An Incremental and Iterative Process for Ontology Building

Andre Menolli[1,3], H. Sofia Pinto[2], Sheila Reinehr[1], Andreia Malucelli[1]

[1]Programa de Pós-Graduação em Informática (PPGIa), Polytechnic School, Pontifícia Universidade Católica do Paraná – PUCPR, Curitiba, Brazil

[2]Instituto Superior Técnico (INESC-ID), Lisboa, Portugal

[3]Computer Science Department, State University of North Paraná, Bandeirantes, Brazil

menolli@uenp.edu.br, sofia@ontol.inesc-id.pt, sheila.reinehr@pucpr.br, malu@ppgia.pucpr.br,

Abstract. *The ontology development area has received some attention over the years. Methodologies focusing in diverse aspects of ontology development have emerged. Some of these methodologies are consolidated, presenting phases and activities. However, existing methodologies do not fully consider the ontology integration process. Therefore, based on METHONTOLOGY and a methodology for integrating ontologies we proposed an incremental and iterative process. We have used this process to develop an ontology following three iterations, which we present in this paper. Furthermore, we discuss the main features of the proposed process.*

1. Introduction

Knowledge representation through ontologies aims at capturing static domain knowledge in a generic way and provide a common agreement upon understanding of that domain, which may be reused and shared across applications and groups [Chandrasekar *et al*. 1999].

Ontologies can describe a hierarchy of concepts connected by subsumption relationships, a concept more aligned with taxonomies; or a structure where axioms are added to express relationships among concepts and to limit their intentional interpretations [Guarino 1998]. Axioms and subsumptions relationships allow the use of inference mechanism. Therefore, an ontology is a complex knowledge representation object, whose development requires the use of some methodology.

In this context, there are several and diverse methodologies focusing in various aspects of ontology development. The most representative ontology building methodologies are by [Uschold 1996], [Uschold and Grüninger 1996] and [Fernández et al. 1997]. Nevertheless, these methodologies present some limitations, as for instance they do not address ontology integration [Pinto 2000]. Therefore, specific methodologies for ontology integration were proposed, as [Gangemi et al. 1998] and [Pinto and Martins 2001]. Nevertheless, these methodologies focus on ontology integration, and despite of them enable work with other methodologies for development ontology, they do not detail how. Furthermore, all work mentioned above are methodologies, thus are more comprehensive than a process.

With the growing number of existing knowledge representation sources, a process to build new ontologies taking full advantage of existing sources is needed. Thus, in this paper we propose an iterative and incremental process for ontology development. This process considers the acquisition and use of external sources to develop each increment, and is concerned with the integration of ontologies developed in each increment.

The proposed process is based on METHONTOLOGY [Fernández et al. 1997] and in the methodology for integrating ontologies proposed by Pinto and Martins (2001), which describes a process of ontologies integration.

2. The Incremental and Iterative Process

The iterative process reduces the complexity of ontology development, since it divides it into small parts, and the incremental life cycle solves some problems, allowing the partial specification of requirements and makes the ontology grow by layers, allowing the inclusion of new definitions only when a new version is planned. Figure 1 shows the life cycle process, and each phase is described following.

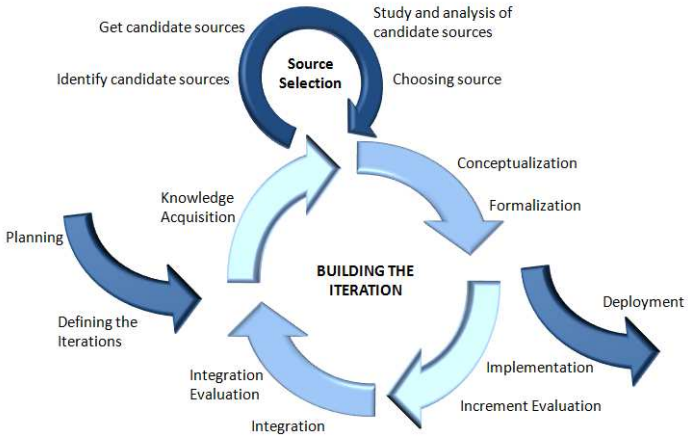

Figure 1. The Incremental and Iterative Life Cycle for Ontology Building

Planning

The planning phase is the first phase of ontology development. In this phase, the planning of whole ontology is done and the main goals are [Fernández et al. 1997]: (i) define the purpose of the ontology, including its intended uses, scenarios of use, end-users; (ii) define the level of formality of the implemented ontology, depending on the formality that will be used to codify the terms and their meaning; and (iii) define the ontology scope.

Defining the Iterations

The ontology, usually, is composed of several parts, which are aggregated to form the whole. So, it is important to define how many iterations will be needed to build the ontology, and the purpose of each one. This phase is extremely important, since the iterations defined here will guide the ontology development process.

Knowledge Acquisition

This phase was first defined by Fernández et al. (1997). In this phase, all knowledge about the domain must be acquired. However, instead of acquiring all knowledge to the whole ontology, we propose to divide and perform this phase for each increment. Thus in our process the knowledge acquisition is made incrementally, which facilitates the understanding of the subject.

Source Selection

Source Selection aims to select external sources that can be reused as base to develop the current increment. In this incremental and iterative process, each increment can be based in ontologies or other kinds of documents. This phase is composed of diverse activities, described following.

- *Identify candidate source:* the candidate sources should not be just ontologies, but any kind of knowledge representation. Among the main kind of knowledge representation, we suggest to use: catalog/id, terms/glossary, thesauri, frames, ontologies, and metadata specifications. This activity is subdivided into: *(1) finding available sources*, and *(2) choosing from the available sources which ones are possible candidates to be used*. To find possible sources, it is recommended to search in different locations, like ontology libraries and repositories of

standards organizations. To choose candidate sources one analyzes all available sources according to a series of features [Pinto and Martins 2001].

- *Get candidate source:* getting candidate sources includes not only their representations, but also, all available documentation. In some cases, this representation can be found in the literature (technical reports, books, thesis, etc.), or at least parts of it [Pinto and Martins 2001]. However, in most cases, only the implementation level representation of a source is available. Therefore, the reengineering process may be applied using the particular technique, according to the source chosen.

- *Study and analysis of candidate sources:* at this phase, it is important to study and analyze the sources to choose the best one. So, some criteria need be used according to Pinto and Martins (2000): (1) what knowledge is missing (concepts, relations, etc); (2) what knowledge should be removed; (3) which knowledge should be relocated; (4) which knowledge sources changes should be performed; (5) which documentation changes should be performed; (6) which terminology changes should be performed; (7) which definition changes should be made; and (8) which practices changes should be made.

- *Choosing source:* at this stage, and given the study and analysis of candidate sources performed by domain experts and ontologists, the final choices must be made. The source to be chosen and reused may lack knowledge, may require that some knowledge is removed, etc., that is, it may not exactly be what is needed. The best candidate source is the one that can better (more closely) or more easily (using less operations) be adapted to become the needed ontology [Pinto and Martins 2001].

Conceptualization

In this phase, the knowledge acquired is organized and structured using an independent knowledge representation. It is recommended that the knowledge domain is structured in a conceptual model that describes the problems and solutions in terms of the identified domain vocabulary [Fernández et al. 1997]. If an external source was selected as initial point to build the iteration, two additional activities are needed: adaptation and preparation to integration.

Adaptation focus on adapt the data from the external source to new domain. Many times an external source provides diverse concepts and attributes that are not need to the ontology that will be built. Preparing to integration it is needed to identify the assumptions and ontological commitments [Gruber 1995] that each increment should comply to.

Formalization

Transforms the conceptual model into a formal or semi-computable model, defining formal axioms. These axioms are introduced to constrain their interpretation and well-formed use [Pretorius 2004].

Implementation

In this phase, the increment is codified in a formal language such as OWL (Web Ontology Language).

Increment Evaluation

After implement the increment, the result ontology of the increment should be evaluated and analyzed. Furthermore, having an adequate design [Gruber 1995] and compliance with evaluation criteria [Gomez-Perez *et al*. 1995] the ontology should have a regular level of detail all over.

Integration

After the first iteration, the resulting ontology of the increment must be integrated with the ontologies created by the previous iterations. For that, one needs integration operations and integration oriented design criteria. Integration operations specify how knowledge from an

integrated ontology is going to be included and combined with knowledge in the resulting ontology, or modified before its inclusion [Pinto and Martins 2001].

Integration Evaluation

If it is not the first iteration, the integrated ontology should be evaluated and analyzed. None of the parts should have less level of detail than the required one or else the ontology would be useless, since it would not have sufficient knowledge represented. The resulting ontology should be consistent and coherent all over (although composed of knowledge from different ontologies).

3. Using the Process

This process was used to create the Unit of Organizational Learning Ontology (UOLO), and bellow are described the execution of all process stage to create the ontology.

Planning: This ontology aims at helping organize the content created in the company, specifically software development companies in units of organizational learning. It is based on educational units of learning, however considers organizational features.

Defining the Iterations: The UOLO was developed into three iterations: (1) organizational learning objects; (2) learning design; and (3) content package. The development of each increment was done following the activities outlined in Figure 1.

The first iteration generated the Ontology for Organizational Learning Object (OOLO) [Menolli *et al.* 2012].

- *Knowledge Representation*: in this phase the main Learning Objects Metadata were studied. From this study, the Learning Object Metadata (LOM) [IEEE 2002] was chosen as the base source to start developing the ontology proposed in this iteration, because it is a standard that facilitates search, acquisition, evaluation and use of LOs [Menolli et al. 2012].
- *Source Selection:* LOM Ontologies, and the complete documentation of LOM [IEEE. 2002] were gotten. Furthermore, FOAF (Friend of Friend) ontology also was gotten in this phase.
- *Conceptualization:* In this phase, all concepts and their properties were defined. This definition was done according to LOM standard, adapting it to our need and considering organizational features.
- *Formalization:* It was created a formal model that facilitates visualizing the taxonomy, covering axioms and properties.
- *Implementation:* The increment was implemented using the Protégé ontology editor and it was represented in OWL.

The other two iterations followed all the phases described in Figure 1. The second iteration implemented a learning design to help organizing materials previously produced in a manner that can enhance their understanding. So, it was based on IMS LD specification, that is a meta-language that describes all the elements of the design of a teaching-learning process, and drawn up by the IMS/LDWG work group [IMS 2003]. After implement this increment, it was integrated with the ontology created in the first iteration.

The third iteration created an ontology for Content Package concept. Content package describes the physical structure of the course defined by learning design. To define the content package concepts the IMS Content Packaging Specification [IMS 2004] was used. This increment was integrated with the ontologies produced in the first two iterations. The complete UOLO was generated as shown by Figure 2. Figure 2 (A) indicates the first increment, Figure 2 (B) the second increment and the Figure 2 (C) the third increment.

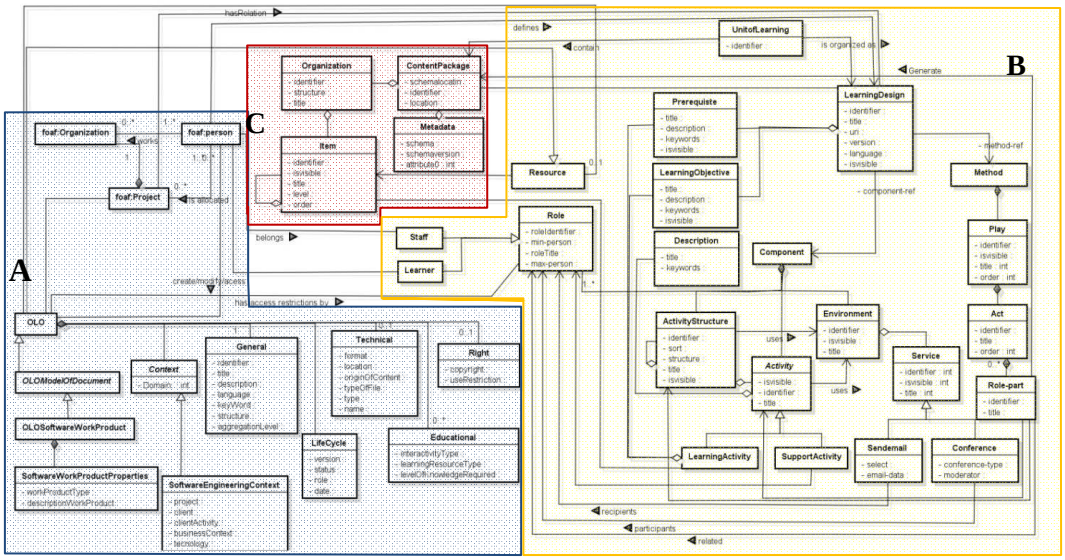

Figure 2. UOLO Concepts, Taxonomy and Relations

4. Discussion

In general, the phases that compose the life cycle tend to be performed following the order by which they were presented. If this order is performed, using the proposed life cycle the effort is divided between the phases.

The knowledge acquisition together with selection phase can require more effort than other approaches, since it is needed find and study several kinds of materials that can be used as base for the ontology; however, this effort should help to reduce the effort in the next phases. Using external sources to help modeling a concept model can reduce the effort of the conceptualization phase. Furthermore, finding external candidate sources, getting them, their evaluation and assessment for reuse purposes, and the choice of the most adequate one remain essential activities to be performed. This helps to create a more concise and consolidated model, since it is based on consensus knowledge.

The integration starts in the knowledge acquisition phase, and it continues for all other phases. Therefore, the integration is planned during all the increment, and if it is well performed, in integration phase, the ontology is just implemented together with the ontology created previously, and in the next phase, the integrated ontology is evaluated. Each increment must be evaluated individually, and after that it must be integrated with the ontology, and at the end evaluate the resulting ontology.

This process facilitates to find external sources to be reused. Moreover, the ontologist is forced to focus on the most critical issues, reducing risks during development; furthermore, the iterative and incremental development enables a continuous assessment of the project status. Finally, develop each increment is simpler than develop the whole ontology. As main limitation, the domain must be known and the scope limited, facilitating the iterations identifications.

5. Final Considerations

In this paper we describe an incremental and iterative process to ontology building. Furthermore, we describe the process life cycle and its phases. An incremental ontology was created using the proposed process, and as main advantages we identified the ease of use external sources, focusing on the most critical issues and the continuous and objective assessment of the project status. However, this process should be used only when the ontologist knows the domain, and he/she is sure that the ontology has more than one iteration.

The proposed process instantiate a particular integration process, using the phases and activities proposed by other ontology methodologies. The process reuses external material to build

each increment. For this, we used and adapted the activities defined by [Pinto and Martins 2001], that help to evaluate and choose the best sources from the identified sources. Furthermore, it integrates the activities to reuse sources with the phases proposed in the METHONTOLOGY. The process puts special emphasis to the quality of the final ontology, since we propose to evaluate each increment as well as the whole ontology.

Acknowledgments

The authors thank the Fundação Araucária of Paraná, Brazil and the Portuguese Fundação para a Ciência e Tecnologia through the financial support of INESC-ID funding under project PEst-OE/EEI/LA0021/2013.

References

Chandrasekaran, B., Josepheson, J. and Benjamins. V.R. (1999) “Ontologies: What are they? Why do we need them?”, In *IEEE Intelligent Systems*, v. 14, n. 1, pages 20-26.

Fernández, M., Gómez-Pérez, M. and Juristo, N. (1997) “Methontology: From Ontological art Towards Ontological Engineering”. Ontological Engineering, Stanford, Califórnia.

Gangemi, A., Pisanelli D, and Steve. G. (1998) “Ontology Integration: Experiences with Medical Terminologies. In N. Guarino (ed.), Formal Ontology in Information Systems, pages 163-178.

Gómez-Pérez, M., Juristo, N. and Pazos J. (1995) “ Evaluation and Assessment of the Knowledge Sharing Technology”, In *N. Mars (ed.), Towards Very Large Knowledge Bases*, pages 289-296. IOS Press.

Gruber. T. (1995) “Towards Principles for the Design of Ontologies for Knowledge Sharing”, In *International Journal of Human Computer Studies*, v. 43, n. 5-6, pages, 907-928.

Guarino, N. (1998) “Formal Ontology and Information Systems”, In *Proceedings of Formal Ontology in Information System*, pages. 3-15.

IEEE (2002) “Draft Standard for Learning Object Metadata”, Available on http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf July, 2002.

IMS (2003) “IMS Learning Design Information Model”, Version 1.0 Final Specification. Available on: http://www.imsglobal.org/learningdesign/ldv1p0/imsld_infov1p0.html.

IMS (2004) “IMS Content Packaging Information Model”, Version 1.1.4, IMS Global Learning. Available on: http://www.imsglobal.org/content/packaging/cpv1p1p2/imscp_infov1p1p2.html.

Menolli, A. L., Reinehr., S. and Malucell, A. (2012) “Ontology for Organizational Learning Objects based on LOM Standard”, In *Proceedings of the Latin American Conference en Informática* (Colombia, 2012).

Pinto H. S. and Martins, J. P. (2000) “Reusing Ontologies”, In *Proc. of AAAI2000 Spring Symposium Series, Workshop on Bringing Knowledge to Business Processes*, pages 77-84.

Pinto H. S. and Martins, J. P. (2001) “A methodology for ontology integration”, In Proceedings of the 1st international conference on Knowledge capture, pages 131-138, Canada.

Pretorius, A. J. (2004) “Ontologies - Introduction and Overview”, In *Mosaic A Journal For The Interdisciplinary Study Of Literature*, pages. 1-13.

Uschold, M. (1996) “Building Ontologies: Towards A Unified Methodology”, In *Expert Systems*, v. 96.

Uschold, M. and Grüninger, M. (1996) “Ontologies: Principles Methods and Applications”, In *Knowledge Sharing and Review*, V. 2.