Applied Ontology 4 (2009) 245–265
DOI 10.3233/AO-2009-0070
IOS Press

A proposal to evaluate ontology content

Mauricio Barcellos Almeida
Information Theory and Management Department, School of Information Science,
UFMG – Universidade Federal de Minas Gerais, Minas Gerais, Belo Horizonte, Brasil
E-mail: mba@eci.ufmg.br

Abstract. Research into ontologies has received much attention and has been the subject of study in different scientific communities. The current growth observed in ontology use and availability increases the importance of the evolution of methods to evaluate ontologies. There are formal methodologies available, but the evaluation of ontology content does not seem to be a main concern. We understand the evaluation of ontology content to be the activity of verifying whether the ontology really represents the knowledge of a domain, according to the assessment of experts. Our objective is to contribute to this sense. In order to do so, we present qualitative research about the content evaluation of an organizational ontology, developed within a large Brazilian energy utility company. The evaluation process consists of a set of questionnaires, which are based on a multidisciplinary approach, and of a prototype system. We believe that this proposal can contribute towards improvements in both modeling and evaluation processes in ontology development.

Keywords: Ontologies, modeling, user-centered evaluation

1. Introduction

Ontologies have been studied since the seventies in Artificial Intelligence research. In the nineties, Semantic Web research increased the demand for ontologies for some kinds of applications, both to solve interoperability problems and to provide a common information structure. The evaluation of ontologies is an important subject as there are many of these structures available, covering multiple domains, with different formats, sizes, goals, degrees of formalism and using different representational languages.

In general, questions about evaluation are related to mechanisms for interacting with an ontology, to the formalism of knowledge representation and to the appropriateness of documentation. The evaluation process is related to a technical judgment, consisting of two steps: verification and validation. The verification process consists of ensuring consistency between the ontology and the related software environment and documentation, throughout the ontology life-cycle; the validation process ensures that the ontology, software and documentation correspond to the system designed.

Proposals for evaluating ontologies are available (Gómez-Pérez, 1994, 2004; Maedche & Staab, 2002; Guarino & Welty, 2002; Porzel & Malaka, 2004; Brewster et al., 2004; Lozano-Tello & Gómez-Pérez, 2004; Velardi et al., 2005; Gangemi et al., 2006; Ding et al., 2004), but a comprehensive and consensual standard methodology does not seem to exist. According to Obrst et al. (2006), the lack of consensus regarding the ontology evaluation process is a result of the different perspectives provided by the several academic fields involved, including but not limited to: Computer Science, Artificial Intelligence, Databases, Software Engineering; Information Science and Librarianship; Logic and Philosophy, and so on. Moreover, the development of ontologies is more a craft than a scientific activity (Jones et al., 1998) and there are no unified proposals available, since different groups have different approaches (Fernández, 1999). This diversity may be a factor which makes the development of unified methodologies a difficult task. In the absence of a standard, the existing proposals try to meet different evaluation possibilities.

From the viewpoint of modeling, our focus is on the process of evaluating ontology content. By ontology content evaluation, we mean the process of verifying whether the knowledge acquired corresponds to that which is present in the environment where we accomplish the knowledge acquisition process. This verification is user-centered, in other words, it is centered on the expert who contributes to the modeling process and is responsible for verifying whether the domain knowledge is properly represented by the ontology as a result of the modeling process. In this context, the present study proposes a multidisciplinary approach to ontology content evaluation.

We introduce here an ontology of an organization, presenting both its development and subsequent evaluation. This research was conducted with the use of a prototype and of a group of questionnaires. We expect our contribution to be twofold: (i) to provide an alternative proposal for ontology content evaluation; (ii) to provide improvements in human communication processes, which are commonly found in conceptual modeling activities.

The body of this article is organized as follows: Section 2 presents ontology evaluation methodologies and a description of the theoretical bases that we used to devise the content evaluation proposal. Section 3 describes the research, which consists of an ontology development process in a large Brazilian energy utility company as well as the corresponding evaluation process. Section 4 presents the results of the evaluation and discusses the use of ontology in the organization. Finally, in Section 5, we offer our conclusions regarding the feasibility of the proposal.

2. Methodologies for ontology evaluation

Since the nineties, various approaches to evaluating and selecting ontologies have been developed. In general, evaluation approaches seek to compare ontologies with regard to some criteria. In contrast, the selection approaches have to do with choosing an appropriate ontology out of a collection of those available. Some authors do not make this distinction and consider the selection process to be a particular kind of evaluation process, in which multiple criteria are used. Indeed, there is no single set of criteria to classify ontology evaluation methodologies.

The remainder of Section 2 is divided as follows: Section 2.1 presents a survey of ontology evaluation initiatives; Section 2.2 describes the theoretical bases for the establishment of content evaluation criteria based on a multidisciplinary approach including theories from different research fields.

2.1. A survey of ontology evaluation initiatives

In order to facilitate understanding of evaluation proposals, we may organize them using certain classifications. However, the diversity of initiatives poses difficulties to creating a consensual classification. Category features may overlap and also initiatives may belong to more than one category. In the present study, we adopt the classification provided by Brank et al. (2005), which outlines two sets of criteria to classify ontology evaluation initiatives: the first being related to the kind of approach; and the second to the object of evaluation. Let us begin by pointing out that the survey is not an exhaustive one, and its goal is only to exemplify the main types of research available on the subject.

2.1.1. The type-of-approach criteria

The type-of-approach criteria are related to the way the evaluation is conducted, specifying the nature of the procedures occurring throughout the evaluation process. Figure 1 depicts an overview of the type-of-approach criteria.

Maedche & Staab (2002) present an approach based on similarity measures for comparing ontologies. The similarity measures are proposed at conceptual and lexical levels. At the lexical level, two lexical entries are compared according to a method for weighting the difference between two strings (Levenshtein's edit distance). This method measures the number of insertions, substitutions and deletions required to transform a set of strings into another set of strings, considering the use of a dynamic programming algorithm. At the conceptual level, the method investigates the existing conceptual relations between terms. It provides a string-matching measure based on the comparison of taxonomies and on the verification of the accuracy level of the overlap between two relations or concepts.

Gangemi et al. (2006) propose two ontology evaluation models: a meta-ontology and an ontology for the selection of other ontologies. From the O^2 meta-ontology it is possible to identify three main kinds of metrics for evaluating ontologies: (i) structural measures, which emphasize syntax and formal semantics in ontologies represented as graphs; (ii) functional measures, related to the intended use of the ontology; (iii) usability measures, related to the ontology annotation level. The O^2 meta-ontology is complemented by the *oQual* ontology, which provides diagnostics about elements, processes and attributes of the ontology under evaluation. The diagnostic is based on ontology descriptions, which describe essential data for evaluation and selection.

Brank et al. (2006) present a proposal for the automatic assessment of ontologies based on an analogy between the ontology learning task and the traditional unsupervised clustering techniques generally applied to text documents. Regarding clustering techniques, the method consists in partitioning a set of instances in a series of disjoint subsets and comparing these partitions. Considering the ontology as a hierarchical way of partitioning, the authors propose an adaptation of the method as a means to make possible the comparison between an ontology acquired via learning and the standard ontology. For this purpose, they also propose a set of similarity metrics based on the distance between concepts pertaining to that ontology, called Onto-Rand Index. The validity of the proposal is verified by presenting the process outputs to a user in several different situations, so as to enable the user to evaluate whether the results are reasonable.

Maynard et al. (2006) propose metrics for an ontology-based evaluation of information extraction. The comparison standard is composed of a text corpus, called OntoNews, enriched by annotations provided by a general purpose ontology. The authors describe traditional information extraction metrics: Precision

Type-of-approach	comparison with a standard	evaluation of use in an application	comparison with data sources	conformity with requirements
Description	Syntactical comparison between an ontology and a standard, which may be another ontology	Use of an ontology in an application followed by evaluation of the results	Comparison with a data source covered by the ontology proper	Evaluation conducted by people who seek to verify the adherence of an ontology to criteria and patterns
Example	(Maedche & Staab, 2002; Gangemi et al.2006; Brank, Grobelnik & Mladenić, 2006; Maynard, Peters & Li, 2006)	(Porzel & Malaka, 2004; Kalfoglou & Hu, 2006)	(Patel et al., 2004; Brewster *et al.* 2004)	(Lozano-Tello & Gómez-Pérez, 2004)

Fig. 1. Examples of initiatives classified by kind of approach.

and Recall (PR), Cost-based Evaluation (CE), Learning Accuracy (LA); and present a new proposal, in fact an LA extension, called Balanced Distance Metric (BDM). Three metrics are used to evaluate the standard: LA, DBM and a combination of PR and CE. The results indicate that both DBM and LA metrics perform better in extracting information from hierarchical structures, however problems are found when humans evaluate them.

Porzel & Malaka (2004) provide a quantitative approach related to specific tasks for evaluating ontologies. It presupposes the existence of a sufficiently complex task important in the context of the ontology or set of ontologies under evaluation. Also, there is an application or an algorithm which uses the ontology to execute a specific task, and a pre-defined standard of comparison. The ontology evaluation results show two kinds of errors: insertion errors and deletion errors. The former indicates the presence of unimportant concepts and relations, and the latter indicates the presence of ambiguous concepts and relations. The performance measures evaluate ontologies in terms of their effects on the execution of the task.

Kalfoglou & Hu (2006) discuss the need to evaluate ontologies available from the Semantic Web, from technological, strategic and political viewpoints. From their experience in building the Computer Science AKTive Space (Shadbolt et al., 2004), which is a portal to explore semantically-annotated domains, the authors emphasize problems with and suggest solutions for ontology evaluation. Problems include verifying appropriateness, validity, consistency, and so on. Solutions provided over the medium term are: the participation of experts in communities of practice (Lave & Wenger, 1991; Brown & Duguid, 2000) and ontology classification according to use. Similarly, solutions provided over the long term are related to the evolution of standardization and certification of ontology processes.

OntoKhoj is a portal featuring search, ranking, aggregation and classification services, focused on ontologies available on the Web (Patel et al., 2003). In order to determine the ontology subject and then classify it, textual data are extracted, such as the names of concepts and relations. Afterwards, these data are used as entries in a classification text model, trained with standard machine-learning algorithms.

Brewster et al. (2004) propose ontology evaluation through comparison with a data corpus, aiming to measure the structural proximity between ontologies and the domain. In order to accomplish this comparison, two alternative strategies are provided: (i) the automatic extraction of terms present in a corpus; (ii) the creation of a vectorial representation of terms presented in texts and in ontologies. After the identification and the extraction of terms, a query expansion aided by WordNet (Miller, 1995), as well as a mapping of the ontology terms to the corpus terms, are executed. Finally, a probabilistic approach is used to identify from among the ontologies selected which will best represent the domain.

Lozano-Tello & Gómez-Pérez (2004) introduce an approach to ontology evaluation based on a taxonomy of 160 features that is delivered in five dimensions: content, language, methodology, tool and cost. These dimensions are defined through a set of factors, which are the elements required for the analysis to obtain the value of the dimension proper. Each factor, in turn, is defined through a set of features, which permit the calculation of a representative value with respect to its appropriateness. Features may be defined recursively by other sub-features. Dimensions, factors and features are named *criteria*, and gathered in a reference ontology. The method, called Ontometric, tries to obtain each criterion value from formulas that account for weights assigned to terms. The goal is to select the best option from among alternatives and decide on the suitability of a certain ontology for a project.

2.1.2. The object of evaluation criteria

The second set of criteria proposed, object of evaluation, considers ontologies as complex structures. Therefore it is more practical to evaluate parts of the structure separately (Brank et al., 2005). Figure 2 depicts a brief description of the criteria classified as object of evaluation.

Object of evaluation	Lexicon and vocabulary	Taxonomy	Semantic relations	Context or application	Syntax	Structure and architecture
Description	Emphasizes the handling of concepts and instances, and the vocabulary used to identify them.	Emphasizes taxonomic relations (is-a relation).	Evaluates other relations, which are not taxonomic relations.	Evaluates ontologies in their context of use and in the context of application of which the ontology itself is part.	Evaluates ontology conformity to syntactical requirements of formal language in which the ontology was developed.	Evaluates ontology conformity to pre-defined structural requirements.
Example	(Aguado et al., 1998; Velardi et al., 2005; Spyns, 2005)	(Kozaki et al., 2006; Guarino & Welty, 2002; Sleeman & Reul, 2006)	(Gangemi et al., 2001)	(Ding et al., 2004; Supekar, 2005; Lewen et al, 2006; Alani & Brewster, 2006; Fernández, Cantador & Castells, 2006; Sabou et al., 2006)	(Gómez-Pérez, 1994)	(Gómez-Perez, 2004)

Fig. 2. Examples of research classified by object of evaluation.

According to Aguado et al. (1998), people have difficulty in understanding the codification used in formal ontologies and an alternative is to translate the ontology content to natural language. Techniques from ontological engineering and from automatic generation of natural language are combined in order to build a prototype called OntoGeneration, which allows users to use Spanish to access the knowledge in a chemical domain ontology. In the experiment, the authors use an ontology building methodology to construct a domain ontology which is then integrated with a linguistic ontology. The proposal allows ontology evaluation based on natural language texts whose content can aid in the comprehension of the knowledge represented in the structure.

Velardi et al. (2005) qualify the development of domain ontologies as an activity characterized by discussion and consensus within a group of domain experts, regarding concepts and relations which make up the conceptualization. In order to make this time-consuming process easier, the authors propose a natural language comment automatic generation method utilizing formal specifications that are assigned to concepts by a system (Ontolearn). Experts make comparisons between what they know and the description provided by this system, concept by concept, evaluating the adequacy of the domain conceptualization that is automatically produced.

Spyns (2005) proposes an initiative named EvaLexon, developed as part of the OntoBasis Project. In order to build ontologies, this initiative uses techniques from information extraction and semi-automated human language technology. The author creates a simple procedure to manage results obtained from a text corpus, and then, verify its validity. The corpus is unique, and there is neither a comparison standard nor an annotated corpus. The results of the experiment are obtained through a quantitative evaluation method based on lexical metrics, and then, applied in outputs of a mining algorithm for ontology learning.

Kozaki et al. (2006) discuss criteria based on the use of role-concepts. A role-concept is a description of a role associated with an entity within a certain context. This role is altered as a function of the context in which the entity participates. This feature, called Context-Dependency, can be used as a criterion to evaluate ontologies. The authors investigate the conceptualization of role-concepts in languages such as the Ontology Web Language (OWL) and the Resource Description Framework (RDFS). Inadequate role modeling will have an influence on the semantics of the hierarchy (is-a relations), compromising the competency of the ontology to represent the domain. Through criteria based on Context-Dependency, developers are able to evaluate and modify the ontology.

The OntoClean methodology (Guarino & Welty, 2002) aims to provide evaluation guidelines based on formal ontological principles obtained from metaphysical ontology. The methodology considers four formal and generic notions for evaluating ontologies: rigidity, identity, dependency and unity. These notions are related to meta-properties which may be used to understand the nature of relations in ontologies and to detect inconsistent modeling in taxonomies. The steps for error-correction are: (i) creating labels for each property and assigning meta-properties to them; (ii) selecting rigid properties; (iii) evaluating taxonomy according to these meta-properties; (iv) completing taxonomy with concepts and relations.

Sleeman & Reul (2006) present a proposal, which is similar to OntoClean, but does not use meta-properties for evaluation of inconsistencies in taxonomies. Their alternative, called CleanOnto, considers definitions to describe each concept. According to the authors, the key concept "definition" corresponds to the path from the concept to the root node of the ontology. Therefore, the definition of a concept C consists of C itself, its parent, the parent of that parent, and so on, reaching the root node, named Entity. The definitions, that is, the data used to create the path to the ontology root, are obtained through queries made to WordNet (Miller, 1995).

Collections, libraries or repositories of ontologies provide various evaluation options through, for example, cross references among classes belonging to different structures. The Swoogle search engine aims to find Semantic Web documents (Ding et al., 2004). The application uses an algorithm named Ontology Rank, which assigns weights to on-line documents as a function of hyperlinks, generating a graph of arcs defined by semantics. This functionality allows for a comparison between gathered structures, and defines whether each of them is indeed an ontology. According to Supekar (2005), this kind of characterization is not easy to obtain, because a reasonable hyperlink net between ontologies cannot yet be found. The author proposes a system allowing the ontology user to provide metadata about it with subjective information. Such a system, called Knowledge Zone (Lewen et al., 2006), is a repository that stores ontologies, which are then characterized by means of metadata proposed by the users themselves. The evaluation based on Open Rating System (ORS) extension is performed via ranking obtained though peer-reviews regarding several dimensions of an ontology.

Alani & Brewster (2006) present a prototype of an ontology ranking system. The first phase of the evaluation process aims to allow the system to classify ontologies gathered by search engines, according to their capacity to conform to specific metric standards. The system applies four types of metrics: (i) a class match measure, corresponding to the ontology coverage concerning the search terms provided; (ii) a density measure, which is related to the capacity to represent a concept through subclasses, sibling-classes and attributes; (iii) a semantic similarity measure, which calculates the proximity of compatible classes in relation to search terms in an ontology and (iv) a betweenness measure, which measures the centrality of classes in an ontology.

Fernández et al. (2006) present a system for selecting which ontology is the most appropriate to represent a domain, based on an informal description. The environment consists of three components: the first receives the problem description based on a set of terms; the second applies automatic evaluation criteria to ontologies in order to determine which one fits the problem better; the third component uses manual ontology evaluation, which incorporates the users' collaborative assessment. These components correspond to three steps of system use: problem definition, system recommendations and collaborative assessment.

Sabou et al. (2006) discuss aspects of evaluation-related ontology selection and selection requirements that impact future evaluations. The proposed process is characterized by the following steps: (i) ontology identification, in relation to a need for information; (ii) selection, based on an algorithm which evaluates a set of ontologies and selects those classified according to criteria; (iii) selection results, which may be a list of ontologies classified according to certain criteria, or a combination of ontologies that satisfies a need for information. The requirements for the selection of ontologies, defined according to usage scenarios, are: automation, performance, heterogeneity, modularity, combination of knowledge sources, relations and semantic match.

Among earlier initiatives to establish systematic ways of evaluating ontologies, it is worth noting the criteria presented by Gómez-Pérez (1994), which emphasize concepts and those definitions that compose these concepts. The proposal suggests the verification of the ontology structure and definitions' syntax, relating the content issue to three problems: (i) consistency, meaning that the entry itself will not simultaneously obtain contradictory conclusions; (ii) completeness, which concerns the degree to which an ontology is able to represent the real world and (iii) conciseness, which concerns the need for consistency among all the information gathered in the ontology. Gómez-Pérez (2004) describes errors in the frame-based ontology building process, based on her evaluation experience with the Ontolingua Server library (Farquhar et al., 1996). Within the scope of these libraries, the primitives are: subclass-of, instance-of, partition-class, partition-subclass, where partition is a set of disjoint classes. Three types of errors are described: consistency errors, completeness errors and redundancy errors.

In spite of the great variety of methodologies presented above, according to Supekar (2005) there are no qualitative methodologies for evaluating ontology content. We understand that there are, as yet, no existing initiatives designed to verify whether the acquired knowledge is indeed present in the domain in question, from an expert's point of view. In the approach adopted for our study, the ontology studied is not evaluated using a set of criteria given by developers. Instead, we permit experts to browse through the ontology using a prototype system, and then to verify it as to content adequacy and to record their impressions through questionnaires about information quality.

2.2. *Foundation of ontology content evaluation criteria*

The development of an ontology content evaluation proposal consists of the creation of questionnaires and the development of a prototype. This section describes multidisciplinary approaches adopted in order to establish foundational criteria for questionnaires. The prototype usage, in addition to building and evaluating processes, is presented in Section 3.

We have combined a set of approaches, using concepts from distinct research fields to establish a content evaluation framework. The concepts include: (i) information quality, an approach used in Information Science to study user information needs and information uses; (ii) competency questions, which are used to define the scope of an ontology and (iii) learning educational objectives, which are used in the educational field to verify whether a person learns properly. We have selected criteria from these approaches that are related to content assessment.

Information Quality is an approach mentioned by several authors in the Information Science field. The Health Summit Working Group (HSWG),[1] for example, defines the following categories for evaluating the quality of information provided on the Web: credibility, content, disclosure, links, design, interactivity and caveats. In relation to the content category, the HSWG defines the following evaluation criteria: accuracy, related to the accuracy of information sources; evidence hierarchy, in which precision is based on evidence supporting the information; announcements, which describe limitations, goals, coverage and

[1]See: Quality of health information on the Internet. Retrieved June 16, 2009, from http://www.ieee.org/organizations/pubs/newsletters/npss/march2000/health.htm.

authority of the information; and completeness, which considers facts, negative results and statements about the subject that must be included.

According to Olaisen (1990), factors related to information quality recorded on digital media are separable into four categories: (i) cognitive quality, which includes credibility indicators, relevance, authority, validity and meaning over time; (ii) information project quality, which incorporates factors related to shape, flexibility and selectivity; (iii) information product, related to accuracy and precision and (iv) quality of transmission, which includes the accessibility criterion.

The Kahn et al. (1997) model accounts for the social dimension of information quality, from the user's point of view, and defines it along four dimensions: relevance, interpretability, credibility and reputation. The authors introduce a methodology for evaluating information quality named Assessment Information Quality Methodology (AIQM) (Lee et al., 2002), which consists of an information quality model, questionnaires for measuring information quality and techniques for metric interpretation. In order to develop these instruments, the methodology includes contributions from academics and practitioners. The result, another model named PSP/IQ, is a matrix comprising several cross-dimensional evaluation categories: (i) conformity to specifications, (ii) the ability to meet or to exceed user expectations, (iii) the quality of information products and (iv) the quality of information services.

Lancaster (1989) defines a set of specific criteria used by users to evaluate information retrieval systems in libraries, including those related to information quality: coverage, retrieval, accuracy, novelty, authority. Parasuraman et al. (1988) propose a scale named SERVQUAL, which aids in the evaluation of information quality services. Through this scale, users evaluate service quality by comparing their expectations with what is really obtained. The model demonstrates the gap between expectations and what the service offers.

The information quality approach seems to be a feasible alternative for evaluating ontology content because it provides objective criteria for the verification of whether the knowledge in the modeled domain was acquired by the structure. This evaluation indicates whether or not the ontology represents the knowledge which is relevant in a context and achieves its stated goals.

The competency questions are used within the scope of the Toronto Virtual Enterprise Ontology Project (TOVE) to specify tasks and problems that may be solved by an ontology, even before one is built (Fox, 1992). They aid in defining scopes and features of an ontology under construction. According to Kim et al. (1999), for each domain explored, the validity of an ontology-based system lies in its capacity to answer competency questions posed. The ontology's competency is determined by the following steps: (i) scenario statement, consisting of a narrative about business questions and about problems which ontology-based systems must be able to address; (ii) scope statement, consisting of a hypothesis about a domain in order to clarify its limits; (iii) problem statement, consisting of the recognition of the general problem that justifies the necessity for building an ontology; (iv) user competency questions statement, consisting of specific competency questions, motivated by scenario statements and prepared according to a user's needs and (v) developer requirements statement, consisting of questions characterizing the ontology project requirements.

Competency questions are suitable to be used in the process of building an ontology. Moreover, they also seem to be a feasible alternative to a posterior validation of ontologies. Once the structure is ready, it is possible to make relevant queries to users. Analysis of the responses will therefore indicate that the ontology is able to acquire knowledge in the domain under analysis.

In the educational field, it is important to assess whether specific content was learned by a person during the teaching process. An approach to this kind of evaluation is the Bloom Taxonomy (Bloom, 1956), which establishes learning educational objectives. This hierarchy, consisting of six categories, represents the learning process and the knowledge about a subject through a spectrum ranging from a simple behavior to a complex one. The categories of the Bloom Taxonomy, ranging from a low level of knowledge acquisition (called *knowledge*) to a higher level (called *evaluation*), are: (i) knowledge: defined as the capacity to remember concepts previously learnt, or to retrieve the proper information; (ii) understanding: defined as the ability to understand the meaning of knowledge, for example, explaining or summarizing; (iii) application: defined as the ability to use knowledge gained in new situations; (iv) analysis: corresponds to the ability to divide knowledge into parts in order to understand its structure; (v) synthesis: corresponds to the ability to gather pieces of knowledge in order to obtain a whole and (vi) evaluation: relates to the ability to judge the value of knowledge for a specific purpose.

The Bloom Taxonomy, when trying to evaluate whether specific content was acquired, represents an alternative for the evaluation of ontology content. We believe adaptations are necessary to keep the basic foundations of the theory which are expressed in the six categories presented.

These three approaches (information quality, competency questions and educational objectives) give rise to three kinds of questionnaires employed in the company where our present research was conducted. The use of the prototype by the employees and their responses to the questionnaires made the evaluation of ontology content possible. In the following section, we present the procedures conducted in the company in which we built and evaluated an ontology representative of organizational processes.

3. Research: Building and evaluating organizational ontology

This study presents research conducted in a large Brazilian energy utility company where we built an ontology representative of a corporate management system that is based on international quality standards. The building process is described briefly in Section 3.1. More attention is paid to the last step of the research, presented in Section 3.2, related to the content evaluation of the resultant ontology.

3.1. Ontology building

The subject of the research is an organization named *Companhia Energética de Minas Gerais* (CEMIG), a Brazilian energy utility with nearly 11,000 employees, serving around 6 million consumers and operating the longest energy distribution line in Latin America. CEMIG has 46 hydroelectric power plants, 2 thermal plants and 1 wind power plant. The company's quality guidelines are put into practice by the Management System (MS). Designed over a period of 5 years, the MS complies with international quality standards (ISO-9001, ISO-14001 and OHSAS-18001) and includes three kinds of systems: Quality-MS, Environment-MS and Occupational Health & Safety-MS. These systems represent a corporate investment in quality, the environment and health & safety policies. They made it possible for the company to be listed in 2007, for the eighth consecutive time, in the Dow Jones Sustainability Index,[2] which includes companies from around the globe.

The organizational ontology was developed as a model specification for the MS. The necessary expertise was obtained from employees from the corporate area responsible for directing the MS. The research methodology consists of four main phases: (i) determining the research tools; (ii) data collection for the design of the ontology; (iii) building the ontology; (iv) ontology content evaluation. The ontology building process was based on experiments of Fernandez et al. (1997) and Gandon (2002), as well as complementary contributions from Uschold & Gruninger (1996) and Noy & McGuinness (2001).

[2]See: Dow Jones Sustainability Index. Retrieved June 16, 2009, from http://www.sustainability-index.com.

In the remaining part of this section, we briefly describe the first three steps. The last step, evaluation, is presented in Section 3.2.

The research tools adopted were: (i) data collection methods; (ii) forms for recording data collected in order to design the ontology; (iii) forms for building the ontology. The choice of the data collection method was based on knowledge acquisition proposals found in popular methodologies. The procedures selected originated from: Joint Application Design (JAD), scenario design and analysis (Benner et al., 1993) and subject analysis (Lancaster et al., 1989). For recording data, we created forms based on the survey on knowledge acquisition methods. These forms and details of their use are outside the scope of this paper, whose emphasis is on the evaluation process.

The data collection process conducted for the research was based on: (i) the formats and goals of the interviews; (ii) the use of system and document analysis techniques; and (iii) the analysis of other ontologies, which provided terms for the ontology under construction. The first meetings enabled us to choose the processes and the experts who were to take part in the research. We selected eight specialized MS employees, with a university-level education and different roles in the system context. Afterwards, we selected critical processes, chosen by employees and management staff, for MS deployment in company areas.

The main goal of the interviews was to obtain candidate-terms for ontology concepts. There were, however, some specific goals. The interviews occurred in four stages, and are therefore identified as Interviews 1, 2, 3 and 4 with the objectives described as follows: (i) Interview 1 was planned in order to clarify the role of each employee in the company; (ii) Interview 2 obtained narratives about organizational context; (iii) Interview 3 obtained definitions of identified terms and (iv) Interview 4 aimed at obtaining a consensus on the definitions from Interview 3. Each interview was recorded and transcribed with a text editor.

In order to organize the data collected, we adopted the following forms: *Summary of Interview* (Interview 1), *Scenario Report* (Interview 2); *Semi-Informal Terminology* form, *Intensions & Extensions Forms* (Interviews 3 and 4). We also used the *Instances, Relations & Concepts Semi-formal* form (Fig. 3),

Instances, Relations & Concepts Semi-Formal Form				**FORM. nº**
Date: 05/03/06 **Project:** CEMIG	**Data source:** Intensional and extensional forms	**This Table:** 1 form 2 v1	C/A = concrete or abstract; I/C = instance or concept; NA = abstract level.	

Nº	**View**	**Parent concept**	**Concept**	**C/A**	**Relation**	**Instance or Class**	**I/C**
C1	CO	Institution	Company	C	Name	Cia Energética de MG	I
					Abbreviation	CEMIG	I
					Has-One	Management System	C
					Invests-In	Quality Management	I
					Invests-In	Environmental Management	I
					Invests-In	Health and Security Manag.	I
					Has-One	Strategic planning	I
					Has-One	Strategic View	C
					Has-One	Mission	C
					Primary services	Power Sector	CC

Fig. 3. Example of a Semi-Formal Concepts Form.

to organize data and insert it into the tool selected for implementation. The content of the last form was entered into Protégé (Noy et al., 2001) and then, the Protégé results were exported to RDFS code. The objective was to use the forms to obtain, gradually and throughout the process, a new formal knowledge structure. Once the forms had been completed, syntheses, candidate terms and reports and tables were selected to aid the building of the ontology based on subject analysis premises.

As the final activity in this stage, general and organizational terms were extracted from high-level ontologies (Sowa, 2000; Niles & Pease, 2001; CyCorp, 2005) and organizational ontologies (Gruninger & Fox, 1995; Uschold et al., 1998; Gandon, 2001). Terms and definitions were selected to build the high-level layers of the new ontology. We chose to extract terms manually, in order to better evaluate their meaning in the original ontology.

The resultant ontology was divided into three layers, so that two of them, the high-level layers, could be reused: the abstract layer had general concepts, which could be reused in several other initiatives; the organizational layer had concepts, which could be reused in any organization. The third layer, namely the specific layer, was built according to peculiarities found in the company, which restricted its widespread reuse. The final structure contained 251 classes, with 109 being related to the high-level layers and 143 being related to the specific layer, in addition to 409 slots, including binary relations and attributes that were defined.

3.2. Ontology evaluation

Having designed and constructed the ontology we proceeded to content evaluation. The goal, as we have mentioned, was to confirm whether the ontology was representative of organization knowledge acquired in the delimited scope. The evaluation process consisted of two phases: prototype use and questionnaire answers.

The prototype was developed as an Extended StyleSheet Language Transformation (XSLT) and it was in fact a simple search engine running in a Web browser. The goal was to allow searches for terms corresponding to concepts, stored in an RDFS file produced by Protégé. The search results were obtained by exploring RDFS tags. There were four search functionalities implemented in the prototype. Firstly, a search and browse interface allowed users to search for concepts, relations, available relations to a concept and available concepts to a relation. Secondly, a concept hierarchy allowed users to access concepts and relations through hyperlinks arranged in a taxonomy. Thirdly, a hyperbolic style view of concepts enabled the user to better visualize the structure as a whole and understand the context of concepts during a search. Fourthly, a management interface allowed different ontologies to be loaded into the prototype.

The prototype was not planned as a tool for end users, but as an application to aid in the evaluation process after the conceptual modeling activity. After a brief pre-test, we presented the prototype to company experts involved in the research. We asked them to use the prototype for one week, and then to answer the evaluation questionnaires. Moreover, they were given a manual with explanations about the prototype. Prototype screens are depicted in Fig. 4(a–c).

Following the use of the prototype, participants answered questionnaires, aiming to prove that the ontology was representative of organizational knowledge. The questionnaires were prepared based on three different dimensions, as mentioned in Section 2.2: competency questions, information quality and educational objectives taxonomy. They were named Questionnaires 1, 2 and 3, respectively.

(a)

(b)

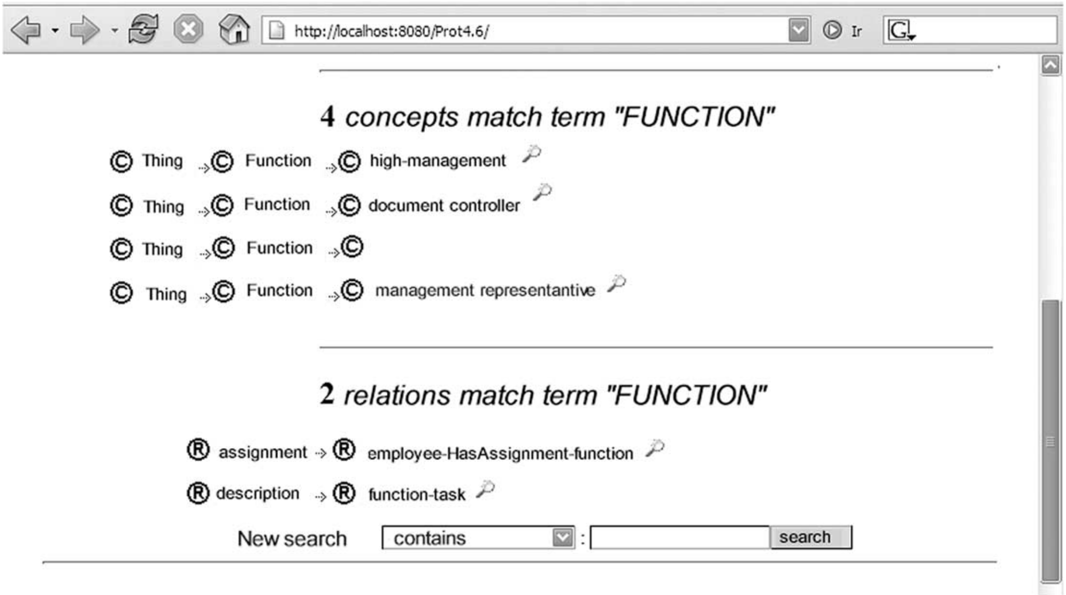
Fig. 4. (a) Concept search interface initial screen. (b) Taxonomy of concepts screen. (c) Results screen, search by term "function".

Questionnaire 1 (Fig. 5) presented employees with competency questions which the ontology was able to answer, and asked them to evaluate whether those questions were compatible with their expectations. Results would indicate whether the ontology was able to represent relevant organizational knowledge.

Questionnaire 2 (Fig. 6) used only criteria related to content evaluation: volume appropriateness, credibility, completeness, correctness, interpretation, objectivity, updating, relevance and understanding.

The educational objectives taxonomy was adapted to make Questionnaire 3 (Fig. 7). With this approach it is possible to evaluate whether the model contains organizational knowledge.

Participants answered the questionnaires simultaneously using the prototype within the agreed period of one week. Then, the results were organized into tables, two for each kind of questionnaire. In the first table, we registered the total number of responses related to each criterion on a scale (1–5) subsequently calculating a weighted mean of the results. In the second table, we registered the responses of each employee for each form question, and then we calculated an arithmetical mean for the question. Afterwards, we calculated a weighted mean of all the previous means. Therefore, in the cases of Questionnaires 2 and 3, we also registered the dimensions of quality criteria used.



(c)

Fig. 4. (*Continued.*)

Read the possibilities below and answer to what extent they meet your needs at work, according to the scale on the right. In this scale, 1 corresponds to “it does not meet the needs” and 5 corresponds to “fully meets the needs”.	
A list of all certified core sectors, by management system kind	- □□□□□ + 1 2 3 4 5
A list of all core sectors in certification process, by management system	- □□□□□ + 1 2 3 4 5
A list of all third party audits conducted by a consultant specialist in certification	- □□□□□ + 1 2 3 4 5
A list of dates to certification renewals, by system, by month and by year	- □□□□□ + 1 2 3 4 5

Fig. 5. Portion of Questionnaire 1 (from a total of 20 questions).

The volume of information is enough for the needs of the corporation	- □□□□□ + 1 2 3 4 5
The information is trustworthy.	- □□□□□ + 1 2 3 4 5
The information features inconclusive credibility.	- □□□□□ + 1 2 3 4 5

Fig. 6. Portion of Questionnaire 2, credibility (from a total of 36 questions).

The information enables judging adequacy of conclusions.	- □□□□□ + 1 2 3 4 5
The information enables judging facts based on internal parameters.	- □□□□□ + 1 2 3 4 5
The information does not enable proper conclusions.	- □□□□□ + 1 2 3 4 5

Fig. 7. Portion of Questionnaire 3, evaluation criterion (from a total of 25 questions).

4. Evaluation results

This section discusses the results obtained from the questionnaires. As described in Section 3.2, the questionnaires were answered by eight participants responsible for different functions related to the MS system. Sections 4.1–4.3 present the results for three kinds of questionnaires and Section 4.4 discusses the evaluation results.

4.1. Competency questions dimensions results

Figure 8(a) depicts a weighted mean (4.16) related to values assigned by participants according to the 1-to-5 scale in Questionnaire 1. This value indicates that the results obtained were positive. This means the model was able to answer the common questions in the staff's work environment. Therefore, we observed that 77% of the answers corresponded to high values on the scale (4 or 5).

Figure 8(b) depicts the following: a sample of assertions from Questionnaire 1, the arithmetical mean of responses according to a 1-to-5 scale and a general mean for all responses. The arithmetical mean calculated for the questionnaire assertions (4.10) indicates that the needs of personnel were met by the ontology content.

4.2. Information quality dimension results

Figure 9(a) depicts the total number of responses for each value using a 1-to-5 scale in Questionnaire 2, where number 1 corresponds to "I do not agree with the statement" and number 5 corresponds to "I agree with the statement".

Figure 9(a) depicts a weighted mean (4.26) related to values assigned by participants according to the 1-to-5 scale in Questionnaire 2. The value obtained indicates that results were positive from an information quality viewpoint. That is, 82% of the responses corresponded to high values in the scale.

Figure 9(b) depicts the following: a sample of assertions from Questionnaire 2, the arithmetical mean of responses according to a 1-to-5 scale and a general mean of all statements. The arithmetical mean calculated (4.26) indicates positive results. In relation to the dimensions, the means are: 3.83 for proper volume; 4.61 for credibility; 3.97 for completeness; 4.33 for correctness; 3.94 for interpretation; 4.22 for objectivity; 4.61 for updating; 4.50 for relevance; and 4.56 for understanding. These results indicated that the ontology was able to acquire knowledge that was relevant for participants and to present it properly according to the criteria and dimensions.

Scale	Meaning	Total of response per value
1	It does not comply	1
2	↓	11
3	↓	17
4	↓	35
5	It complies	62
Weighted mean of scale criteria ……………………………4.16		

(a)

Statement of questionnaire 1	Mean
A list of all core sectors in the certification process for each management system.	4.83
A list of all external audits accomplished by a consultant specialized in certification.	4.50
A list of dates by month and year when each system's certifications must be renewed.	4.50
A list of open Non-Conformity Reports for each core sector.	4.00
A list of costs necessary to meet quality requirements for each core sector.	4.33
A list of employees who participate in internal certification processes per date.	3.50
A list of obsolete and invalid documents per date.	3.17
General mean of responses…………………………………………………………… 4.10	

(b)

Fig. 8. (a) Answers by criterion and criteria mean. (b) Portion of data obtained in Questionnaire 1.

4.3. Educational objectives dimension results

Figure 10(a) depicts the following: the total number of responses of each value according to a 1-to-5 scale from Questionnaire 3, where the number 1 corresponds to "I do not agree with the statement" and the number 5 corresponds to "I agree with the statement".

Figure 10(a) depicts a weighted mean (4.09) related to the values assigned by participants according to the 1-to-5 scale. The value obtained indicates that the results were positive. The ontology was relevant to the staff's work. Therefore, 79% of the responses corresponded to high values in the scale (4 and 5).

Figure 10(b) depicts the following: a sample of assertions from Questionnaire 3, the arithmetical mean of responses according to a 1-to-5 scale and a general mean for all assertions. The arithmetical mean of the statements indicates positive results. In relation to dimensions, the means are: 4.42 for knowledge, 4.17 for comprehension, 4.5 for application, 4.13 for analysis, 3.88 for synthesis and 3.71 for evaluation. These results indicate that the ontology was able to acquire knowledge according to the criteria and dimensions considered.

4.4. Ontology use and evaluation in the organization

Since we have presented the evaluation results, it is also worth analyzing three relevant facts in order to better characterize the feasibility of this proposal. First, we consider the organizational context in which the ontology was produced and evaluated. Two, there is the relation between the ontology and the

Scale	Meaning	Total of responses per value
1	I do not agree	3
2	↓	4
3	↓	32
4	↓	72
5	I agree	105
Weighted mean of scale criteria 4.26		

(a)

Direction	Statement of questionnaire 2*	Mean**
proper volume	The information volume is enough for your needs.	4.00
	The information volume does not attend to your needs (I).	4.33
	The information volume is not enough for your needs (I).	3.83
	The information volume is neither large nor small.	3.17
	...	
mean for "proper volume" direction 3.83		
understanding	The information is easy to understand.	4.67
	The meaning of the information is hard to understand. (I)	4.50
	The information is enough for your work.	4.83
	The information is easy to apprehend.	4.17
	The meaning of the information is easy to understand.	4.50
	...	
mean for "comprehension direction" 4.56		
General mean of responses 4.26		

(b)

Fig. 9. (a) Criterion mean and answers by criteria for Questionnaire 2. (b) Portion of data obtained from Questionnaire 2.
* The statements tagged with (I) are inverse in relation to other statements and inserted in order to check the consistency of the answers.
** To calculate means, inverse statements were considered as inverse weight because one negative answer to this statement was considered a positive result. For example, an inverse statement with a grade of 1 is considered as having a grade of 5.

organizational business needs. Finally, we consider how the employees and experts related to the ontology, that is, our understanding of their activities and their level of competency as required to evaluate content. The remaining part of this section discusses these three topics and the related activities that took place during the research.

In order to better understand the organizational context in which the research occurs, some details on the company's operations have been provided. Nearly 6,000 employees work directly or indirectly with the certified Management System (MS) processes. Over the last 6 years, a gradual effort has been made to extend the MS to the organization as a whole. The MS is established in an independent functional structure that overlaps the organizational chart. In order to use the system, the employees of a unit must understand the structure and know the functions defined in this structure and their respective responsibilities.

The organizational unit responsible for the creation, establishment, control and maintenance of MS is a supervisory unit called Quality Control and Environmental Supervision (QC). They must establish the

Scale	Meaning	Total of responses per value
1	I do not agree	2
2	↓	4
3	↓	25
4	↓	67
5	I agree	52
Weighted mean of scale criteria 4.09		

(a)

Direction	Statements of questionnaire 3	Mean
knowledge	The principles of the subject considered are present.	4.17
	Terms and concepts used in your work are present.	4.50
	Procedures of your work are present.	4.17
	...	
"Knowledge" direction mean 4.42		
evaluation	The information allows you to judge the adequacy of a conclusion.	3.83
	The information allows you to judge facts via internal guidelines.	3.33
	The information allows you to judge via external guidelines.	3.17
	...	
"Evaluation" direction mean... 3.71		
General mean of responses 4.13		

(b)

Fig. 10. (a) Criteria mean and answers by criteria. (b) Portion of data obtained from Questionnaire 3.

MS in other units, control the periodical certifications and maintain the necessary conditions to renew these certifications over time. As resources to accomplish these tasks, they have a set of documents that specify the system through standard procedures and a team of 20 employees, out of which 8 are allocated to specific activities related to MS.

According to the QC supervisor, the MS has been gradually becoming a kind of "organizational language" used by more than 5,000 employees to communicate. Soon, those employees who have little or no understanding of this language will face difficulties in corporate communication. Moreover, as the MS expands geographically, it is becoming evident how difficult it is to maintain consistency in the terminology used. In this scenario, the design of the ontology was planned and delivered via the corporate intranet, allowing employees to access the agreed terminology of the MS. Plans are being made to use the ontology to train new employees and also employees participating in internal auditing of MS processes. Thus, the ontology demonstrably meets the organizational business needs related to the MS.

In this context, the ontology content evaluation came to be relevant as the knowledge represented in the ontology would be delivered across the organization. Primarily, this evaluation could be performed only by the experts who took part in the ontology building process to ensure that the knowledge acquired was equivalent to that which was available throughout the modeling process. If the domain experts could not evaluate the ontology content, it would be possible to spread incorrect, incomplete, inconsistent or low quality knowledge throughout the company. These experts participated in the creation of the MS and are also responsible for its maintenance, thus assuring that they have the necessary expertise to perform the evaluation process.

The ontology is an integral part of several of those tasks pertaining to the experts' workflow. Among these tasks are: (i) reaching a better understanding of the domain; (ii) obtaining improvements in communication processes; (iii) reaching a unique and shared general agreement among the experts and users, related to the terms in use; (iv) correcting inconsistencies in original documents that specify the MS; (v) obtaining a common terminology to develop automated information systems, thus improving interoperability. The last task mentioned is noteworthy as it involved standardization of terms and because it became a routine after the building and evaluation of the ontology.

Besides the use of ontology, the evaluation process conducted during the research became a part of the daily maintenance activities of the experts. The ontology is updated every time the language dynamics in the organization requires purposeful changes. During each update, the ontology content is evaluated again by the experts, proving its adequacy before it is delivered across the organization. Our expectation is that soon this interactive process of ontology improvement will allow us to accomplish the content evaluation with end users and not only with experts.

5. Summary and conclusions

In this study, we have presented a proposal for ontology evaluation which emphasizes issues related to content. Content evaluation is considered to be an important part of the modeling process. This process verifies whether a model specification produced from a knowledge acquisition process corresponds, from the viewpoint of experts, to the organizational knowledge. In order to contextualize this proposal, we have been conducting a survey of ontology evaluation initiatives, classifying them according to the type of approach and the object of evaluation. We concluded that the issues regarding content evaluation are not always considered important; furthermore, we noticed a lack of formal approaches related to this kind of evaluation. We presented information about the construction of an organizational ontology related to the quality management system of an energy utility company. Afterwards, we described the content evaluation process of that ontology and reached a conclusion about the feasibility of the proposal.

In order to highlight the significance of the content evaluation process, it is important to consider the practical viewpoint adopted for systems development. The proposal is supported by possibilities for improvements in modeling related to human communication processes. An ontology is a specification of a model which represents a conceptualization. In order to verify the adequacy of the specification, it is a good idea to aid the domain experts in verifying the quality of the knowledge present in the ontology according to scientific criteria. The proposal presented was conducted within this scope. In this study conducted in CEMIG, the ontology was planned to be available on the corporative intranet in order to aid the enlargement of MS and to become the cornerstone for software development, used by internal and external development teams.

It is also worth emphasizing the well-established theories provided by other research fields applied to this ontology content evaluation proposal. As described, we have borrowed premises from Information Science (information quality), from ontology building methodologies (competency questions) and from the Educational field (taxonomy of educational objectives). These three approaches gave rise to questionnaires which were shown to be simple and efficient when applied in a corporate setting. Throughout the study, the participants did not encounter difficulties in understanding the questions. Furthermore, they reported that the task promoted motivating discussions and reflection.

The prototype system developed, although simple, shows effectiveness in an organizational environment, and we regard it as a contribution. However, improvements in the prototype should be considered in order to make it more effective in different situations. Some areas needing improvement that have been identified include: more flexibility in queries, which could be made possible with the addition of more options to match search terms and ontology terms; improvements in the query ranking algorithm, since the algorithm used allows only alphabetical order and is based on the existence of the "label", "comment" and "about" RDFS tags; modification of client-server application loading, since moving some instructions to the server-side could reduce application constraints.

In spite of the simplicity of the prototype and the questionnaires, the proposal is feasible from a theoretical and practical point of view and does not lack theoretical foundation. Following the presentation of the results of the study, future research will now be considered. Perhaps the most important need is to conduct additional research in other companies, using the proposal presented. This study is a case study conducted in one organization, with decidedly positive results. In order to characterize the proposal as a generic methodology and allow for its systematic use, it is necessary to undertake similar research in other companies.

References

Aguado, G., A. Bañón, A., Bateman, J., Bernardos, S., Fernández, M., Gómez-Pérez, A. Nieto, E., Olalla, A., Plaza, R. & Sánchez, A. (1998). Ontogeneration: Reusing domain and linguistic ontologies for Spanish text generation. In *Workshop on Applications of Ontologies and Problem-Solving Methods at The European Conference on Artificial Intelligence (ECAI'98)*, Brighton, UK. Retrieved 1 October 2008, from http://fuencarral.dia.fi.upm.es:8080/145/1/ ONTOGENERATION._Reusing_Domain_and_Linguistic_Ontologies_for_Spanish_Text_Generation.pdf.

Alani, H. & Brewster, C. (2006). Metrics for ranking ontologies. In *International EON Workshop located at the 15th International World Wide Web Conference*, Edinburgh, UK. Retrieved 10 January 2008, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006alanietal.pdf.

Benner, K.M., Feather, M.S., Johnson, L. & Zorman, L.A. (1993). Utilizing scenarios in the software development process. In N. Prakash, C. Rolland & B. Pernici (eds), *Information System Development Process* (pp. 117–134). North-Holland: Elsevier Science Publishers.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longman.

Brank, J., Grobelnik, M. & Mladenić, D. (2005). A survey of ontology evaluation techniques. In *Conference on Data Mining and Data Warehouses*, Ljubljana, Slovenia. Retrieved 20 July 2007, from http://kt.ijs.si/dunja/sikdd2005/Papers/BrankEvaluationSiKDD2005.pdf.

Brank, J., Grobelnik, M. & Mladenić, D. (2006). Gold standard based ontology evaluation using instance assignment. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, UK. Retrieved 3 July 2007, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006branketal.pdf.

Brewster, C., Alani, H., Dasmahapatra, S. & Wilk, Y. (2004). Data driven ontology evaluation. In *International Conference on Language Resources and Evaluation*, Lisbon, Portugal. Retrieved 12 December 2007, from http://eprints.aktors.org/337/02/BrewsterLREC-final.pdf.

Brown, J.S. & Duguid, P. (2000). *The Social Life of Information*. Cambridge: Harvard Business School.

Cycorp, Inc. (2005). What's in Cyc? Retrieved 20 March 2005, from http://www.cyc.com/cyc/technology/whatiscyc_dir/whatsincyc.

Ding, L., Pan, R., Joshi, A., Kolari, C., Java, A. & Peng, Y. (2004). Swoogle: a search and metadata engine for the semantic web. In *13th ACM Conference of Information and Knowledge Management*, Washington, DC, USA. Retrieved 11 December 2007, from http://www-scf.usc.edu/~csci586/paper/p652-liding.pdf.

Farquhar, A., Fikes, R. & Rice, J. (1996). The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human–Computer Studies*, *46*(6), 707–727. Retrieved 15 December 2001, from http://citeseer.ist.psu.edu/farquhar96ontolingua.html.

Fernández, M. (1999). Overview of methodologies for building ontologies. In *Workshop on Ontologies and Problem-Solving Methods*, Stockholm, Sweden, Retrieved 19 July 2005, from http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf.

Fernandez, M., Gomez-Perez, A. & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In A. Farquhar & M. Gruninger (eds), *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering* (pp. 33–40). Menlo Park: AAAI Press.

Fernández, M., Cantador, I. & Castells, P. (2006). Core: A tool for collaborative ontology reuse and evaluation. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, UK. Retrieved 3 June 2007, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006fernandezetal.pdf.

Fox, M.S. (1992). The TOVE Project: towards a common-sense model of the enterprise. In *5th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, London, UK. Retrieved 5 July 2005, from http://www.eil.utoronto.ca/enterprise-modelling/papers/index.html.

Gandon, F. (2001). Engineering an ontology for a multi-agents corporate memory system. In S.Tsuchiya, B.Ramond & J.-P. Barthès (eds), *Proceedings of The International Symposium on The Management of Industrial and Corporate Knowledge (ISMICK 2001)* (pp. 209–228). Compiègne: Université de Technologie.

Gandon, F. (2002). Distributed artificial intelligence and knowledge management: ontologies and multi-agent systems for a corporate semantic web. PhD Thesis, School of Sciences and Technologies of Information and Communication, INRIA and University of Nice, Nice, France.

Gangemi, A., Catenacci, C., Ciaramita, M. & Lehmann, J. (2006). Qood grid: A metaontology-based framework for ontology evaluation and selection. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference*, Edinburgh, UK. Retrieved 3 June 2007, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006gangemietal.pdf.

Gómez-Pérez, A. (1994). Some ideas and examples to evaluate ontologies. In *11th Conference on Artificial Intelligence Applications*, Los Angeles, CA, USA. Retrieved 29 August 2002, from http://citeseer.ist.psu.edu/gomez-perez94some.html.

Gómez-Pérez, A. (1996). Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications*, *11*(4), 519–529.

Gómez-Pérez, A. (1999). Evaluation of taxonomic knowledge in ontologies and knowledge bases. In *Workshop on Knowledge Acquisition, Modeling and Management*, Alberta, Canada. Retrieved 12 August 2005, from http://sern.ucalgary.ca/KSI/KAW/KAW99/review_papers/Gomez-Perez1/index.html.

Gómez-Pérez, A. (2004). Ontology evaluation. In S. Staab & R. Studer (eds), *Handbook on Ontologies* (pp. 251–274). Berlin: Springer-Verlag.

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: AAAI Press.

Guarino, N. & Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, *45*(2), 61–65. Retrieved 24 March 2005, from http://citeseer.ist.psu.edu/guarino02evaluating.html.

Jarke, M. & Vassiliou, Y. (1997). Data warehouse quality: a review of the DWQ Project. In *2nd International Conference on Information Quality*, Cambridge, MA, USA. Retrieved 21 October 2001, from http://citeseer.ist.psu.edu/jarke97data.html.

Jones, D., Bench-Capon, T. & Visser, P. (1998). Methodologies for ontology development. In *IT&KNOWS Conference, XV IFIP World Computer Congress*, Budapest, Hungary. Retrieved 23 February 2001, from http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-onto-dev.pdf.

Kahn, B.R., Strong, D.M & Wang, R.Y. (1997). A model for delivering quality information as product and service. In D.M. Strong & B.K. Kahn (eds), *Proceedings of Conference on Information Quality* (pp. 80–94). Cambridge: MIT Press.

Kalfoglou, Y. & Hu, B. (2006). Issues with evaluating and using publicly available ontologies. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference*, Edinburgh, UK. Retrieved 5 June 2007, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006kalfoglouetal.pdf.

Kim, H.M., Fox, M.S. & Gruninger, M. (1999). An ontology for quality management: enabling quality problem identification and tracing. *BT Technology Journal*, *17*(4), 131–140.

Kozaki, K., Sunagawa, E., Kitamura, Y. & Mizoguchi, R. (2006). Fundamental consideration of role concepts for ontology evaluation. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, UK. Retrieved 3 October 2008, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006kozakietal.pdf.

Lancaster, F.W. (1989). *The Measurement and Evaluation of Library Systems*. Arlington: Information Resource.

Lancaster, F.W., Elliker, C. & Connel, T.H. (1989). Subject analysis. *Annual Review of Information Science and Technology*, *24*, 35–84.

Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University.

Lee, Y., Strong, D.M., Kahn, B.K. & Wang, R.Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, *40*(2), 133–146.

Lewen, H., Supekar, K., Noy, N.F. & Musen, M.A. (2006). Topic-specific trust and open rating systems: An approach for ontology evaluation. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International*

World Wide Web Conference (WWW 2006), Edinburgh, UK. Retrieved 3 October 2008, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006lewenetal.pdf.

Lozano-Tello, A. & Gómez-Pérez, A. (2004). ONTOMETRIC: A method to choose the appropriate ontology. *Journal of Database Management*, *15*(2), 1–18.

Maedche, A. & Staab, S. (2002). Measuring similarity between ontologies. In *13th European Conference on Knowledge Acquisition and Management (EKAW 2002)*, Madrid, Spain. Retrieved 13 September 2004, from http://www.aifb.uni-karlsruhe.de/~sst/Research/Publications/ekaw2002-compare.pdf.

Maynard, D., Peters, Y. & Li, Y. (2006). Metrics for evaluation of ontology-based information extraction. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, UK. Retrieved 3 October 2008, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006maynardetal.pdf.

Miller, G.A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39–41.

Niles, I. & Pease, A. (2001). Towards a standard upper ontology. In N. Guarino, B. Smith & C. Welty (eds), *Proceedings of the International Conference on Formal Ontology in Information Systems* (pp. 2–9). New York: ACM Press.

Noy, N.F. & McGuinness, D. (2001). Ontology Development 101: a guide to creating your first ontology. Retrieved 23 July 2003, from http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf.

Noy, N., Sintek, M., Decker, S., Crubezy, M., Fergerson, R.W. & Musen, M.A. (2001). Creating semantic Web contents with Protegé-2000. *IEEE Intelligent Systems*, *16*(2), 60–71.

Obrst, L., Hughes, T. & Ray, S. (2006). Prospects and possibilities for ontology evaluation: The view from NCOR. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference*, Edinburgh, UK. Retrieved 3 June 2007, from http://www.nist.gov/msidlibrary/doc/prospects.pdf.

Olaisen, J. (1990). Information quality factor and the cognitive authority of electronic information. In I. Wormell (ed.), *Information Quality: Definitions and Dimensions* (pp. 91–121). Los Angeles: Taylor Graham.

Parasuraman, A., Berry, L.L. & Zeitham, V.A. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–29.

Patel, C., Supekar, K., Yugyung, L. & Park, E.K. (2003). OntoKhoj: a semantic web portal for ontology searching, ranking and classification. In R. Chiang, A.H.F. Laender & E.P. Lim (eds), *Proceedings of the 5th ACM International Workshop on Web Information and Data Management* (pp. 58–61). New York: ACM Press.

Porzel, R. & Malaka, R. (2004). A task-based approach for ontology evaluation. In *Workshop on Ontology Learning and Population at the 16th European Conference on Artificial Intelligence ECAI*, Valencia, Spain. Retrieved 7 June 2007, from http://olp.dfki.de/ecai04/final-porzel.pdf.

Reed, S.L. & Lenat, D.B. (2002). Mapping ontologies into Cyc. In *AAAI 2002 Conference Workshop on Ontologies for The Semantic Web*, Edmonton, Canada. Retrieved 20 November 2003, from http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf.

Sabou, M., Lopez, V., Mota, E. & Uren, V. (2006). Ontology selection: Ontology evaluation on the real semantic Web. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference*, Edinburgh, UK. Retrieved 3 June 2007, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006sabouetal.pdf.

Shadbolt, N., Gibbins, N., Glaser, H., Harris, S. & Schraefel, M. (2004). CS Aktive space or how we learned to stop worrying and love the Semantic Web. *IEEE Intelligent Systems*, *19*(3), 41–47.

Sleeman, D. & Reul, R. (2006). CleanONTO: Evaluating taxonomic relationships in ontologies. In *4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, UK. Retrieved 3 October 2008, from http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006sleemanetal.pdf.

Spyns, P. (2005). EvaLexon: assessing triples mined from texts. Technical report n. STAR-2005-0, STAR Lab, Vrije Universiteit Brussel, Brussels, Belgium. Retrieved 20 September 2008, from http://wise.vub.ac.be/ontobasis/Publications/STAR-2005-09ps.ps.

Sowa, J.F. (2000). Ontoloy. Retrieved 20 March 2005, from http://www.jfsowa.com/ontology/.

Supekar, K. (2005). A peer-review approach for ontology evaluation. In *8th International Protégé Conference*, Madrid, Spain. Retrieved 20 November 2003, from http://protege.stanford.edu/conference/2005/submissions/abstracts/accepted-abstract-supekar.pdf.

Uschold, M. & Gruninger, M. (1996). Ontologies: principles, methods an applications. *Knowledge Engineering Review*, *11*(2), 93–155.

Uschold, M., King, M., Moralee, S. & Zorgios, Y. (1998). Enterprise ontology. *The Knowledge Engineering Review*, *13*, 1–69. Retrieved 20 July 2005, from http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html.

Velardi, P., Navigle, R., Cucchiarelli, A. & Neri, F. (2005). Evaluation of OntoLearn, a methodology for automatic learning of domain ontologies. In P. Buitelaar, P. Cimiano & B. Magnini (eds), *Ontology Learning from Text: Methods, Evaluation and Applications* (pp. 92–106). Amsterdam: IOS Press.