

Francisco José Zamith Guimarães

Utilização de ontologias no domínio B2C

DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE INFORMÁTICA
Programa de Pós-Graduação em Informática

Rio de Janeiro, novembro de 2002



Francisco José Zamith Guimarães

Utilização de ontologias no domínio B2C

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. Carlos José Pereira de Lucena

Rio de Janeiro, novembro de 2002



Francisco José Zamith Guimarães

Utilização de ontologias no domínio B2C

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. Carlos José Pereira de Lucena

Orientador
PUC-Rio

Prof. Arndt von Staa

PUC-Rio

Prof. Hugo Fuks

PUC-Rio

Profa. Simone Diniz Junqueira Barbosa

PUC-Rio

Prof. Ney Dumont

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

Rio de Janeiro, 19 de dezembro de 2002

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

Francisco José Zamith Guimarães

Graduou-se em Engenharia da Computação na Pontifícia Universidade Católica do Rio de Janeiro em 2000. Desenvolve pesquisas sobre ontologias e Web semântica.

Ficha Catalográfica

Guimarães, Francisco José Zamith

Utilização de ontologias no domínio B2C / Francisco José Zamith Guimarães ; orientador: Carlos José Pereira de Lucena. – Rio de Janeiro : PUC, Departamento de Informática, 2002.

[14], 195 f. : il. ; 30 cm

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

Inclui referências bibliográficas.

1. Informática – Teses. 2. Ontologias. 3. B2C. 4. Web semântica. I. Lucena, Carlos José Pereira de. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

Para meus pais, Martha e Francisco
Pelo apoio e incentivo.

Agradecimentos

Aos meus pais, pelo apoio e incentivo.

Aos meus avós, pelo carinho.

A minha irmã, pelo companheirismo.

A Danielle, pelo amor e paciência.

Ao meu orientador Professor Carlos José Pereira de Lucena pelo apoio e auxílio para a realização deste trabalho.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus amigos da PUC-Rio.

Aos professores que participaram da Comissão examinadora.

A todas as pessoas do Departamento de Informática.

A todos os meus familiares e amigos que de uma forma ou de outra me estimularam ou me ajudaram.

Resumo

Guimarães, Francisco José Zamith; Lucena, Carlos José Pereira. **Utilização de ontologias no domínio B2C**. Rio de Janeiro, 2002. 195p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A principal dificuldade dentro do domínio B2C está em aumentar a utilidade da WWW para o comércio eletrônico através da melhoria das possibilidades oferecidas ao consumidor. Apesar de a WWW permitir ao comprador ter acesso a uma grande quantidade de informação, obter a informação do fornecedor certo que venda o produto desejado a um preço razoável, pode ser uma tarefa muito custosa. Uma das formas de melhorar essa situação é através do uso de agentes inteligentes de busca de informação, isto é, agentes de compra, que auxiliam os compradores a encontrar produtos de seu interesse. Para que isso ocorra esbarra-se em uma dificuldade inerente à própria WWW: a mistura da linguagem natural, imagens e informação de layout de HTML são uma das maiores barreiras para a automatização do comércio eletrônico, pois a semântica da informação é somente compreensível por seres humanos. Desta forma espera-se conseguir agentes de compra mais eficientes quando associados ao uso de ontologias, e lojas virtuais que tenham anotações especiais que sigam uma ontologia. Nessa dissertação fazemos um estudo sobre as principais tecnologias envolvidas no desenvolvimento de ontologias em Ciência da Computação. Fazemos também um estudo de caso sobre a aplicação de ontologias dentro do domínio de B2C, visando assim avaliar o potencial e as dificuldades existentes para o desenvolvimento desse tipo de aplicação.

Palavras-chave

Informática; Ontologias; B2C; Web Semântica.

Abstract

Guimarães, Francisco José Zamith; Lucena, Carlos José Pereira (Advisor). **Ontologies use in B2C domain**. Rio de Janeiro, 2002. 195p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The main difficulty associated with the B2C domain is increasing the usefulness of WWW for the electronic trade through the improvement of the services provided to the consumer. Even though the WWW allows the buyer to have access to a great amount of information, to obtain the information from the right supplier that sells the desired product by a reasonable price can be a very expensive task. One of the ways of improving the web functionality is through the use of intelligent agents for search of information, that is, the introduction of purchase agents that aid the buyers to find products of their interest. For that to happen we need to overcome an inherent difficulty of the WWW: the mixture of natural language, images and layout information in HTML is one of the greatest barriers for the automation of the electronic trade, because the semantics of the information is only comprehensible for human beings. To solve this problem we hope to produce more efficient purchase agents by associating them to the use of ontologies, and virtual stores that have special annotations that follow ontologies. In the present dissertation we make a study of the main technologies related to ontologies development in computer science. We also develop a case study about the ontologies application to the B2C domain, seeking in this way to evaluate potential and existing difficulties for the development of this type of application.

Keywords

Computer science; Ontology; B2C; Web Semantics.

Sumário

1	Introdução	19
2	Business to Consumer	23
2.1.	Situação atual	23
2.2.	Comportamento do consumidor	27
2.3.	Relacionamento Comprador – Vendedor	32
3	Ferramentas de busca	34
3.1.	Máquina de busca genérica	35
3.1.1.	Percorrendo a Internet	35
3.1.2.	Mantendo um índice	37
3.1.3.	Construindo a busca	37
3.2.	Meta máquina de busca	40
3.3.	Softbots	43
3.3.1.	Shopbots	43
3.3.2.	Wrappers e a Web	46
4	Ontologia	49
4.1.	O que é ontologia	49
4.2.	Vantagens do uso de ontologias	53
4.3.	O que compõe uma ontologia	55
4.4.	Classificação das ontologias	56
4.5.	Principais áreas de aplicação das ontologias	58
4.5.1.	Processamento de linguagem natural	58
4.5.2.	Gestão do conhecimento	58
4.5.3.	Web semântica	59
4.5.4.	Comércio eletrônico	59
4.6.	Abordagens para o desenvolvimento de uma ontologia	61
4.6.1.	Abordagem partindo de uma inspiração	61

4.6.2. Abordagem indutiva	61
4.6.3. Abordagem dedutiva	62
4.6.4. Abordagem sintética	62
4.6.5. Abordagem colaborativa	63
4.7. Metodologias para o desenvolvimento de uma ontologia	64
4.7.1. Metodologia proposta por Mike Uschold e Martin King (Uschold & King, 1995)	65
4.7.2. Metodologia proposta por Michael Grüninger e Mark S. Fox (Grüninger & Fox, 1995)	65
4.7.3. Metodologia proposta por Mariano Fernández, Asunción Gómez-Pérez e Natalia Juristo (METHONTOLOGY) (Fernández et al., 1997)	67
4.7.4. Conclusão a respeito das metodologias	72
5 Tecnologias estudadas	73
5.1. XML	74
5.1.1. Visão geral de um documento XML (Martin et al., 2001; Holzner, 2001; Bray et al., 1998)	75
5.1.2. Partes que formam um documento XML	78
5.1.3. Tecnologias auxiliares a XML	80
5.1.3.1. DTD e XML Schema	80
5.1.3.2. Namespace	82
5.2. RDF (Martin et al., 2001; Holzner, 2001; Swick & Lassila, 1999)	85
5.2.1. Modelo básico	85
5.2.2. Sintaxe básica	88
5.2.2.1. Sintaxe serializada básica	88
5.2.2.2. Sintaxe abreviada básica	90
5.2.3. Coleções	92
5.2.4. Declaração sobre declaração RDF	95
5.3. RDF Schema	97
5.3.1. Vocabulário RDF Schema	98
5.4. DAML+OIL	103
5.4.1. Documento DAML+OIL	104
5.4.1.1. Cabeçalho do documento DAML+OIL	105

5.4.1.2. Elemento de classe	106
5.4.1.3. Restrições de propriedade	107
5.4.1.4. Elemento de propriedade	110
5.4.1.5. Instâncias	112
5.5. SHOE	113
5.5.1. Vocabulário SHOE	113
5.6. Comparação entre as linguagens	115
6 Estudo de caso	119
6.1. Ontologia proposta para a aplicação	120
6.1.1. 2BuyNet (Fortunato, 1999)	122
6.2. Aplicação	124
6.2.1. Máquina de inferência	125
6.2.2. Vantagens obtidas com o uso de ontologias	127
6.2.2.1. A descrição dos produtos de forma estruturada, e a separação das tags de layout da informação útil	128
6.2.2.2. Padronização da descrição de uma família de produtos	131
6.2.2.3. A capacidade de fazer inferências sobre as informações que seguem a ontologia	133
6.2.3. Dificuldades encontradas	134
6.2.3.1. Desempenho da aplicação	135
Expressividade da máquina de inferência	136
7 Conclusão	138
7.1. Principais contribuições	138
7.2. Trabalhos futuros	140
8 Referências Bibliograficas	141
Apêndice A – Telas do sistema	147
Apêndice B – Documentação ontologia	156
Documento sobre o objetivo da ontologia com breve descrição do domínio	157

Árvores de classificação de conceitos	158
Diagramas de relações binárias	159
Dicionário de conceitos	160
Tabela de relações binárias	163
Tabela de atributos de instância	164
Tabela de axiomas	167
Tabela de instâncias	169
Apêndice C – Diagramas de classe da aplicação	171
Pacote responsável pela interface com o usuário (Interface)	172
Pacote responsável pela interface com a base de conhecimento (BaseConhecimento)	177
Pacote responsável por controlar as classes das ontologias (OntologiaControle)	179
Pacote responsável por formular as perguntas às ontologias(Pergunta)	181
Pacote responsável pela fábrica de uma família de produtos (Fabricas)	185
Apêndice D – Diagramas de seqüência da aplicação	187

Lista de Figuras

Figura 1 – Estimativas de vários órgãos de pesquisa sobre o volume de transações B2C a serem realizadas nos Estados Unidos(OECD, 2001b).	24
Figura 2 – Estimativas de vários órgãos de pesquisa sobre o volume de transações B2C a serem realizadas na Europa (OECD, 2001b).	24
Figura 3 – Volume de transações B2C na América Latina (Wenrich & Becerra, 2001).	25
Figura 4 – Colocação das vendas, em valores, no segmento B2C por várias categorias no ano de 2000, desconsiderando-se a categoria de leilões virtuais (OECD, 2002).	25
Figura 5 – Colocação das vendas no segmento de B2C na América Latina por várias categorias no ano de 2001 (Wenrich & Becerra, 2001).	26
Figura 6 – Relação entre Indivíduos que usam a Internet versus Indivíduos que compra pela Internet em diversos países (OECD, 2001b).	26
Figura 7 – Processo de decisão de consumidor.	29
Figura 8 – Máquina de busca genérica.	39
Figura 9 – Meta máquina de busca genérica.	41
Figura 10 – Agente de compra (Shopbot) passivo de livros.	45
Figura 11 – Níveis de organização da informação na Web.	46
Figura 12 – Comparação entre ontologia e base de conhecimento.	52
Figura 13 – Diferentes tipos de ontologias e seus relacionamentos.	57
Figura 14 - Procedimento para o desenvolvimento e avaliação de ontologia.	66
Figura 15 – Visão geral do processo de desenvolvimento de ontologias (Fernández et al., 1997).	69
Figura 16 – Artefatos gerados na metodologia proposta em (Fernández et al., 1997; López et al., 1999).	72
Figura 17 - Organização em camada das linguagens.	73
Figura 18 – Representação gráfica de uma declaração RDF.	87
Figura 19 –Declaração RDF com relacionamento entre recursos.	87

Figura 20 – Declaração RDF que utiliza coleções.	93
Figura 21 – Declaração sobre declaração RDF.	96
Figura 22 – Funcionamento do 2BuyNet.	123
Figura 23 – Arquitetura da aplicação.	125
Figura 24 – Exemplo de inferência.	127
Figura 25 - Código 17 apresentado pelo Internet Explorer.	129
Figura 26 – Mapeamento entre ontologias.	132
Figura 27 – Relacionamento entre livro, CD e produto.	134
Figura 28 – Aplicação com arquitetura distribuída.	136
Figura 29 – Tela inicial do sistema.	149
Figura 30 – Formulário do perfil da pessoa a ser presenteada.	149
Figura 31 – Livros indicados ao perfil de pessoa.	150
Figura 32 – Formulário para busca de livros.	151
Figura 33 – Resposta do sistema a uma busca de livros.	151
Figura 34 – Detalhes de um Livro.	152
Figura 35 – Formulário de uma busca por CDs.	152
Figura 36 – Resposta do sistema a uma busca por CDs.	153
Figura 37 – Detalhes de um CD.	153
Figura 38 – Formulário de uma busca de produtos.	154
Figura 39 – Resposta do sistema a uma busca por produtos.	155
Figura 40 - Árvores de classificação de conceitos.	158
Figura 41 - Diagramas de relações binárias.	159
Figura 42 - Classes responsáveis pela interface principal do sistema.	173
Figura 43 - Classes responsáveis pela interface da opção de listar as triplas da máquina de inferência.	173
Figura 44 - Classes responsáveis pela interface de construção de formulários.	174
Figura 45 - Classes responsáveis pela interface da listagem dos produtos.	175
Figura 46 - Classes responsáveis pela interface da apresentação dos detalhes dos produtos.	176
Figura 47 - Classe responsável pela interface com a base de conhecimento.	178

Figura 48 - Classes responsáveis por controlar as classes das ontologias.	180
Figura 49 - Classes responsáveis por formular as perguntas de busca por produto.	182
Figura 50 - Classes responsáveis por formular a pergunta para indicação de produto a um perfil de consumidor.	183
Figura 51 - Classes responsáveis por formular a pergunta para obtenção de todas as triplas da máquina de inferência.	184
Figura 52 – Classes responsáveis pela fábrica de uma família de produtos.	186
Figura 53 - Formulário Busca CD.	188
Figura 54 - Lista Busca CD.	189
Figura 55 - Buscar Atributos Exclusivos Tipo Produto CD.	190
Figura 56 - Buscar Atributos Comuns Todos Produtos – parte 1	191
Figura 57 - Buscar Atributos Comuns Todos Produtos – parte 2	192
Figura 58 - Buscar Atributos Comuns Todos Produtos – parte 3	193
Figura 59 - Obter Descrição Produtos CD – parte 1.	194
Figura 60 - Obter Descrição Produtos CD – parte 2.	195

Lista de Tabelas

Tabela 1 – Classes RDF/RDF Schema predefinidas na especificação (Brickley & Guha, 2002).	99
Tabela 2 – Propriedades RDF/RDF Schema predefinidas na especificação (Brickley & Guha, 2002).	100
Tabela 3 – <i>Tags</i> utilizadas para a declaração de ontologias em SHOE.	114
Tabela 4 – <i>Tags</i> utilizadas para a declaração de instâncias em SHOE.	114
Tabela 5 – Comparação das funcionalidades para descrição de conceitos.	115
Tabela 6 – Comparação das funcionalidades para descrição de relacionamentos e funções.	116
Tabela 7 – Comparação das funcionalidades para descrição de axiomas.	116
Tabela 8 – Comparação das funcionalidades para descrição de instâncias.	116
Tabela 9 - Documento sobre o objetivo da ontologia com breve descrição do domínio.	157
Tabela 10 - Dicionário de conceitos - Taxinomia Produto.	160
Tabela 11 - Dicionário de conceitos - Taxinomia Gênero Produto.	160
Tabela 12 - Dicionário de conceitos - Taxinomia Pessoa.	161
Tabela 13 - Dicionário de conceitos - Taxinomia Loja.	162
Tabela 14 - Dicionário de conceitos - Taxinomia Local.	162
Tabela 15 - Dicionário de conceitos - Taxinomia Unidade Monetária.	162
Tabela 16 – Dicionário de conceitos - Taxinomia Forma Pagamento.	162
Tabela 17 - Tabela de relações binárias - Taxinomia Produto.	163
Tabela 18 - Tabela de relações binárias - Taxinomia Gênero Produto.	163
Tabela 19 - Tabela de relações binárias - Taxinomia Loja.	163
Tabela 20 - Tabela de atributos de instância - Conceito Produto.	164
Tabela 21 – Tabela de atributos de instância - Conceito Livro.	164
Tabela 22 - Tabela de atributos de instância - Conceito CD.	165

Tabela 23 - Tabela de atributos de instância - Conceito Local.	165
Tabela 24 - Tabela de atributos de instância - Conceito Loja.	165
Tabela 25 – Tabela de atributos de instância - Conceito GeneroProduto.	165
Tabela 26 – Tabela de atributos de instância - Conceito Pessoa.	166
Tabela 27 – Tabela de atributos de instância - Conceito UnidadeMonetaria.	166
Tabela 28 – Tabela de atributos de instância - Conceito FormaPagamento.	166
Tabela 29 - Tabela de axiomas - Conceito Pessoa.	168
Tabela 30 – Tabela de axiomas - Conceito Produto.	168
Tabela 31 - Tabela de instâncias - Conceito GeneroLiterario.	169
Tabela 32 – Tabela de instâncias - Conceito GeneroMusical.	169
Tabela 33 - Tabela de instâncias - Conceito UnidadeMonetaria.	170
Tabela 34 - Tabela de instâncias - Conceito FormaPagamento.	170

Lista de Códigos

Código 1 – Características principais de um documento XML.	77
Código 2 – Partes de um documento XML. Documento possui somente prólogo e corpo.	78
Código 3 – Exemplo de utilização de namespaces.	84
Código 4 – Prólogo e elemento raiz de um documento RDF típico.	89
Código 5 – Documento RDF da figura 19.	90
Código 6 – Versão serializada e versão abreviada de um código.	91
Código 7 – Versão serializada e versão abreviada de um código.	91
Código 8 – Versão serializada e versão abreviada de um código.	92
Código 9 – Documento RDF que instância uma <i>Bag</i> de literais.	95
Código 10 – Documento RDF que instância uma <i>Bag</i> de recursos.	95
Código 11 – Documento RDF Schema que especifica a ontologia seres.	101
Código 12 – Documento RDF que instância conceitos do esquema do código 11.	102
Código 13 – Cabeçalho de um documento DAML+OIL.	105
Código 14 – Elementos de classe.	107
Código 15 – Restrições de propriedade.	110
Código 16 – Elementos de propriedade.	112
Código 17 – Documento HTML que descreve um livro da Livraria Dirigida.	129
Código 18 – Documento RDF que descreve o livro da Livraria Dirigida.	130
Código 19 – Definição da propriedade ehRelacionaodo na ontologia de produto.	134
Código 20 – Documento RDF que descreve o livro da Livraria Dirigida.	134