Semantic Lattes and VIVO Project

Alexandre Rademaker*   Edward Hermann Haeusler†

May 11, 2013

Abstract

We present a project for data integration about researchers resumes using open data standards languages such as RDF and OWL. We also use standard ontologies and vocabularies like FOAF, BIBO, SKOS etc. The data is collected from the Brazilian CNPq Lattes Platform and will be made available online as SPARQL endpoint and in a VIVO Application instance after the mapping from our data model to VIVO ontology.

Keywords: Lattes, VIVO, RDF, Semantic Web, data integration, data consistence and data integrity constraints.

In 2010, we started a project at Getulio Vargas Foundation (FGV) that aims to integrate the information about publications, interests and all academic activities of their researchers. Fundação Getulio Vargas (FGV), founded in 1944, is a world renowned center for quality education dedicated to promoting Brazil's economic and social development. With eight schools, two research institutes, technical assistance projects and a publishing unit, FGV is ranked one of the top 25 Think Tanks and Top 100 higher education institutions in the World with 750 researchers.

As part of the project, we also aim to provide a interface for browser and query that data making it useful for: (1) researchers and students interested to find scientific articles, research projects, academic courses and colleagues for sharing ideas; (2) FGV's president and administration board that want to evaluate the return of the investments in the research projects and the performance of their researchers.

The data was collected from the Brazilian Lattes Platform [1]. The Lattes Platform is a online system used by almost all researchers in Brazil to maintain their curriculum vitae maintained by CNPq (National Council for Scientific and Technological Development). Created by CNPq in the mid-80s, the platform is an instrument for guide investments in research in Brazil and evaluate the brazilian research community. Every researcher in Brazil that want to submit research proposals for public investments must keep his Lattes curriculum updated since this is a eligibility precondition for proposal submissions.

The Lattes data is available for the institutions from a Web Service and a standard XML format defined by a community of researchers called "Conscientias". In the first part of the project we developed an XSLT stylesheet for transform a curriculum vitae in XML to RDF using standard ontologies as FOAF, Bibo, SKOS etc. This transformation is released as an open source project called "Semantic Lattes" [2]. In the second phrase of the project we developed routines for data integration and consistence check. In the next phrase, we started to evaluate alternatives for web interfaces. We considered starting from scratch or using existing projects like VIVO. In this presentation we aim to present our project and share ideas about the benefits of integrating our results with VIVO project ideas.

---

*School of Applied Mathematics/FGV and IBM Research, Rio de Janeiro, RJ, Brazil.

†PUC-Rio, Department of Informatics, Rio de Janeiro, RJ, Brazil.

[1] http://lattes.cnpq.br

[2] github.com/arademaker/slattes/