A model for the construction of an inter-domain ontology: Corporate Sustainability Index and the G4 Guidelines of the Global Reporting Initiative

Tâmara Batista Reis[1], Paulo Caetano da Silva[1]

[1]Master Program in Systems and Computing -Salvador University -Brazil

tamarabr38@hotmail.com, paulo.caetano@pro.unifacs.br

***Abstract**: Ontologies are tools for knowledge representation that can help solve the diversity of representation of concepts that have similar meaning. Therefore, this paper proposes a method for building an ontology for the representation of a common knowledge base among the Corporate Sustainability Index, adopted in Brazil, with the G4 Guidelines of the Global Reporting Initiative, an international standard.*

1. Introduction

When choosing a sustainably responsible posture, an organization needs to measure, monitor and report organization's sustainability performance data, this gave rise to the concept of *Sustainable and Responsible Investment (SRI)*. Therefore, it is necessary to choose a methodology that provides the parameters through which the organization can compare its results achieved with the objectives pursued. In this sense a variety of indices and methodologies have been proposed.This diversity has caused the problem of lack of consensus and makes communication difficult between organizations that adopt different processes to manage and report their sustainability performance. Another problem concerns the lack of information standardization, because the documents requested by various stakeholders (e.g. shareholders, government) using different software, can only be obtained if there is an integration of information from heterogeneous systems. This generates costs and waste of resources for mapping this information between systems, without adding value to information.

This is the scenario in which the ontology is presented as an instrument for the representation of knowledge. This work propose a construction of an ontology, which enhances the integration the sustainability indices most used by Brazilian companies, the *Corporate Sustainability Index - ISE,* main representative of *SRI* in Brazil (BM&FBOVESPA 2014), with highly indicators worldwide, through alignment with the *Global Reporting Initiative GRI G4 Guidelines* (which are aligned to the UN Global Compact, the OECD and UNGC) wich provides a methodology of the most used worldwide [GRI 2014].The development of a taxonomy structure semantics between the relevant concepts common to the ISE / GRI will be able to provide the information quickly, efficiently, and independent methodologies. Such features can help overcome the limitations caused by the diversity indices and methodologies, and provide the integration of information, helping to overcome the computational problems faced across stakholders using heterogeneous systems.

The ontology construction activity requires the adoption of a methodology to structure the construction process [Rautenberg 2010; Luna et al., 2012]. However, by the finding of lack of consensus among the proposed methodologies for building ontologies and the particular needs of the domains addressed, this work presents a proposal of a model for building a domain ISE / GRI ontology.

2. A proposal for the construction of the ISE/GRI ontology

The methodologies presented by Fernandez et al., (1997), and Uschold Gruninger (1996), and Noy and MacGuinnes (2001), consider ontologies like a software products and demonstrated that the development stages, are equivalent to thesoftware *life cycle phase*. These phases were adapted by the extracted processes of the IEEE-1074 standard (1997) and characteristics that are particular to ontologies, i.e. *formalization and integration*.Therefore, the IEEE-1074 (1997) was used as standard quality for the development process, describing a structured process for software development that includes all life cycle stages, described as: *project management, pre-development, development process, postdevelopment and integral processes*. Thus, based on the analyzed methodologiesand on the IEEE-1074 standard (1997), a model process has been defined, whose development phases are described as shown in Figure 1.

➢ **Phase 1: Project Management:** having observed the suggestion of Methontology methodology [Fernandez et al.1997], for this phase were adopted related activities beginning at the planning and project management throughout its life cycle.On the activity **Definition of the life cycle process of the ontology** is proposed that the development process is based on the evolution of prototypes [Fernandez et al. 1997].

➢ **Phase 2: Ontology Pre-development:** at this stage it is recommended to search the domain knowledge and the identification of problems in order to propose possible solutions through the ontology. The sources for the pursuit of knowledge can be the literature, sites, knowledge experts, etc. [Fernandez et al. 2004]. In support is recommended to perform the following activities:

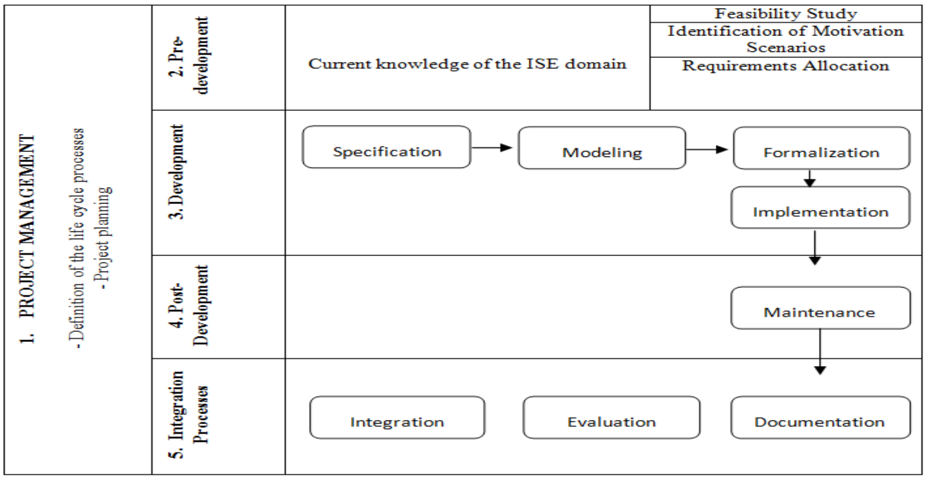

Figure 1. Phases of the development process of the Ontology

a)Feasibility Study: [ANSI/NISO Z39.19-2005]**:** is based on supplementary questions, assuggested by Silva (2008), in order to analyze the importance of building the ontology.Such questions are: 1. Why build the ontology? 2. What would happen if the ontologywas not built? 3. What are the problems with the current knowledge? 4. Howcan theproposed ontology help? 5. Will some existing ontology be reused, or be integrated? 6. Will any resources or technologies that are different from the onesalready used within the domains be needed? 7. What skills are required?

b) Identification of motivation scenarios:the motivation scenario analysis technique proposed by Gruninger and Fox (1995) helps detect the ontology domain problems and to present alternative solutions. The description of motivation scenarios are based on the identified initial requirements (which will be detailed in the requirements specification).

c) Requirements Allocation [Silva 2008]**:** this activity requires the choice of the artifacts needed to build the ontology, such as tools, *software* and *hardware*. The recommendations of this proposal for this step are: c1) for the ontology development the use of *Protégé* tool is recommended, for the following reasons : i) it has a friendly interface; ii) it documents objects; iii) it is in the public domain; iv) it has a modular architecture, which allows the inclusion of new features; v) it has a research community that contributes to its development and update; and vi) it has documentation; c2) for the ontology representation and formalization, OWL-DL is the recommended language, based on the following reasons: i) it is considered the *World WideWeb Consortium* (W3C) standard language, which enables its integration with ontologies implemented in standard Web technologies; ii) it supports axioms; iii) it provides inference mechanisms that allow to submit the ontology to evaluation; iv) it is used in *Protégé* tool, which assists in the process of implementation and formalization of the ontology; c3) for the conceptual modeling task it is suggested to use *Microsoft Visio* tool (*Microsoft*, 2014), for its usability features, user-friendly interface, and the fact that it allows the preparation of diagrams necessary for modeling the ontology.

- **Phase 3: Ontology development process:** this stage is the beginning of ontology construction process, comprising the activities described in the following.

a) Requirements specification: according to Gruninger and Fox (1995), from the observation of *Motivation Scenarios* it is possible to draw up a set of *competency questions*. These questions and their answers allow identifying information in real situations in the domain of the ontology in question. By analyzing *the questions* that the ontology will have to answer it is possible to determine the domain that the ontology should cover and delimit the ontology scope. It is recommended the documentation of this process for preparing the Scope Document Ontology, which includes information about: its purpose, its usefulness, who can use and maintain the ontology, degree of formality, responsible for the construction, sources of knowledge used, process adopted for the development, quality assurance, used tools, languages used for the representation and formalization, and the products generated.

b) Conceptual Modeling: to identify the ontology components the contribution of Silva (2008) was adopted, which reports the following elements: *conceptual classes;class attributes; instances; instance attributes; relations among the classes; constants;terms; formal axioms;* and *rules*. For best results in the *conceptual modeling* activity it is recommended to treat *the terms and concepts* involved, and only then organize them in the *taxonomic structure*. The activities flow for the conceptualization of the ISE-GRI ontology is illustrated in Figure 2.

To identify *relevant terms* it was adopted the Noy and McGuinness (2001) *proposed* which suggests questions related to *competency questions*. Such questions inquire: i) which are the terms that are relevant?;ii) what are the properties of these *terms*?; iii) what is necessary to say about these *terms*? Another contribution was taken from the ANSI/NISO Z39.19-2005 standard for the construction of controlled vocabularies, and it proposes the analysis of the domain through consultation with several knowledge sources, according to criteria based on: i) *literary warranty* (specialized literature); ii) *structural warranty*; iii) *warranty of use*. To assist in the construction of knowledge it is suggested to use the *documentanalysis method* [Dalhberg 1978], applying a technique used in the fields of Library and Information Science, the *subject analysis technique*, recommended by Silva (2008), which assists in the identification and selection of *concepts* that represent the essence of documents,. The application of these techniques has allowed the *identification ofrelevant terms* which represent the knowledge of the ISE-GRI ontology domain, and they were recorded in the *Glossary of Terms* document, proposed by Fernandez *et al*. (1997).

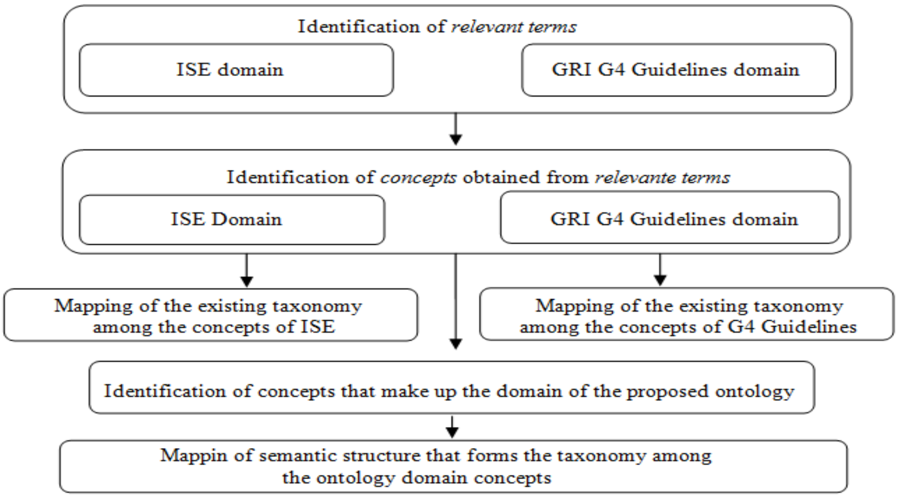

Figura 2. Fluxo de Atividades para a conceitualização da ontologia.

The next task comprises the definition of *domain concepts* and the principles adoptedare explained in the *Concept Theory* [Dalhberg 1978], which were summarized as: i) identification of the object or reference item in the domain; ii) analysis of the intrinsic and extrinsic features of the object, to define the *concept* and *relationships* among *concepts*, which allowed to form sentences about the object; iii) identification of the existing *taxonomy* among the *concepts* from the principle of contextualization, in which the definitions of *concepts* and their positions in the semantic structure are directly related to the *domain* in which the terminology is being built; iv) selection of *terms* to express the *concepts* present in the ontology.

After *defining the concepts* should be sought to know the nature of *concepts* in order to *rank these concepts into categories*, which identifies the *classes, attributes* and *relationships*. From the ANSI/NISO Z39.19-2005 standard, for the *classification of terms intocategories*, certain *facets* are determined, based on categories and subcategories of the *Concept Theory* [Dalhberg 1978], which determines the formal-categorical relationship to classify *concepts* of the same nature into a category. Thus, for each *term* the *category* to which it belongs is identified, as:*dimensions*; *activities*; *properties* and *entities*. The comparison task among the *concepts* for *classification* should consider the domain characteristics, seeking the most appropriate definition, i.e. one that meets the ontology purposes, identifying the *concepts* through the establishment of clear and unambiguous textual descriptions, defined by observing the semantic match of the meanings of terms and their relationships with each other, not in isolation or independently, as in a classical dictionary . The application of these procedures and the elaboration of the *faceted structure* helped identify the *concepts, attributes, constants* and *relationships* of the ISE-GRI ontology domain.

The *mapping of the semantic taxonomic structure* among *domain concepts* requires analysis first ofhow the *concepts* of the same kind relate, establishing two types of relationships: a) *hierarchical*); b) *partitive*: [Dalhberg 1978]. To organize the concepts in the taxonomy and identify the levels of classes, this work suggests the combined use of methods arising from ANSI / NISO Z39.19-2005: a) top-down: identifies generic concepts, high level; b) middle-out: identifies mid-level concepts; c) bottom-up: identifies low-level concepts.As an aid in decision-making are recommended principles proposed by Noy and McGuinness (2001), which help to: a) distinguish disjoint classes; b) identifying a transitive property; c) decide by inserting / or not of new sub-classes; d) decide to create a new class or getting a property; e) decide between creating a new class or identification of an instance; f) design relations types, "is a" or "type". To ensure that the methodological process of construction of knowledge about the ontology conceptualization is correct in order to avoid distortions in the semantic meanings of the concepts, it is recommended to carry out detailed descriptions of binary relations, class attributes, the instance attributes and constant, beyond the definition of relevant concepts instances, based from models for the *intermediate representation* proposed by Fernandez et al. (2004).

c) Ontology Formalization: the formalization activity follows the suggestion by Fernandez *et al*. (2004), indicating that it can be configured through tools that generate the code (e.g. generated in the specification of axioms) by exporting the ontology specification in the representation language used by the tool. We suggested the use ofOWL-DL language in the *Protégé* tool, which is based on *descriptive logic*. This process enables the definition and formalization of *axioms* and *rules* that restrict possible deviations of domain interpretation. A formalization example of the ISE/GRI ontology is illustrated in Figure 3.

Figure 3. Existencial restriction formalization.

Figure 3 shows that *General_Standard_Disclousure* class was selected and shows the formalization of the restriction created for this class, which has the following meaning: for an individual wich is a *General_Standard_Disclousure* class member, it is necessary that this individual belongs to *G4_Guidelines_GRI* class and has at least one type of relationship with *Indicator* class, through *hasIndicator* property.

d) Implementation of Ontology: This activity aims to transform the ontology written in natural language in a computable model, capable of meeting certain requirements defined in the conceptualization phase. The terminology designed for intermediate representation models must be mapped to the constructors and axioms of OWL-DL language, Protégé tool associated with getting the same concepts, attributes, relationships and described instances. This process used for implementation of the ISE / GRI allowed the construction of ontology classes, properties, and constraints creating instances of the ontology.

➢ **Phase 4: Post-development process: t**his phase comprises the *maintenance* required to the ontology, after the completion of the *development* and *evaluation* processes, in which the necessary procedures are performed, given the identified needs [Uschold e Gruninger 1996].

➢ **Phase 5: Integration process**

a)Integration: This step includes the evaluation of high-level ontologies for the reuse of terms relevant to the conceptualization of the ontology being built.

b)Ontology Evaluation: this activity comprises technical inspections of productsthat are generated at each stage of the process, reporting the product to maintenance whenever a need for changes is detected. Otherwise, the product is documented. Gruninger and Fox (1995) suggest that the *evaluation* process to investigate the consistancy of ontology after *implementation*using the *competency questions* to observe if the ontology is able to satisfactorily respond to these questions. This paper proposes the use of OWL-DL inference engine to perform these queries to the ontology. This procedure was applied to evaluate the consistency of the ISE / GRI ontology, indicating satisfactory results, as shown in Figure 4.

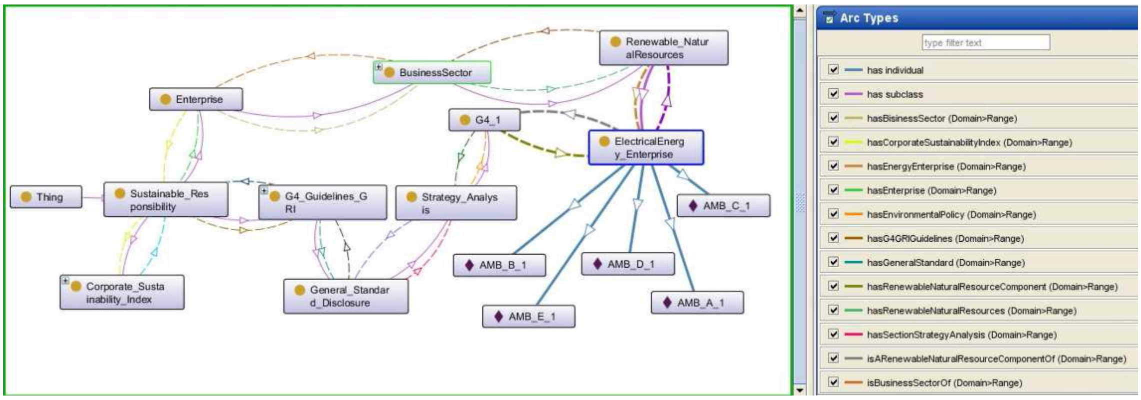

Figure 4. Evaluation of ontology in relation to *Competency Questions*

For instance, *Competency Question* inquired: Is there a relationship between environmental indicators of a company belonging to the electricity sector with the GRI indicators? "The ontology showed that a company of the electricity sector, represented by the *ElectricalEnergy_Enterprise*class, has concepts adopted by ISE (AMB_A_1 instances, AMB_B_1 , etc.) relating to the class *GRI_G4_1*.

c) Documentation: the documentation activity must be observed at all stages of the ontology *life cycle* and the generated documents should be recorded in the scope of the ontology, as suggested by Metonthology. In ISE-GRI ontology all documents were properly organized, and are available at: <http://xbrlframework.com/wiki/csa_gri/>.

3. Results obtained and future work

The results obtained by the tests performed during the *evaluation phase* showed that this proposal for the construction of the ontology attended its pre-determined purposes because the built ontology responds satisfactorily to the queries regarding the *competency questions,* as demonstrated in the evaluation section of the ontology. The proposal of this work allowed the construction of the ontology that relates the concepts of ISE with their counterparts in the G4 Guidelines of the GRI, using the construction of a common *semantic taxonomic structure*, which represented an alignment between their indicators. This semantic environment may facilitate the manipulation of information and integration of Information Systems using these concepts.

For future work, we suggest: a) the use of this model for the construction of other domain ontologies; b) the extension of this model in more detail for the formalization and integration processes.

References

ANSI/NISO Z39.19-2005. "Guidelines for the construction, format andmanagemenet of monolingual controlled vocabularies". Bethesda: NISO Press, 2005. 176 p.

BM&FBOVESPA (2014).“Índice de Sustentabilidade Empresarial (ISE)”. Available at:<http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=ISE&Idioma=pt-BR>.

Dalhberg, I. (1978). "Teoria do Conceito". Ciência da Informação, Rio de Janeiro, v.7, n.2, p.102-107.

Fernandez, M., Gomez-Perez, O., Juristo, H. (1997). “Methontology: from Ontological Art Towards Ontological Engineering”.

Fernandez, M.; Gomez-Perez, A., Corcho, O. (2004).“Methodologies and Methods for Building Ontologies”. In: Gomez-Perez, A., Fernandez- Lopes, M, Corcho, O. Ontological Engineering. London: Springer. pp. 107-153.

GRI - *Global Reporting Iniciative* (2014)."For the Guide Lines and Standard Setting - G4". Available at: <https://www.globalreporting.org/Pages/default.aspx>.

Gruninger, M.; Fox, M. S. (1995).“Methodology for the Design and Evaluation of Ontologies”.Department of Industrial Engineering University of Toronto, Toronto, Canada M5S 1A4.

Luna, J. A. G., Bonilla, M. L., Torres, I. D. (2012). "Methodologies and methods for building ontologies." Ciência e Técnica. Año XVII, No 50, Abril de 2012. Universidade Tecnológica de Pereira. ISSN 0122-1701.

MICROSOFT (2014). Microsoft Visio - Suporte. Available at:<http://support.microsoft.com/kb/943589>.

Noy, N. F. e McGuinness, D. L. (2001).“Ontology development 101: A guide to creating your first ontology”. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Rautenberg, S., Filho, A., Todesco, J. and Ostuni-Gauthier, F. (2010) “Ferramenta ontoKEN: Uma Contribuição à Ciência da Informação para o Desenvolvimento de Ontologias”.Perspectivas em Ciência da Informação, v.15, n.1, p. 239-258.

Silva, D., (2008). “Princípios Metodológicos para Construção de Ontologias: uma abordagem interdisciplinar entre as Ciências da Informação e da Computação”. Universidade Federal de Minas Gerais.

Uschold, M. and Gruninger, M. (1996) “Ontologies: Principles, Methods an Applications”. KnowledgeEngineering Review, v. 11, n. 2.