Ontological Representation and Learning of Enterprise Artifacts

Fernanda Baião, Flavia Santoro, Kate Revoredo, Sean Siqueira,
Felipe Leão, Juliana França, Lucia Castro, Mauro Lopes

NP2Tec – Research and Practice Group in Information Technology,
Department of Applied Informatics, Federal University of the State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil
{fernanda.baiao, flavia.santoro, katerevoredo, sean,
felipe.leao, juliana.franca, lucia.castro,
mauro.lopes}@uniriotec.br

Abstract. This paper presents the ontology-related research themes being conducted at NP2Tec/UNIRIO. The main focuses that are being investigated handle the ontological representation of Enterprise artifacts, including structural conceptual models, process models, and business rules. Besides, maintaining a consistent representation of a Universe of Discourse is also a challenge since new (valid) facts are constantly being observed, and may not be consistent with the existing ontological representation.

1 Introduction

Ontologies are commonly used in computer science either as a reference model to support semantic interoperability, or as an artifact that should be efficiently represented to support tractable automated reasoning and semantic search. This duality poses a tradeoff between expressivity and computational tractability that should be addressed in different phases of an Ontology engineering process, using different Ontology representation languages, as advocated in our previous work [11]. In fact, in order for an ontology to be able to adequately serve as a reference model, it should be constructed using an approach that explicitly takes foundational concepts into account; this is, however, typically neglected for the sake of computational complexity.

From the perspective of an Enterprise, the adequate representation of its artifacts (including its structural conceptual models, process models, business rules, etc) is a very important issue, since it contributes to prevent several kinds of problems, such as misinterpretation of a piece of information by different business stakeholders throughout an Enterprise, inconsistent data integration procedures, and incorrect information exchange between applications.

Moreover, as with any specification, the dynamic nature of several domains makes it hard to keep up with a consistent representation. When new (valid) knowledge about the domain is known, previous rules specified in Enterprise models may no longer be true, and therefore need to be revised. In real scenarios where the

number of instances and rules may become arbitrarily large, manually detecting and fixing those inconsistencies becomes unfeasible, and is still a challenge.

2 Research Themes

In this section we sketch the main research topics being addressed in our group on the representation and learning of Enterprise artifacts supported by Ontologies, and point to further results obtained.

2.1 Semantic Oriented Method for Conceptual Data Modeling in OntoUML Based on Linguistic Concepts

Enterprise conceptual data models, as means of communication, must have semantic quality. Such quality relies on the model's completeness and validity in relation to the concepts it is supposed to represent. Since the modeler acquires such concepts mostly from texts created in a natural language, we argue that a semantic-oriented linguistic approach should be adopted for building unambiguous conceptualizations. Also, the chosen modeling language must offer enough constructs for the creation of a faithful representation, such as OntoUML. Such languages, however, may require a learning period that modelers hardly can afford. This research topic proposed a a semantic-oriented method for conceptual data modeling that consists of a set of systematic steps which activities promote the understanding of the concepts inherent to the domain to be modeled. The method makes use of the theories of semantic types proposed by Dixon [8], as well as linguistic concepts. Figure 1 illustrates Dixon´s semantic types assocaiated with nouns. The modeling language adopted is OntoUML, described and documented in [10], [11], [12] and [13].

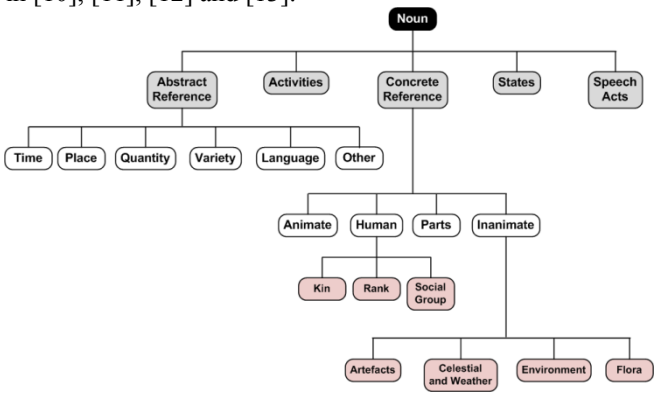

Figure 1. Semantic types associated with Nouns

This method was tested through a case study [1][2][3], which results indicated that it makes the modeling less complex by allowing for modeling choices to be dealt with within the realm of the natural language.

2.2 Expressing Business Rules in a Foundational-based Domain Ontology

Despite all the research efforts in the last decades, information integration is a problem yet to be solved in organizations, especially when it involves semantic issues. A complete and precise conceptual model of a Universe of Discourse is required to prevent several kinds of problems, such as misinterpretation of a piece of information by different business stakeholders throughout an Enterprise, inconsistent data integration procedures, and incorrect information exchange between applications. A key goal of any conceptual data model is to provide the best possible understanding of its subjacent domain. Ontologically well-founded conceptual models present themselves as a solution to represent a domain in a more correct and complete scheme. Current well-founded conceptual modeling representation languages, however, focus on the structural perspective, such as OntoUML [4]. Business rules describe both the structural and behavioral perspectives of a UoD. There are several business rules representation languages, such as URML [7] and RuleML [6]; however, they lack a precise semantics and may raise ambiguous interpretations from the same representation. This work-in-progress addresses this issue and intends to produce more precise business rules models by proposing an Ontologically well-founded business rule representation language. Preliminary ideas of this representation were presented in [5], and illustrated in figure 2.

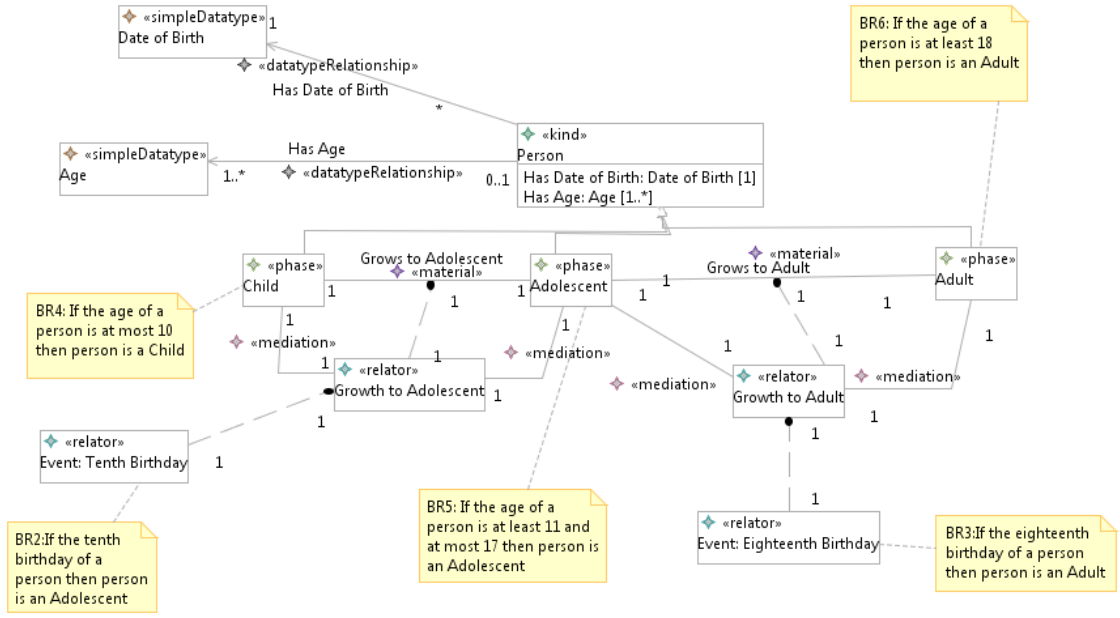

Figure 1. Initial proposal for the representation of business rules into ontologically well-founded conceptual models.

2.3 Towards the Ontological Representation of Knowledge Intensive Processes

Enterprises are increasingly focusing on technological innovations to remain competitive. Business Process Management continuously supports the improvement of Business Processes, increasing the quality of organizational services. However, existing methods for discovering and representing Business Processes do not adequately support available knowledge. Knowledge-Intensive Processes (KIP) focus on the dynamic conversion of knowledge between the participants involved in the execution of business processes, and therefore typically relies on the use of collaborative technology. This research-in-progress proposes an ontologically well-founded representation of KIP. Partial results comprise a set of proposed characteristics required to represent KIP (including contingency structural elements, innovation agents, impact agents, informal knowledge exchange and knowledge artifacts), and a method for automatically discovering KIP from unstructured data collected from collaborative tools.

2.4 Ontology Refinement through Theory Revision Techniques

The use of Ontologies for several purposes in Computer Science has increased, due to their support for concise and unambiguous descriptions of concepts and relationships in a domain of interest. The usefulness of an ontology depends on its consistency with regard to the set of instances representing this domain, that is, how precisely the ontology structure – and its axioms – represents instances of the domain.

In this topic, we focus on lightweight ontology representation languages. Moreover, while previous research topics focused on the Ontological representation of Enterprise artifacts, this topic handles the evolution of this representation since, as with any specification, the dynamic nature of several domains makes it hard to keep up with a consistent representation. When new (valid) instances of the domain are known, previous rules specified in the Ontology may no longer be true, and therefore need to be revised. In real scenarios where the number of instances and rules may become arbitrarily large, manually detecting and fixing those inconsistencies becomes unfeasible. We address this problem by providing a way to automatically learn a lightweight ontology representation from a set of instances, through the use of theory revision techniques from the ILP research area.

The theory revision task involves changing the answer set of the given theory, i.e., improving its inferential capabilities by adding previously missing answers, generalization or by removing incorrect answers, specialization [9]. Figure 3 presents a schema for theory revision.

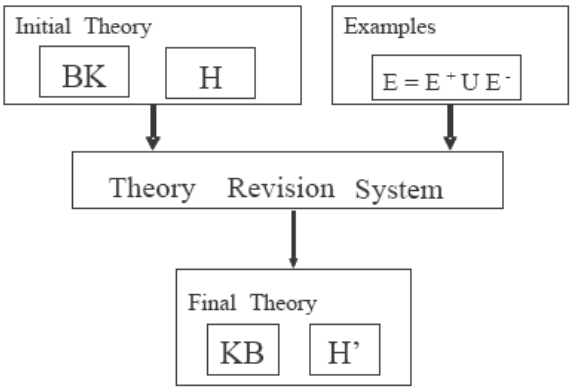

Figure 3. Revision Theory Schema

A theory revision system receives an initial theory and a set of examples. This initial theory includes two components: an invariant component, named background knowledge (BK), and one component that can be modified (H).

An ontology is defined as a set $O = \{G; \_; \mathcal{R}; \mathcal{I}\}$, where G is a directed graph, _ a set of concepts, $\mathcal{R}$ a set of relations between these concepts and $\mathcal{I}$ a set of axioms providing semantics to these relations. In this research, we consider the set of axioms as a set of function-free Horn clauses, and propose to revise them using the theory revision system FORTE [10]. Therefore, the set of axioms $\mathcal{I}i$ of the initial ontology Oi is revised by system FORTE finding a final ontology Of with the set of revised axioms $\mathcal{I}f$. We define the ontology revision task as:

Given:an ontology $O_i = \{G; \_; \mathcal{R}; \mathcal{I}i\}$ and a set of positive and negative examples E = $E_+ \cup E_-$.
Find: an ontology $O_f = \{G; \_; \mathcal{R}; \mathcal{I}f\}$

In [8], we sketched how the ontology elements are mapped to the theory revision system arguments, and illustrated the applicability of our proposal in a scenario using the family domain, through the use of the FORTE system.

3 Research Team

NP2Tec is the Research and Practice Group in Information Technology in the Department of Applied Informatics at the federal University of the State of Rio de Janeiro (UNIRIO). NP2Tec gathers several researches and students working on fostering Business-IT alignment in Enterprise through Business Process Modelling, Enterprise Architecture and Ontologies.

Fernanda Baião is a Professor of the Department of Applied Informatics of the Federal University of the State of Rio de Janeiro (UNIRIO) since 2004, where she is one of the coordinators of the NP2Tec research group. She received the Doctor of Science degree from the Federal University of Rio de Janeiro (COPPE/UFRJ) in 2001. During the year 2000 she worked as a visiting student at the University of

Wisconsin, Madison (USA). Her current research interests include conceptual modelling, well-founded representation languages, information architecture, data management in scientific workflows. She participates in research projects in those areas, with funding from several Brazilian government agencies, including CNPq, CAPES and FAPERJ. She participates in several program committees of national and international conferences and workshops, and is a member of ACM and of the Brazilian Computer Society.

Flavia Santoro is a Professor of the Department of Applied Informatics of the Federal University of the State of Rio de Janeiro (UNIRIO) since 2002, where she is one of the coordinators of the NP2Tec research group. She received the Doctor of Science degree from the Federal University of Rio de Janeiro (COPPE/UFRJ) in 2001. Her current research interests include collaborative learning, CSCW, business process management, software engineering and requirements, and knowledge management.

Kate Revoredo is a Professor at Department of Applied Informatics at Federal University of the State of Rio de Janeiro (UNIRIO). She received her PhD from the Department of Systems Engineering and Computer Science (PESC), filiated to the Alberto Luiz Coimbra Graduate and Engineering Research Institute (COPPE) at the Federal University of Rio de Janeiro (UFRJ). Her main research areas are Machine Learning and Data Mining, relational mining (inductive logic programming), relational statistical learning, theory revision, predicate invention and Bayesian networks

Sean Siqueira is a Professor of the Department of Applied Informatics of the Federal University of the State of Rio de Janeiro (UNIRIO). He received the Doctor of Science degree from PUC-Rio in 2005. His current research interests include semantic web, ontologies, "technology-enhanced learning", "learning objects", data integration, data warehousing, information retrieval, CRM, knowledge management, complex object modelling, musical information systems, data, text and web mining, social networks and web 2.0.

4 References

[1] CASTRO, L. ; BAIÃO, F. ; GUIZZARDI, G. . A Linguistic Approach to Conceptual Modeling with Semantic Types and OntoUML. In: International Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE 2010), 2010, Vitoria. EDOC 2010 Workshops, 2010. v. 1. p. 1-1.

[2] CASTRO, L. ; BAIÃO, F. ; GUIZZARDI, G. . Conceptual Modeling: the Linguistic Approach. Revista de Informática Teórica e Aplicada, v. 16, p. 103-104, 2009.

[3] CASTRO, L. ; GUIZZARDI, G. ; BAIÃO, F. . Conceptual Modeling: the Linguistic Approach. In: International Conference on Conceptual Modeling (ER 2009), 2009, Gramado. Journal of Theoretical and Applied Informatics. Porto Alegre : Instituto de Informática da UFRGS, 2009. v. 1. p. 1-1.

[4] GUIZZARDI, G. . Ontological Foundations for Structural Conceptual Models. Universal Press, 2005. 410 p.

[5] LOPES, M ; BAIÃO, F. ; SIQUEIRA, S. . Expressing Business Rules in a Foundational-based Domain Ontology: Towards Higher-quality Conceptual Models. In: International Conference on Information Integration and Web-based Applications & Services, 2010, Paris. 12th International Conference on Information Integration and Web-based Applications & Services (iiWAS2010), 2010. v. 1. p. 1-1.

[6] Halpin, T.A., 'Business Rules and Object-Role Modeling', Database Prog. & Design, vol. 9, no. 10, Miller Freeman, San Mateo CA, pp. 66-72, 1996.

[7] Wagner, G.; Antoniou, G.; Tabet, S.; Boley, H., The Abstract Syntax of RuleML - Towards a General Web Rule Language Framework 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI'04). Beijing, China. Setembro 20-24, 2004. pp. 628-631. NRC 48059.

[8] LEAO, F. ; REVOREDO, K. ; BAIÃO, F. . Ontology Refinement through Theory Revision Techniques. In: Workshop on Computational Intelligence, 2010, São Bernardo do Campo. Anais do III Workshop on Computational Intelligence, 2010. v. 1. p. 1-1.

[9] [Wrobel, 1996] Wrobel, S. (1996). First-order theory refinement. In Raedt, L. D., editor, Advances in Inductive Logic Programming, pages 14–33. IOS Press.

[10] [Richards and Mooney, 1995] Richards, B. L. and Mooney, R. J. (1995). Automated refinement of first-order Horn-clause domain theories. *Machine Learning*, 19:95–131.

[11] GUIZZARDI, G. ; BAIÃO, F. ; LOPES, M ; FALBO, R. . The Role of Foundational Ontologies for Domain Ontology Engineering: An Industrial Case Study in the Domain of Oil and Gas Exploration and Production. International Journal of Information System Modeling and Design, v. 1, p. 1-22, 2010.