Pattern for Re-engineering a Term-based Thesaurus, Which Follows the Record-based Model, to a Lightweight Ontology

http://ontologydesignpatterns.org/wiki/Submissions:Term-based_-_record-based_model_-_thesaurus_to_lightweight_ontology

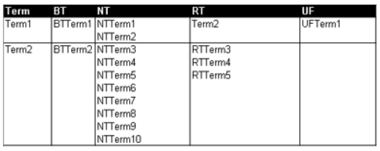
Boris Villazón-Terrazas[1], Mari Carmen Suárez-Figueroa[1], and Asunción Gómez-Pérez[1]

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 term-based thesaurus components into ontology representational primitives. This pattern comes from the experience of ontology engineers in developing ontologies using thesauri in several projects (SEEMP[1], NeOn[2], and Knowledge Web[3]). 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

<table>
<tr><th colspan="2">Problem</th></tr>
<tr><td colspan="2">Re-engineering a term-based thesaurus, which follows the record-based model, to design a lightweight ontology.</td></tr>
<tr><th colspan="2">Non-Ontological Resource</th></tr>
<tr><td>A non-ontological resource holds a term-based thesaurus which follows the record-based model. A thesaurus represents the knowledge of a domain with a collection of terms and a limited set of relations between them.
The record-based data model [4] is a denormalized structure, uses a record for every term with the information about the term, such as synonyms, broader, narrower and related terms.</td><td>
<table>
<tr><th>Term</th><th>BT</th><th>NT</th><th>RT</th><th>UF</th></tr>
<tr><td>Term1</td><td>BTTerm1</td><td>NTTerm1
NTTerm2</td><td>Term2</td><td>UFTerm1</td></tr>
<tr><td>Term2</td><td>BTTerm2</td><td>NTTerm3
NTTerm4
NTTerm5
NTTerm6
NTTerm7
NTTerm8
NTTerm9
NTTerm10</td><td>RTTerm3
RTTerm4
RTTerm5</td><td></td></tr>
</table></td></tr>
<tr><th colspan="2">Applicability</th></tr>
<tr><td colspan="2">The semantics of the relation between narrower and broader terms are <i>subClassOf</i>.</td></tr>
</table>

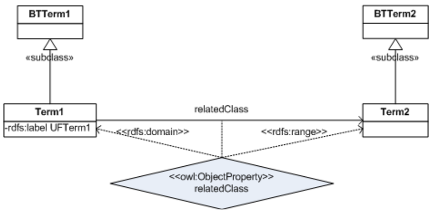
[1] http://www.seemp.org

[2] http://www.neon-project.org

[3] http://knowledgeweb.semanticweb.org

Ontology Generated

The ontology generated will be based on the lightweight ontology architectural pattern (AP-LW-01) [5].
Each thesaurus term is mapped to a class. A *subClassOf* relation is defined between the new classes for the BT/NT relation. A *relatedClass* relation is defined between the new classes for the RT relation. For the UF/USE relations the SynonymOrEquivalence (SOE) pattern [1] is applied.



Process - Solution

1. Identify the records that contain thesaurus terms without a *broader term*.

2. For each one of the above identified thesaurus terms t_i:

2.1. Create the corresponding ontology class, C_i class, if it is not created yet.

2.2. Identify the thesaurus term, t_j, which are narrower terms of t_i. They are referenced in the same record that contains t_i.

2.3. For each one of the above identified thesaurus term t_j:

2.3.1. Create the corresponding ontology class, C_j class, if it is not created yet.

2.3.2. Set up the *subClassOf* relation between C_j and C_i

2.3.3. Repeat from step 2.2 for c_j as a new t_i

2.4. Identify the thesaurus term, t_r, which are related terms of t_i. They are referenced in the same record that contains t_i.

2.5. For each one of the above identified thesaurus term t_r:

2.5.1. Create the corresponding ontology class, C_r class, if it is not created yet.

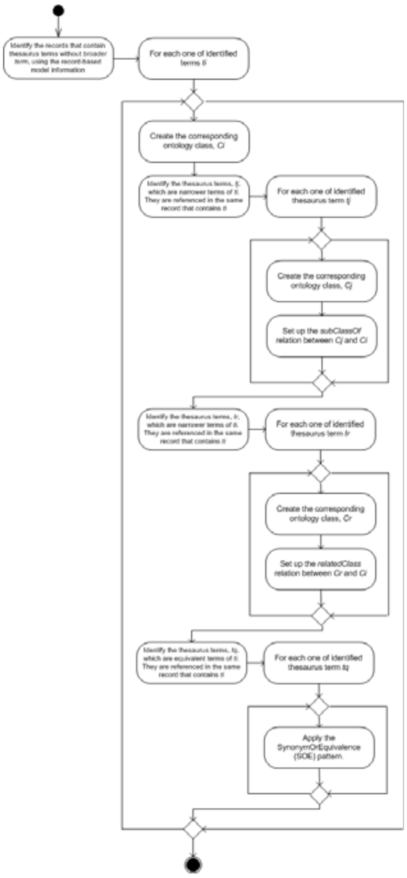
2.5.2. Set up the *relatedClass* relation between C_r and C_i

2.5.3. Repeat from step 2.4 for t_r as a new t_i

2.6. Identify the thesaurus term, t_q, which are equivalent terms of t_i. They are referenced in the same record that contains t_i.

2.7. For each one of the above identified thesaurus term t_q:

2.7.1. Apply the SynonymOrEquivalence (SOE) pattern.

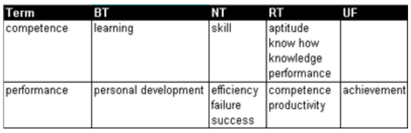

Identify the records that contain thesaurus terms without broader term, using the record-based model information
For each one of identified terms ti
Create the corresponding ontology class, Ci
Identify the thesaurus terms, tj, which are narrower terms of ti. They are referenced in the same record that contains ti
For each one of identified thesaurus term tj
Create the corresponding ontology class, Cj
Set up the subClassOf relation between Cj and Ci
Identify the thesaurus terms, tr, which are narrower terms of ti. They are referenced in the same record that contains ti
For each one of identified thesaurus term tr
Create the corresponding ontology class, Cr
Set up the relatedClass relation between Cr and Ci
Identify the thesaurus terms, tq, which are equivalent terms of ti. They are referenced in the same record that contains ti
For each one of identified thesaurus term tq
Apply the SynonymOrEquivalence (SOE) pattern.

Example

Suppose that someone wants to build a lightweight ontology based on the European Training Thesaurus (ETT), which is a term-based thesaurus and it follows the record-based model.

Non-Ontological Resource

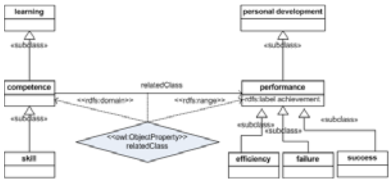
The European Training Thesaurus (ETT) constitutes the controlled vocabulary of reference in the field of vocational education and training (VET) in Europe. The relation semantics between the sub-ordinate and the super-ordinate concepts is *subClassOf*. This classification scheme is available at http://libserver.cedefop.europa.eu/ett/en/



Term	BT	NT	RT	UF
competence	learning	skill	aptitude know how knowledge performance	
performance	personal development	efficiency failure success	competence productivity	achievement

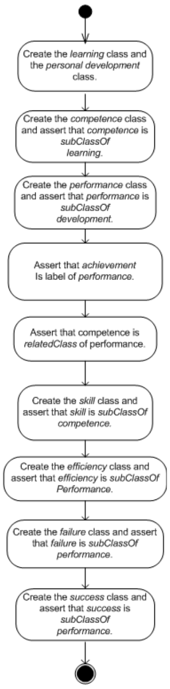
Ontology Generated

The ontology generated will be based on the lightweight ontology architectural pattern (AP-LW-01) [5].
Each thesaurus term is mapped to a class. A *subClassOf* relation is defined between the new classes for the BT/NT relation. A *relatedClass* relation is defined between the new classes for the RT relation. For the UF/USE relations the SynonymOrEquivalence (SOE) pattern [1] is applied.



Process - Solution

1. Create the `learning` class and the `personal development` class.
2. Create the `competence` class and assert that `competence` is *subClassOf* `learning`.
3. Create the `performance` class and assert that `performance` is *subClassOf* `development`.
4. Assert that `achievement` is label of the `performance` class.
5. Assert that `competence` is *relatedClass* of `performance`.
6. Create the `skill` class and assert that `skill` is *subClassOf* `competence`.
   - 6.1. Create the `efficiency` class and assert that `efficiency` is *subClassOf* `performance`.
   - 6.2. Create the `failure` class and assert that `failure` is *subClassOf* `performance`.
   - 6.3. Create the `success` class and assert that `success` is *subClassOf* `performance`.


Create the learning class and the personal development class.
Create the competence class and assert that competence is subClassOf learning.
Create the performance class and assert that performance is subClassOf development.
Assert that achievement Is label of performance.
Assert that competence is relatedClass of performance.
Create the skill class and assert that skill is subClassOf competence.
Create the efficiency class and assert that efficiency is subClassOf Performance.
Create the failure class and assert that failure is subClassOf performance.
Create the success class and assert that success is subClassOf performance.

Related Resources

This pattern is related to the architectural pattern AP-LW-01 [5] for modelling a lightweight ontology.

3 Pattern Usage

This pattern is being applied to re-engineer the European Training Thesaurus (ETT)[4] into a Education Ontology[5], within the context of the SEEMP project. It contains over 2500 terms (1550 are descriptors, and 950 non descriptors). This term-based thesaurus is modelled following the record-based data model.

4 Summary and Future Work

We have presented a pattern for transforming a term-based thesaurus, which is modelled following a record-based data model, into a lightweight ontology. 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

1. C. Roussey and O. Corcho. SynonymOrEquivalence (SOE) Pattern. *http://ontologydesignpatterns.org*, 2009.
2. 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.
4. D. Soergel. Data models for an integrated thesaurus database. *Comatibility and Integration of Order Systems*, 24(3):47–57, 1995.
5. 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.

[4] http://libserver.cedefop.europa.eu/ett/en/

[5] The ontology will be available at http://droz.dia.fi.upm.es/hrmontology/