conjunction with machine reasoning for process model verification. The approach is demonstrated using real-life administrative process models taken from a capital city.

Whereas the two previous approaches aim at improvements regarding the construction and validation or verification of conceptual models and hence stick to more traditional ways of conceptual modeling, a more fundamental shift in the way of modeling is accompanied by the idea of a fully automated model construction. In the domain of process modeling, first results are shown by Bernd Heinrich, Mathias Klier, and Steffen Zimmermann, who present an algorithm and accompanying method for the automated planning of process models.

At the end of this section, the use of semantic technologies from a practical perspective is reported by Hans-Georg Fill and Ilona Reischl. Their chapter describes how semantic technologies can be combined with conceptual models to support management executives in the distribution of knowledge and the analysis of compliance. The approach is based on a stepwise semantic enrichment of conceptual models with formal semantic schemata and has been implemented on the ADONIS meta-modeling platform in the context of a real-life project with the Austrian competent authority in regard to safety in healthcare.

The fourth section of the book, *Semantic Process Description,* is dedicated specifically to the semantically enhanced representation and annotation of semi-formal process models. A detailed model how to annotate semi-formal process models both with lexical and semantic labels bridges the gap between human understandability and machine interpretability. This is introduced by Andreas Bögl, Michael Karlinger, Michael Schrefl, and Gustav Pomberger. Amongst other purposes, it can be applied for the automated refactoring of model elements and automated semantic annotation. The latter is also addressed by Yun Lin and Darijus Strasunskas, albeit they suggest annotating processes on a more coarse-grained level of process model templates. The annotation consists of the three basic parts: meta-model, domain, and goal annotation. It may be used to facilitate the retrieval and reuse of process knowledge.

In the last chapter of this section, the semantically enhanced business process modeling notation is introduced by Witold Abramowicz, Agata Filipowska, Monika Kaczmarek and Tomasz Kaczmarek. This "ontologized" version of the Business Process Modeling Notation is formalized in the sBPMN ontology, which is consequently suggested as a serialization format for BPMN modeling tools. In such a way, annotations can be created invisible to the users and directly embedded into the models leading to advanced machine interpretability which facilitates and mechanizes the task of transforming semi-formal process models into executable workflows.

The last section of the book provides insights on how to use semantic technologies to support *Services and Workflows*. The first contribution in this part by Tariq Mahmoud, Jorge Marx Gómez and Timo von der Dovenmühle offers a Semantic Web services based reference model. This model relies on the idea of applying lightweight semantics to web services targeted at improved service advertisement, service composition, and service validation. While this chapter mainly focuses on the description and composition of services, the check whether such processes comply with regulations or policies is the main concern of the contribution by Rainer Telesko and Simon Nikles. They describe a concept for the semantics-based configuration of service packages with respect to service level agreements thereby capitalizing on the principles and use cases of the EU-project plugIT and additionally reporting on the economic benefits. With similar intentions, semantic policies for modeling regulatory process compliance are envisioned by Marwane El Kharbili and Elke Pulvermueller. They also motivate the need for automation in compliance management and propose the use of policies as a modeling concept for regulations. The authors introduce the CASE model and the corresponding policy modeling ontology. Both are used to support automated compliance checking of enterprise processes to regulations. The utilization of the CASE method as well as the policy ontology is showcased using an example of resource access control in business processes.

In the last chapter of the book, Barbara Thönssen and Daniela Wolff take a broader view on context models with the intent to support business process agility. They present dimensions of change concentrating on a specific ability of an enterprise to deal with change and propose a semantically enriched context model based on a well-known enterprise architecture. Finally, a context aware workflow engine is presented which leverages these concepts and rules which trigger process adaptations during run time.

We hope that this book will receive widespread recognition both from practitioners and the scientific community.

Stefan Smolnik
EBS Business School, Germany

Frank Teuteberg
University of Osnabrueck, Germany

Oliver Thomas
University of Osnabrueck, Germany

Acknowledgment

First and foremost, the editors wish to thank all authors for their excellent contributions to this volume and for shouldering the bulk of the efforts it took to realize this project. Without their effort and cooperation, we could not have finished this book. The high quality of all papers included here was assured by a double blind review process (each submitted chapter was reviewed at least twice). All articles are part of the rapid progress that we are currently witnessing in scientific research on Semantic Technologies for Business Information Systems Engineering.

Furthermore, we would like to thank IGI Global, especially Beth Ardner, and Mike Killian, Editorial Content Department, who have both supported this project at all phases of development in a constructive, professional, dedicated, friendly, and open manner.

This book could not have been completed without professional preparation and organization. Therefore, we would particularly like to thank Ms. Anja Grube for her assistance in coordination and communication as well as for proofreading and translating services.

We hope that this book provides valuable insights for the reader and will encourage the adoption of semantic technologies for business and information systems engineering.

As for the contents of this volume, the editors are always open to any suggestions for improvement and are looking forward to the readers' comments and fruitful discussions.

Stefan Smolnik
EBS Business School, Germany

Frank Teuteberg
University of Osnabrueck, Germany

Oliver Thomas
University of Osnabrueck, Germany

Section 1
Models and Methods

Table of Contents

Section 4
Semantic Process Description

Section 5
Services and Workflows

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-