

Parallelization and Distribution Techniques for Ontology Matching in Urban Computing Environments

Axel Tenschert¹, Matthias Assel¹, Alexey Cheptsov¹, Georgina Gallizo¹, Emanuele Della Valle², Irene Celino²

¹ HLRS – High-Performance Computing Center Stuttgart, University of Stuttgart,
Nobelstraße 19,
70569 Stuttgart, Germany
 {tenschert, assel, cheptsov, gallizo}@hlrs.de

² CEFRIEL - ICT Institute, Politecnico of Milano,
Via Fucini 2,
20133 Milano, Italy
 {emanuelle.dellavalle, irene.celino}@cefriel.it

Abstract. The usage of parallelization and distribution techniques in the field of ontology matching is of high interest for the semantic web community. This work presents an approach for managing the process of extending complex information structures as used in Urban Computing system by means of ontology matching considering parallelization and distribution techniques.

Keywords: Ontology Matching, Semantic Content, Parallelization, Distribution

Ontology Matching through Distribution and Parallelization

Current ontology matching approaches [1] require a high amount of compute resources with the aim to meet the requirements of the matching and merging methods. Hence, several issues have to be considered such as the selection of a suitable ontology, scalability and robustness, matching sequence and identification of the ontology repositories. Approaches for partitioning selected ontologies with the aim to execute matching processes independently from other parts of the ontology are considered to solve this challenge [2]. However, a local ontology matching is a risk for these approaches in terms of scalability and performance issues. Therefore, local ontology matching could be extended by making use of distribution methods as well as parallelization techniques allowing overcoming existing limitations and improving the overall performance.

Within the LarKC project¹, respective techniques for processing large data sets in the research field of the semantic web are investigated and developed. In particular, distribution methods and parallelization techniques are evaluated by executing the matching processes concurrently on distributed and diverse compute resources. A

¹ LarKC (abbr. The Large Knowledge Collider): <http://www.larkc.eu>

dedicated use case in LarKC deals with the application of these techniques for Urban Computing problems [3].

Common ontology matching algorithms often perform computation intensive operations and thus being considerably time consuming. That poses a number of challenges towards their practical applicability for complex tasks and efficient utilization of the computing architectures that best fit the requirements in order to achieve maximal performance and scalability of the performed operations [4]. Distributed ontology matching enables the use of diverse computing resources, from users' desktop computers to heterogeneous Grid/Cloud infrastructures. Parallelization is the main approach for the effective ontology matching, especially when time characteristics are settled to the point. When thinking of matching several parts of an ontology in parallel in a cluster environment, the matching processes needs to be partitioned. After processing the data, the parts of the ontology have to be merged together again and an extended ontology is generated [5].

Several techniques can be recognized for the parallel implementation of distributed ontology matching.

- **Single Code Multiple Data (SCMD workflow)**
In this case the data that is being processed in the code region can be constructed of subsets that have no dependencies between them. The same operation is performed on each of these subsets.
- **Multiple Code Single Data (MCSD workflow without conveyer dependencies)**
For this workflow, several different operations are performed on the same dataset. Herewith, no dependencies between processed data sets exist. This is typical for a transformation of one dataset to another one according to rules, which are specific for each subset of the produced data.
- **Multiple Code Multiple Data (MCMD workflow)**
This type of workflow is the combination of both previous workflows (SCMD and MSCD).

The presented approach is an effective method to solve the challenge of matching large ontologies in a scalable, robust and timesaving way. Within the LarKC project, these parallelization and distribution techniques for processing semantic data structures are deeply analyzed and further developed.

Acknowledgments. This work has been supported by the LarKC project (<http://www.larkc.eu>), partly funded by the European Commission's IST activity of the 7th Framework Program. This work expresses only the opinions of the authors.

References

1. Alasoud, A., Haarslev, V., Shiri N.: An empirical comparison of ontology matching techniques. *Journal of Information Science* 35, 379--397 (2009)
2. Hu W., Cheng G., Zheng D., Zhong X., Qu Y.: The Results of Falcon-AO in the OAEI 2006 Campaign. *Ontology Alignment Evaluation Initiative* (2006)
3. Kindberg, T., Chalmers, M., Paulos E.: Introduction: Urban Computing. *IEEE Pervasive Computing* 6, 18--20 (2007)
4. Shvaiko, P., Euzenat, J.: Ten Challenges for Ontology Matching. In: *Proceedings of ODBASE, LNCS 5332*, pp. 1164--1182, Springer (2008)
5. Dean, J., Ghemawat, S.: Simplified Data Processing on Large Clusters. *OSDI'04: Sixth Symposium on Operating System Design and Implementation*, San Francisco (2004)