# Role of the Ontologies in the context of Grid computing and application for the human disease studies

Maja Hadzic, Elizabeth Chang

Curtin University of Technology, School of Information Systems,
GPO Box U1987 Perth,
Western Australia 6845, Australia
{hadzicm, change}@cbs.curtin.edu.au

## 1 Introduction

We describe the increasing impact of ontologies in the context of Grid computing for obtaining, comparing and analyzing distributed heterogeneous scientific data. The inherently autonomous and heterogeneous nature of the information resources forces applications to share data and services often without prior knowledge of their structure and functionality, respectively. Ontologies are needed to provide a way to capture and present in the computer, knowledge shared by all people in a certain community. Computer based ontologies may be seen as shared formal conceptualizations of domain knowledge and therefore constitute an essential resource for enabling interoperation in an open environment such as the Web on the Grid. We illustrate how ontologies can be developed for the knowledge domain of biomedical and bioengineering research. We chose the application domain of human disease research and control since it necessarily involves resources of phenotypic, genetic, environmental and treatment data.

## 2 Importance of Grid in the biomedical environment

Medical researchers consist of teams with heterogeneous members with different capabilities. There does not exist a unique organization that has all the required resources or skills and team members to be distributed around the globe. The Grid should enable resources sharing and usage co-ordination in dynamic, virtual, multi-institutional organizations [2]. Grid middleware provide the required distributed collaborative platform as well as easy access to resources. The source information covers different areas of interest with respect to human diseases in order to allow different user categories, each having specific intentions, to query the system. Another major advantage of using the Grid is that it respects complete autonomy of the existing ontology nodes. Each of the existing nodes can withdraw or join the Grid whenever it is necessary. This is very important when generating on request Specific Human Disease Ontologies.

## 3 Generic vs. Specific Human Disease Ontology, using DOGMA methodology

Our Generic Human Disease Ontology has four main branches: *(1) phenotype*, describing symptoms of a disease; *(2) treatment*, giving an overview of all treatments possible for that particular disease as well as treatment's efficiency; *(3) causes* responsible for that disorder, which can be environmental and/or genetical; *(4) types*, describing different types of a disorder

By combining grid services with a prototype of Generic Human Disease Ontology (GHDO), we extract and align the relevant information requested by an user from publication and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc. This allows generating Specific Human Disease Ontologies (SHDO) for simple disorders such as SCA (sickle cell anemia) but, also for complex disorders such as MDD (manic-depressive disorder). The Generic Human Disease ontology stands here central as a link between multiple heterogeneous information resources on one side and the users on the other side. The specific information requested by user is accessed by using Grid services. Merging and aligning of this information to the Generic Human Disease Ontology results in Specific Human Disease Ontologies.

For the ontology design, we adopted the methodology of DOGMA [1]. In this approach database schema elements, as well as linguistical elements are represented as lexons combining the knowledge domain. Knowledge about their usage (such as constraints, rules, etc.) is kept rigorously separate and is specified as part of the formal commitment of an application to these lexons. This so-called double articulation [3] permits a high degree of scalability, an essential requirement for Grid-based computing. A second fundamental aspect of DOGMA is that it distinguishes data models (which are embedded in specific applications) from proper ontologies (this should be application-independent). The mapping of a data model to an ontology precisely constitutes its formal semantics, in fact reified as part of commitment.

## 4 Use cases

We give some examples for the purpose of illustration of use and application of the Human Disease Ontologies. The Generic Human Disease Ontology with its four main branches (symptoms, treatments, causes and types) serves as a template. Grid services then "feed" applications committed to this ontology with relevant data required by a user which results in Specific Human Disease Ontologies. Researchers and physicians are strongly connected because they are working towards the same goal, but on different knowledge levels. The examples are intended to show typical, common problems researchers and physicians encounter.

## 5 Conclusion

We show how the combination of two different but complementary techniques, namely Grid computing and ontology, results in a dynamic and intelligent information system. The two approaches together, being complementary, enable the system as a whole.

# References

1. Deray T., Verheyden P.: Towards a semantic integration of medical relational databases by using ontologies: a case study. In: Meersman R.: Semantic Ontology Tools in Information System Design. In: Ras Z., Zemankova M. (eds.): Proceedings of the ISMIS 99 Conference, Lecture Notes in Computer Science Vol. 1609, Springer-Verlag Heidelberg, (1999) 30-45.
2. Carole Goble: The Grid Needs you! Enlist Now. In: Robert Meersman, Zahir Tari, Douglas C. Schmidt (eds.): On the Move to Meaningful Internet Systems 2003, Confederated International Conferences DOA, CoopIS and ODBASE 2002 Irvine, California, USA, Proceedings. Lecture Notes in Computer Science, Springer-Verlag Heidelberg, ISBN 3-540-20498-9 (2003) 589-600.
3. Spyns P., Meersman R., Jarrar M.: Data modelling versus Ontology engineering. SIGMOD Record, Vol. 31, No. 4 (2002) 12-17.