



**Carla Gonçalves Ourofino**

**Materialização e Manutenção de Ligações owl:sameAs**

**DISSERTAÇÃO DE MESTRADO**

Dissertação apresentada ao Programa de Pós-Graduação em Informática da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Informática.

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro  
Agosto de 2013



**Carla Gonçalves Ourofino**

## **Materialização e Manutenção de Ligações owl:sameAs**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Marco Antonio Casanova**

Orientador

Departamento de Informática - PUC-Rio

**Prof. Vania Maria Ponte Vidal**

UFC

**Prof. Antonio Luz Furtado**

Departamento de Informática – PUC-Rio

**Prof. Giseli Rabello Lopes**

Departamento de Informática - PUC-Rio

**Prof. José Eugênio Leal**

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 23 de agosto de 2013

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Carla Gonçalves Ourofino**

Graduou-se em Sistemas de Informação pela PUC-Rio em 2010. Trabalha como Analista de Sistemas na Coordenação Central de Planejamento e Avaliação (CCPA) da PUC-Rio. Ingressou no programa de mestrado do Departamento de Informática em 2011. Atualmente suas áreas de pesquisa são Bancos de Dados, Web Semântica e Linked Data.

#### Ficha Catalográfica

Ourofino, Carla Gonçalves

Materialização e manutenção de ligações owl:sameAs / Carla Gonçalves Ourofino ; orientador: Marco Antonio Casanova. – 2013.

58 f. : il. (color.) ; 29,7 cm

Dissertação (mestrado)—Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2013.

1. Informática – Teses. 2. Web semântica. 3. Linked data. 4. Banco de dados. 5. Materialização. 6. Manutenção. 7. Ligações. 8. owl:sameAs. I. Casanova, Marco Antonio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

## Agradecimentos

Ao professor Marco Antonio Casanova, que me orientou e motivou para a realização desta pesquisa.

Ao meu noivo (e quase esposo) Raphael, que me apoiou emocionalmente, colaborou com o meu planejamento, e compreendeu as dificuldades que precisamos superar durante este tempo.

Aos meus pais, que sempre me incentivaram a vencer através da dedicação aos estudos.

Aos meus amigos da CCPA, que colaboraram com suas opiniões e sugestões sobre esta pesquisa.

## Resumo

Ourofino, Carla Gonçalves; Casanova, Marco Antonio. **Materialização e Manutenção de Ligações owl:sameAs**. Rio de Janeiro. 2013. 58p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A Web de Dados cresceu significativamente nos últimos anos, tanto em quantidade de dados, quanto em fontes responsáveis por esses. A partir desse aumento no número de fontes de dados, ligações owl:sameAs têm sido cada vez mais utilizadas para conectar dados equivalentes e publicados por fontes distintas. Com isso, torna-se necessário haver uma rotina de identificação e manutenção dessas conexões. Com o objetivo de automatizar essa tarefa, desenvolvemos o "Framework MsA – Materialização de sameAs" para materializar e recomputar ligações do tipo owl:sameAs entre bancos de dados locais e dados publicados na Web. Essas ligações, uma vez identificadas, são materializadas juntamente aos dados locais e recomputadas apenas quando necessário. Para isso, a ferramenta monitora as operações (cadastramento, remoção e atualização) realizadas nos dados locais e remotos e, para cada tipo, implementa uma estratégia de manutenção das ligações envolvidas.

## Palavras-chave

Web semântica; Linked Data; Banco de Dados; Materialização; Manutenção; Ligações; owl:sameAs.

## Abstract

Ourofino, Carla Gonçalves; Casanova, Marco Antonio. **Materialization and Maintenance of owl:sameAs Links**. Rio de Janeiro. 2013. 58p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The Web of Data has grown significantly in recent years, not only in the amount of data but also in the number of data sources. In parallel with this tendency, owl:sameAs links have been increasingly used to connect equivalent data published by different sources. As a consequence, it becomes necessary to have a routine for the identification and maintenance of these connections. In order to automate this task, we have developed the "MsA Framework - sameAs Materialization" to materialize and recompute owl:sameAs links between local databases and data published on the Web. These connections, once identified, are materialized along with the local data and recomputed only when necessary. To achieve this goal, the tool monitors the operations (insertion, update and deletion) performed on local and remote records, and for each type of operation it implements a maintenance strategy on the links involved.

## Keywords

Web of Data; Semantic Web; Linked Data; Database; Materialization; Maintenance; Links; owl:sameAs.

# Sumário

1	Introdução	11
1.1	Motivação	11
1.2	Estrutura do Documento	12
2	Fundamentos	13
2.1	Linked Data	13
2.2	Web de Dados	14
2.3	OWL e ligações owl:sameAs	16
2.4	Uma crítica ao uso da propriedade owl:sameAs	17
3	Trabalhos Relacionados	19
3.1	Silk Framework	19
3.2	Limes	20
3.3	Sindice.com	20
3.4	Reconciliação de entidades entre fontes da Web de Dados	21
4	O Framework MsA – Materialização de sameAs	22
4.1	Motivação do Framework MsA	22
4.1.1.	O problema da manutenção de ligações owl:sameAs	22
4.1.2.	Identificação	23
4.1.3.	Materialização	23
4.1.4.	Manutenção	25
4.2	Requisitos e Casos de Uso	25
4.3	Arquitetura do Framework MsA	27
4.3.1.	Visão Geral da Arquitetura	27
4.3.2.	Resolução de Entidades	28
4.3.3.	Tabelas resolver, sameas e sameas_uri	28
4.3.4.	Triggers	30
4.3.5.	Monitor	31
4.4	Monitoramento das Bases de Dados	34
4.4.1.	Alterações na Base Local (SGBD)	34

4.4.2. Alterações na Base Externa (Web de Dados)	37
4.5 Instalação e Inicialização	39
4.5.1. Infraestrutura	39
4.5.2. Instalação do Framework	40
4.5.3. Inicialização do Monitor	41
4.6 Comentários sobre as Etapas de Desenvolvimento do Framework	42
4.6.1. Tecnologias	42
4.6.2. Especificação de Requisitos	42
4.6.3. Planejamento da Estratégia	42
4.6.4. Modelagem de Dados	43
4.6.5. Desenvolvimento da Arquitetura	43
4.6.6. Implementação do Monitor	43
4.6.7. Testes	44
4.6.8. Documentação	44
5 Estudo de Caso	45
5.1 Bases de Dados utilizadas	45
5.1.1. RPA@PUC	45
5.1.2. DBLP	45
5.2 Montagem	46
5.3 Inicialização	49
5.4 Execução	50
5.4.1. Alterações na Base Local	51
5.4.2. Alterações na Base Externa	53
5.5 Publicação do Resultado	53
6 Conclusão	55
7 Referências Bibliográficas	56



## Lista de Figuras

Figura 1: Exemplo de tripla RDF contendo a propriedade owl:sameAs	17
Figura 2: Exemplo de uso alternativo da propriedade owl:sameAs	17
Figura 3: Diagrama de Casos de Uso do Framework MsA	26
Figura 4: Visão geral da arquitetura do Framework MsA	27
Figura 5: Modelagem de dados das tabelas de controle do Framework MsA	28
Figura 6: Diagrama de Classes do Monitor	33
Figura 7: Tabelas do RPA@PUC referentes a dados de professores	47
Figura 8: Execução do script de instalação do Framework MsA	47
Figura 9: Criação do mapeamento entre a tabela e o módulo de resolução de entidades	48
Figura 10: Modelo do banco de dados após a instalação e o mapeamento do Framework MsA	49
Figura 11: Inicialização do Monitor	50
Figura 12: Mapeamento do D2R Server para publicação de ligações owl:sameAs	54
Figura 13: Uso conjunto do D2R Server com o Framework MsA	54

## Lista de Tabelas

Tabela 1: Especificações técnicas das tabelas de controle	29
Tabela 2: Descrição do uso de cada coluna das tabelas de controle	29