A Framework for Session-based Ontology Alignment

Patrick Lambrix

Department of Computer and Information Science
Linköping University, 581 83 Linköping, Sweden

In this abstract we tackle the problem of aligning large ontologies where the mappings suggested by the ontology alignment system need to be validated. Although we focus on an ontology alignment framework, the ideas may be used and extended for community-based or collaborative ontology alignment.

In contrast to the case of small ontologies, the computation of mapping suggestions can take a long time and therefore, we would like to be able to start the validation before every mapping suggestion is computed. Further, it is clear that for large ontologies, in general, there are too many mapping suggestions to validate in one time. Therefore, we want a system that allows to partially validate the mapping suggestions and resume the validation later. However, whenever validation decisions have been made, they increase our knowledge about the ontologies and mappings and this knowledge can be used to provide better mapping suggestions. In the remainder of the abstract we propose an iterative ontology alignment framework that deals with these issues.

Framework. Our framework is presented in figure 1. The input to the system are the ontologies that need to be aligned, and the output is an alignment between the ontologies. When starting an alignment process the user starts a computation session. When a user returns to an alignment process, she can choose to start or continue a computation session or a validation session.

During the *computation sessions* mapping suggestions are computed. The computation may involve preprocessing of the ontologies, matching, and combination and filtering of matching results. Auxiliary resources such as domain knowledge and dictionaries may be used. Users may be involved in the choice of algorithms. This is similar to what most ontology alignment systems do. However, in this case the algorithms may also take into account the results of previous validation and recommendation sessions. The output of a computation session is a set of mapping suggestions. The computation sessions can be stopped and partial results can be delivered.

During the *validation sessions* the user validates the mapping suggestions generated by the computation sessions. The output of a validation session is a set of mapping decisions (accepted and rejected mapping suggestions). The accepted mapping suggestions form a partial reference alignment (PRA) and are part of the final alignment. The mapping decisions can be used in future computation sessions (e.g. PRA-based preprocessing and filtering [1]) as well as in recommendation sessions. Validation sessions can be stopped and resumed at any time. It is therefore not neccesary for a domain expert to validate all mapping suggestions in one session. The user may also decide not to resume the validation but start a new computation session, possibly based on the results of a recommendation session.

The input for the *recommendation sessions* consists of a database of algorithms for the preprocessing, matching, combination and filtering in the computation sessions.

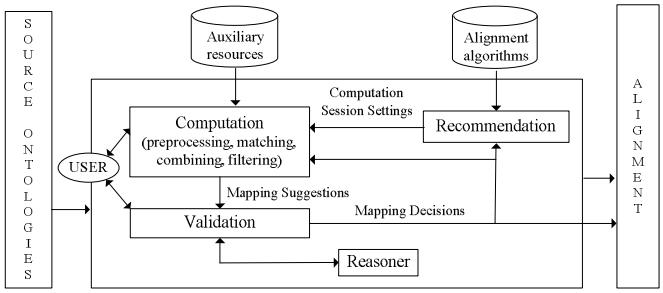

Fig. 1. Framework.

During the recommendation sessions the system computes recommendations for which (combination) of those algorithms may perform best for aligning the given ontologies. When validation results are available these may be used to evaluate the different algorithms, otherwise an oracle may be used. The output of this session is a recommendation for the settings of a future computation session. These sessions are normally run when a user is not validating and results are given when the user logs in into the system again.

Current implementation. We have implemented a prototype based on the framework described above. Regarding the computation sessions, when a PRA is available, the ontologies are preprocessed to partition the ontologies into corresponding mappable parts [1]. For the matching we use the linguistic, WordNet-based, structural and instance-based algorithms from the SAMBO system [2], a weighted sum approach for the combination, and the single and double threshold filter approaches. When a PRA is available we also use the filter approaches from [1]. Users can choose which algorithms, weights and tresholds to use, or use default values. For the validation we use the user interface of SAMBO where a user can accept, reject or modify mapping suggestions as well as annotate the decisions. A reasoner is used to detect conflicts in the decisions and notify the user. Validation sessions can be stopped at any time and resumed later on (if so desired - the user may also start a new computation session). The recommendation algorithm is based on the algorithm described in [3]. Currently, the performance of the different alignment algorithms is evaluated based on how well they do on small pieces of the ontologies already aligned by an oracle. In the future, we will also take the validation decisions of the user into account and adapt the recommendation.

Acknowledgements. We thank Qiang Liu, Muzammil Zareen Khan and Shahab Qadeer for their implementation work in this project.

References

1. P Lambrix and Q Liu. Using partial reference alignments to align ontologies. In *6th ESWC, LNCS 5554*, pages 188–202, 2009.
2. P Lambrix and H Tan. SAMBO - a system for aligning and merging biomedical ontologies. *Journal of Web Semantics*, 4(3):196–206, 2006.
3. H Tan and P Lambrix. A method for recommending ontology alignment strategies. In *6th ISWC and 2nd ASWC, LNCS 4825*, pages 494–507, 2007.