Pontifícia Universidade Católica
do Rio de Janeiro

Lívia Couto Ruback Rodrigues

LDC Mediator: A Mediator for Linked Data Cubes

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática da PUC-Rio as partial fulfillment of the requirements for the degree of Mestre.

Advisor: Prof. Marco Antonio Casanova

Rio de Janeiro
September 2013

Pontifícia Universidade Católica
do Rio de Janeiro

Lívia Couto Ruback Rodrigues

LDC Mediator: A Mediator for Linked Data Cubes

Dissertation presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre.

Prof. Marco Antonio Casanova
Advisor
Departamento de Informática – PUC-Rio

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

Prof. Luiz André Portes Paes Leme
Departamento de Informática – UFF

Profa. Giseli Rabello Lopes
Departamento de Informática - PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de Janeiro, September 12th, 2013

Lívia Couto Ruback Rodrigues

Graduated in Computer Science from Universidade Federal de Juiz de Fora (UFJF), Minas Gerais - Brazil in 2009. She joined the Master in Informatics at Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2011. Her main research interests cover Web Semantic and Linked Data.

Bibliographic data

Couto Ruback Rodrigues, Lívia

LDC Mediator: a Mediator for Linked Data Cubes / Lívia Couto Ruback Rodrigues; advisor: Marco Antonio Casanova. – 2013.

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

Dissertação (Mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2013.

Inclui bibliografia

1. Informática – Teses. 2. Dados Estatísticos. 3. Dados ligados. 4. Arquitetura de mediação. 5. Triplificação. 6. RDF. 7. Cubo de Dados OLAP. 8. REST. I. Casanova, Marco Antonio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Acknowledgments

I would like to say a special thank you to Marco Antonio Casanova, the best advisor I could ever ask for. His wisdom, knowledge, refined sense of humor and kindness were essential to motivate me to keep going and finish this work.

To Profa. Simone for her classes and for the best brownies recipe.

To PUC-Rio for giving me the opportunity to make such important lifelong friends.

To my family and friends. In particular to my parents Rubens and Izabel for their support raising me. To my sisters Flávia and Bianca for all the wonderful memories.

To my boyfriend and partner Adriano, for still being there no matter what. Even taking up dance classes to dance with me.

To CAPES for funding my research.

Abstract

Ruback, Lívia; Casanova, Marco Antonio (Advisor). **LDC Mediator: A Mediator for Linked Data Cubes.** Rio de Janeiro, 2013. 69p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A statistical data set comprises a collection of observations made at some points across a logical space and is often organized as what is called a *data cube*. The proper definition of the data cubes, especially of their dimensions, helps to process the observations and, more importantly, helps to combine observations from different data cubes. In this context, the Linked Data Principles can be profitably applied to the definition of data cubes, in the sense that the principles offer a strategy to provide the missing semantics of the dimensions, including their values. This work introduces a mediation architecture to help consume linked data cubes, exposed as RDF triples, but stored in relational databases. The data cubes are described in a catalogue using standardized vocabularies and are accessed by HTTP methods using REST principles. Therefore, this work aims at taking advantage of both Linked Data and REST principles in order to describe and consume linked data cubes in a simple but efficient way.

Keywords

Statistical Data, Linked Data, Mediation Architecture, Triplification, RDF, OLAP Data Cube, REST.

Resumo

Ruback, Lívia; Casanova, Marco Antonio. **Mediador LDC: Um mediador de Cubos de Dados Interligados.** Rio de Janeiro, 2013. 69p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Um banco de dados estatístico consiste de um conjunto de observações feitas em pontos de um espaço lógico, e, muitas vezes, são organizados como cubos de dados. A definição adequada de cubos de dados, em especial de suas dimensões, ajuda a processar as suas observações e, mais importante, ajuda a combinar observações de cubos de dados diferentes. Neste contexto, os princípios de dados interligados podem ser proveitosamente aplicados à definição de cubos de dados, oferecendo uma estratégia para fornecer a semântica das dimensões, incluindo seus valores. Este trabalho introduz uma arquitetura de mediação para auxiliar no consumo de cubos de dados, expostos como triplas RDF e armazenados em bancos de dados relacionais. Os cubos de dados são descritos em um catálogo usando vocabulários padronizados e são acessados por métodos HTTP usando os princípios de REST. Portanto, este trabalho busca tirar proveito tanto dos princípios de dados interligados quanto dos princípios de REST para descrever e consumir os cubos de dados interligados de forma simples e eficiente.

Palavras-chave

Dados estatísticos, dados ligados, arquitetura de mediação, triplificação, RDF, Cubo de Dados OLAP, REST.

Table of Contents

List of Figures

List of Tables