

Samur Felipe Cardoso de Araújo

**Explorator: uma ferramenta para
exploração de dados RDF baseado
em uma interface de manipulação
direta**

DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE INFORMÁTICA
Programa de Pós-Graduação em Informática

Rio de Janeiro
Fevereiro de 2009



Samur Felipe Cardoso de Araújo

**Explorator: uma ferramenta para
exploração de dados RDF baseado em
uma interface de manipulação direta**

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: Daniel Schwabe

Rio de Janeiro, 02 de fevereiro de 2009



Samur Felipe Cardoso de Araújo

**Explorator: uma ferramenta para
exploração de dados RDF baseado em
uma interface de manipulação direta**

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. Daniel Schwabe
Orientador
PUC-Rio

Prof. Marco Antonio Casanova
PUC-Rio

Profa. Simone Diniz Junqueira Barbosa
PUC-Rio

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

Rio de Janeiro, 02 de fevereiro de 2009

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

Samur Felipe Cardoso de Araújo

Graduou-se em Ciências da Computação (2002) pela Universidade Federal de Viçosa - MG. Atuou como analista de sistema em companhias de grande porte que atuam no setor de Petróleo e Telecomunicações. Atuou como empreendedor, construindo uma empresa que hoje se encontra incubada no Instituto Genesis na PUC-Rio. Possui interesse acadêmico e profissional nas áreas de Hipertexto e Multimídia, Interação Humano-Computador e Web Semântica.

Ficha Catalográfica

Araújo, Samur Felipe Cardoso de

Explorator: uma ferramenta para exploração de dados RDF baseado em uma interface de manipulação direta / Samur Felipe Cardoso de Araújo ; orientador: Daniel Schwabe. – 2009.

127 f. : il. (col.) ; 30 cm

Dissertação (Mestrado em Informática)– Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Informática – Teses. 2. Web semântica. 3. Navegação. 4. Exploração. 5. Navegação por conjuntos. 6. Modelo de informação. 7. RDF. 8. WWW. I. Schwabe, Daniel. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Este trabalho é dedicado à minha família
e ao Prof. Daniel, por tudo.

Agradecimentos

Ao meu orientador, Prof. Daniel Schwabe, pela sua atenção e por me proporcionar à experiência de conviver com um ser brilhante.

A todos os professores, funcionários e alunos do Departamento de Informática pelas boas experiências trocadas e pela excelente convivência .

Ao meus amigos e colegas de curso pelo companheirismo e por enriquecerem meus conhecimentos: Carlos Laufer, Francisco Santana, Guilherme Szundy, Luiz André Leme, Marcelo Cohen e Maurício Bomfim.

A todos os colaboradores que tornaram este projeto viável.

Ao Eduardo Velloso pelas contribuições.

Ao meu amigo Leo Lodi por acompanhar minhas conquistas, tropeços e vitórias.

A todos os meus amigos que torcem por mim.

Ao meu pai José Clovis e minha mãe Haydée; aos meus irmãos Clovis, André, Nicole, Maira e à minha avó Helena Guadagnin, pelo amor incondicional.

Resumo

Araújo, Samur Felipe Cardoso; Schwabe, Daniel. **Explorator: uma ferramenta para exploração de dados RDF baseado em uma interface de manipulação direta.** Rio de Janeiro, 2009. 127p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Nessa dissertação propomos uma ferramenta destinada à exploração dos dados na Web Semântica. Nosso objetivo foi desenvolver um modelo de exploração que permitisse o usuário explorar uma base de dados RDF sem deter qualquer conhecimento prévio de seu domínio. Para tanto, apresentamos um modelo de operações que suportadas por uma interface baseada no paradigma de manipulação direta e query-by-example, nos permite explorar uma base de dados RDF semi-estruturada para ganhar conhecimento e responder questões específicas sobre o domínio, através de navegação, busca e outros mecanismos de exploração. Também desenvolvemos um modelo de especificação e geração automática de facetas que pode ser utilizado na construção de mecanismos de navegação facetada sobre dados RDF. O produto final desse trabalho é a ferramenta Explorator, que propomos como um ambiente para Exploração dos dados na Web Semântica.

Palavras-chave

Web semânticas, navegação, exploração, navegação por conjuntos, modelo de informação, RDF

Abstract

Araújo, Samur Felipe Cardoso; Schwabe, Daniel. **Explorator: a tool for exploring RDF data through direct manipulation.** Rio de Janeiro, 2009. 127p. Msc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

In this dissertation we propose a tool for Semantic Data exploration. We developed an exploration model that allows users without any a prior knowledge about the data domain to explore an RDF database. So that, we presented an operation model, that supported by an interface based on the direct manipulation and query-by-example paradigm, allows users to explore an RDF base to both gain knowledge and answer questions about a domain, through navigation, search and others exploration mechanisms. Also, we developed a facet specification model and a mechanism for automatic facet extraction that can be used in the development of facet navigation systems over RDF. The final product of this work is a tool called Explorator that we are proposing as an environment for Semantic Web data exploration.

Keywords

Semantic Web, navigation, exploration, set-based navigation, information model , RDF

Sumário

1	Introdução	18
1.1.	Ontologia na Web Semântica	19
1.2.	RDF (Resource Description Framework)	20
1.3.	Explorando a Web Semântica	23
1.4.	Organização da dissertação	25
2	Explorando dados na Web Semântica	26
2.1.	O que é exploração de dados?	26
2.2.	Trabalhos Relacionados	27
2.2.1.	Navegadores ou Browsers	27
2.2.2.	Navegação Facetada	30
2.2.3.	Linguagem de Consulta	33
2.2.4.	Modelos de navegação RDF	35
2.3.	Modelo de Exploração para Dados RDF	36
2.3.1.	Conjuntos	37
2.3.2.	Operações de Consulta	39
2.3.3.	Operações sobre conjuntos	41
2.3.4.	Exemplo de uso	42
3	Navegação Facetada	44
3.1.	O que é uma faceta?	44
3.2.	Sistemas existentes	45
3.2.1.	Flamenco	46
3.2.2.	BrowseRDF	47
3.2.3.	FacetMap	47
3.2.4.	Longwell	48
3.2.5.	Avaliação dos Modelos Existentes	49
3.3.	Casos de Uso	58
3.3.1.	Cenário 1	59

3.3.2. Cenário 2	59
3.3.3. Cenário 3	60
3.3.4. Cenário 4	61
3.4. Modelo de Especificação das Facetas	61
3.4.1. Visão Geral do Modelo	62
3.4.2. Características do Modelo	63
3.4.3. Camada lógica	63
3.4.4. Camada do Modelo	64
3.5. Vocabulário Faceto	64
3.5.1. FacetGroup	65
3.5.2. Facet	66
3.5.3. Esquema Faceto	71
3.6. Geração Automática de Facetas	73
3.6.1. Definições Gerais	73
3.6.2. Algoritmo	76
3.7. Conclusão	77
4 A Ferramenta Explorator	78
4.1. Manipulação direta	78
4.2. Query-by-example	79
4.3. Recursos, triplas e conjuntos	80
4.4. Classes, propriedades, literais e recursos	82
4.5. Modelo de seleção	82
4.6. Operações do modelo na interface	84
4.6.1. Operações sobre conjunto	85
4.6.2. Operações de consulta SPO	85
4.7. Repositórios	89
4.8. Navegação de-referenciada	90
4.9. Navegação Facetada	91
4.10. Outras operações na interface.	93
4.11. Resumo das operações do Explorator	94
4.12. Cenário de Exemplo	95
5 Detalhes da Implementação	102

5.1. Arquitetura Geral	102
5.1.1. Dados RDF	103
5.1.2. ActiveRDF	103
5.1.3. Ruby-on-Rails	103
5.2. Explorator	104
5.3. Diagramas de Pacotes de Classes	105
5.3.1. Diagrama de Classes	106
5.4. Interface	111
5.4.1. Anotações Semânticas	111
5.4.2. CSS	112
5.4.3. Comportamento	113
5.5. Desempenho	115
6 Avaliação do Explorator	117
6.1. Primeiro Estudo	117
6.2. Segundo Estudo	119
6.3. Resultados	120
7 Conclusão e Trabalhos Futuros	123
7.1. Contribuições	123
7.2. Trabalhos Futuros	124
8 Referências bibliográficas	126

Lista de figuras

Figura 1 - Representação em grafo do RDF.	23
Figura 2 – Nuvem exemplificando a interconexão das bases RDF do projeto LOD. (nuvem de setembro de 2008).	25
Figura 3 – Exemplo da visualização das triplas do arquivo RDF http://www.ivan-herman.net/foaf.rdf na interface do Tabulator.	28
Figura 4 – Exemplo da de-referenciação da URI http://www.w3.org/2000/01/combined-ns-translation.rdf.fr na interface do Disco.	29
Figura 5 - Interface do FacetMap exibindo uma lista de elemento e suas facetas.	31
Figura 6 – Esquema representando os conjuntos de triplas, sujeitos, predicados e objetos.	39
Figura 7 – Exemplo da interface facetada do Flamenco	45
Figura 8 – Interface do Flamenco	46
Figura 9 – Interface do BrowseRDF	47
Figura 10 – Interface do FacetMap	48
Figura 11 – Interface do Longwell	49
Figura 12 – Figura 6 referida na citação de (Oren et al., 2006, p.8)	51
Figura 13 – Estruturas de uma interface de navegação facetada	51
Figura 14 - Tela do FacetMap demonstrando as facetas e a cardinalidade.	52
Figura 15 – Tela do Flamenco evidenciando os 3 grupos de informação de uma interface de navegação facetada.	53
Figura 16 – Resultado da aplicação das formatações Fresnel sobre os dados RDF.	58
Figura 17 – Visão da arquitetura de um modelo de navegação facetada.	62
Figura 18 – Diagrama representando o vocabulário faceto.	71
Figura 19 - Representa o recurso Product na interface.	80
Figura 20 – Representação das duas triplas (Product, comment, ‘A product supported by Forum Nokia’) e (Product, label, ‘Product’)	

, na interface.	81
Figura 21 - Representa o conjunto dos recursos Wireless Terminal Model, Product e Operating Band na interface.	81
Figura 22 - tipos de recursos na interface. A cor azul indica recursos do tipo Class, verde do tipo Property e marrom do tipo literal.	82
Figura 23 - Seleção de um conjunto que possui três recursos (Caspian Sea, Ozero Chanka, Ozero Pskovskoje).	83
Figura 24 - Seleção de um único recurso (Caspian Sea).	83
Figura 25 – Modelo de seleção de predicado e objetos. A fig. A representa a seleção de propriedades das triplas, e a fig. B a seleção de objetos das triplas.	84
Figura 26 - Operadores de união, interseção e diferença.	85
Figura 27 - Operação SPO na interface do Explorator.	86
Figura 28 - Barra de ferramenta do Explorator.	86
Figura 29 – Recursos locatedin e kazakstan.	86
Figura 30 – Seleção do recurso locatedIn.	87
Figura 31 – Atribuição do recurso locatedin ao parâmetro P da operação SPO.	87
Figura 32 - Atribuição do recurso kazakstan ao parâmetro O da operação SPO.	88
Figura 33 – Resultado da operação SPO	88
Figura 34 – Menu de acesso a interface de configuração de repositórios.	89
Figura 35 – Interface de configuração de repositórios do Explorator.	90
Figura 36 – Exemplo da adição do Sparql Endpoint do DBpedia.	90
Figura 37 – Barra do Explorator evidenciando a barra de endereço.	91
Figura 38 – Destaque das operações F e F* para facetar um conjunto de recursos.	91
Figura 39 – Interface de navegação facetada do Explorator. O conjunto da esquerda representa as facetas e o conjunto da direita os elementos sendo facetados.	92
Figura 40 - Cabeçalho da caixa de conjunto	93
Figura 41 - Exemplos das diferentes visualizações de uma tripla.	93
Figura 42 – Menu de atalhos de um recurso.	94

Figura 43 – Pesquisa pelos lagos.	96
Figura 44 – Instancias de lagos	96
Figura 45 – Pesquisa por Russia	96
Figura 46 – Detalhes do recurso Rússia	97
Figura 47 – Exemplo de uso da operação SPO	97
Figura 48 – Exemplo de uso da operação SPO	98
Figura 49 – Lista de lagos da Russia.	98
Figura 50 – Lista de vizinhos da Rússia.	98
Figura 51 – Obtendo os lagos dos vizinhos.	99
Figura 52 – Obtendo o lagos dos vizinhos da Rússia com a operação SPO.	99
Figura 53 – Lista de lagos contidos nos países vizinhos da Rússia e na Rússia.	100
Figura 54 – Obtendo a interseção entre os lagos da Rússia e dos vizinhos.	100
Figura 55 – Lagos contidos exclusivamente no território Russo.	101
Figura 56 – Arquitetura do Explorator	102
Figura 57 – Visão da chamada AJAX feita pela interface do Explorator.	104
Figura 58 – Diagrama de Pacotes	105
Figura 59 – Diagrama de Classes do pacote <i>Controller e Model</i>	106
Figura 60 – Diagrama de classes do pacote <i>Helper</i> .	107
Figura 61 – Classes do pacote Lib	109
Figura 62 – Classes do pacotes ExploratorUI	110
Figura 63 – Visão geral da arquitetura da interface do Explorator	111
Figura 64 – Esquema exemplificando uso de anotações semânticas na interface do Explorator.	114

Lista de quadros

Quadro 1 – Exemplo da notação N3.	21
Quadro 2 – Exemplo de um recurso com duas propriedades, em N3.	21
Quadro 3 – Exemplo do uso de , e ; da notação N3.	22
Quadro 4 - Exemplo de RDF na sintaxe N3	22
Quadro 5 - conjunto de dados RDF representados na sintaxe N3.	33
Quadro 6 - Exemplo de consulta na linguagem SPARQL.	33
Quadro 7 - Resultado da consulta SPARQL.	34
Quadro 8 – Definição de conjunto de recursos.	38
Quadro 9 – Definição de conjunto de sujeito, predicado e objeto.	38
Quadro 10 - Hierarquia utilizada para exemplificar a linguagem XFML	55
Quadro 11 – Exemplo do uso de XFML	55
Quadro 12 – Exemplo de uso de XFML, associando conteúdo com termos.	55
Quadro 13 – Dados RDF utilizados no exemplo da linguagem Fresnel.	56
Quadro 14 – Exemplo de formatações definidas na linguagem Fresnel.	57
Quadro 15 - Exemplo de um grupo de facetas	65
Quadro 16 – Exemplo da definição de uma faceta com termo literal.	66
Quadro 17 – Exemplo da definição de um termo derivado de propriedade.	66
Quadro 18 – Exemplo de uma faceta derivada de um predicado RDF.	67
Quadro 19 – Exemplo de uma faceta computada.	67
Quadro 20 – Exemplo de uma faceta com valores escalares.	68
Quadro 21 - Exemplo da definição de uma faceta com sinônimo.	69
Quadro 22 – Exemplo de uma faceta com hierarquia de valores.	71
Quadro 23 – Vocabulário Faceto.	73
Quadro 24 – Definição de ocorrência	74
Quadro 25 – Definição de ocorrência total.	74
Quadro 26 – Definição de ocorrência de um objeto.	75
Quadro 27 – Definição da probabilidade de ocorrência de um valor.	75
Quadro 28 – Fórmula da entropia.	75
Quadro 29 – Exemplo de um trecho de código HTML com	

anotação semântica.	112
Quadro 30 – Exemplo de um trecho de código Javascript que adiciona comportamento.	112
Quadro 31 – Triplas utilizadas no exemplo de aplicação de CSS e RDF.	112
Quadro 32 – Exemplo da representação de um recurso RDF em HTML usando anotações CSS.	113

Lista de tabelas

Tabela 1 – Lista de funcionalidades dos modelos de navegação facetada.	61
Tabela 2 - Resultado tabulado do experimento 1.	118