

Chapter 9

Grid Computing for Ontology Matching

Axel Tenschert

High Performance Computing Center Stuttgart (HLRS), Germany

ABSTRACT

This chapter examines the challenge of ontology matching in a grid environment in a scalable and high efficient way. For this, ontology matching approaches as well as grid computing are considered with the aim to present an approach for ontology matching on various resources. Hence, related approaches and tools are presented and discussed in order to provide an adequate background. Through this, a distributed ontology matching as it is required for ontology matching in a distributed environment such as the grid becomes usable. However, a novel ontology matching approach which meets the requirements of a grid architecture is considered in this chapter.

INTRODUCTION

Nowadays, lots of different ontologies as well as tools and frameworks that support matching are available. However, the amount of available semantic data structures is significantly growing. This issue faces the challenge of ontology matching in a scalable way considering large scale ontologies. For this, a strategy for handling big data sizes in an efficient way is required. Further,

the field of bioinformatics is a beneficial use case because of the amount of available domain ontologies such as from the official NCBO BioPortal website (The National Center for Biomedical Ontology, 2009) and the need to examine several ontologies to solve a specific question.

The complexity of matching large-scale ontologies and ontology matching in urgent computing use cases entails the problem of matching in a scalable way. Hence, distribution techniques such as grid computing are used to increase scalability by executing it by the use of distributed

DOI: 10.4018/978-1-61350-113-9.ch009

heterogeneous computing resources. Hence, the aim of this work is to support ontology matching strategies with a grid architecture to provide required computing resources for compute resource intensive and urgent computing use cases. The ontologies that are considered for this work are mostly OWL ontologies.

Furthermore, the LarKC project (The LarKC project, 2010) in which new techniques for processing large datasets are developed for the usage of concrete use cases such as “Semantic Integration for Early Clinical Development” or “Carcinogenesis Reference Production” is a beneficial basis for this work. Within LarKC project ontologies are used as well and new techniques usage of large-scale data sets are developed and used for real time applications such as an urban city use case that requires continuously semantic data with the aim to analyze such data in a time-saving way. For use cases as they are mentioned above, sophisticated techniques are required to run several processes at same time in a distributed fashion as it is done in the grid. At this, the new idea is to set one ontology from a given set as the priority ontology which is enhanced by matching the concepts of the priority ontology with the concepts of other selected ontologies from a given set.

This work supports an adequate matching procedure as well as a mapping of similar concepts or properties of the ontologies in a grid. For this, the first step is to define an adequate architecture which meets the requirements of ontology matching in a grid. After the matching is executed the ontology parts and the matching results are merged together in the priority ontology with the aim to extend this ontology. However, the matching strategy will be aligned to the grid architecture. The presented idea for matching ontologies in a grid environment is an effective method to solve the challenge of matching in a scalable, robust and time-saving way. Further, it is of interest to examine the principles and design issues for this topic.

Currently, lots of ontology matching strategies have been published so far. Hence, a clear distinction between this work and current approaches is required. This work aims to map concepts, properties as well as relations between entities of several ontologies in a semi-automatic manner by considering well known developments in this field and extension of these approaches. Further, the ontology matching is supported by a distributed architecture in order to use several compute resources. Beyond the distribution of ontology matching processes, ontologies are selected with the aim to enhance one ontology to a priority ontology of a given set. In order to achieve an adequate mapping of similarities for the enhanced priority ontology similarity values for entity pairs within the matched ontologies are calculated. However, the calculation of similarity values for entity pairs encounters the problem of evaluating one measurement out of lots of values such as similarity of properties or relations and relevance to neighboring entity pairs and much more.

Further, as mentioned the matching of large data sets requires a high amount of compute resources as well. For this, processing algorithms at same time on several nodes (e.g. cluster architecture) is a beneficial solution. However, this raises the challenge of distributing the jobs in a high effective way as well as aligning algorithms to avoid latencies or conflicts. Therefore, two main issues are considered in this work: (i) Ontology matching based on similarity values, and (ii) Distributed ontology matching on several resources by usage of a grid architecture.

RESEARCH BACKGROUND

At present the number of different semantic information is growing significantly because of a huge amount of different available resources. This trend raises the problem of managing required information in order to increase heterogeneity

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/grid-computing-ontology-matching/58747

Related Content

IoT and Big Data Technologies for Monitoring and Processing Real-Time Healthcare Data

Abdelhak Kharbouch, Youssef Naitmalek, Hamza Elkhouchi, Mohamed Bakhouya, Vincenzo De Florio, Moulay Driss El Ouadghiri, Steven Latreand Chris Blondia (2019). *International Journal of Distributed Systems and Technologies* (pp. 17-30).

www.irma-international.org/article/iot-and-big-data-technologies-for-monitoring-and-processing-real-time-healthcare-data/240251

Energy Efficient Resource Allocation During Initial Mapping of Virtual Machines to Servers in Cloud Datacenters

Nimisha Patel and Hiren Patel (2018). *International Journal of Distributed Systems and Technologies* (pp. 39-54).

www.irma-international.org/article/energy-efficient-resource-allocation-during-initial-mapping-of-virtual-machines-to-servers-in-cloud-datacenters/196266

Fuzzy Allocation of Fine-Grained Compute Resources for Grid Data Streaming Applications

Wen Zhang, Junwei Cao, Yisheng Zhong, Lianchen Liu and Cheng Wu (2012). *Evolving Developments in Grid and Cloud Computing: Advancing Research* (pp. 283-294).

www.irma-international.org/chapter/fuzzy-allocation-fine-grained-compute/61998

Deep Analysis of Enhanced Authentication for Next Generation Networks

Mamdouh Gouda (2010). *International Journal of Grid and High Performance Computing* (pp. 37-52).

www.irma-international.org/article/deep-analysis-enhanced-authentication-next/43883

Analyzing the Impact of Machine Learning on Cancer Treatments

Victor Chang, Gunji Srilikhita, Qianwen Ariel Xu, M. A. Hossain and Mohsen Guizani (2022). *International Journal of Distributed Systems and Technologies* (pp. 1-22).

www.irma-international.org/article/analyzing-the-impact-of-machine-learning-on-cancer-treatments/304429