Percy Enrique Rivera Salas

StdTrip: An a priori design process for publishing Linked Data

Dissertação de Mestrado

DEPARTAMENTO DE INFORMÁTICA
Postgraduate Program in Informatics

Rio de Janeiro
April 2011

Pontifícia Universidade Católica
do Rio de Janeiro

Percy Enrique Rivera Salas

StdTrip: An a priori design process for publishing Linked Data

Dissertação de Mestrado

Dissertation presented to the Postgraduate Program in Informatics of the Departmento de Informática , PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor: Prof. Karin Koogan Breitman

Rio de Janeiro
April 2011

Pontifícia Universidade Católica
do Rio de Janeiro

Percy Enrique Rivera Salas

StdTrip: An a priori design process for publishing Linked Data

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Informática.

Prof. Karin Koogan Breitman
Advisor
Departmento de Informática — PUC–Rio

Prof. Marco Antonio Casanova
Departamento de Informática — PUC–Rio

Prof. José Viterbo Filho
Departamento de Ciência e Tecnologia — UFF

Prof. Antonio Luz Furtado
Departamento de Informática — PUC–Rio

Prof. José Eugenio Leal
Coordinator of the Centro Técnico Científico da PUC–Rio

Rio de Janeiro — April 01, 2011

Percy Enrique Rivera Salas

Graduated in Systems Engineering from Universidad Católica de Santa Maria (UCSM), Arequipa – Peru in 2009. He conducted training courses on Publishing Open Government Data for the W3C Brazil. His current research areas are Semantic Web and Linked Data. He is also a member of the W3C RDB2RDF Working Group.

Bibliographic data

Rivera Salas, Percy Enrique

StdTrip: An a priori design process for publishing Linked Data / Percy Enrique Rivera Salas; advisor: Karin Koogan Breitman – 2011.

74 f.: il. (color.) ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2011.

Inclui bibliografia

1. Informática – Teses. 2. Linked Data. 3. Triplification. 4. Alinhamento de Ontologias. 5. Reutilização de Ontologias. 6. Interoperabilidade. I. Koogan Breitman, Karin. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Acknowledgments

I wish to thank, first and foremost, my advisor Professor Dr. Karin Breitman, whose encouragement, guidance and everyday kindness made possible the realisation of this work.

I also would like to make a special reference to Dr. José Viterbo for his friendship, valuable insights and recommendations.

I owe my deepest gratitude to my beloved family, for encouraging me to face the challenges and for giving me the strength to carry on.

I remain indebted to many colleagues and professors at PUC–Rio for providing me the means to learn and understand.

To CAPES, for the financial support, without which this work would not have been possible.

Abstract

Salas, Percy; Koogan Breitman, Karin. **StdTrip: An a priori design process for publishing Linked Data**. Rio de Janeiro, 2011. 74p. MSc Dissertation — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Open Data is a new approach to promote interoperability of data in the Web. It consists in the publication of information produced, archived and distributed by organizations in formats that allow it to be shared, discovered, accessed and easily manipulated by third party consumers. This approach requires the triplification of datasets, i.e., the conversion of database schemata and their instances to a set of RDF triples. A key issue in this process is deciding how to represent database schema concepts in terms of RDF classes and properties. This is done by mapping database concepts to an RDF vocabulary, used as the base for generating the triples. The construction of this vocabulary is extremely important, because the more standards are reused, the easier it will be to interlink the result to other existing datasets. However, tools available today do not support reuse of standard vocabularies in the triplification process, but rather create new vocabularies. In this thesis, we present the StdTrip process that guides users in the triplification process, while promoting the reuse of standard, RDF vocabularies.

Keywords

Linked Data. Triplification. Ontology Matching. Ontology Reuse. Interoperability.

Resumo

Salas, Percy; Koogan Breitman, Karin. **StdTrip: Um processo de projeto a priori para publicação de "Linked Data"**. Rio de Janeiro, 2011. 74p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A abordagem de Dados Abertos tem como objetivo promover a interoperabilidade de dados na Web. Consiste na publicação de informações em formatos que permitam seu compartilhamento, descoberta, manipulação e acesso por parte de usuários e outros aplicativos de software. Essa abordagem requer a triplificação de conjuntos de dados, ou seja, a conversão do esquema de bases de dados relacionais, bem como suas instâncias, em triplas RDF. Uma questão fundamental neste processo é decidir a forma de representar conceitos de esquema de banco de dados em termos de classes e propriedades RDF. Isto é realizado através do mapeamento das entidades e relacionamentos para um ou mais vocabulários RDF, usados como base para a geração das triplas. A construção destes vocabulários é extremamente importante, porque quanto mais padrões são utilizados, melhor o grau de interoperabilidade com outros conjuntos de dados. No entanto, as ferramentas disponíveis atualmente não oferecem suporte adequado ao reuso de vocabulários RDF padrão no processo de triplificação. Neste trabalho, apresentamos o processo StdTrip, que guia usuários no processo de triplificação, promovendo o reuso de vocabulários de forma a assegurar interoperabilidade dentro do espaço da Linked Open Data (LOD).

Palavras–chave

Linked Data. Triplification. Alinhamento de Ontologias. Reutilização de Ontologias. Interoperabilidade.

Contents

List of Figures

List of Tables

Ad Majorem Dei Gloriam

San Ignacio de Loyola, *(1491 – 1556)*.