OntoFarm: Towards an Experimental Collection of Parallel Ontologies[*]

Ondřej Šváb, Vojtěch Svátek, Petr Berka, Dušan Rak and Petr Tomášek
University of Economics, Prague, CZ-13067, Praha 3, Czech Republic
xsvao06@vse.cz, svatek@vse.cz, berka@vse.cz

Abstract

The number of semantic web tools assumed to operate on multiple ontologies from the same domain is growing far ahead of availability of suitable testing material. We suggest a domain and principles for building a widely-usable collection of parallel ontologies. An initial fragment of the collection has been created and processed with two among available tools.

1 Introduction

Although the OWL standard has been around for a while, it seems that the development of ontology-based tools aiming at interoperability on the semantic web [Kalfoglou *et al.*, 2004] is far ahead of the availability of solid material on which those techniques could be tested. By [Tempich and Volz, 2003], most ontologies with OWL *syntax* available from the DAML repository amount to either: (1) small and contrived artefacts developed for demonstration of reasoners or ontology editors, (2) converted database schemata with very few object properties, or (3) converted linguistic taxonomies, thesauri or topic hierarchies, with very few properties at all. Larger ontologies are very poor in axioms, which reduces their usability by reasoners. In addition, most ontologies are devoted each to a different domain, and even if the domain is similar at the first sight, the conceptual match is low: this further reduces their amenability to *ontology mapping/merging* or *distributed reasoning*. It seems that the community would benefit from the existence of a collection of OWL ontologies that would

- share the same, generally understandable domain
- be built by different groups, thus reflecting slightly different conceptualisations, but at the same time sharing some core commitments
- be rich in various types of axioms.

Within this paper, we describe the principles and very first accomplished steps of a collaborative initiative nicknamed *OntoFarm*, which aims to grow such a collection.

2 Choice of Domain

We believe that the domain of *conference organisation* could be most suitable, among other, for the following reasons:

- Most ontology engineers are academics who themselves submit and review papers and organise conferences: there is zero overhead of acquiring the *domain expertise*.
- Organisation of a conference shares some aspects with (heavier-weighted) *business activities*: access restrictions, hard vs. soft constraints, temporal dependencies among events, evolution of the meaning of concepts in time etc. There is also a wide range of supporting *software tools* covering various aspects of conference organisation. Their domain assumptions can also be captured using ontologies (specific for each system). The process of matching the requirements of conference organisers with the capacities of such tools is analogous with that of matching the requirements of a business with the capacities of an off-the-shelf enterprise information system.
- In many cases, even the underlying *instance data* could be obtained, since legal restrictions are typically not as strong as e.g. in business or medicine.

Although an ontology *biased* by a particular event or supporting tool does not fully comply with the definition of ontology in narrow sense, we believe that it could mimic true semantic-web ontologies reasonably well.

3 Development of Start-Up Collection

Four ontologies were developed so far. A medium-sized[1] ontology was designed based on the structure of the well-known *EKAW conference* (by one of authors, who is Co-Chair of the next edition). The remaining three (smaller) ontologies were derived from different conference *support tools*, using their documentation and experiments with installed tools. The designers were students of the course on Knowledge Modelling at the University of Economics, Prague, in the 2005 Spring term. The students were instructed to work in isolation from each other, to include as many (meaningful) axioms as possible, to avoid describing the tool itself (thus focusing on the content semanics), and were given a list of sub-areas possibly covered by the tool/s (such as 'CFP distribution', 'reviewing process' or 'participant registration') as informal guidance.

[*] The project is supported by the Internal Grant Agency of University of Economics, Prague, under no.26/05, and by the Knowledge Web Network of Excellence (IST FP6-507482).

[1] It contains about 80 classes and 30 properties.

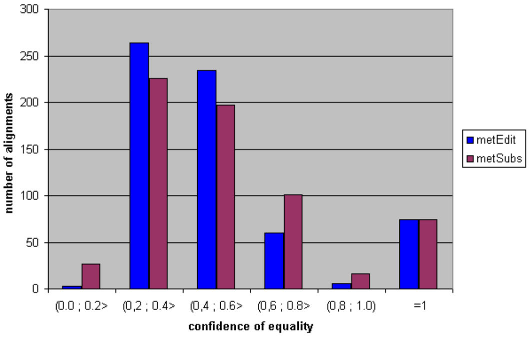

Figure 1: Distribution of alignment confidence

4 Subsequent Experiments

As a proof of concept, this tiny initial collection of ontologies was submitted to some freely available tools; most experiments were done using the DRAGO *distributed reasoner* [Serafini and Tamilin, 2005] and the tools accessible via the *Ontology Alignment API* [Euzenat, 2005].

DRAGO allows to specify one-way mappings (so-called bridge rules using $\sqsubseteq$ and $\sqsupseteq$ operators) among ontology concepts and applies usual description logic (subsumption and satisfiability) reasoning in a distributed manner. For pairs of ontologies i, j, we specified bridge rules and arrived at some meaningful structural updates of one ontology based on the other one. As a very simple example: the concept of ProgramCommitteeChair was subclass of ProgramCommitteeMember in ontology i, while Chair_PC was not subclass of Member_PC in ontology j. However, upon specifying bridge rules in the form i:ProgramCommitteeChair $\sqsupseteq$ j:Chair_PC, i:ProgramCommitteeMember $\sqsubseteq$ j:Member_PC, the subclass relationship between Chair_PC and Member_PC was derived in j via distributed tableau reasoning.

Ontology alignment API gives access to several simple methods for mapping between ontology entities (classes/properties). We used two among the four available methods: *EditDistNameAlignment* uses an editing (or Levenstein) distance between (downcased) entity names; *SubsDistNameAlignment* computes a substring distance on the (downcased) entity name. In our case, we arrived at more than 600 individual alignments for each method, with varying degree of confidence (computed from inverse distance); the results decomposed to confidence intervals are at Figure 1. We have not yet managed to compute the overall accuracy of alignments (which may actually be quite difficult due to some ambiguous cases). Examples of correct non-exact matches are e.g.: Chairman=Chair_PC with confidence of 0.625 and (only!) 0.344, by the two methods, respectively, or title=hasTitle with confidence 0.625 and 0.769.

Experiments with both tools indicated that even a small initial collection of parallel conference ontologies may serve as usable material for testing ontology processing tools. Obviously, the utility of the collection would grow with its size.

5 Related Work and Future Plans

Parallel ontologies are examined within *Ontology Alignment Contests*[2]; those are however large taxonomies (without much additional structure) or simple 'calibration' models derived from one another in a domain-irrelevant fashion. Within *OntoFarm*, we plan to arrive at a collection of realistic and axiomatised ontologies, which could further stimulate the development of advanced ontology engineering tools. A problem identified by the organisers of OA Contests is the difficulty of obtaining an a priori 'gold standard' for alignments. We suggest to get around this problem via relaxing the competitive aspect of such a multi-party activity in favour of the *collaborative* aspect. Our own experience with organising a similar activity in the area of Knowledge Discovery from Databases[3] (from 1999 annually) is quite encouraging: multiple teams analysed the same collection of data and submitted papers with results, which were peer-reviewed and presented within a workshop and eventually brought better insight into strengths and weaknesses of each technique.

A possible bottleneck of our initiative could be the fact that organisers of conferences related to ontology engineering may not be sufficiently motivated to take the additional pain of building an ontology of the conference itself, despite their excellent capacity for this task. In order to provide them with *instant gratification*, we envisage an HTML-based application over the gradually growing collection. Its functionality could be e.g. to identify un/usual aspects of individual events, to suggest the most appropriate support tool, or even (assuming the availability of instances and A-box reasoning) to explore thematical or personal overlaps.

Acknowledgments

We thank Andrei Tamilin for his assistance with DRAGO. Feedback from Jerôme Euzenat and Heiner Stuckenschmidt is also deeply acknowledged.

References

[Euzenat, 2005] Jerôme Euzenat. *An API for ontology alignment (version 1.3)*. Online at http://co4.inrialpes.fr/align/align.pdf.

[Kalfoglou *et al.*, 2004] Yannis Kalfoglou, Marco Schorlemmer, Amit Sheth, Steffen Staab, Mike Uschold (eds.). *Proceedings of the Dagstuhl Seminar 04391 on Semantic Interoperability and Integration*. Online at http://www.dagstuhl.de/04391/index.en.phtml.

[Serafini and Tamilin, 2005] Luciano Serafini and Andrei Tamilin. DRAGO: Distributed Reasoning Architecture for the Semantic Web. In *Proc. 2^{nd} European Semantic Web Conference (ESWC'05)*, Springer-Verlag, 2005.

[Tempich and Volz, 2003] Christoph Tempich and Raphael Volz. Towards a benchmark for Semantic Web reasoners – an analysis of the DAML ontology library. In *Proc. EON2003 Workshop at ISWC 2003, Florida.*

[2] http://co4.inrialpes.fr/align/Contest, http://www.kcap05.org/integrating-ontologies.html

[3] http://lisp.vse.cz/challenge

Explanation of Demo

By the time of the conference, we plan to have a demo application, which will actually be the very initial version of the ‘gratification’ application mentioned in the end of the poster abstract. The application will address two different tasks.

The first task is to *sort* the ‘software tools’ ontologies according to *similarity* with a given ontology. The input will be an ontology presumably derived from an actual conference series (such as EKAW). Concept-level similarity will be computed using the tools available through the OWL Alignment API (or others). The overall similarity among the ontologies will then be computed from it. The output will be a sorted list of ‘software tools’ ontologies. Ideally, the user could use it as guidance for choosing an adequate tool for supporting the organisation of his/her conference.

The second task of demo application aims at discovering *differences* between a given pair of ontologies. The input will be two ontologies and the threshold for concept similarity. The whole procedure consists of two phases. In the first phase, individual concepts of these two ontologies will be *automatically mapped* (again using tools available through the OWL Alignment API) every time their similarity is lower than the given threshold. The mappings will be saved in C-OWL format. Both ontologies (one considered as target and on as source ontology) will be uploaded into two Drago Reasoning Peers with obtained mappings. In the second phase, both the ontologies will be checked for *consistency*, and also local and distributed *taxonomy* will be inferred. The output will be the listing of both local and distributed taxonomy. Potentially, we also consider to implement *automatic* discovery of differences between the two ontologies. Ideally, this could help e.g. suggest potentially useful enhancements of conference support tools in terms of coverage.

The primary target of the application is, however, to demonstrate the utility of a collection of parallel ontologies for evaluation of ontology engineering tools rather than to provide immediate benefits to the end user.