A Semantically-Enabled System for Clinical Trials

Zhisheng Huang Annette ten Teije Frank van Harmelen

Department of Computer Science, VU University Amsterdam, The Netherlands

Abstract

In this demo we present *SemanticCT*, a semantically-enabled systems for clinical trials. SemanticCT is built on the top of LarKC, a platform for scalable semantic data processing. SemanticCT has been integrated with large-scale trial data and patient data, and provided various automatic services for clinical trials, which include automatic patient recruitment service (i.e., identifying eligible patients for a trial), trial finding service (i.e., finding suitable trials for a patient), and trial feasibility service (i.e., analyzing possible recruitment at design time).

1 Introduction

Clinical trials provide tests which generate safety and efficacy data for health interventions. Clinical trials usually involve large-scale and heterogeneous data. The lack of integration and of semantic interoperability among the systems of clinical trials and the systems of patient data is the main source of inefficiency of clinical trial systems. Enhancing clinical trial systems with semantic technology to achieve the semantic interoperability of large-scale and heterogeneous data would improve the performance of clinical trials significantly. Those semantically-enabled systems would achieve efficient and effective reasoning and data processing services in various settings of clinical trials systems.

We have developed a semantically-enabled system for clinical trials, which is called *SemanticCT*[2]. SemanticCT has been semantically integrated with various data, which include various trial documents with semantically annotated eligibility criteria and large amount of patient data with structured EHR and clinical medical records. Well-known medical terminologies and ontologies, such as SNOMED, LOINC, etc., have been used for the semantic interoperability. SemanticCT is built on the top of LarKC, a platform for scalable semantic data processing[1]. With the built-in reasoning support for large-scale RDF/OWL data of LarKC, SemanticCT is able to provide various reasoning and data processing services for clinical trials, which include faster identification of eligible patients for recruitment service and efficient identification of eligible trials for patients, and a trial feasibility service. The trial feaibility service provides functionality to change eligibility criteria and their parameters, and to support the process of designing the eligiblity criteria by calculating the absolute and relative feasibility.

2 Semantic Data Integration

Many existing trial data are usually represented as XML data with the standard fields. For example, the clinical trial service in the U.S. National Institutes of Health[2] provides the structured CDISC 20 fields of XML-encoded trial data. We can convert those XML data into standard semantic data, like RDF NTriple

[1] http://www.larkc.eu

[2] http://www.clinicaltrials.gov

data with the annotations of medical ontologies or terminologies, like SNOMED, LOINC, MESH and others. Those ontologies can be used individually, or in a group with the ontology alignments which are provided by the BioPortal ontology service[3] or other alignment tools. LinkedCT[4] provides large-scale semantic data of clinical trials with the standard formats of Linked Open Data in the Semantic Web. The total loaded RDF NTriple data are over 6 million triples.

3 Implementation

The architecture is shown in Figure 1. SemanticCT Management launches a web server which serves as the application interface, so that the users can use a web browser to access the system locally or remotely. SemanticCT Management manages SPARQL endpoints which are built as SemanticCT workflows for several tasks like semantic search, patient recruitment, trial feasibility. A generic reasoning plug-in in LarKC provides the basic reasoning service over large-scale semantic data, like RDF/RDFS/OWL data. SemanticCT Management interacts with the SemanticCT Prolog component which provides the rule-based reasoning[1].

SemanticCT provides the following services: i)semantic search, i.e., use SPARQL queries to search over semantic data, ii) keyword search, i.e., use keywords to search over eligibility criteria, iii) annotated criteria browsing: the service for browsing semantically annotated eligibility criteria of trials, iv) patient recruitment: the service for identifying eligible patients with the support of rule-based reasoning, v) trial finding: the service for finding suitable trials for a patient, and vi)trial feasibility: the service for the feasibility analysis on a trial at design time.

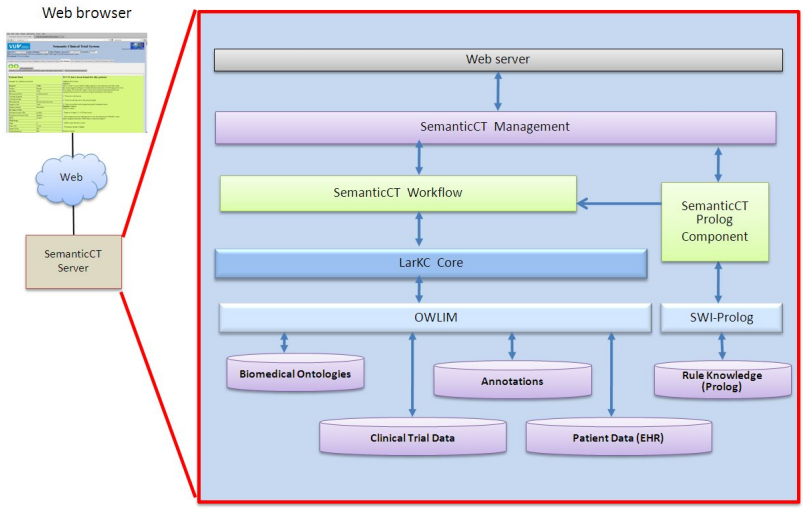

Figure 1: The architecture of SemanticCT.

References

[1] Zhisheng Huang, Annette den Teije, and Frank van Harmelen. Rule-based formalization of eligibility criteria for clinical trials. In *Proceedings of the 14th Conference on Artificial Intelligence in Medicine(AIME 2013)*, 2013.

[2] Zhisheng Huang, Annette ten Teije, and Frank van Harmelen. SemanticCT: A semantically enabled clinical trial system. In R. Lenz, S. Mikszh, M. Peleg, M. Reichert, D. Riano, and A. ten Teije, editors, *Process Support and Knowledge Representation in Health Care*. Springer LNAI, 2013.

[3] http://bioportal.bioontology.org/
[4] http://linkedct.org/