Semantic Data Integration in the European ACGT project

Luis Martín¹, Alberto Anguita¹, Guillermo de la Calle¹, Miguel García-Remesal, PhD¹, Jose Crespo, PhD¹, Manolis Tsiknakis, PhD², Víctor Maojo, MD, PhD¹,
¹Biomedical Informatics Group, Artificial Intelligence Lab./Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain; ²Biomedical Informatics Laboratory. Ins. of Computer Science. Foundation for Research and Technology - Hellas. Heraklion, Crete, Greece

Abstract

ACGT is an IST-FP6 Integrated Project, funded by the European Commission, for the development of services to support clinico-genomic trials on cancer in a grid-based environment. In these trials, physicians and researchers need to access heterogeneous and disparate data sources. Semantic access to these data and the possibility to integrate them seamlessly are issues that ACGT aim to solve.

Introduction

Advancing Clinico Genomic Trials on Cancer (ACGT) is a European research project devoted to the development of tools and services for supporting multicenter, post-genomic trials on cancer. ACGT is focused on providing a portal of services that will help researchers in carrying out their investigations and sharing their data, methods and tools.

The integration of multilevel clinical and genomic data is a key topic in the ACGT project. Within a post-genomic trial, both clinicians and researchers need to access federated data sources of different nature. In ACGT, a middleware called the “ACGT Mediator” is under development to solve this issue.

The ACGT Mediator

Integration of heterogeneous and distributed data is not a trivial problem. Different types of heterogeneity, both at the schema and instance level, must be solved. Furthermore, the solution must allow to integrate new sources as seamlessly as possible.

In the ACGT platform, the semantic mediator is responsible for the integration of distributed and heterogeneous data. The mediation middleware offers to client layers, in the ACGT architecture (end users interfaces and analytical tools), uniform access to a set of heterogeneous sources, containing all kind of data regarding the specific clinical trial. To address this task, the semantic mediator is supported by the ACGT Master Ontology, developed by researchers at IFOMIS (a member of the consortium), to model the cancer trials domain.

The selected approach for the semantic mediator is based on a LaV² (Local as View) approach, opposed to GaV (Global as View). It seems particularly adequate in the ACGT context, given the nature of data in the selected clinical trials, where new databases can be continuously added. This data-driven approach allows maintaining the mediation software while introducing changes in the data set.

Conclusion

Preliminary experiments are already proving the viability of this work. User requirements have already been specified and will continue during the whole project timeframe. In the first year of the project, the work has been centered on research on semantic mediation and the development of a demonstrator which is near completion. This includes a subset of the features of the semantic mediator (including the integration of images and clinical and genetic data sources). Together with this mediator, suitable interfaces to both end users and analytical tools, a Clinical Report Form (CRF) manager and an ontology navigator are being developed.

Acknowledgments

The present work has been funded by the European Commission (FP6, IST Thematic area) through ACGT (FP6-2005-IST-026996).

References


AMIA 2007 Symposium Proceedings Page - 1042