Exploiting patterns in Ontology Mapping

Ondřej Šváb

Department of Information and Knowledge Engineering,
University of Economics, Prague, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic
e-mail: `svabo@vse.cz`

Abstract. Unsatisfactory performance of ontology mapping methods on many tasks seems to indicate the need for employing some background knowledge. My PhD work focuses on exploiting pattern-like structures for this purpose. Design patterns are related to different styles of modelling ontology structure and/or naming of concepts. On the other hand, mapping patterns emerge as result of the mapping process. Preliminary results and schedule of future work are presented.

1   Introduction

Current ontology mapping methods are typically based on generic methods and/or ad hoc heuristics. However, ontologies themselves are not randomly assembled graph structures but conceptual models of reality, and thus reflect some explicit or implicit patterns that have already been identified in ontological engineering research.

We examine two different though interrelated notions of patterns that are relevant for ontology mapping. One is that of *mapping pattern*: a frequently appearing structure involving semantic links both across the mapped ontologies (i.e. alignments resulting from the mapping process) and within each of them. The other is that of *design pattern* in the general sense: modeller's choices such as *naming conventions* or *structural patterns*. While the analysis of mapping patterns can provide insight into the functionality of mapping methods a posteriori, the prior analysis of design patterns can help select or tune the mapping methods a priori.

Section 2 of the paper briefly surveys the experiments with mapping patterns. Section 3 discusses the role of design patterns in general, and reports on preliminary evaluation of sensitivity of a popular graph-based mapping method (similarity flooding) to the presence of some simple structural design patterns. Finally, the paper includes a survey of related work (section 4) and a schedule of future work (section 5).
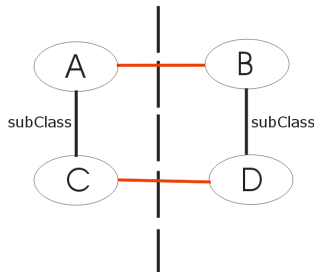
2   Mapping Patterns

Mapping patterns reflect the *internal structure* of ontologies as well as *mappings* between elements of (typically, two) ontologies. A mapping pattern is a graph structure, where nodes are classes, properties or instances. Edges represent mappings, relations between elements (eg. domain and range of properties) or structural relations between classes (eg. subclasses or siblings). In our experiments we employed three simple patterns. The *first* one is depicted in Figure 1. The left-hand side (class A) is from ontology

O1 and the right-hand side (class B and its subclass C) is from ontology O2. There is a mapping between A and B and at the same time between A and C. The *second* pattern is depicted in Figure 2. It is quite similar to the previous one, but now we consider a child and a parent from each ontology and simultaneous mappings between parents and between children. The *third* pattern is depicted in Figure 3. It consists of simultaneous mappings between class A from O1 and two sibling classes C and D from O2.
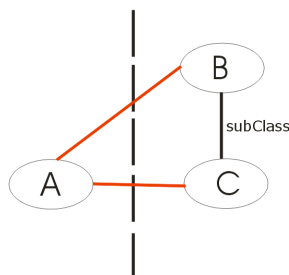
Fig. 1. Pattern 1 – 'Parent-child triangle'

Fig. 2. Pattern 2 – 'Mapping along taxonomy'

First experiments have been done over mapping results from participants of 'Conference track' [1] within OAEI-2006 [2] where I was co-organiser. The goal of these experiments was to find out about features of different OM systems[2] via a Data Mining (DM) approach using mapping patterns. An example of a hypothesis discovered is: "Mappings output by the HMatch tool with medium validity (between 0,5 and 0,8) are more likely to connect a child with a class that is also connected (with high va-

---

[1] http://nb.vse.cz/~svabo/oaei2006/

[2] This work also follows with my previous research in exploring interdependencies between mapping methods using Bayesian Networks, see [7].
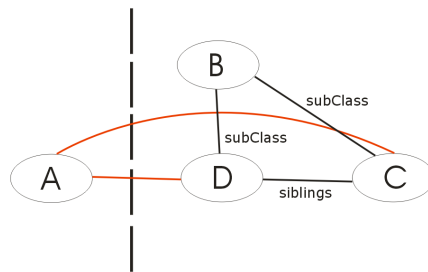
Fig. 3. Pattern 3 – 'Sibling-sibling triangle'

lidity) with a parent (Pattern 1) than such correspondences with all validity values (on average)." Detailed description of the approach used and more results are in [8].

3   Ontology Design Patterns

3.1   Design Patterns Types

The structure and naming in ontologies reflect diverse modelling styles of their creators. Typically, their aim is to ease the comprehensibility of the ontology for themselves and for others. We can observe symptoms of modelling styles as (design) patterns, which have been used either consciously, e.g. based on recommendations provided by the SWBPD group[3], or unconsciously. We are particularly interested in two types of ontology design patterns: *naming conventions* and *structural patterns*.
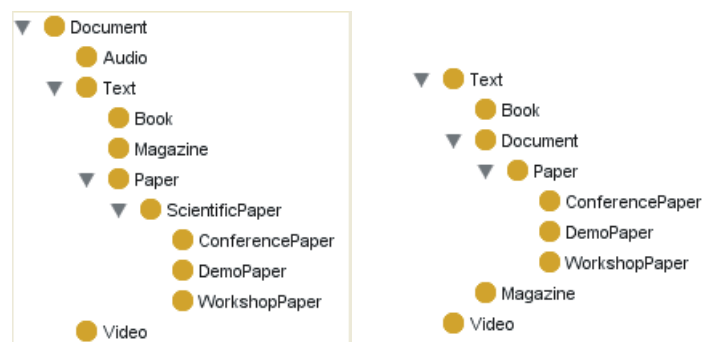

Fig. 4. Fragments of ontologies exemplifying design patterns

Naming conventions are related to naming classes, properties and/or instances. The way concepts are named can reveal their character. For example, in Figure 4 classes named with an appositive and noun (ConferencePaper) are subclasses of the class named with the noun (Paper). Another example could be the naming of inverse properties, e.g. *hasAuthor* vs. *authorOf*. Such heuristics are often globally used by the mapping tools, our aim however is to assess the degree in which each such pattern is a part of the author's modelling style and then seek it systematically in the ontology in question.

Structural patterns concern the modelling choices in using certain ontology entities and connecting them together. An example is a situation when some part of an ontology can be either 'class-centric' or 'property-centric', e.g. one designer may prefer to use a property 'holdsSeasonalTicket' (of which the domain is 'Person'), while another uses the class 'SeasonalTicketHolder' (as subclass of 'Person'). Another example could be the situation where one ontology is more granular than other. E.g. in the left-hand ontology in Fig. 4 there is an intermediate class 'ScientificPaper' in contrast with the right-hand ontology where this concept is missing.

3.2 Initial experiments: impact of patterns on Similarity Flooding

I just started to work on exploring how particular methods tackle with occurences of patterns. The Similarity Flooding (SF) algorithm [5] is based on the intuition that elements of two distinct models are similar when their adjacent elements are similar. At the beginning, initial mappings are typically set; the similarity of two elements is then iteratively (partly) propagated to their respective neighbors.

For initial setting, we used the string matcher distributed together with the SF code, which is based on comparing common prefixes and suffixes of literals. The question was whether SF would improve the initial string-based mapping specifically in the surroundings of patterns. The ontologies to be mapped were artificial variations of real ontologies from the OAEI 'conference' track. Patterns were systematically introduced to them, both naming conventions such as the couple 'ConferencePaper' and 'Paper' (helping the initial string matching) and structural patterns such as having 'Scientific-Paper' as additional intermediate class in the left-hand ontology in Fig. 4. SF algorithm improves the result of string comparison of 'ScientificPaper' and 'Paper' (which could presumably be equal) from 0.03 to 0.89. In the case of 'Paper' (left) and 'Paper' (right) concepts, SF algorithm decreases the similarity from 1 to 0.31, which could perhaps be desirable, as in the right-hand ontology the semantics of 'Paper' seems to be connected to that of its subclasses (types of scientific paper), while in the left-hand ontology it is broader than 'Scientific Paper' and can thus also include non-scientific literature. On the other side, SF algorithm yields the similarity of concepts 'ConferencePaper' from the left-hand ontology and 'ConferencePaper' from the right-hand ontology as 0.25, while string method yields it correctly as 1.

There thus seems to be a trade-off between the gains/losses of the string-based and graph based decisions. However, if we could automatically recognise an instance of pattern consisting of co-occurence of three siblings in both ontologies (which are even compound from an appositive and noun), we could prevent the graph-based method from reverting the correct decision yielded by the string-based method.

4 Related Work

Mapping patterns are implicitly considered in [4]; however, they focus on 'heterogeneous mappings' (class to property) as special kind of pattern. We also considered this, but it appeared too infrequently to allow for meaningful data mining. Our planned work on *design patterns* shares the general underlying idea with e.g. projects presented by Aleksovski [1] and Sabou [6], which also try to overcome inefficiencies of mapping methods using some kind of additional ontological knowledge. However, both these projects exploit the content of additional ontologies, while we focus on detecting pre-defined patterns in the original ontologies to be mapped.

5 Schedule of Future Work

The main direction of my PhD work in the next cca 6 months will consist in evaluating different mapping methods in terms of sensitivity to diverse design patterns, as suggested in section 3. Based on the results of this first step, I will try, in another 6 months, to automatically recognise these patterns and exploit them in the mapping process.

As secondary topic, I would also like to consider a richer variety of ontology mapping patterns as input to the data mining process as posterior evaluation of mappings (cf. section 2).

References

1. Aleksovski, Z., ten Kate, W., van Harmelen, F.: Exploiting the Structure of Background Knowledge Used in Ontology Matching In: Workshop on Ontology Matching at ISWC-2006.
2. Euzenat J., Mochol M., Shvaiko P., Stuckenschmidt H., Šváb On., Svátek V., van HAGE W. R., Yatskevich M. Results of the Ontology Alignment Evaluation Initiative 2006. In: Workshop on Ontology Matching at ISWC-2006.
3. Gangemi, A.: Ontology Design Patterns for Semantic Web Content. In: Proc. ISWC 2005.
4. Ghidini, C., Serafini, L.: Reconciling concepts and relations in heterogeneous ontologies. In: Proc. ESWC 2006, Budva, Montenegro, 2006.
5. Melnik S., Garcia-Molina H., Rahm E. Similarity flooding: A versatile graph matching algorithm. ICDE 2002.
6. Sabou, M., d'Aquin, M., Motta, E. Using the Semantic Web as Background Knowledge for Ontology Mapping. In: Workshop on Ontology Matching at ISWC-2006.
7. Šváb O., Svátek V. Combining Ontology Mapping Methods Using Bayesian Networks. In: Workshop on Ontology Matching at ISWC-2006.
8. Šváb O., Svátek V., Stuckenschmidt H. A Study in Empirical and 'Casuistic' Analysis of Ontology Mapping Results. Accepted for ESWC-2007.