



Edgard Luiz Marx

**Babel: Um Framework Extensível para a
publicação de RDF de Várias Fontes de
Dados Utilizando Templates**

Dissertação de Mestrado

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

Orientador: Prof.^a Karin Koogan Breitman



Edgard Luiz Marx

Babel: Um Framework Extensível para a publicação de RDF de Várias Fontes de Dados Utilizando Templates

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.^a Karin Koogan Breitman

Orientador
Departamento de Informática - PUC-Rio

Prof. José Viterbo Filho

Departamento de Informática - UFRJ

Prof.^a Vânia Maria Ponte Vidal

Departamento de Informática - UFC

Prof. José Eugênio Leal

Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 17 de Fevereiro de 2012

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

Edgard Luiz Marx

Edgard Luiz Marx é formado em administração pela escola de negócios do SEBRAE. Graduou-se em Informática pela PUC-RIO em 2006 despertando interesse pelo tema Governo Eletrônico. Desde 2005 trabalha para o laboratório de pesquisa em Computação Gráfica (Tecgraf) no desenvolvimento de Sistemas de Informações Geográficas destacando a participação no desenvolvimento cooperativo de bibliotecas de código aberto como Geotools. Em 2010 juntou-se ao grupo de pesquisas em Publicação de Dados Abertos ministrando cursos e tutorias junto ao W3C Brasil.

Ficha Catalográfica

Marx, Edgard

Babel: Um Framework Extensível para a publicação de RDF de Várias Fontes de Dados Utilizando Templates / Edgard Luiz Marx; orientador: Karin Breitman – 2012.

107 f : il. ; 30cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2012.

Inclui bibliografia

1. Informática – Teses. 2. RDF. 3. Banco de Dados Relacionais. 4. Plugin. 5. Arquitetura Extensível. 6. Template. 6. Conversor. 7. Linked Data. 8. XML.

CDD:004

Agradecimentos

Agradeço e dedico este trabalho a todas as pessoas que contribuíram de alguma forma para a minha formação, a minha família, aos meus amigos e mestres pela sorte de ter tido a oportunidade de aprender com vocês.

Resumo

Edgard Luiz Marx; Karin Breitman. **Babel: Um Framework Extensível para a Publicação de RDF de Várias Fontes de Dados Utilizando Templates.** Rio de Janeiro, 2012. 107p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A grande maioria dos dados que se encontram hoje na Web não estão preparados para a Web Semântica. Para facilitar e promover a conversão de dados, armazenados em bancos de dados relacionais e planilhas em particular, nós introduzimos a abordagem do Babel. Diferentemente das abordagens existentes, nomeadamente RDB2RDF, Babel e promove a conversão de dados em uma ampla variedade de formatos, que incluem OWL, RDFa, RSS e (X)HTML, além de RDF. A principal contribuição de Babel, no entanto, é sua facilidade de uso. Babel suaviza a curva de aprendizado, eliminando a necessidade de se familiarizar com técnicas de mapeamento complexas, que são substituídas pelo uso de templates.

Palavras-chave

RDF, Banco de dados relacionais, Plugin, Arquitetura Extensível, Template, Conversor, Linked Data, XML.

Abstract

Edgard Luiz Marx; Karin Breitman (Advisor). **Babel: An Extensible Framework for Easy RDF Publication from Multiple Data Sources Using Templates.** Rio de Janeiro, 2012. 107p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The vast majority of data on the Web today is not Semantic Web ready. To facilitate and promote the conversion of data, stored in relational databases and spreadsheets in particular, we introduce the Babel approach. Differently from existing approaches, notably RDBtoRDF, Babel outputs data in a wider range of formats, that include OWL, RDFa, RSS and (X)HTML, in addition to RDF. The main contribution of Babel, however, is its ease of use. Babel smoothes the learning curve by altogether eliminating the need of getting acquainted with complex mapping techniques, that are substituted by the use of templates.

Keywords

RDF, Relational Databases, Plugin, Extensible Architecture, Template, Convertor, Linked Data, XML.

Sumário

1	INTRODUÇÃO	14
1.1	MOTIVAÇÃO	15
1.2	OBJETIVO	26
1.3	CONTRIBUIÇÕES	26
1.4	ORGANIZAÇÃO DO TRABALHO.....	26
2	CONCEITOS BÁSICOS.....	27
2.1	A WEB SEMÂNTICA.....	28
2.2	RDF.....	30
2.3	R2RML.....	33
2.4	TEMPLATES.....	34
2.5	XML	35
3	TRABALHOS RELACIONADOS	38
3.1	FERRAMENTAS DE CONVERSÃO RDF	38
3.1.1	<i>Triplify</i>	38
3.1.2	<i>Wrapper Babel</i>	39
3.1.3	<i>RDF123</i>	39
3.1.4	<i>SWAML</i>	40
3.1.5	<i>STDTrip</i>	41
3.2	FRAMEWORKS DE ARMAZENAMENTO, PUBLICAÇÃO E MANIPULAÇÃO DE RDF	42
3.2.1	<i>Jena</i>	42
3.2.2	<i>Plataforma D2RQ</i>	43
3.2.3	<i>Virtuoso</i>	43
3.2.4	<i>Sesame</i>	44
3.2.5	<i>StrixDB</i>	45
3.3	FERRAMENTAS DE INTEROPERABILIDADE	45
3.4	FRAMEWORKS DE DESENVOLVIMENTO WEB BASEADOS EM TEMPLATES	46
3.4.1	<i>JavaServer Pages</i>	47
3.4.2	<i>JavaServer Faces</i>	47
3.4.3	<i>Tapestry</i>	48
3.4.4	<i>Less</i>	49
3.5	OUTRAS TECNOLOGIAS	50
3.5.1	<i>Suporte a estruturas XML do SQL Server</i>	50

3.5.2	<i>XML- Especificações Relacionadas (SQL/XML)</i>	51
3.5.3	<i>Tecnologias de Web Semântica da Oracle</i>	51
3.5.4	<i>Clio</i>	52
4	TML (TEMPLATE MODEL LANGUAGE)	54
4.1	VISÃO GERAL	54
4.2	URI	58
4.3	LIMITAÇÕES	59
4.4	UTILIZAÇÃO DE TEMPLATES EM ESTRUTURAS SENSÍVEIS AO CONTEXTO	61
4.5	ESCREVENDO UM COMPACTADOR DE TEXTO SIMPLES UTILIZANDO TML	62
4.6	MAPEAMENTO DIRETO UTILIZANDO TML	63
5	BABEL	66
5.1	CONTRIBUIÇÕES	66
5.2	CONCEITOS	68
5.3	ARQUITETURA	69
5.3.1	<i>O template</i>	70
5.3.2	<i>O Arquivo de Configuração</i>	71
5.3.3	<i>Máquina de processamento de Templates (Template Engine)</i>	74
5.3.4	<i>Interface por Comando Texto (Textual Command User Interface)</i>	74
5.3.5	<i>Interface Web</i>	76
5.3.6	<i>Inteface gráfica</i>	77
5.4	IMPLEMENTAÇÃO	78
5.4.1	<i>Máquina de Processamento de Templates (Template Engine)</i>	82
5.4.2	<i>DataStore</i>	87
6	CONCLUSÃO	91
6.1	CONTRIBUIÇÕES	91
6.2	CONSULTA DE BASES ATRAVÉS DE MAPEAMENTOS	92
6.3	MASHUPS	93
6.4	TRABALHOS FUTUROS	94
7	REFERÊNCIAS BIBLIOGRÁFICAS	97
	ANEXOS	103

Lista de Figuras

FIGURA 1. SISTEMA ELETRÔNICO DE TABULAÇÃO DESENVOLVIDO POR HOLLERITH EM 1989, A FIM DE FACILITAR O CENSO NORTE AMERICANO.....	14
FIGURA 2. O “BOLO DE ANIVERSÁRIO” DA WEB SEMÂNTICA PROPOSTA POR BERNERS-LEE	29
FIGURA 3. REPRESENTAÇÃO SIMPLES DE UMA DECLARAÇÃO RDF NA FORMA DE GRAFO.....	31
FIGURA 4. EXEMPLO DE TEMPLATE UTILIZANDO LETL PARA GERAÇÃO DE CONTEÚDO.	49
FIGURA 5. INTERFACE DE DEFINIÇÃO DE MAPEAMENTOS DO CLIO.	53
FIGURA 6. REPRESENTAÇÃO DE UM VETOR EM UMA ESTRUTURA DE DADOS UTILIZANDO ELEMENTOS SIMPLES E COMPLEXOS.....	55
FIGURA 7. REPRESENTAÇÃO DE UMA ESTRUTURA FORMADA POR DOIS ELEMENTOS COMPOSTOS, ONDE A TABELA CIDADE (<i>CITY</i>) REPRESENTA UMA PROPRIEDADE COMPOSTA DA ESTRUTURA ESTADO (<i>STATE</i>).	55
FIGURA 8. REPRESENTAÇÃO DOS DADOS A PARTIR DE UM TEMPLATE GENÉRICO UTILIZANDO TML.	57
FIGURA 9. RESULTADO DA CONVERSÃO DOS DADOS CONTIDOS EM UMA COLEÇÃO <i>STATE</i> PARA DIFERENTES FORMATOS DE SERIALIZAÇÃO (JSON, XML E N3) UTILIZANDO TML.	58
FIGURA 10. EXEMPLO DE CRIAÇÃO DE URI DINÂMICA UTILIZANDO TML.	59
FIGURA 11. POSSÍVEIS RESULTADOS DE UM MAPEAMENTO UTILIZANDO TML EM UMA BASE	60
FIGURA 12. UTILIZANDO TEMPLATES EM ESTRUTURAS SENSÍVEIS AO CONTEXTO.	62
FIGURA 13. EXEMPLO DE COMPACTAÇÃO DE TEXTO UTILIZANDO TML.	63
FIGURA 14. DEMONSTRAÇÃO DO MAPEAMENTO DA COLEÇÃO <i>STATE</i> PARA TRIPLAS UTILIZANDO A ESTRATÉGIA DE MAPEAMENTO DIRETO.....	64
FIGURA 15. TEMPLATE DE TRIPLAS GERADO A PARTIR DA COLEÇÃO <i>STATE</i> , UTILIZANDO A TÉCNICA DE MAPEAMENTO DIRETO.....	65
FIGURA 16. ILUSTRAÇÃO DE UM <i>DATASTORE</i> CONTENDO DUAS COLEÇÕES DE DADOS.	68
FIGURA 17. DIAGRAMA QUE ILUSTRA OS COMPONENTES PRINCIPAIS DA ARQUITETURA DO <i>BABEL</i> E SUAS RELAÇÕES.	70
FIGURA 18. DIAGRAMA DE MODELO DO ESQUEMA DO ARQUIVO DE CONFIGURAÇÃO.....	72
FIGURA 19. ILUSTRAÇÃO DO MODELO REQUISIÇÃO E RESPOSTA HTTP UTILIZADO PELO <i>ENDPOINT</i> DO <i>FRAMEWORK BABEL</i>	77
FIGURA 20. INTERFACE GRÁFICA DO <i>FRAMEWORK</i> PROPOSTO	78
FIGURA 21. PUBLICANDO DADOS ATRAVÉS DA INTERFACE GRÁFICA.....	78
FIGURA 22. MÓDULOS PRINCIPAIS DO <i>FRAMEWORK</i>	79
FIGURA 23. PARÂMETROS DA PLANILHA.....	90

FIGURA 24. ILUSTRAÇÃO DA CONVERSÃO DE UMA BASE PARA RDF E DE UMA CONSULTA SPARQL PARA UMA CONSULTA SQL.	93
FIGURA 25. FIGURA ILUSTRATIVA DE UM MASHUP REALIZADO A PARTIR DE DUAS BASES DE DADOS UTILIZANDO BABEL.	94

Lista de Gráficos

GRÁFICO 1. CRESCIMENTO ESTIMADO WEB E DA WEB SEMÂNTICA NOS ÚLTIMOS DOIS ANOS, UTILIZANDO O ÍNDICE GYBA DO WORLD WIDE WEB SIZE E O ÍNDICE DE DOCUMENTOS INDEXADOS PELO SWOOGLE.	20
GRÁFICO 2. COMPARAÇÃO ENTRE O TAMANHO DA WORLD WIDE WEB E A WEB SEMÂNTICA EM 2011 UTILIZANDO OS ÍNDICES APROXIMADOS DO WORLD WIDE WEB SIZE E SWOOGLE ENCONTRADOS EM JANEIRO DE 2011, COM O TAMANHO ESTIMADO DE DOCUMENTOS NA LOD (UMA ESTIMATIVA UTILIZANDO A RELAÇÃO DO NÚMERO DE TRIPLAS POR DOCUMENTO EXTRAÍDO DO SWOOGLE E APLICADO À LOD).	22
GRÁFICO 3. COMPARAÇÃO DE DESEMPENHO ENTRE TRIPLIFY 1.0B E D2RQ 0.7 NA GERAÇÃO DE INDIVÍDUOS DO TIPO FOAF CONTENDO APENAS UM ATRIBUTO NOME .	23
GRÁFICO 4. COMPARAÇÃO DE DESEMPENHO ENTRE BABEL, TRIPLIFY E D2R.....	88

Lista de Diagramas

DIAGRAMA 1. DIAGRAMA DE PACOTES.	80
DIAGRAMA 2. DIAGRAMA DE CLASSES DO PACOTE <i>DATSTORE</i> , DO DIAGRAMA 1.....	81
DIAGRAMA 3. DIAGRAMA DE CLASSES DO PACOTE <i>PROCESS</i> , DO DIAGRAMA 1.....	81
DIAGRAMA 4. DIAGRAMA DE CLASSES DO PACOTE <i>PRINTER</i> , DO DIAGRAMA 1.	81
DIAGRAMA 5. DIAGRAMA DE CLASSES DO PACOTE <i>TEMPALTE</i> , DO DIAGRAMA 1.	82
DIAGRAMA 6. EXEMPLO DE UM DIAGRAMA DE EXECUÇÃO HIPOTÉTICO DE UM TEMPLATE PELA MÁQUINA DE PROCESSAMENTO DE TEMPALTES	84

“Quem deseja aprender a voar, deve primeiro aprender a caminhar, a correr, a escalar e a dançar. Não se aprende a voar voando.”

Friedrich Nietzsche