Pattern for Re-engineering a Classification Scheme, Which Follows the Path Enumeration Data Model, to a Taxonomy

http://ontologydesignpatterns.org/wiki/Submissions:Classification_scheme_-path_enumeration_model_-_to_Taxonomy

Boris Villazón-Terrazas 1, Mari Carmen Suárez-Figueroa 1, and Asunción Gómez-Pérez 1

Ontology Engineering Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Spain {bvillazon,mcsuarez,asun}@fi.upm.es,

WWW home page: http://www.oeg-upm.net/

1 Introduction

This pattern for re-engineering non-ontological resources (PR-NOR) fits in the Schema Re-engineering Category proposed by [3]. The pattern defines a procedure that transforms the classification scheme components into ontology representational primitives. This pattern comes from the experience of ontology engineers in developing ontologies using classification schemes in several projects (SEEMP¹, NeOn², and Knowledge Web³). The pattern is included in a pool of patterns, which is a key element of our method for re-engineering non-ontological resources into ontologies [2]. The patterns generate the ontologies at a conceptualization level, independent of the ontology implementation language.

2 Pattern

Problem Re-engineering a classification scheme, which follows the path enumeration model, to design a Non-Ontological Resource A non-ontological resource holds a classification scheme which follows the path enumeration model A classification scheme is a rooted tree of concepts Path Category Category in which each concept groups entities by some par Name Description ticular degree of similarity. Category1 Category1Desc The semantics of the hierarchical relation between Category11 Category11Desc parents and children concepts may vary depend-Category111 Category111Desc ing of the context. The path enumeration data 12 Category12 Category12Desc model [1] for classification schemes take advantage 121 Category121Desc Category121 of that there is one and only one path from the Category2 Category2Desc root to every item in the classification. The path enumeration model stores that path as string by concatenating either the edges or the keys of the classification scheme items in the path

http://www.seemp.org

² http://www.neon-project.org

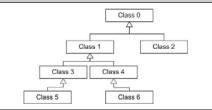
³ http://knowledgeweb.semanticweb.org

Applicability

The semantics of the relation between parent and children items are $sub\,Class\,Of$. There is not multi-inheritance nor cyclic relations.

Ontology Generated

The ontology generated will be based on the taxonomy architectural pattern (AP-TX-01) [4]. Each category in the classification scheme is mapped to a class, and the semantics of the re-lationship between children and parent categories are mapped to *subClassOf* relations.



Process - Solution

- 1. Identify the classification scheme items whose their path enumeration values have the shortest length, i.e. classification scheme items
- without parents.

 2. For each one of the above identified classification scheme items ce_i :
 - 2.1. Create the corresponding ontology class, C_i class.
 - 2.2. Identify the classification scheme items, ce_j , which are children of ce_i , by using the path enumeration values.
 - For each one of the above identified classification scheme items ce_i :
 - 2.3.1. Create the corresponding ontology
 - class, C_j class. 2.3.2. Set up the subClassOf relation between C_j and C_i . 2.3.3. Repeat from step 2.2 for ce_j as a
- new ce_i . 3. If there are more than one classification
 - scheme items without parent ce_i 3.1. Create an ad-hoc class as the root class
 - of the ontology.

 3.2. Set up the subClassOf relation between C_i class and the root class.

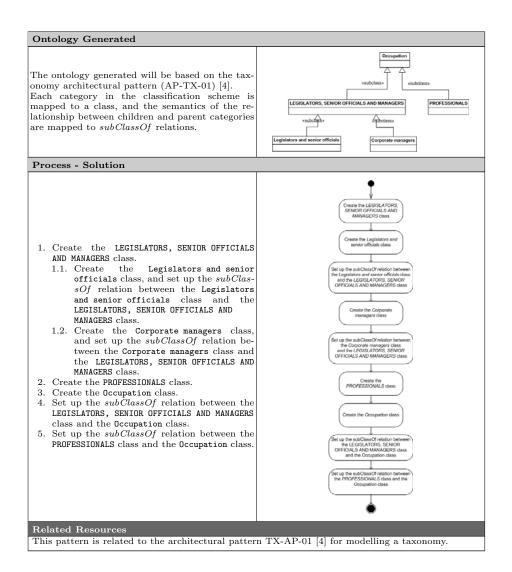


Suppose that someone wants to build an ontology based on the International Standard Classification of Occupations (for European Union purposes) ISCO-88 (COM). This classification scheme follows the path enumeration data model.

Non-Ontological Resource

The International Standard Classification of Occupations (for European Union purposes), 1988 version: ISCO-88 (COM) published by Eurostat is modelled with the path enumeration data model. This classification scheme is available at http://ec.europa.eu/eurostat/ramon/

Code	Level	Name
1	1	LEGISLATORS, SENIOR OFFICIALS AND MANAGERS
11	2	Legislators and senior officials}
111	3	Corporate managers
2	1	PROFESSIONALS



3 Pattern Usage

This pattern was applied to re-engineer the ISCO-88(COM)⁴, International Standard Classification of Occupations (for European Union purposes), into a Occupation Ontology⁵, within the context of the SEEMP project. This standard is a classification scheme which consists of 520 occupations. ISCO-88(COM) is modelled following the path enumeration data model. Because of the number of occupations of the ISCO-88(COM) standard, it was not practical to create the

⁴ Available at http://ec.europa.eu/eurostat/ramon/

⁵ The ontology is available at http://droz.dia.fi.upm.es/hrmontology/

ontology manually. Therefore, we created an *ad-hoc* wrapper, implemented in Java, that reads the data from the resource implementation and automatically creates the corresponding elements of the new ontology following the suggestion given by the pattern.

4 Summary and Future Work

We have presented a pattern for transforming a classification scheme, which is modelled following the path enumeration data model, into a taxonomy. The pattern is included in a pool of patterns, which is a key element of our method for re-engineering non-ontological resources into ontologies [2].

We plan to develop software libraries within a framework that implement the transformation process suggested by the pattern. Moreover, we will include external resources to improve the quality of the resultant ontologies. Finally, we need to calculate how much effort do we save re-engineering classification schemes using patterns compared with re-engineering classification schemes without them.

Acknowledgments. This work has been partially supported by the European Comission projects NeOn(FP6-027595) and SEEMP(FP6-027347), as well as by an R+D grant from the UPM.

References

- D. Brandon. Recursive database structures. Journal of Computing Sciences in Colleges, 2005.
- A. García, A. Gómez-Pérez, M. C. Suárez-Figueroa, and B. Villazón-Terrazas. A
 Pattern Based Approach for Re-engineering Non-Ontological Resources into Ontologies. In Proceedings of the 3rd Asian Semantic Web Conference (ASWC2008).
 Springer-Verlag, 2008.
- 3. V. Presutti, A. Gangemi, S. David, G. Aguado de Cea, M. C. Surez-Figueroa, E. Montiel-Ponsoda, and M. Poveda. NeOn Deliverable D2.5.1. A Library of Ontology Design Patterns: reusable solutions for collaborative design of networked ontologies. In NeOn Project. http://www.neon-project.org, 2008.
- M. C. Suárez-Figueroa, S. Brockmans, A. Gangemi, A. Gómez-Pérez, J. Lehmann, H. Lewen, V. Presutti, and M. Sabou. Neon modelling components. Technical report, NeOn project deliverable D5.1.1, 2007.