

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE
ÁREA DE CIÊNCIA DA COMPUTAÇÃO

Roberval Gomes Mariano

**Desenvolvimento de uma família de sistemas de
recomendações baseados na tecnologia da Web
Semântica e seu reuso na recomendação de
instrumentos jurídico-tributários**

São Luís-MA

2008

ROBERVAL GOMES MARIANO

**DESENVOLVIMENTO DE UMA FAMÍLIA DE SISTEMAS DE RECOMENDAÇÕES
BASEADOS NA TECNOLOGIA DA WEB SEMÂNTICA E SEU REUSO NA
RECOMENDAÇÃO DE INSTRUMENTOS JURÍDICO-TRIBUTÁRIOS.**

Dissertação de Mestrado apresentada ao curso de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão, como parte dos requisitos para a obtenção do título de Mestre em Engenharia de Eletricidade, na área de Ciência da Computação.

Orientadora: Prof^a. Dra. Rosário Girardi

São Luís-MA
2008

Mariano, Roberval Gomes

Desenvolvimento de uma família de sistemas de recomendações baseados na tecnologia da web semântica e seu reuso na recomendação de instrumentos jurídicos tributários / Roberval Gomes Mariano. – São Luís, 2008.

152f.

Orientador: Rosário Girardi.

Impresso por computador (Fotocópia).

Dissertação (Mestrado) – Universidade Federal do Maranhão, Curso de Engenharia de Eletricidade na área de Ciência da Computação. São Luís, 2008.

1. Sistemas de recomendações. 2. Web semântica. 3. Reuso. I. Girardi, Rosário, orientador. II Título.

CDU 004.41

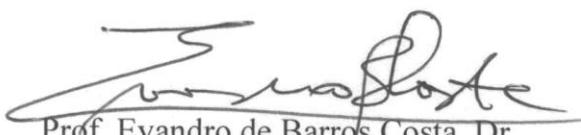
**DESENVOLVIMENTO DE UMA FAMÍLIA DE SISTEMAS DE RECOMENDAÇÕES
BASEADOS NA TECNOLOGIA DA WEB SEMÂNTICA E SEU REUSO NA
RECOMENDAÇÃO DE INSTRUMENTOS JURÍDICO-TRIBUTÁRIOS.**

ROBERVAL GOMES MARIANO

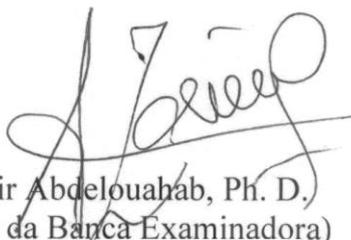
Dissertação aprovada em 05 de dezembro de 2008.



Profa. Maria del Rosario Girardi, Ph. D.
(Orientadora)



Prof. Evandro de Barros Costa, Dr.
(Membro da Banca Examinadora)



Prof. Zair Abdelouahab, Ph. D.
(Membro da Banca Examinadora)

A minha esposa, a qual foi a pessoa mais importante, para que eu retomasse este sonho, embora sacrificados muitos dos nossos momentos.

Aos meus filhos, aos quais prometi nunca me virem desistir de nada, não importando o quanto fosse difícil e para quem mais uma vez cumpri a minha promessa.

A minha mãe já falecida e a meu pai, os quais foram e sempre serão meus exemplos.

A meus irmãos e a minha irmã, com os quais queria ter convivido mais e estudado e trabalhado menos, mas aí não seria o Mariano que eles conhecem.

AGRADECIMENTOS

A todos que acreditaram que era possível e aos que mesmo acreditando ser impossível, ajudaram a ser possível.

A Akio Valente Wakiyama, o qual soube ser amigo e direcionador para que se chegasse ao fim, mesmo não explicando os meios.

A José Monteiro, o qual mim orientou na vida e foi amigo num momento fundamental em minha vida.

A todos os colegas do mestrado e do grupo de pesquisa, em especial a Lucas Drumond, Adriana Leite e à Professora Girardi pela paciência e apoio.

Ao Djefferson Smith Santos Maranhão e Geraldo Abrantes Sarmiento Neto, cujos trabalhos compuseram esta dissertação.

“Os sonhos só são impossíveis quando não trabalhamos neles todos os dias. Mesmo que para isso tenhamos de abrir mão de coisas muito importantes e ser cegos, surdos e mudos.”

RESUMO

A grande quantidade de dados disponíveis na Web e a sua natureza dinâmica criam uma demanda por aplicações de filtragem de informação, tais como os sistemas de recomendação. A falta de estruturação semântica dos dados disponíveis na Web é uma barreira para a melhoria da efetividade desta família de aplicações. Este trabalho apresenta a análise, projeto, implementação e avaliação de agentes de filtragem híbrida baseados na tecnologia da Web Semântica. Estes agentes foram integrados na ONTOSERS, uma família de aplicações para o desenvolvimento de sistemas de recomendações baseados na tecnologia da Web Semântica. Os agentes implementados foram testados e tiveram seus resultados comparados com os resultados de agentes utilizando filtragem colaborativa e baseada em conteúdo. As técnicas de filtragem híbrida apresentaram resultados melhores do que os obtidos com as outras técnicas nos experimentos realizados. As técnicas de filtragem híbrida testadas foram a ponderada e a alternada. O feedback explícito foi utilizado para validar as recomendações, apresentando uma melhor correlação com as técnicas de filtragem híbrida. Os agentes desenvolvidos foram ainda avaliados através do reuso da família de sistemas ONTOSERS na construção do INFOTRIB, um sistema multiagente de recomendações no domínio tributário brasileiro.

Palavras-chave: Sistema de recomendações, Reuso, Web Semântica, filtragem de Informação, Ontologia Tributária, Engenharia de Domínio Multiagente, Engenharia de Aplicação Multiagente.

ABSTRACT

The huge amount of data available on the Web and its dynamic nature create a demand for information filtering applications such as recommender systems. The lack of semantic structure of data available on the Web constitutes a barrier for increasing the effectiveness of such applications family. This work discusses the analysis, design, implementation and evaluation of Semantic Web based hybrid filtering agents. Such agents were integrated in ONTOSERS, an application family for the development of recommender systems based on the Semantic Web technology. The implemented agents were evaluated and their results were compared with the results of collaborative and content-based filtering agents. The hybrid filtering techniques presented better results than the other approaches in the conducted experiments. The tested hybrid filtering approaches were the weighted and switched ones. The explicit feedback was used to validate the recommendations, presenting a better correlation with the hybrid filtering techniques. The developed agents were also evaluated through the reuse of the ONTOSERS systems family, a multi-agent recommender system in the Brazilian tributary domain.

Keywords: Recommender Systems, Reuse, Semantic Web, Information Filtering, Tributary Ontology, Multi-agent Domain Engineering, Multi-agent Application Engineering.

LISTA DE FIGURAS

Figura 1	Classificação dos Sistemas de Recomendação (Schafer 2001)	24
Figura 2	Camadas da Web Semântica (ANTONIOU, HARMELEN, 2004).....	30
Figura 3	Processos da Engenharia de Domínio e da Engenharia de Aplicações	33
Figura 4	Complementariedade das metodologias MADEM e MAAEM (LINDOSO 2006).....	39
Figura 5	Modelo de Conceitos da ONTOSERS-DM.....	44
Figura 6	Modelo de Objetivos da ONTOSERS-DM	45
Figura 7	Pontos de Variabilidade e variantes no Modelo de Objetivos do ONTOSERS-DM.....	47
Figura 8	Modelo de Papéis do Objetivo Específico “Modelar Usuário”	48
Figura 9	Modelo de Papéis do Objetivo Específico “Filtrar Informação”	49
Figura 10	Modelo de Papéis do Objetivo Específico “Entregar Recomendação”	50
Figura 11	Modelo de Interações entre Papéis do Objetivo Específico “Modelar Usuário” ..	51
Figura 12	Modelagem de Variabilidade do Modelo de Interações entre Papéis	51
Figura 13	Modelo de Interações de Papéis do Objetivo Específico “Filtrar Informações” ..	52
Figura 14	Variabilidade do Modelo de Interações entre Papéis do Objetivo Específico “Filtrar informação”	52
Figura 15	Modelagem de Interações entre Papéis do Objetivo Específico “Entregar recomendações”	53
Figura 16	Interface de Login.....	53
Figura 17	Interface de Identificação	54
Figura 18	Interface de Categorias de Interesse	54
Figura 19	Interface de Recomendação.....	54
Figura 20	Interface de Avaliação da Recomendação.....	55
Figura 21	Modelo de Conhecimento da Sociedade Multiagente	55
Figura 22	Modelo da Sociedade Multiagente relacionadas ao agente “Interface do Usuário”	58
Figura 23	Variabilidade relacionada ao Agente “Interface do Usuário”	59
Figura 24	Modelo da Sociedade Multiagente relacionadas ao Agente “Filtrador”	60
Figura 25	Modelo de Interação dos Agentes	61
Figura 26	Modelo de Atividade	63
Figura 27	Modelo do Mecanismo de Coordenação e Cooperação	64
Figura 28	Modelo de Conhecimento do Agente “Interface do Usuário”	64
Figura 29	Variabilidade do Modelo de Conhecimento do Agente “Interface do Usuário” ..	65
Figura 30	Modelo de Conhecimento do Agente “Filtrador”	65
Figura 31	Modelo de Estado do Agente “Interface do Usuário”	66
Figura 32	Variabilidade do Modelo de Estado do Agente “Interface do Usuário”	67
Figura 33	Modelo de Estado do Agente “Filtrador”	67
Figura 34	Ponto de Variação do Modelo de Estado do Agente “Filtrador”	68
Figura 35	Modelo de Agentes e Comportamentos do agente “Interface de Usuário”	69
Figura 36	Modelo de Agentes e Comportamentos do agente “Filtrador”	69
Figura 37	Ponto de Variação do Behavior do Agente “Filtrador”	70
Figura 38	Modelo de Atos de Comunicação.....	71
Figura 39	Estrutura do Código-fonte da família ONTOSERS	72
Figura 40	Os diferentes Experimentos Realizados	75
Figura 41	As subclasses da ONTOTRIB	78

Figura 42	Os relacionamentos da classe “Instrumento Normativo Tributário”	78
Figura 43	Classificação dos Tributos	79
Figura 44	Instâncias da Classe “Usuário Legal Tributário” da ONTOTRIB	80
Figura 45	Principais classes, relacionamentos e instâncias da ONTOTRIB, em termos dos elementos da relação tributária.	81
Figura 46	Detalhamentos dos relacionamentos entre instâncias tributárias da ONTOTRIB	82
Figura 47	Resumo do Modelo de Análise de Similaridade Baseado em Casos Semânticos	83
Figura 48	Parte da Hierarquia da ONTOTRIB	85
Figura 49	Gráfico do recall e da precisão da FBC em função do limiar utilizado.....	95
Figura 50	Influência do Parâmetro K, no recall, para os limiares de 10% a 100%	97
Figura 51	Influência do Parâmetro K, na precisão, para os limiares de 10% a 100%	97
Figura 52	Gráfico do recall e da precisão da FC, em função do limiar utilizado, para $k=3$	99
Figura 53	Gráfico do recall e da precisão na FH alternada, em função do limiar utilizado	100
Figura 54	Gráfico do recall e da precisão da FH ponderada por intersecção, em função do limiar utilizado.....	101
Figura 55	Gráfico do recall e da precisão da FH ponderada, por pesos, com o peso do FBC = 0.4 e FC = 0.6, em função do limiar utilizado	102
Figura 56	FH PP, FBC 0.1 e FC 0.9	102
Figura 57	FH PP, FBC 0.2 e FC 0.8	103
Figura 58	FH PP, FBC 0.3 e FC 0.7	103
Figura 59	FH PP, FBC 0.4 e FC 0.6	103
Figura 60	FH PP, FBC 0.5 e FC 0.5	103
Figura 61	FH PP, FBC 0.6 e FC 0.4	103
Figura 62	FH PP, FBC 0.7 e FC 0.3	103
Figura 63	FH PP, FBC 0.8 e FC 0.2	103
Figura 64	FH PP, FBC 0.9 e FC 0.1	103
Figura 65	Gráfico do recall da FH ponderada, por pesos, com o peso de FBC = 0.1 e FC = 0.9 a FBC = 0.9 e FC = 0.1, em função do limiar utilizado.....	106
Figura 66	Gráfico da precisão da FH ponderada, por pesos, com o peso de FBC = 0.1 e FC = 0.9 a FBC = 0.9 e FC = 0.1, em função do limiar utilizado.....	107
Figura 67	Recall das filtragens FBC, FC, FHC, FHI e FHW	108
Figura 68	Precisão das filtragens FBC, FC, FHC, FHI e FHW	109
Figura 69	Correlação dos resultados das filtragens com o feedback dos usuários	111
Figura 70	Modelo de Conceitos	115
Figura 71	Modelo de Objetivos	116
Figura 72	Modelo de Papéis do INFOTRIB, do Objetivo Específico “Modelar Usuário”	116
Figura 73	Modelo de Papéis do Objetivo Específico “Filtrar Informação”	117
Figura 74	Modelo de Papéis do Objetivo Específico “Entregar Informação”	117
Figura 75	Modelo de Interação entre Papéis do Objetivo Específico “Modelar Usuário”	118
Figura 76	Modelo de Interação entre Papéis do Objetivo Específico “Filtrar Informação”	118
Figura 77	Modelo de Interação entre Papéis do Objetivo Específico “Entregar Recomendação”	119
Figura 78	Tela Apresentação	119
Figura 79	Tela Login.....	120
Figura 80	Tela Opções	120
Figura 81	Tela Perfil do Usuário.....	120
Figura 82	Tela Identificação	121

Figura 83	Tela Tipos Normativos	121
Figura 84	Tela Ramos Jurídicos.....	121
Figura 85	Tela Opções Extendida.....	122
Figura 86	Tela Elementos Tributários	122
Figura 87	Tela Tributos.....	122
Figura 88	Tela Tributos Federais	123
Figura 89	Tela Tributos Estaduais	123
Figura 90	Tela de Tributos Municipais.....	123
Figura 91	Configurações do Administrador.....	124
Figura 92	Modelo de Conhecimento da Sociedade Multiagente	124
Figura 93	Modelo da Sociedade Multiagente do Agente “Interface do Usuário”	125
Figura 94	Modelo da Sociedade da Multiagente do Agente “Filtrador”	126
Figura 95	Modelo de Interação entre Agentes	126
Figura 96	Modelo de Atividade da Sociedade Multiagente.....	127
Figura 97	Filtragem híbrida ponderada (BURKE 2007)	128
Figura 98	Modelo de Atividades da filtragem híbrida ponderada	129
Figura 99	Formas de se fazer filtragem híbrida alternada	129
Figura 100	Modelo de Atividade da filtragem híbrida alternada.....	130
Figura 101	Modelo do Mecanismos de Coordenação e de Cooperação	131
Figura 102	Modelo de Conhecimento do Agente “Interface do Usuário”	131
Figura 103	Modelo de Conhecimento do Agente “Filtrador”.....	132
Figura 104	Modelo de Estado do Agente “Interface do Usuário”	132
Figura 105	Modelo de Estado do Agente “Filtrador”	133
Figura 106	Modelo de Comportamento do Agente “Interface do Usuário”	134
Figura 107	Modelo de Comportamento do Agente “Filtrador”	134
Figura 108	Modelo de Atos de Comunicação dos Agentes no INFOTRIB FH	135
Figura 109	Índice de Acertos por Instrumento Normativo (IAin)	149
Figura 110	Índice de Acertos por Usuário (IAus).....	150
Figura 111	Comparativo entre os diversos índices	150
Figura 112	Correlação das filtragens com o feedback do usuário	152

LISTA DE TABELAS

Tabela 1	Técnicas de Recomendação (BURKE 2002).....	25
Tabela 2	Prós e contra das técnicas de recomendação (BURKE 2002)	26
Tabela 3	Insumos, tarefas e produtos da fase DM, da MADEM	35
Tabela 4	Insumos, tarefas e produtos da fase DD, da MADEM	37
Tabela 5	Insumos, tarefas e produtos da fase DI, da MADEM.....	37
Tabela 6	Mapeamento dos Conceitos de Projeto para implementação no JADE	38
Tabela 7	Insumos, tarefas e produtos da fase PER, da MADEM.....	38
Tabela 8	Fase Eng. Requisitos da Aplicação.....	40
Tabela 9	Fase Projeto da Aplicação	41
Tabela 10	Fase Implementação da Aplicação	42
Tabela 11	Classe “Instrumento Normativo”	77
Tabela 12	Subclasse “Instrumento Normativo Tributário” sem os slots herdados	77
Tabela 13	Classe “Usuário Legal”	77
Tabela 14	Subclasse “Usuário Legal Tributário”, sem os slots herdados	77
Tabela 15	Instâncias de Perfis de Usuários de Teste.....	87
Tabela 16	Instâncias de Instrumentos Normativos de Teste	89
Tabela 17	Matriz Resultados Esperados E(us,in).....	91
Tabela 18	Matriz de Similaridade Obtida na FBC	94
Tabela 19	Matriz de Resultados Obtidos O(us,in), na FBC	94
Tabela 20	Índice de Acertos por Instrumento Normativo(IAin)	96
Tabela 21	Índice de Acertos por Usuário(IAus).....	96
Tabela 22	Avaliações Realizadas pelos Usuários	98
Tabela 23	Matriz de Similaridade Obtida na FC, com K-3.....	98
Tabela 24	Matriz Similaridade Obtidos na FH alternada	99
Tabela 25	Matriz de Similaridade Obtida na FH Intersecção	100
Tabela 26	Matriz de Similaridade Obtida na FH ponderada.....	101
Tabela 27	Matriz Feedback Explícito dos Usuários.....	110
Tabela 28	Matriz Resultados Esperados E(us,in).....	148
Tabela 29	Matriz Resultados Obtidos O(us,in)	148
Tabela 30	Valores Obtidos das filtragens e do Feedback.....	151
Tabela 31	Valores Obtidos das filtragens e do Feedback.....	152

LISTA DE SIGLAS E ABREVIATURAS

DDEMAS	Domain Design technique for Multi-Agent Systems
FBC	Filtragem baseada no conteúdo
FC	Filtragem colaborativa
FH	Filtragem híbrida
FI	Filtragem de Informação
GESEC	Grupo de Pesquisa em Engenharia de Software e Engenharia do Conhecimento
GRAMO	Generic Requirement Analysis Method based on Ontologies
GUI	Graphical User Interface
HTML	Hypertext Markup Language
IDE	Integrated Development Environments
JADE	Java Agent Development Framework
MaAE	Multi-agent Application Engineering (Projeto do grupo GESEC)
MADEM	Multi-agent Domain Engineering Methodology
ONTOINFO	Ontologia para o domínio da recuperação de filtragem de informação
ONTOMADEM	Ontologia que suporta a aplicação da MADEM
RDF	Resource Description Framework
TF-IDF	Total Frequency-Inverse Document Frequency
XML	Extensible Markup Language
ICMS	Imposto sobre Circulação de Mercadoria e Prestação de Serviços de Comunicação e de Transporte Interestadual e Intermunicipal de Transporte
IPVA	Imposto sobre a Propriedade de Veículo Automotor
ITCD	Imposto Causa Mortis e Doações
IE	Imposto de Exportação
II	Imposto de Importação
IGF	Imposto sobre Grandes Fortunas
IOF	Imposto sobre Operações de Crédito, Câmbio e Seguro, e Sobre Operações Relativas A Títulos e Valores Mobiliários
IPI	Imposto sobre Produtos Industrializados
IR	Imposto sobre a Renda e Proventos de Qualquer Natureza

ITR	Imposto sobre A Propriedade Territorial Rural
ISSQN	Imposto sobre Serviços de Qualquer Natureza
ITBI	Imposto Sobre A Transmissão De Bens Imóveis E De Direitos A Eles Relativos

SUMÁRIO

1. INTRODUÇÃO	17
1.1 Relevância e motivação	19
1.2 Objetivos.....	20
1.2.1 <i>Objetivo Geral</i>	20
1.2.2 <i>Objetivos Específicos</i>	20
1.3 Estrutura da Dissertação	20
2. SISTEMAS DE RECOMENDAÇÕES E WEB SEMÂNTICA	22
2.1 Sistemas de recomendações.....	22
2.1.1 <i>Técnicas de recomendação</i>	24
2.1.2 <i>A filtragem híbrida</i>	28
2.2 A Web Semântica	30
2.2.1 <i>Ontologias</i>	30
2.2.2 <i>Sistemas de recomendações baseados na tecnologia da Web Semântica</i>	31
2.2.2.1 <i>A Ontologia ONTOJURIS, ONTOTRIB e o sistema INFONORMA</i>	31
2.3 Considerações finais do capítulo	32
3. AS METODOLOGIAS MADEM E A MAAEM.....	33
3.1 MADEM.....	34
3.1.1 <i>A Fase Análise de Domínio</i>	34
3.1.2 <i>A Fase Projeto de Domínio</i>	35
3.1.3 <i>A Fase Implementação de Domínio</i>	37
3.1.4 <i>A Fase Extração e Representação de Padrões</i>	38
3.2 MAAEM.....	39
3.2.1 <i>Fase Engenharia dos Requisitos da Aplicação</i>	39
3.2.2 <i>Fase Projeto da Aplicação</i>	40
3.2.3 <i>Fase Implementação da Aplicação</i>	41
3.3 Considerações finais	42
4. ONTOSERS: UMA FAMÍLIA DE SISTEMAS DE RECOMENDAÇÕES BASEADA NA TECNOLOGIA DA WEB SEMÂNTICA	43
4.1 Análise de Domínio da ONTOSERS-DM.....	43
4.1.1 <i>A Modelagem de Conceitos</i>	43
4.1.2 <i>A Modelagem de Objetivos</i>	44
4.1.3 <i>Modelagem de Papéis</i>	47
4.1.3.1 <i>Modelagem de Papéis Relacionada ao Objetivo Específico “Modelar Usuários”</i> 47	
4.1.3.2 <i>Modelo de Papéis Relacionado ao Objetivo Específico “Filtrar Informação”</i> 48	
4.1.3.3 <i>A Modelagem de Papéis Relacionados ao Objetivo Específico “Entregar recomendações”</i>	49
4.1.4 <i>Modelagem de Interações entre Papéis</i>	50
4.1.4.1 <i>Modelagem de Interações entre Papéis referentes ao Objetivo Específico “Modelagem do Usuário”</i>	50
4.1.4.2 <i>Modelagem de Interações de Papéis referentes ao Objetivo Específico “Filtrar Informação”</i>	52
4.1.4.3 <i>Modelagem de Interações de Papéis referentes ao Objetivo Específico “Entregar Recomendações”</i>	53
4.1.5 <i>Protótipo das Interfaces do Usuário</i>	53

4.2	Projeto de Domínio.....	55
4.2.1	<i>Modelagem do Conhecimento da Sociedade Multiagente</i>	55
4.2.2	<i>Projeto Arquitetural</i>	56
4.2.2.1	Modelagem da Sociedade Multiagente.....	56
4.2.2.1.1	O Agente “Interface do Usuário” na Modelagem da Sociedade Multiagente.....	56
4.2.2.1.2	O Agente “Filtrador” na Modelagem da Sociedade Multiagente.....	59
4.2.2.2	Modelagem das Interações dos Agentes.....	61
4.2.2.3	Modelagem das Atividades	62
4.2.2.4	Modelagem dos Mecanismos de Coordenação e Cooperação.....	63
4.2.3	<i>Projeto dos Agentes</i>	64
4.2.3.1	Modelagem do Conhecimento do Agente ”Interface do Usuário”	64
4.2.3.2	Modelagem de Conhecimento do Agente ”Filtrador”	65
4.2.3.3	Modelagem do Estado do Agente “Interface do Usuário”	66
4.2.3.4	Modelagem do Estado do Agente “Filtrador”	67
4.3	Implementação do Domínio	68
4.3.1	<i>Mapeamento do Projeto para a Implementação dos Agentes e dos Comportamentos</i>	69
4.3.2	<i>Mapeamento das Interações entre Agentes para Atos de Comunicação</i>	70
4.3.3	<i>Implementação dos Agentes</i>	71
4.3.3.1	O Modelo Genérico do Agente de interface do Usuário	72
4.3.3.2	O Modelo Genérico do Agente Filtrador.....	73
4.4	Extração e Representação de Padrões	73
4.5	Considerações Finais	73
5.	AVALIAÇÃO DOS AGENTES DE FILTRAGEM.....	75
5.1	A Ontologia Tributária ONTOTRIB	76
5.1.1	<i>Estrutura das classes principais da ONTOTRIB</i>	76
5.2	A Medida Semântica de Similaridade	83
5.3	A Medida Semântica de Similaridade e o KNN	86
5.4	Experiências Off-line.....	87
5.4.1	<i>Medidas de Avaliação</i>	91
5.4.2	<i>Resultados</i>	93
5.4.2.1	Os Resultados na FBC	93
5.4.2.2	Os Resultados na FC.....	96
5.4.2.3	Os Resultados na FH alternada.....	99
5.4.2.4	Os Resultados na FH, com a técnica ponderada, por Intersecção	100
5.4.2.5	Os Resultados na FH ponderada por pesos	101
5.5	Análises dos resultado das experiências off-line	104
5.6	Experiências On-line	109
5.6.1	<i>O Feedback dos Usuários</i>	109
5.7	Comparativo com os resultados de outros autores	111
5.8	Considerações Finais	112
6.	ESTUDO DE CASO: O SISTEMA INFOTRIB.....	114
6.1	Introdução.....	114
6.2	Engenharia de Requisitos do INFOTRIB	115
6.2.1	<i>Modelo de Conceitos</i>	115
6.2.2	<i>Modelo de Objetivos</i>	115
6.2.3	<i>Modelo de Papéis do INFOTRIB, do Objetivo Específico “Modelar Usuário”</i>	

6.2.4	<i>Modelo de Papéis do Objetivo Específico “Filtrar Informação”</i>	116
6.2.5	<i>Modelo de Papéis do Objetivo Específico “Entregar Informação”</i>	117
6.2.6	<i>Modelo de Interações entre Papéis do Objetivo Específico “Modelar Usuário”</i> 117	
6.2.7	<i>Modelo de Interações entre Papéis do Objetivo Específico “Filtrar Informação”</i>	118
6.2.8	<i>Modelo de Interações entre Papéis do Objetivo Específico “Entregar Recomendações”</i>	118
6.2.9	<i>Protótipo das Interfaces</i>	119
6.3	Projeto de Aplicação	124
6.3.1	<i>Modelo de Conhecimentos da Sociedade Multiagente</i>	124
6.3.2	<i>Modelo da Sociedade Multiagente do Agente “Interface do Usuário”</i>	125
6.3.3	<i>Modelo da Sociedade Multiagente do Agente “Filtrador”</i>	125
6.3.4	<i>Modelo de Interação entre os Agentes da INFOTRIB</i>	126
6.3.5	<i>Modelo de Atividade do INFOTRIB na FH ponderada</i>	127
6.3.5.1.1	<i>Modelo de Atividade, na FH, com método ponderado</i>	129
6.3.5.1.2	<i>Modelo de Atividade, na FH, com método Alternado</i>	129
6.3.6	<i>Projeto do Agente</i>	130
6.3.6.1	<i>Modelo do Mecanismo de Coordenação e Cooperação</i>	130
6.3.6.2	<i>Modelo de Conhecimento do Agente “Interface do Usuário”</i>	131
6.3.6.3	<i>Modelo de Conhecimento do INFOTRIB do Agente “Filtrador”</i>	131
6.3.6.4	<i>Modelo de Estado do INFOTRIB do Agente “Interface do Usuário”</i>	132
6.3.6.4.1	<i>Modelo de Estado do INFOTRIB do Agente Filtrador</i>	133
6.3.6.5	<i>Implementação da Aplicação INFOTRIB na FH</i>	133
6.3.6.5.1	<i>Modelo de Comportamento do Agente Interface do Usuário no INFOTRIB FH</i>	133
6.3.6.5.2	<i>Modelo de Comportamento do Agente Filtrador no INFOTRIB FH</i> ...	134
6.3.6.5.3	<i>Modelo de Atos de Comunicação dos Agentes no INFOTRIB FH</i>	135
6.4	<i>Considerações Finais</i>	135
7.	CONCLUSÕES	136
7.1	<i>Resultados e contribuições da pesquisa</i>	136
7.2	<i>Trabalhos futuros</i>	137
	REFERÊNCIAS	138
	ANEXO I – EXEMPLO DE CÁLCULO DAS MEDIDAS DE AVALIAÇÃO UTILIZADAS NOS EXPERIMENTOS OFF-LINE	148
	ANEXO II – EXEMPLO DE CÁLCULO DAS MEDIDAS DE AVALIAÇÃO UTILIZADAS NOS EXPERIMENTOS ON-LINE	151

1. INTRODUÇÃO

O grande volume de informação disponível aos usuários, na Web, cria uma demanda por mecanismos efetivos de acesso à informação. Outro fator complicador é a adequação semântica do termo ao contexto desejado. A existência de polissemia¹, onde uma palavra pode ter mais de um significado, dificulta a localização da informação desejada.

Em relação à polissemia tem-se, por exemplo, que a consulta, no Google, da palavra “manga”, retorna 10 links em 0,03 segundos, mas informa que o conjunto total de links é de aproximadamente 190.000.000 de links, nos quais se tem: manga de roupa, manga de fruta, manga de tirar chacota, etc.

Encontrar o que se deseja em grandes domínios não é uma tarefa trivial, considerando, por exemplo, que no caso da Internet, em setembro de 2003 existiam 3,3 bilhões de páginas na Web (POKORNÝ, 2004), o que torna a consulta de artigos na Web tradicional um problema para os usuários. (POKORNY 2004) informa também que 23% de todas as páginas Web mudam diariamente e que a parte não indexada da Web é muito menor que a existente (indexada, não indexada² e escondida³).

Mesmo quando considerados domínios menores que a Web ainda há problemas. O usuário nem sempre sabe o que quer, mas sabe exatamente o que não quer. Tratar isso, à nível de pesquisa de informação, é preocupar-se com o problema da Sobrecarga de Informação (“Information Overload”). A Sobrecarga de Informação ocorre quando o usuário tem a sua disposição uma quantidade de itens maior do que aquela com a qual ele pode lidar devido às suas limitações cognitivas. Outra preocupação é quanto à relevância da informação retornada, a qual é mensurada em termos de métricas como o recall⁴ e a precisão⁵ (SALTON, MCGILL,

¹ Polissemia - multiplicidade de sentidos de uma palavra ou locução. Fonte: HOUAISS, A. Dicionário Houaiss da língua portuguesa. Rio de Janeiro: <http://pt.wikipedia.org/wiki/Polissemia>

² Uma página ou site não indexado é aquele que ainda não foi registrado por nenhum mecanismo de busca ou que não possui referências para ele. Em função disso, ele existe, em muitos casos é público, mas ninguém sabe da sua existência, a não ser um grupo muito restrito de usuários.

³ Escondidas são as páginas e site existentes internamente. Necessitando de usuário e senha, protegidas por firewall e outras medidas de segurança e que só são acessíveis a um pequeno grupo de usuários, por ter informações restritas, sigilosas e/ou confidenciais.

⁴ recall é a razão entre o número de elementos relevantes recuperados (A) e a quantidade total de elementos relevantes existentes (R⁺). $\text{recall} = |A| / |R^+|$

⁵ precisão é a razão entre o número de elementos relevantes recuperados (A) e a quantidade total de elementos relevantes recuperados (r). $\text{precisão} = |A| / |r|$

1986). Tais métricas podem evidenciar problemas como ruído (itens indesejáveis), ou silêncio (ausência dos itens desejados).

O problema da Sobrecarga de Informação é observado mesmo em domínios específicos. Por exemplo, no domínio jurídico tributário brasileiro, abordado especificamente neste trabalho, a complexidade advém da existência de 3 esferas jurídicas competentes para emitir a respectiva legislação, com 5 espécies tributárias. Só em termos de impostos se tem: União, com IPI, IE, II, IR, IOF, ITR, IGF, IRes. IEG; os 26 Estados, com ITCD, ICMS, IPVA; os 5651 municípios com ITBI, IPTU, ISSQN e o Distrito Federal com os impostos estaduais e municipais. Só o Estado do Maranhão emitiu nos últimos 5 anos um Código Tributário Estadual, o qual foi alterado 11 vezes, bem como 38 leis tributárias estaduais e mais de 300 decretos, neste período.

A Web Tradicional é focada na apresentação para o ser humano e na atividade artesanal de busca (DIAS, SANTOS, 2003). Em função disso foi proposta uma nova geração da Web, a Web Semântica. Esta não está completa, mas já possui um conjunto de promissoras tecnologias associadas (DRUMOND, GIRARDI, 2006a). Na Web Semântica o conteúdo dos documentos é representado através de ontologias, que são uma especificação formal e explícita de uma conceitualização. As ontologias são compreensíveis por agentes de software, que podem realizar as atividades de identificação e seleção dos itens de interesse do usuário, o qual terá de se preocupar na análise e interpretação. A Aprendizagem de Máquina (WEBB, PAZZAINI, BILLSUS, 2001)(RUSSEL, NORVIG, 2004), a Mineração de Dados na Web (EIRINAKI, 2004)(MOBASHER, COOLEY, SRIVASTAVA, 2000) e o uso da tecnologia da Web Semântica (DRUMOND, GIRARDI, 2006) têm sido utilizados no contexto da filtragem da informação, inclusive de forma conjunta (MENCZER, 2004), para melhorar a efetividade dos sistemas de recomendações.

Sistemas de recomendações são sistemas que têm por objetivo sugerir itens de interesse de um usuário em um determinado domínio. Isto pode ser feito através do uso de técnicas de filtragem de informação (ADOMAVICIUS, TUZHILIN, 2005), que podem ser classificadas em:

- Filtragem baseada no conteúdo (FBC), em que um item é recomendado em termos de sua similaridade com outros itens avaliados positivamente pelo usuário no passado.

- Filtragem colaborativa (FC), em que é feita uma automatização do processo de divulgação de boca a boca, com os itens sendo recomendados, em função das preferências passadas de outros usuários com perfis similares.
- Filtragem híbrida (FH), em que são utilizados os dois tipos de filtragem anteriores.

1.1 Relevância e motivação

O domínio jurídico é um domínio complexo do conhecimento. Buscando a transparência⁶ e a eficiência das Administrações Públicas, diversas esferas, órgãos e países, publicam os seus instrumentos normativos na Internet, mas este ato não possibilita ao cidadão a satisfação da sua necessidade de informação, dada a dificuldade de busca de informação na Web.

Para sanar esta dificuldade houve a convergência entre metadados e ontologias, na Web Semântica, a qual tem exemplos no domínio jurídico, como: Serviço Localizador do Governo Australiano (AGLS) (CUNNINGHAM, 1998), Ontologias Lexical para Compartilhamento de Informações Legais (LOIS), da Itália (BERTOLDI, CHRISHMAN, ALVES. 2006), Jur-WordNet (Jur-WN), o qual também é desenvolvido pelo ITTIG-CNR, da Itália (BERTOLDI, CHRISHMAN, ALVES. 2006) e o EULEGIS, da Comunidade Econômica Européia (CEE).

Seja no âmbito de um país (Itália, Austrália, etc.), ou de vários (CEE) é importante a substituição dos mecanismos de busca de informação na Web por sistemas de recomendações baseados na Web Semântica, como forma de efetivar a transparência e a eficiência da Administração Pública, em tempestivamente, informar aos cidadãos a existência de instrumentos normativos de seus interesses.

No Brasil, exemplos de sistemas de recomendações, baseados na Web Semântica, na área jurídica são o INFONORMA (LINDOSO 2006) e o INFOTRIB (DRUMOND, GIRARDI et al.. 2006).

⁶ Além da legalidade é imposta à Administração Pública do Brasil, no caput do artigo 37, da CF/88, com a redação da Emenda Constitucional 19, os princípios da: impessoalidade, moralidade, publicidade e da eficiência. A publicidade refere-se a publicação oficial dos atos administrativos, a fim de que possam produzir os seus efeitos externos. Outro aspecto da publicidade é a exigência de transparência da atividade administrativa (inciso XXXIII, artigo 5º, da CF88).

1.2 Objetivos

1.2.1 Objetivo Geral

Contribuir para a melhoria da efetividade dos sistemas de recomendações através da análise, projeto, implementação e avaliação de algoritmos de filtragem de informação híbrida baseados na tecnologia da Web Semântica.

1.2.2 Objetivos Específicos

- Disponibilizar uma família de sistemas para a construção de sistemas de recomendação que integre os agentes desenvolvidos.
- Avaliar a efetividade da família construída através do desenvolvimento de um estudo de caso na construção do sistema INFOTRIB, um sistema de recomendação no domínio tributário.
- Avaliar as metodologias MADEM e MAEEM para a Engenharia de Domínio e de Aplicações Multiagente.

1.3 Estrutura da Dissertação

O capítulo 2 apresenta os conceitos sobre Sistemas de recomendações, Web Semântica e Ontologias, a fim de fornecer a base conceitual necessária ao entendimento desta dissertação. O capítulo 3 descreve a Metodologia de Engenharia de Domínio Multiagente (MADEM) e a Metodologia de Engenharia de Aplicação Multiagente (MAAEM), para que se compreenda o desenvolvimento dos artefatos de software para o reuso e com reuso. O capítulo 4 explica a Ontologia de Sistemas de recomendações baseados na tecnologia da Web Semântica (ONTOSERS), que integra um conjunto de artefatos de software para o reuso. O capítulo 5 define a Ontologia Tributária, ONTOTRIB, bem como a medida semântica de similaridade e realiza a avaliação dos agentes de filtragem. O capítulo 6, formaliza o INFOTRIB, o sistema de recomendação desenvolvido com o reuso, a ser

utilizado nos experimentos. O capítulo 7 expõe as conclusões sobre o trabalho realizado, enfatizando os resultados e os trabalhos futuros.

2. SISTEMAS DE RECOMENDAÇÕES E WEB SEMÂNTICA

Sistemas de recomendações são sistemas que buscam atender às necessidades dos usuários, a partir dos seus interesses, por meio da entrega de itens de informações compatíveis com os mesmos.

As primeiras gerações destes sistemas utilizavam técnicas da área de recuperação da informação (BAEZA-YATES, RIBEIRO-NETO, 1998) (BELKIN, CROFT, 1992), as gerações seguintes utilizavam técnicas de aprendizagem de máquina (PAPATHEODOROU 2001) (PIERRAKOS, PALIOURAS, et al.. 2001) (WEBB, PAZZANI, BILLSUS, 2001) e as gerações atuais integram os sistemas de recomendações com a tecnologia da Web Semântica (MIDDLETON, ALANI, DE ROURE, 2002).

Este capítulo está organizado em 2 seções. A seção 2.1 apresenta uma visão geral dos sistemas de recomendações, dos elementos que os compõem e exemplifica as principais técnicas de recomendações existentes. A seção 2.2 define o que é Web Semântica, especificando as suas camadas, apresentando o conceito de ontologia, a relação entre sistema de recomendações e Web Semântica e correlaciona a ontologia de teste ONTOTRIB e o sistema de recomendação jurídico INFOTRIB.

2.1 Sistemas de recomendações

(ADOMAVICIUS, 2005) define os sistemas de recomendações como:

Sistemas de recomendações são sistemas que procuram maximizar a função utilidade dos itens que podem ser recomendados a um usuário, isto é:

$$\forall u \in U, i \in I \quad i'_u = \arg \max_{i \in I} ut(u,i)$$

onde:

u é um usuário específico ($u \in U$)

U é o conjunto de todos os usuários

i é o conjunto de itens recomendados para um usuário específico ($i \in I$)

I é o conjunto de todos os possíveis itens que podem ser recomendados

ut é uma função de utilidade de um usuário

$$ut: U \times I \rightarrow R,$$

onde:

R é o conjunto ordenado de itens recomendados

(ADOMAVICIUS, 2005) complementa informando que:

As recomendações são tipicamente indicações de itens de informações, pelos quais o usuário deve ter interesse. A indicação de interesse pode ocorrer de diversas formas: pelo histórico de consultas do usuário, pelo comportamento dos usuários similares ao usuário ou pela avaliação expressa do item de informação. Para os itens não avaliados o sistema de recomendação estima (realiza a predição), pela extrapolação dos itens conhecidos para os desconhecidos. A função utilidade mede a utilidade do item s para o usuário c , onde C e S tendem a serem grandes.

Outro detalhe é que cada elemento do espaço do usuário C pode ser definido como um perfil, o qual inclui várias das características dos usuários. Já a função utilidade u não é definida usualmente sobre todo espaço $C \times S$, porém sobre um subconjunto deste e depois o resultado é extrapolado para espaço completo.

Na Figura 1 é apresentado o modelo geral dos sistemas de recomendações, de (SHAFFER 2001). Este modelo possui os seguintes componentes: as entradas do usuário-alvo, o método de recomendação, as entradas da comunidade, as saídas do método de recomendação, as formas de entregar e o grau de personalização. As entradas do usuário-alvo são capturadas pela navegação implícita, navegação explícita, histórico de compras, etc.. O método de recomendação pode ser: seleção manual, sumário estatístico, baseada em atributo, correlação item-a-item ou correlação usuário-a-usuário. As entradas da comunidade são especificadas pela popularidade de um item, histórico de compras, etc.. As saídas do método de recomendação variam entre sugestões, predições, etc.. As formas de entregar a recomendação podem ser: ativas ou passivas. As ativas podem ser da modalidade “push” ou “pull”. A “push” é aquela em que a recomendação é empurrada para o usuário, como ocorre com as janelas “pop-up”. A “pull” é aquela em que o usuário solicita a partir de mensagens do tipo “nós temos recomendações para você.”. A forma de entrega do tipo orgânica, também chamada de passiva, ocorre quando as recomendações são incluídas nos site com as demais informações. O grau de personalização especifica se haverá personalização e de que espécie.

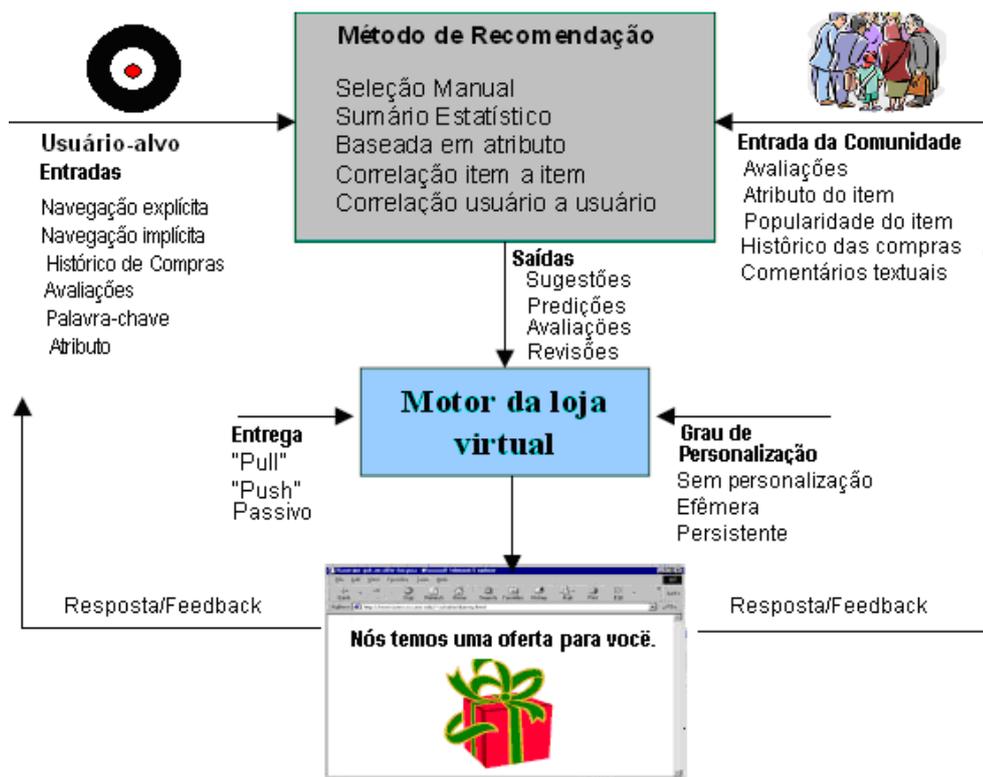


Figura 1 Classificação dos Sistemas de Recomendação (Schafer 2001)

Do modelo geral dos sistemas de recomendações de (SHAFFER 2001), para o foco desta dissertação, serão abordados os componentes: modelagem do usuário, filtragem da informação e a entrega da informação.

2.1.1 Técnicas de recomendação

Existem diversas técnicas de recomendações, como a: colaborativa (FC), a baseada no conteúdo (FBC), a Demográfica (FD), a baseada na Utilidade (FUT) e a baseada no Conhecimento (FKB).

Na filtragem baseada no conteúdo a análise de similaridade é em cima dos itens de informação similares aos que o usuário costuma consultar. De forma que pessoas que leram artigos de futebol tendem a, no futuro, querer ver outros artigos sobre futebol.

Na filtragem colaborativa, a análise de similaridade é realizada entre a preferência dos usuários. Esta abordagem tem a vantagem de ser independente do item de informação, o qual pode ser um filme, um prato em um restaurante, um caso jurídico, um artigo científico ou qualquer outro tipo de item de informação que se

queira avaliar. Por ser realizada a partir das avaliações de diferentes usuários, permite identificar o cruzamento entre interesses de diferentes nichos, com possibilidade de recomendações inesperadas.

A técnica Demográfica consiste no uso do formalismo estatístico para agrupar os usuários e suas preferências. Nesta técnica os usuários são agrupados em classes e associadas a elas freqüências, conforme a distribuição das preferências, a qual pode ser representada num polígono de freqüência. As preferências são “recomendadas” com base em fórmulas estatísticas.

A técnica baseada na Utilidade baseia-se na definição de uma função que seja capaz de representar o interesse (utilidade) do usuário pelo item.

A Tabela 1 apresenta um resumo das principais técnicas de recomendações apresentadas em (BURKE 2002). A coluna pré-requisito representa a exigência a ser satisfeita em relação ao universo de usuários ou de item, enquanto que a coluna de entrada representa o que o usuário deverá fornecer para obter a execução da respectiva técnica de filtragem. A coluna processo especifica como a técnica é executada a partir dos indivíduos e universo de itens e de usuários.

Tabela 1 Técnicas de Recomendação (BURKE 2002)⁷

Técnica	Pré-requisito	Entrada	Processo
Colaborativa	Avaliação dos usuários (U) sobre os itens em I	Avaliação de u dos itens em I	Identificar usuários em U similares a u e extrapolar as avaliações de i
Baseada no conteúdo	Características dos itens em I	Avaliações do usuário u dos itens em I	Gera uma classificação que ajusta o comportamento das avaliações de u sobre o item i
Demográfica	Informações demográficas sobre U e suas avaliações dos itens em I	Informações demográficas sobre u	Identifica usuários que são demograficamente similares a u extrapola suas avaliações para i
Baseada na utilidade	Características dos itens em I	A função utilidade sobre itens em I que descreve as preferências de u	Aplica a função utilidade para os itens e determina o rank de i
Baseada no Conhecimento	Características dos itens em I e conhecimento de como estes itens atendem as necessidades dos usuários	A descrição das necessidades ou interesses de u	Inferir uma combinação entre i as necessidades de u

⁷ **I** é o conjunto de itens sobre os quais recomendações podem ser feitas.

U é o conjunto de usuários, cujas preferências são conhecidas.

u é o usuário para o qual uma recomendação precisa ser gerada.

Há restrições, como as de (TORRES, 2004), a algumas das técnicas acima, em função de não haver personalização, sendo entregue o mesmo conteúdo para todos os usuários, como no caso da demográfica. Neste trabalho não se adentrará nesta seara.

Em (BURKE 2002) são detalhadas as vantagens e desvantagens de cada técnica, os quais estão resumidos na Tabela 2.

Tabela 2 Prós e contra das técnicas de recomendação (BURKE 2002)

Técnica	Prós	Contra
Colaborativa	A: Pode identificar cruzamento de nichos. B: Não é necessário conhecimento do domínio C: Adaptativo: qualidade melhora com o tempo D: Suficiente feedback implícito	I: Problema novos usuários J: Problema novos itens K: Problema do usuário "Ovelha negra" L: Qualidade depende de um enorme conjunto de dados históricos. M: Problema estabilidade x plasticidade
Baseada no conteúdo	B, C, D	I, L, M
Demográfica	A, B, C	N: Ter informações demográficas I, K, L, M, N
Baseada na utilidade	E: Não tem os problemas de novos usuários e itens F: Sensível a mudanças de preferências G: Pode incluir características sem produto.	O: Usuário deve fornecer as entradas para a função utilidade P: Sugere habilidades estáticas (não aprende)
Baseada no Conhecimento	E, F, G H: Pode mapear as necessidades dos usuários para produtos	Q: Requerido motor de conhecimento P

A filtragem colaborativa tem a vantagem de recomendações inesperadas pelo cruzamento de nichos. Pode se ter feedback implícito e explícito, os quais possibilitam melhoria na qualidade da recomendação.

A filtragem baseada no conteúdo tem a vantagem de que não é necessário o conhecimento do domínio, mas necessita que seu conteúdo sofra algum processamento ou descrição, quando se referir a diferentes mídias que não a textual, como som e imagem, a fim de ser possível identificar onde estão as características consideradas importantes para a definição do item de informação. Nela também é possível o uso de *feedback* para que se ajuste as recomendações à realidade do seu universo de usuários.

i é um item para o qual deseja-se prever a preferência que o usuário **u** tem pelo item.

A filtragem demográfica tem a vantagem de ter o formalismo estatístico.

A filtragem baseada na utilidade tem a vantagem de não se ter os problemas dos usuários e itens novos.

A filtragem baseada no conhecimento tem a vantagem de associar as necessidades dos usuários aos produtos, sem a necessidade de reclassificação ou agrupamento dos usuários como ocorre na demográfica.

A filtragem colaborativa tem problemas como: dos novos usuários; dos novos itens de informação; da “ovelha-negra”; de poucas avaliações e da instabilidade do domínio.

O problema dos novos usuários ocorre quando o usuário por não ter ainda realizado suficientes avaliações não receberá a informação de itens de informação compatíveis com o seu perfil, uma vez que, nesta técnica, o perfil é definido a partir das avaliações que o usuário realiza. O problema dos novos itens de informação consiste em que, quando eles são cadastrados, ainda não possuem avaliações, o que impede a sua recomendação a algum usuário. O problema do usuário “ovelha-negra” caracteriza o usuário que apresenta um perfil que não se encaixa em nenhum grupo de usuários. O problema da existência de poucas avaliações é o de que a qualidade da recomendação será diretamente proporcional a quantidade de avaliações existentes. Logo, quanto mais avaliações existirem melhor a capacidade de se recomendar corretamente, o que depende de tempo e do percentual de participação dos usuários. Quanto mais avaliações, melhor a capacidade de adaptação e ajuste das recomendações, principalmente quando existe o feedback do usuário validando as recomendações e possibilitando se fazer um ajuste mais fino.

O problema da instabilidade do domínio dá-se quando as preferências dos usuários mudam muito rápido, de forma que se recomendam itens de informação, que não são mais do interesse dos usuários, por não existir um padrão estável de comportamento.

Outro problema da filtragem colaborativa é o da esparsividade. Como temos um universo de usuários e de itens de informação, os quais tendem a ser grandes, podemos ter a esparsividade do primeiro, do segundo ou de ambos. Ou seja, um número pequeno do universo de usuários realiza o uso do sistema de recomendação, fazendo com que as suas preferências sejam induzidas para os

demais, os quais poderão ter preferências distintas dos avaliadores. Bem como podemos ter este subconjunto dos usuários avaliando apenas um pequeno conjunto dos itens de informação, ficando a grande maioria sem serem referenciados ou avaliados.

A filtragem baseada no conteúdo tem os problemas de: usuários novos, novos itens de informação e da estabilidade x plasticidade.

O problema dos usuários novos já foi abordado na filtragem colaborativa. O problema da estabilidade x plasticidade é sintetizado na frase: “Os sistemas de recomendações precisam refletir as novas preferências (Plasticidade), mas estas não podem ser tão mutáveis, que não se possa definir quais são as reais preferências dos usuários (necessidade de estabilidade)”.

A técnica Demográfica tem os problemas de não poder agrupar itens de informação e usuários que não tenham referências de uso e dos agrupamentos terem de serem estáveis. Esta técnica possibilita o cruzamento de nichos, por extrapolar a informação presente no item, e é independente do domínio (tudo que puder ser tabulado poderá ser analisado). É adaptativa, na medida em que se pode criar mais ou menos faixas de agrupamento dos usuários.

A técnica baseada na Utilidade, por utilizar uma função, não tem o problema dos novos usuários e dos novos itens de informação, além de permitir ajustes para representar as novas preferências e permitir a análise da característica do produto. A dificuldade no uso desta técnica é a necessidade de um conhecimento estatístico, para que se faça as calibrações de forma consistente.

Na filtragem baseada no conhecimento temos os mesmos prós da baseada na utilidade, com o adicional de se poder mapear as necessidades dos usuários para produtos. O problema desta filtragem está na aquisição do conhecimento.

2.1.2 A filtragem híbrida

(BURKE 2002) propõe que as técnicas sejam combinadas, de forma a sanar as deficiências individuais de cada uma.

A literatura científica, na maioria dos casos, cita a filtragem híbrida como sendo aquela resultante do uso da filtragem baseada no conteúdo (FBC) e da

colaborativa (FC), mas para (BURKE 2002), o conceito de filtragem híbrida é qualquer junção coerente de técnicas de filtragem distintas.

Os métodos de recomendações híbridas identificados por (BURKE 2002) e (TORRES 2004a) são:

- Ponderado (“Weighted”) - A similaridade de um item é computada a partir de uma combinação de várias técnicas de recomendação, com diferentes pesos para cada uma delas (Ponderada por pesos) ou pela interseção das recomendações (Ponderada por Intersecção);
- Alternado (“Switching”) - O sistema utiliza um critério para alternar a técnica que gera a recomendação;
- Cascata (“Cascade”) - Uma técnica de recomendação refina as recomendações fornecidas por outra técnica;
- Combinação de características (“Feature Combination”) – Características de diferentes dados são utilizadas em um único algoritmo;
- Aumento de Características (“Feature Augmentation”) - As recomendações geradas por uma técnica de recomendação são utilizadas como entrada para outra técnica;
- Misto (“Mixed”) - recomendações de várias técnicas são apresentadas na mesma lista;
- Meta-nível (“Meta-nível”) - O modelo aprendido por uma técnica de recomendação é utilizado como informação de entrada de outra técnica.

Para o contexto deste trabalho será desenvolvida uma solução computacional utilizando filtragem baseada no conteúdo, outra colaborativa e uma terceira híbrida, com uso de diversos métodos híbridos, a fim de comprovar seus benefícios.

2.2 A Web Semântica

A tecnologia da Web Semântica ainda está em processo de evolução, mas muito dela já está formalizada e já existem várias tecnologias padronizadas. A Web Semântica não é uma nova Web, mas uma extensão da Web tradicional.

A Web Semântica propõe que as máquinas passem a entender o conteúdo dos documentos, a fim de melhorar a efetividade do acesso à informação na Web. Neste sentido foi proposta a arquitetura em camadas, mostrada na Figura 2, onde cada camada tem compatibilidade com a camada imediatamente inferior e entendimento parcial da camada imediatamente superior (DRUMOND 2006).

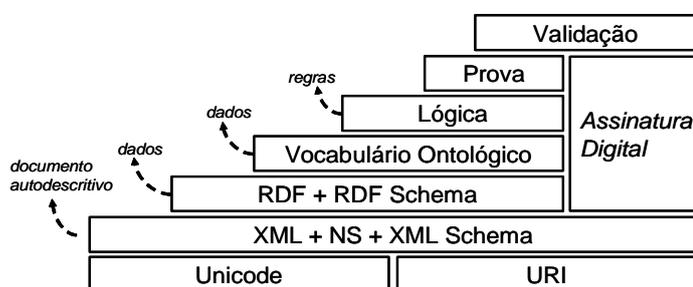


Figura 2 Camadas da Web Semântica (ANTONIOU, HARMELEN, 2004)

A contribuição da Web Semântica, para a melhoria da efetividade dos sistemas de recomendação, advém da capacidade de se adicionar a semântica do respectivo domínio de conhecimento, representado pela ontologia associada. Com isso se dá um formalismo à informação, que passa a ser possível de ser tratada pelos agentes de software.

2.2.1 Ontologias

As ontologias são estruturas de representação do conhecimento particularmente úteis para a especificação de abstrações de alto nível (GIRARDI, LINDOSO, 2006). As ontologias têm as características de serem não ambíguas, formais, explícitas e compartilháveis. Elas são não ambíguas, porque as definições presentes nelas possuem apenas uma possibilidade de interpretação. Formais, por serem capazes de serem processadas pelos computadores. Explícitas, pelos conceitos e restrições relacionados serem prévia e expressamente definidos. Compartilháveis, por serem um conhecimento advindo do consenso do grupo sobre

o significado dos conceitos. O fato de uma ontologia possuir uma terminologia não ambígua, compartilhada entre todos, faz com que ela viabilize a comunicação entre entidades computacionais como agentes de software, os quais se comunicam por meio de expressões formuladas em termos de uma ontologia (LINDOSO, 2006).

2.2.2 Sistemas de recomendações baseados na tecnologia da Web Semântica

Diversos produtos já foram criados e artigos publicados explorando o uso da tecnologia da Web Semântica em Sistemas de recomendações (DRUMOND 2006) (GIRARDI, MARINHO 2007) (LINDOSO 2006).

Em particular, vários trabalhos mostram a aplicação das ontologias em sistemas de recomendações. (FANIZZI, ESPÓSITO, 2005) expressa um exemplo de medida de similaridade semântica baseada em expressões de lógica de descrições e o método híbrido de mapeamento dos pesos, para a pesquisa na Web Semântica. (MIDDLETON, ROURE, 2002) e (ZIEGLER, 2004) desenvolve um método para modelar usuário baseado em inferências ontológicas, representando os perfis dos usuários em termos de uma ontologia de tópicos de artigos de pesquisa, onde eles são classificados usando técnicas de clustering e já apresenta os conceitos do conceito de relevance feedback. (MENCZER, 2004) e (MENCZER, 2005) propõem a combinação das técnicas de mineração com a combinação das análises de conteúdo e de link para estimar uma análise de similaridade semântica que una as duas análises citadas.

2.2.2.1 A Ontologia ONTOJURIS, ONTOTRIB e o sistema INFONORMA

O INFONORMA é um sistema multiagente de recomendações, para o domínio jurídico que recomenda instrumentos jurídico-normativos para usuários de acordo com seus interesses particulares. Os itens de informação do INFONORMA são representados como instâncias da ontologia ONTOJURIS, a qual é escrita em OWL, conforme as recomendações do W3C (DRUMOND, GIRARDI, LEITE, 2007).

O INFONORMA foi modelado seguindo a orientação da Metodologia de Engenharia de Aplicação MultiAgente (MAAEM) (DRUMOND, GIRARDI, 2008).

A ONTOJURIS modela os ramos do Direito Brasileiro, os instrumentos normativos e seus tipos (DRUMOND, GIRARDI, 2008).

A ONTOTRIB especializa a ONTOJURIS no ramo do Direito Tributário (ARAÚJO, et ali., 2008).

2.3 Considerações finais do capítulo

Este capítulo apresentou as diversas gerações de sistemas de recomendações, apresentou os conceitos envolvidos, algumas técnicas de recomendação e aprofundou a filtragem híbrida. A seguir contextualizou a Web Semântica e o uso desta nos sistemas de recomendações.

3. AS METODOLOGIAS MADEM E A MAAEM

A construção de aplicações de software através da reutilização envolve dois processos distintos e complementares, que são o desenvolvimento para o reuso e o desenvolvimento com o reuso. Ambos são também denominados, respectivamente, Engenharia de Domínio e Engenharia de Aplicações, sendo a primeira responsável por fornecer artefatos reusáveis ao segundo, que retroalimenta aquele com novos requisitos e novas soluções (POHL, BÖCKLE, LINDEN, 1998).

Estes processos são complementares na medida em que a Engenharia de Domínio, também conhecida como Desenvolvimento PARA o Reuso, é um processo para criação de abstrações de software reusáveis de alto nível para o desenvolvimento de uma família de aplicações em um domínio e a Engenharia de Aplicações ou Desenvolvimento COM Reuso é um processo para a construção de uma aplicação específica usando abstrações de software reusáveis disponibilizados pela Engenharia de Domínio (GIRARDI, LINDOSO, 2005) (Figura 3).

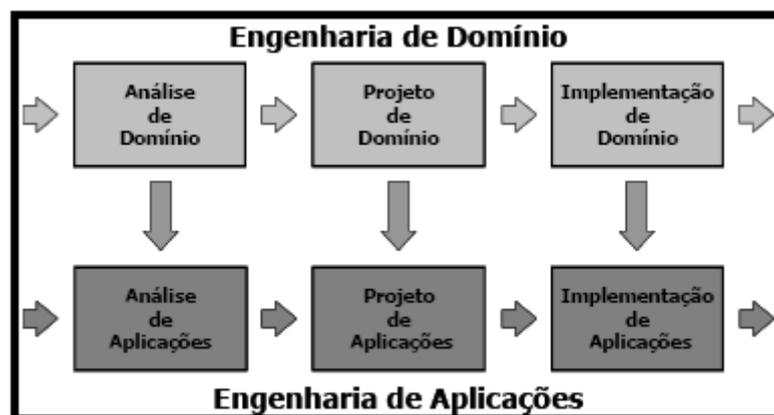


Figura 3 Processos da Engenharia de Domínio e da Engenharia de Aplicações

A MADEM (GIRARDI, LINDOSO, 2005) (GIRARDI, LINDOSO, 2005a) (GIRARDI, MARINHO, 2007) e a MAAEM (DRUMOND, GIRARDI, LEITE, 2007) (DRUMOND, GIRARDI, LEITE, 2007a) são duas metodologias multiagentes desenvolvidas pelo GESEC. Ambas trabalham o reuso, sendo que na MADEM se trabalha *para o reuso* e na MAAEM se trabalha *com o reuso*. Ou seja, a primeira

metodologia focaliza a criação de artefatos de software reutilizáveis e a segunda, o uso destes artefatos.

3.1 MADEM

A MADEM (“Multi-agent Domain Engineering Methodology”) é uma metodologia para a Engenharia de Domínio Multiagente, que utiliza as técnicas: GRAMO (“Generic Requirement Analysis Methodology based on Ontologies”), para a análise de domínio (DM), a DDEMAS (“Domain DEsign of Multi-Agent Systems”), para o projeto de domínio (DD) e a DIMAS (“Domain Implementation of Multi-Agent Systems”), para a implementação de domínio (DI), as quais constituem as fases de: Análise de Domínio, Projeto de Domínio e Implementação de Domínio. Encerrando a MADEM há a fase de Extração e Representação de Padrões de reuso. Cada fase tem o seu conjunto de insumos, tarefas e produtos (GIRARDI, MARINHO, 2007).

3.1.1 A Fase Análise de Domínio

A fase de análise do domínio tem o objetivo de identificar o que a família de aplicações de sistemas multiagentes se propõe a resolver e consiste das tarefas de: modelagem dos conceitos do domínio, modelagem dos objetivos, modelagem dos papéis, modelagem das interações entre papéis e prototipação das interfaces com o usuário (Tabela 3).

Em cada fase é realizada a modelagem das variabilidades, onde variabilidade se refere a capacidade ou habilidade a mudar (POHL, BÖCKLE, LINDEN, 1998), na qual são identificados os pontos de variação, que são a representação de um conceito sujeito a variabilidade dentro de um artefato de domínio enriquecido pela informação contextual, e as respectivas variantes, que são uma representação das partes variantes dentro do artefato de domínio. As variantes podem ser agrupadas em alternativas ou opcionais (POHL, BÖCKLE, LINDEN, 1998). As responsabilidades variantes dos Objetivos Específicos determinam quais possibilidades de interações os papéis terão. De modo que se a variante for do tipo alternativa, teremos grupos de interações diferentes para cada variante e se for do

tipo opcional estes poderão existir, ou não. Os insumos são aqueles elementos necessários, a fim de realizar a tarefa, para se obter o respectivo produto.

A tarefa de modelagem dos conceitos do domínio representa uma tempestade de idéias, a fim de identificar os conceitos principais associados ao domínio em questão. A modelagem dos objetivos procura identificar o objetivo geral e os seus diversos objetivos específicos, sendo que o objetivo geral é alcançado a partir do alcance dos objetivos específicos. Estes, por sua vez, para serem atingidos necessitam que sejam exercidas certas responsabilidades, detalhadas no Modelo de Papéis de cada objetivo específico, no qual são relacionados os papéis, conhecimentos, condições e entidades externas associadas, de forma que um papel realiza determinada responsabilidade. Posteriormente, são identificados os relacionamentos entre os diversos papéis para criar a modelagem de interações entre papéis (GIRARDI, MARINHO, 2007). Encerrando a fase de análise de domínio tem-se a prototipagem da interface com o usuário, cujo foco é o desenvolvimento de interfaces de usuário genéricas da família de sistemas a serem usadas como referência na construção de aplicações específicas.

Tabela 3 Insumos, tarefas e produtos da fase DM, da MADEM

Fases	Insumos	Tarefas	Produtos
Análise de Domínio	Conhecimento de Especialistas, Literatura no Domínio, Aplicações de Software Existente	Modelagem de Conceitos	Modelo de Conceitos
	Requisitos da família de aplicações	Modelagem de Objetivos	Modelo de Objetivos
	Modelo de Objetivos	Modelagem de Papéis	Modelo de Papéis
	Modelo de Objetivos	Modelagem de Interações entre Papéis	Modelo de Interações entre Papéis
	Modelo de Papéis		
	Modelo de Interações entre Papéis	Prototipagem da Interface com o usuário	Protótipo da Interface com o usuário
Modelo de Papéis			

3.1.2 A Fase Projeto de Domínio

A fase de projeto de domínio consiste em prover uma solução computacional ao problema identificado na fase de análise de domínio, para os requisitos de uma família de sistemas de software multiagente especificados no

modelo de domínio. Esta fase é composta das tarefas de modelagem do conhecimento da sociedade multiagente, projeto arquitetural e projeto de agentes (Tabela 4).

A tarefa de modelagem do conhecimento da sociedade multiagente tem a função de representar o significado dos conceitos que os agentes da sociedade multiagente necessitam entender para se comunicarem.

A tarefa de projeto arquitetural é composta da modelagem da sociedade multiagente, modelagem das interações dos agentes, da modelagem das atividades e da modelagem da coordenação e cooperação. Esta tarefa objetiva desenvolver um modelo arquitetural, que represente uma solução ao problema especificado no modelo de domínio (GIRARDI, LINDOSO. 2006).

A modelagem da sociedade multiagente identifica e representa, a partir do modelo de papéis, do modelo de interações entre os papéis, do Modelo de Domínio, os agentes que irão povoar a sociedade multiagente, bem como as destrezas que cada agente necessitará para executar a responsabilidade associada. Nesta etapa aspectos de coesão funcional são considerados, a fim de se determinar se um agente terá um ou mais papéis. A modelagem das interações dos agentes identificará os inter-relacionamentos entre os agentes da sociedade multiagente. A modelagem das atividades dos agentes visa especificar as atividades a serem realizadas por cada agente, o conhecimento necessário para a realização dessa atividade, os conhecimentos produzidos através da atividade, as destrezas e as condições a serem atendidas (COSTA, 2006). A modelagem da coordenação e cooperação é desenvolvida a partir da organização da sociedade multiagente, com base no modelo da sociedade multiagente, do modelo de interações entre agentes, de padrões arquiteturais e de adequados mecanismos de cooperação e de coordenação (GIRARDI, LINDOSO. 2006).

A tarefa de projeto de agentes é composta da modelagem do conhecimento do agente e da modelagem do estado do agente e tem a finalidade de detalhar o projeto arquitetural.

A modelagem do conhecimento do agente é realizada para cada agente, com a seleção de padrões adequados e estruturação dos agentes de acordo com a solução proposta. A modelagem do estado do agente apresenta os possíveis

estados e transições dos estados, a partir do modelo da sociedade multiagente, do modelo de interações entre os agentes e do modelo de coordenação e cooperação.

Tabela 4 Insumos, tarefas e produtos da fase DD, da MADEM

Fases	Insumos	Tarefas	Produtos		
Projeto de Domínio	Modelo de Conceitos	Modelagem do Conhecimento da Sociedade Multiagente	Modelo do Conhecimento da Sociedade Multiagente		
	Modelo de Interações entre Papéis		Modelagem Arquitetural	Modelo da Sociedade Multiagente	Modelo do Framework Multiagente
	Modelo de Papéis	Modelagem da Sociedade de Agentes		Modelo de Interações entre Agentes	
	Modelo da Sociedade Multiagente	Modelagem das Interações entre Agentes		Modelos de Atividades	
	Modelo da Sociedade Multiagente	Modelagem das Atividades		Modelo dos Mecanismos de Cooperação e Coordenação	
	Modelos de Estados dos Agentes	Modelagem dos Mecanismos de Cooperação e Coordenação		Modelos do Conhecimento do Agente	
	Modelo de Objetivos (Requisitos Não-Funcionais)	Projeto do Agente			
	Modelo da Sociedade Multiagente			Modelagem do Conhecimento do Agente	
	Modelo das Interações entre Agentes	Modelagem dos Estados dos Agentes			
	Modelo de Interações entre Papéis				
	Modelo do Conhecimento da Sociedade Multiagente				
	Modelo da Sociedade Multiagente				
	Modelo de Interações entre Agentes				
	Modelos de Atividades				

3.1.3 A Fase Implementação de Domínio

Na fase de implementação os agentes de projeto são mapeados para agentes de implementação, as responsabilidades para comportamentos (JADE Behaviour) e as interações para atos de comunicações (Tabela 5).

Tabela 5 Insumos, tarefas e produtos da fase DI, da MADEM

Fases	Insumos	Tarefas	Produtos	
Implementação de Domínio	Modelos de Atividades	Modelagem de Comportamentos	Modelo de Comportamentos	Modelo de Implementação da Sociedade Multiagente
	Modelo de Interações entre Agentes	Modelagem de Atos de Comunicação	Modelo de Atos de Comunicação	
	Modelo de Comportamentos	Implementação dos Agentes	Agentes de Software Executáveis	
	Modelo de Atos de Comunicação			

A fase Implementação de Domínio é apoiada pela técnica DIMAS (GIRARDI, MARINHO, 2007), que mapeia o Framework da fase de Projeto de

Domínio para atos de comunicação, comportamentos e agentes reativos ou deliberativos, com o uso do Framework JADE, o qual é uma plataforma para desenvolvimento de agentes. No caso dos agentes do tipo deliberativo foi utilizado o JESS, para proporcionar um mecanismo de inferência. O mapeamento destes conceitos de projeto para a implementação é descrito na Tabela 6.

Tabela 6 Mapeamento dos Conceitos de Projeto para implementação no JADE

	Conceitos de Projeto	Conceitos JADE
Agente Reactivo	1 Agente	1 Classe Agente
	1 Responsabilidade para cada agente	1 Classe JADE Behaviour
	1 atividade para cada agente	1 método do JADE Behaviour
	1 Modelo de Conhecimento da Sociedade Multi-agente	1 Ontologia JADE (Compartilhada por todos os agentes da sociedade)
	1 Modelo de Conhecimento de cada agente	1 Atributo da classe JADE
	Conceitos de Projeto	Conceitos JADE/JESS
Agente Deliberativo	1 Agente	1 Classe Agente
	1 Responsabilidade para cada agente	1 Classe JADE Behaviour + arquivo JESS
	1 atividade para cada agente	Regras JESS no arquivo JESS
	1 Modelo de Conhecimento da Sociedade Multi-agente	1 Ontologia JADE (Compartilhada por todos os agentes da sociedade)
	1 Modelo de Conhecimento de cada agente	Fatos JESS no arquivo JESS

3.1.4 A Fase Extração e Representação de Padrões

Encerrando a metodologia MADEM tem-se a fase de extração e representação de padrões, de forma a se identificar padrões: boas soluções a problemas recorrentes úteis para sua utilização no desenvolvimento de sistemas multiagentes (Tabela 7).

Tabela 7 Insumos, tarefas e produtos da fase PER, da MADEM

Fases	Insumos	Tarefas	Produtos
Extração e Representação de Padrões	Experiências de Desenvolvimento e outras fontes disponíveis	Extração e Representação de Padrões	Padrões de Sistema e de Software

produto “Especificação dos Requisitos da Aplicação”. Ou seja, a definição do problema a ser resolvido.

Tabela 8 Fase Eng. Requisitos da Aplicação

Fases	Insumos	Tarefas	Produtos	
Engenharia dos Requisitos da Aplicação	Modelo de Conceitos (MADEM) Conhecimento de Especialistas, Literatura no Domínio, Aplicações de Software Existentes	Modelagem de Conceitos	Modelo de Conceitos	Especificação dos Requisitos da Aplicação
	Modelo de Objetivos (MADEM) Requisitos da família de aplicações	Modelagem de Objetivos	Modelo de Objetivos	
	Modelo de Papéis (MADEM) Modelo de Objetivos	Modelagem de Papéis	Modelo de Papéis	
	Modelo de Interações entre Papéis (MADEM) Modelo de Objetivos	Modelagem de Interações entre Papéis	Modelo de Interações entre Papéis	
	Modelo de Papéis	Prototipagem da Interface com o usuário	Protótipo da Interface com o usuário	
	Protótipo da Interface com o usuário (MADEM)			
	Modelo de Papéis			
	Modelo de Interações entre Papéis			

3.2.2 Fase Projeto da Aplicação

Na fase de Projeto da Aplicação o desenvolvedor irá reutilizar algumas soluções de projeto relacionadas a uma família de aplicações e adaptá-la aos requisitos específicos da aplicação em desenvolvimento. Essa fase consiste de três tarefas: Modelagem do Conhecimento da Sociedade Multiagente, a qual representa os conceitos compartilhados por todos os agentes em sua comunicação; o Projeto Arquitetural, que estabelece um modelo arquitetural da sociedade multiagente incluindo seus mecanismos de coordenação e cooperação; e o Projeto do Agente, que define o projeto interno de cada agente da sociedade, modelando sua estrutura de conhecimento e seus comportamentos.

A fase Projeto da Aplicação (Tabela 9) possui como insumos os produtos gerados pela fase Projeto de Domínio, da metodologia MADEM, representados no Modelo do framework multiagente e complementados pelos da fase de Engenharia dos Requisitos da Aplicação, da MAAEM. A saída desta fase é o produto Arquitetura da Aplicação. Ou seja, uma solução computacional ao problema definido na fase Engenharia dos Requisitos da Aplicação.

Tabela 9 Fase Projeto da Aplicação

Fases	Insumos	Tarefas	Produtos		
Projeto da Aplicação	Modelo do Conhecimento da Sociedade Multiagente (MADEM)	Modelagem do Conhecimento da Sociedade Multiagente	Modelo do Conhecimento da Sociedade Multiagente		
	Modelo de Conceitos				
	Modelo de Interações entre Papéis				
	Modelo da Sociedade Multiagente (MADEM)	Modelagem Arquitetural	Modelagem da Sociedade de Agentes	Modelo da Sociedade Multiagente	Modelo Arquitetural
	Modelo de Papéis				
	Modelo de Interações entre Agentes (MADEM)		Modelagem das Interações entre Agentes	Modelo de Interações entre Agentes	
	Modelo da Sociedade Multiagente				
	Modelos de Atividades (MADEM)		Modelagem das Atividades	Modelos de Atividades	
	Modelo da Sociedade Multiagente				
	Modelos de Estados				
	Modelo dos Mecanismos de Cooperação e Coordenação (MADEM)		Modelagem dos Mecanismos de Cooperação e Coordenação	Modelo dos Mecanismos de Cooperação e Coordenação	
	Modelo de Objetivos (Requisitos Não-Funcionais)				
	Modelo da Sociedade Multiagente				
	Modelo das Interações entre Agentes				
	Modelos do Conhecimento do Agente (MADEM)	Projeto do Agente	Modelagem do Conhecimento do Agente	Modelos do Conhecimento do Agente	Modelos do Agente
	Modelo de Interações entre Papéis				
	Modelo do Conhecimento da Sociedade Multiagente				
	Modelos de Estados do Agente (MADEM)				
Modelo da Sociedade Multiagente					
Modelo de Interações entre Agentes	Modelagem dos Estados dos Agentes		Modelos de Estados do Agente		
Modelos de Atividades					

3.2.3 Fase Implementação da Aplicação

Na fase de Implementação da Aplicação, os agentes são identificados a partir do modelo de atividades da fase anterior, sendo associado a cada uma de suas responsabilidades comportamentos em uma determinada linguagem/plataforma de desenvolvimento de agentes, como o JADE (BELLIFEMINE, CAIRE, et al.. 2003). São ainda identificadas as interações entre agentes presentes no modelo de interações entre agentes, sendo estas mapeadas para uma determinada linguagem de comunicação entre agentes, como a FIPA_ACL (GREENWOOD, 2004). O reuso, se dá através da complementação do código gerado, para a qual o programador utiliza agentes de software anteriormente implementados na fase de Implementação do Domínio. A seleção dos agentes ocorre conforme a semelhança de comportamentos e comunicações. A adaptação torna-se necessária em função dos detalhes que diferem de uma aplicação para outra.

Na fase Implementação da Aplicação (Tabela 10), a entrada consiste nos produtos gerados pela fase “Implementação de Domínio”, da MADEM,

representados no Modelo de Implementação complementados pelos da fase “Arquitetura da Aplicação”, da MAAEM. A saída desta fase é o produto “Modelo de Implementação da Aplicação”. Ou seja, a solução computacional multiagente específica.

Tabela 10 Fase Implementação da Aplicação

Fases	Insumos	Tarefas	Produtos		
Implementação da Aplicação	Modelo de Comportamentos (MADEM)	Modelagem de Comportamentos	Modelo de Comportamentos	Modelo de Implementação da Aplicação	
	Modelos de Atividades		Modelo de Comportamentos		
	Modelo de Atos de Comunicação (MADEM)	Modelagem de Atos de Comunicação	Modelo de Atos de Comunicação		
	Modelo de Interações entre Agentes		Modelo de Atos de Comunicação		
	Agentes de Software Executáveis (MADEM)	Implementação dos Agentes	Agentes de Software Executáveis		
	Modelo de Comportamentos				
Modelo de Atos de Comunicação					

3.3 Considerações finais

A Engenharia de domínio e de aplicações, são processos complementares, cujo relacionamento foi mostrado neste capítulo, dentro das metodologias multiagente que serão utilizadas (MADEM e MAAEM), a fim de que na primeira seja desenvolvido um sistema de recomendação genérico *para o reuso*, cujos artefatos de softwares sejam reutilizados *com o reuso*, no caso específico a ser utilizado nos experimentos.

4. ONTOSERS: UMA FAMÍLIA DE SISTEMAS DE RECOMENDAÇÕES BASEADA NA TECNOLOGIA DA WEB SEMÂNTICA

Este capítulo apresenta o desenvolvimento do ONTOSERS, uma família de sistemas de recomendações baseadas na tecnologia da Web Semântica. Uma família de sistemas é definida como um conjunto de sistemas de software existentes que compartilham algumas características comuns e cada um tem algumas características particulares (CZARNECKI, EISENECKER, 2000). A modelagem foi realizada conforme as diretrizes da metodologia MADEM, apresentada no capítulo anterior.

Na seção 4.1 é tratada a fase Análise de Domínio do ONTOSERS (ONTOSERS-DM). Na seção 4.2 é abordada a fase Projeto de Domínio do Sistema Multiagente (ONTOSERS-DD). Na seção 4.3 é apresentada a fase Implementação de Domínio (ONTOSERS-DI).

4.1 Análise de Domínio da ONTOSERS-DM

Esta seção descreve o desenvolvimento das tarefas, bem como seus respectivos produtos, da fase Análise de Domínio, da metodologia MADEM, para a construção do modelo de domínio ONTOSERS-DM. O ONTOSERS-DM especifica os requisitos comuns e variáveis de uma família de artefatos de software para o desenvolvimento de sistemas multiagente de recomendações que utilizem técnicas de filtragem de informação e tecnologias da Web Semântica.

4.1.1 A Modelagem de Conceitos

O Modelo de Conceitos da ONTOSERS-DM é ilustrado na Figura 5. Nela, “Sistemas de recomendações baseados na Web Semântica” produzem e fazem uso de uma “Ontologia de Domínio” que tem como instâncias “Modelos de Item de Informação” e “Modelos de Usuários”.

Uma “Recomendação Personalizada” pode ser construída baseada em: um “Modelo de Item de Informação”, uma representação interna de um “Item de

Informação“, um “Modelo de Usuário” e na representação interna de um “Perfil do Usuário“. Na filtragem baseada no conteúdo, a recomendação é baseada no conceito “Modelo de Item de Informação“. Na filtragem colaborativa a recomendação é baseada no conceito “Modelo de Grupos de Usuários com Perfis Similares“. Na filtragem híbrida a base são os dois conceitos anteriores.

Uma “Recomendação Personalizada” satisfaz um “Perfil de Usuário” e é composta por “Itens Filtrados”, os quais são entregues aos usuários.

No ONTOSERS-DM, os “Modelos de Itens de Informação baseados em Ontologias” são representados como “Recursos da Web Semântica”.

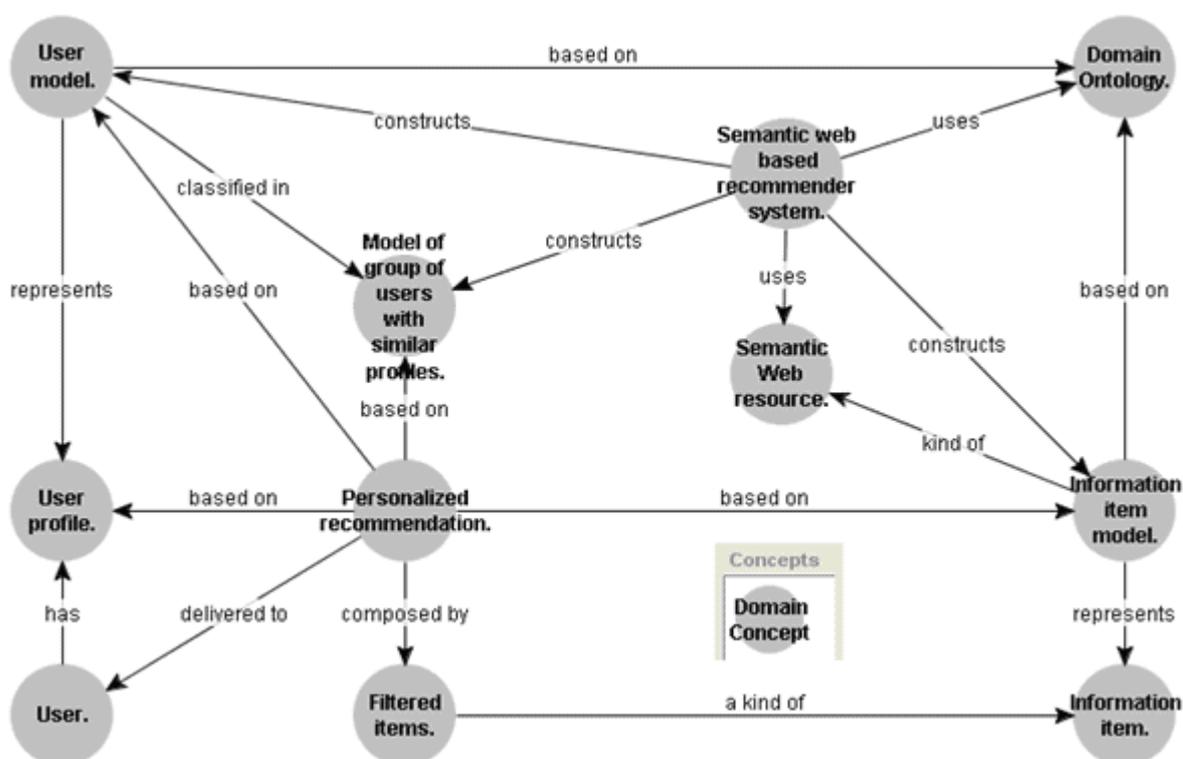


Figura 5 Modelo de Conceitos da ONTOSERS-DM

4.1.2 A Modelagem de Objetivos

A Figura 6 apresenta o Modelo de Objetivos do ONTOSERS-DM. O objetivo geral de “Prover recomendações Usando a Tecnologia da Web Semântica” é atingido através dos objetivos específicos “Modelar Usuários”, “Filtrar Informação” e “Entregar as recomendações”.

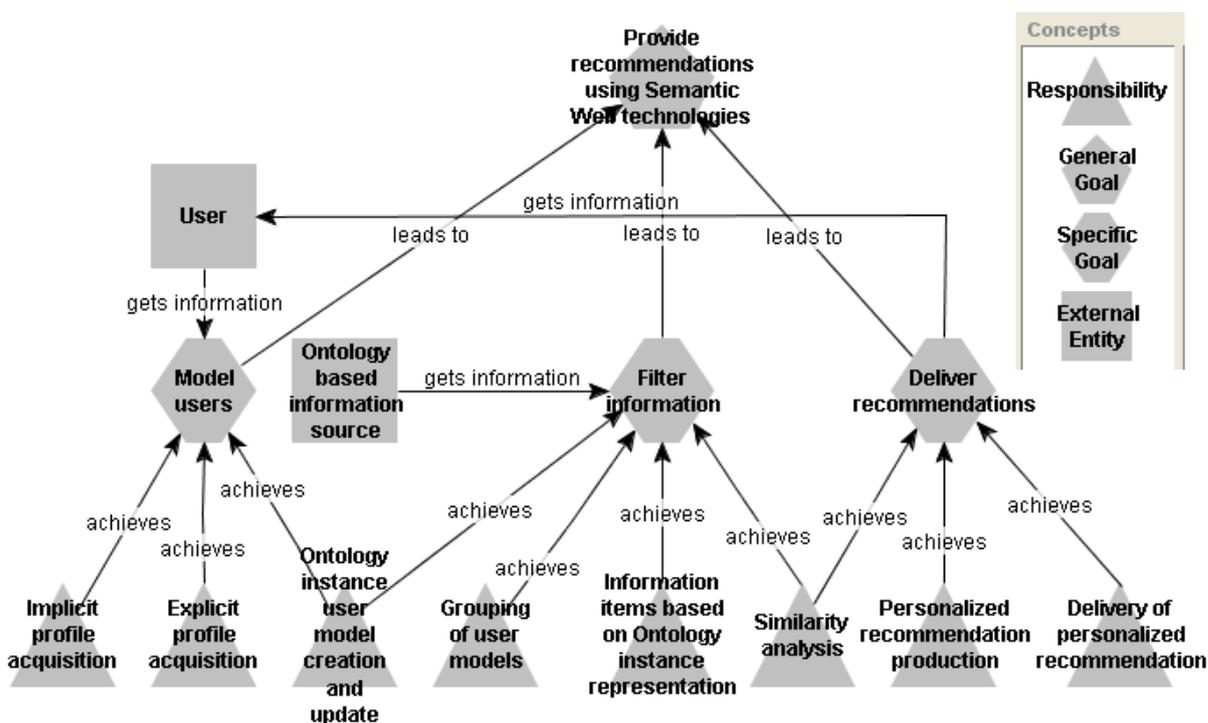


Figura 6 Modelo de Objetivos da ONTOSERS-DM

Para atingir o objetivo específico “Modelar Usuário” é necessária a realização das responsabilidades “Aquisição Implícita dos Perfis”, “Aquisição Explícita dos Perfis” e “Criação e Atualização dos Modelos de Usuários Baseados em Instâncias de Ontologias”. A “Aquisição Implícita do Perfil” corresponde à obtenção do perfil do usuário sem que os seus interesses sejam explicitamente especificados, o que configuraria a “Aquisição Explícita dos Perfis”, a qual é uma forma intrusiva de obtenção do perfil do usuário. A “Criação e Atualização dos Modelos de Usuários Baseados em Instâncias de Ontologias” corresponde à manutenção da representação dos modelos de usuários.

O objetivo específico “Modelar Usuários” tem um ponto de variação com grupo de responsabilidades para aquisição do perfil do usuário. Este ponto possui três variantes alternativas que correspondem às formas de adquirir o perfil do usuário: “Aquisição Implícita do Perfil”, “Aquisição Explícita do Perfil” ou ambas (Figura 7).

Para que o objetivo específico “Filtrar Informação” seja alcançado é preciso que se realize a responsabilidade “Criação e Atualização dos Modelos de Usuários em Instâncias de Ontologias”, que também contribui para atingir o objetivo

específico de “Modelar Usuário”. Além dessa, são necessárias as responsabilidades “Agrupamento dos Modelos de Usuários”, “Representação em Instâncias Ontológicas dos Itens de Informação” e “Análise de Similaridade”. A responsabilidade “Agrupamento dos Modelos de Usuários” permite identificar grupos de usuários com interesses similares. A “Representação em Instâncias Ontológicas dos Itens de Informação” permite representar os itens de informação numa estrutura capaz de ser processada pelos agentes de software. A responsabilidade “Análise de Similaridade” consiste em determinar a relevância de um item para um determinado usuário (MARIANO, GIRARDI, et al.. 2008).

Para alcançar o objetivo específico “Entregar recomendações” é pré-requisito que a responsabilidade “Análise de Similaridade” seja realizada, a fim de que haja itens de informação a serem entregues. Tendo sido realizado o pré-requisito há a execução das responsabilidades de “Produção de recomendações personalizadas” e “Entrega das recomendações Personalizadas”.

Em relação ao objetivo específico “Filtrar Informação” foi colocado um ponto de variação nele com as variantes alternativas nas responsabilidades: “Agrupar modelos de usuários”, “Representação dos Itens de Informação baseados em Instâncias de Ontologia”, e em um grupo representando ambas, relacionado à filtragem híbrida (Figura 7). A responsabilidade “Agrupar modelos de usuários” ocorre em sistemas que usam a filtragem colaborativa (FC). A responsabilidade “Representação dos Itens de Informação baseados em Instâncias de Ontologia” ocorre nos que usam filtragem baseada no conteúdo (FBC). As duas responsabilidades anteriores ocorrem nos sistemas que realizam a filtragem híbrida.

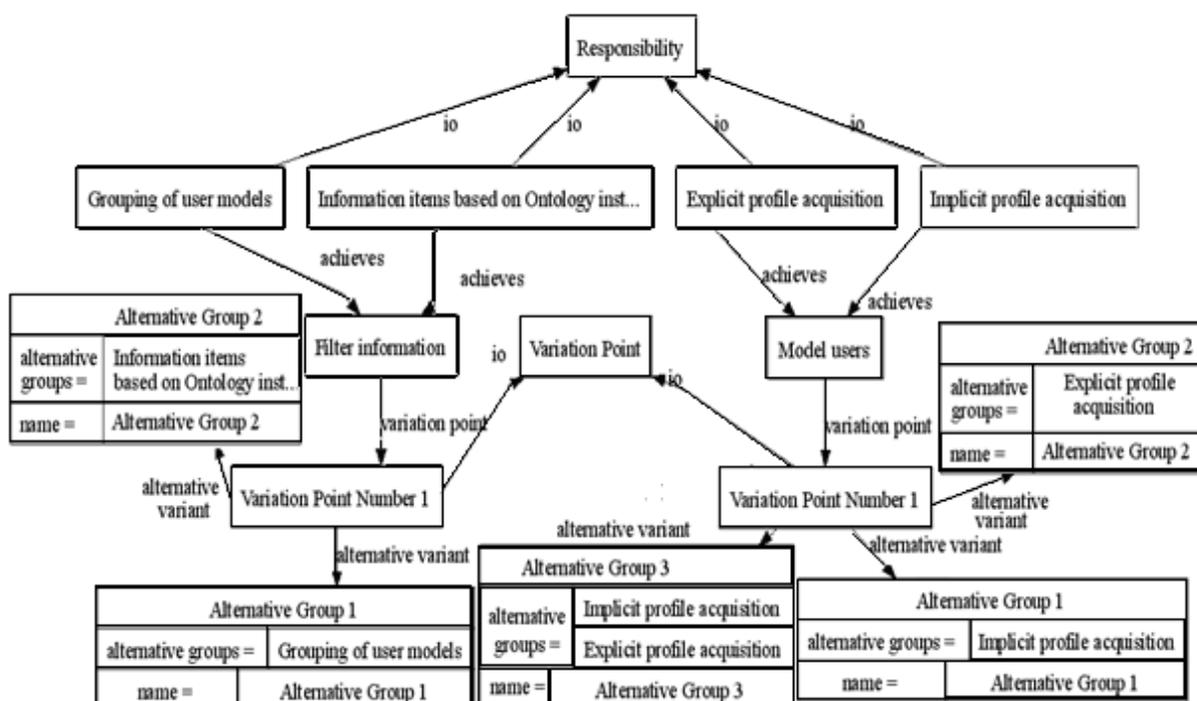


Figura 7 Pontos de Variabilidade e variantes no Modelo de Objetivos do ONTOSERS-DM

4.1.3 Modelagem de Papéis

Na Modelagem de Papéis cada uma das responsabilidades identificadas na Modelagem de Objetivos é associada a um papel. Para cada responsabilidade são identificadas suas pré e pós-condições, os conhecimentos requeridos e produzidos. As entidades externas e os conhecimentos associados a elas também são especificados neste modelo (MARIANO, GIRARDI, et al.. 2008).

4.1.3.1 Modelagem de Papéis Relacionada ao Objetivo Específico “Modelar Usuários”

A Figura 8 apresenta o Modelo de Papéis relacionado ao objetivo específico “Modelar Usuário”, que é atingido através do exercício das responsabilidades: “Aquisição Implícita do Perfil”, “Aquisição Explícita do Perfil” e “Criação e Atualização do modelo de usuário baseado em instâncias de ontologia”.

O papel “Monitor do Usuário” desempenha a responsabilidade “Aquisição Implícita do Perfil”. Esta responsabilidade tem como pré-condição “Usuário estar conectado”.

referendado para o grupo de modelos de usuário. A responsabilidade “Agrupamento dos Modelos de Usuários” tem como pós-condição “Grupos Disponíveis”.

A responsabilidade “Representação dos Itens de Informação em Instâncias da Ontologia” utiliza o conhecimento “Novo Item de Informação” e produz o conhecimento “Modelo do Item de Informação”, a partir do momento que a pré-condição “Novo Item de Informação Está Disponível” for satisfeita e deve satisfazer a pós-condição “Item de Informação está Representado”.

A responsabilidade “Análise de Similaridade” exercida pelo papel “Filtrador” utiliza os conhecimentos “Modelo do Item de Informação” e “Grupos de Modelos de Usuários”, a partir dos quais produz o conhecimento “Itens filtrados”.

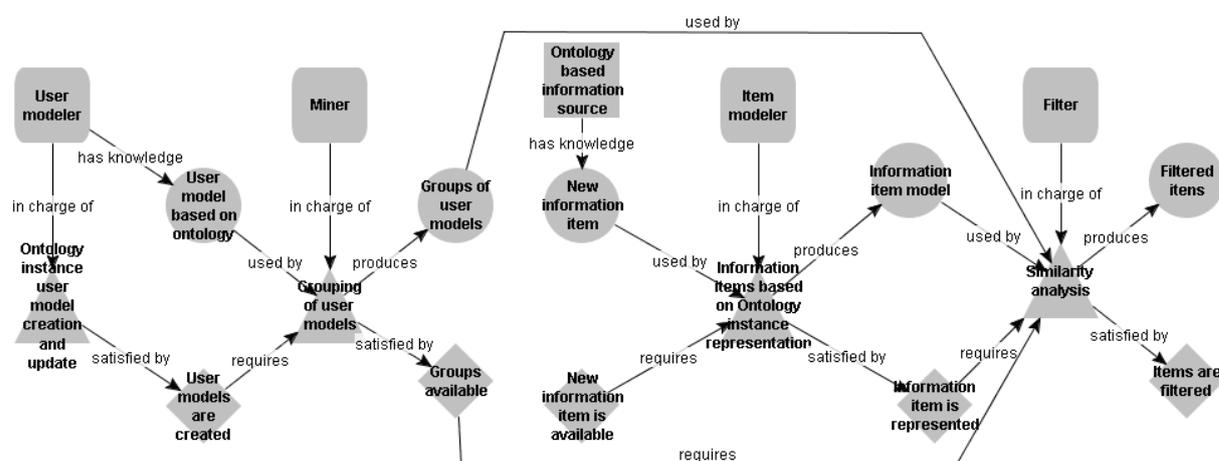


Figura 9 Modelo de Papéis do Objetivo Específico “Filtrar Informação”

4.1.3.3 A Modelagem de Papéis Relacionados ao Objetivo Específico “Entregar recomendações”

O objetivo específico “Entregar recomendações” é alcançado após a execução da responsabilidade “Análise de Similaridade”, executada no objetivo específico “Filtrar Informação”. Tendo ocorrido a responsabilidade “Análise de Similaridade” serão realizadas as responsabilidades “Produção de recomendações Personalizadas” e “Entrega das recomendações Personalizadas” (Figura 10).

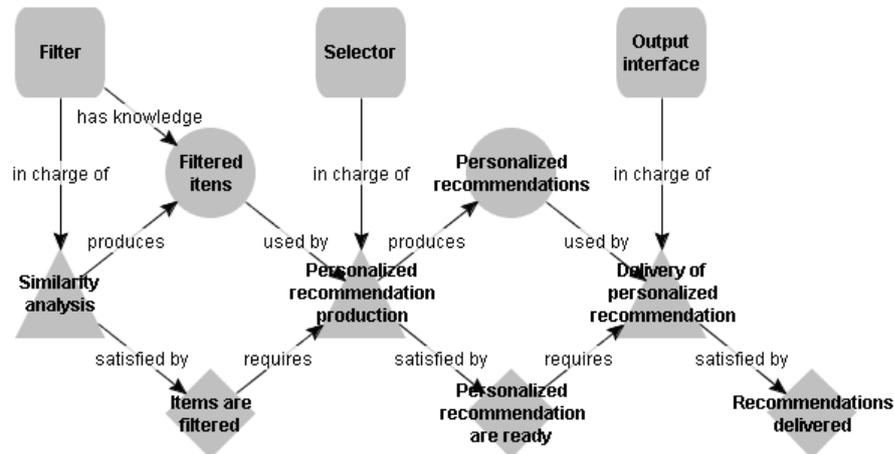


Figura 10 Modelo de Papéis do Objetivo Específico "Entregar Recomendação"

4.1.4 Modelagem de Interações entre Papéis

O Modelo de Interações entre Papéis descreve como cada papel irá se inter-relacionar com os demais e com as entidades externas para atingir um objetivo específico. Cada Modelo de Interações entre Papéis é relacionado a um objetivo específico (MARIANO, GIRARDI, et al.. 2008).

4.1.4.1 Modelagem de Interações entre Papéis referentes ao Objetivo Específico "Modelagem do Usuário"

No Modelo de Interação de Papéis do Objetivo Específico "Modelagem do Usuário", mostrado na Figura 11, tem-se que enquanto o usuário navega, o papel "Monitor de Usuário" captura o seu comportamento navegacional. Este perfil do usuário, adquirido implicitamente será repassado para o "Modelador de Usuário", para que ele possa realizar a criação do modelo de usuário. Outra forma de aquisição do modelo do usuário é a forma explícita, onde o usuário informa ao papel "Interface de entrada" o seu perfil e esta repassa para o "Modelador de Usuário" os dados recebidos.

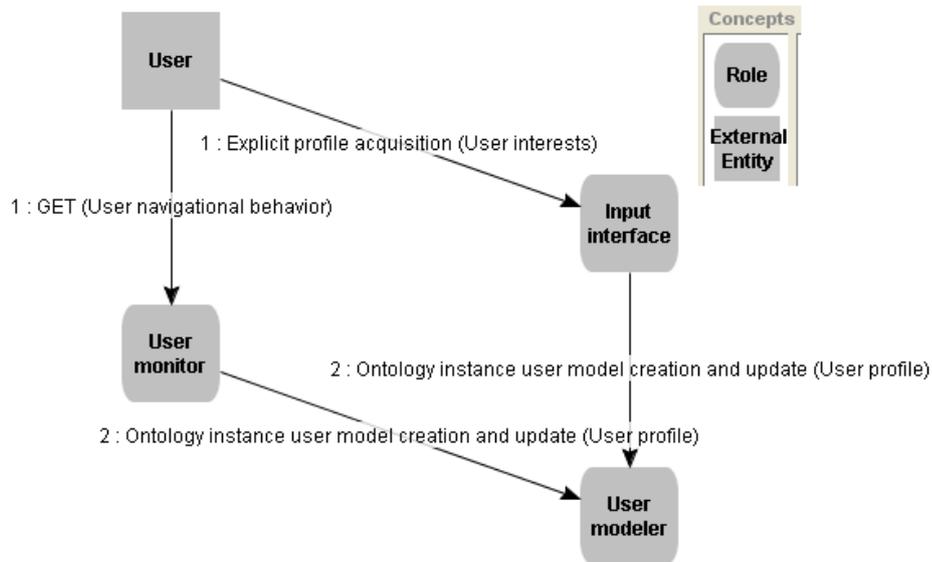


Figura 11 Modelo de Interações entre Papéis do Objetivo Específico “Modelar Usuário”

Durante a Modelagem de Interações entre Papéis foi criado o Ponto de Variação no Objetivo específico contendo como variantes grupos de interações entre papéis. Para o objetivo específico “Modelar Usuário”, que tem como Variantes Alternativas “Aquisição Implícita dos Perfis”, “Aquisição Explícita dos Perfis” e as duas ao mesmo tempo, foi criado um ponto de variação tendo como variantes um conjunto de interações entre papéis específicos para cada uma destas responsabilidades (Figura 12).

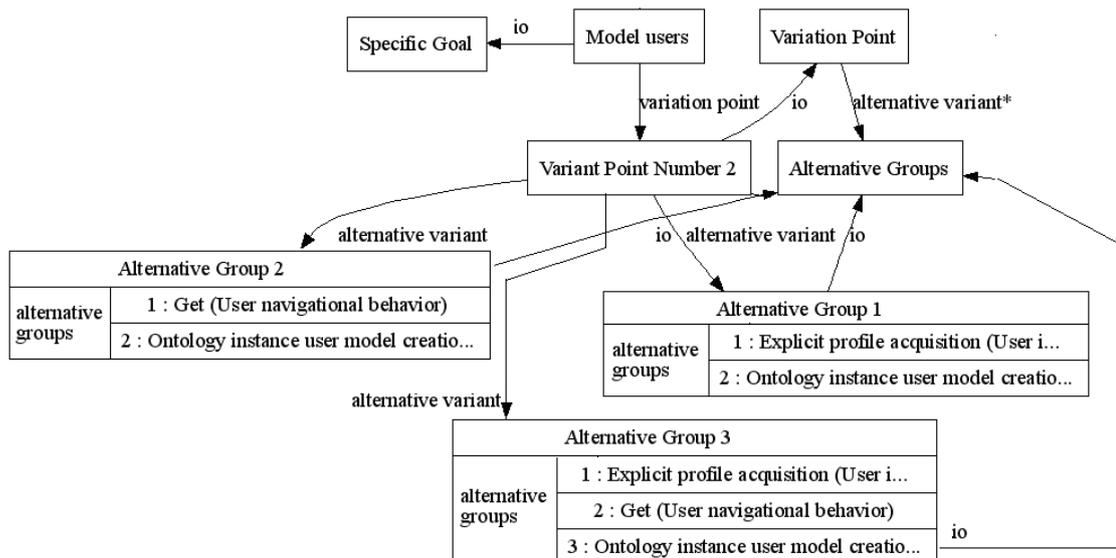


Figura 12 Modelagem de Variabilidade do Modelo de Interações entre Papéis

4.1.4.2 Modelagem de Interações de Papéis referentes ao Objetivo Específico “Filtrar Informação”

Nesta interação de papéis, para a filtragem baseada no conteúdo, Figura 13, o “Filtrador” realiza a responsabilidade “Análise de Similaridade” utilizando o “Modelo de Item de Informação” entregue pelo “Modelador de Item”. Na filtragem colaborativa, por outro lado, o “Filtrador” solicita ao “Minerador” o conhecimento “Grupos de Modelos de Usuários”.

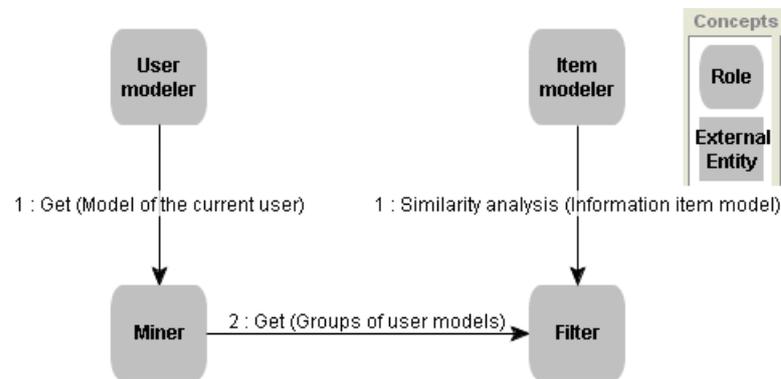


Figura 13 Modelo de Interações de Papéis do Objetivo Específico “Filtrar Informações”

O papel “Filtrador” possui um ponto de variação para indicar que ele pode realizar a filtragem de duas formas, como mostrado na Figura 14, uma baseada nos modelos de itens de informação e outra nos grupos de modelos de usuários.

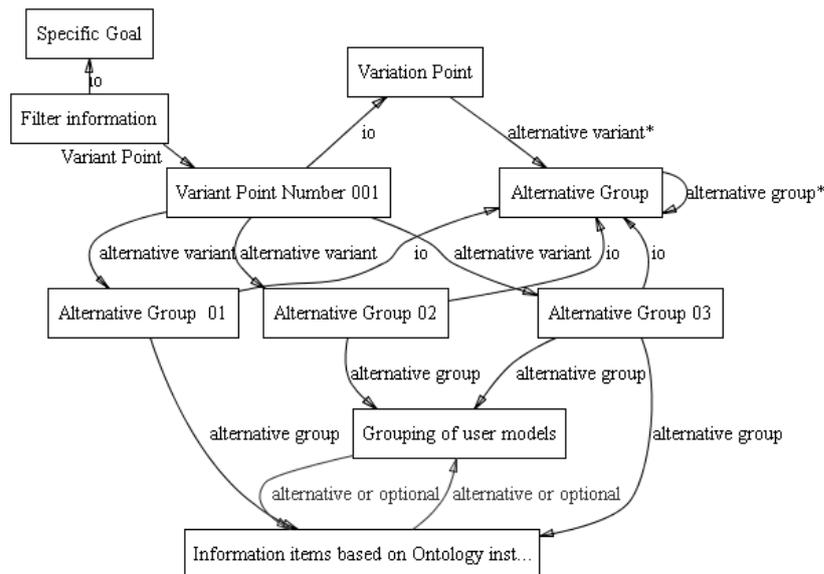


Figura 14 Variabilidade do Modelo de Interações entre Papéis do Objetivo Específico “Filtrar informação”

4.1.4.3 Modelagem de Interações de Papéis referentes ao Objetivo Específico “Entregar Recomendações”

Nesta modelagem, o papel “Interface de saída” entregará ao usuário, os itens personalizados repassados pelo papel “Seleto”, a partir dos itens filtrados (Figura 15).

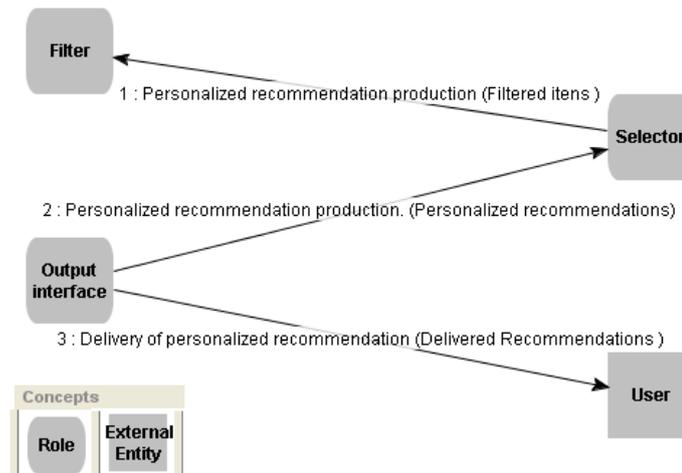


Figura 15 Modelagem de Interações entre Papéis do Objetivo Específico “Entregar recomendações”

4.1.5 Protótipo das Interfaces do Usuário

Da Figura 16 à Figura 20 são mostrados os Protótipos das Interfaces do usuário do ONTOSERS, o que corresponde à última etapa da Análise de Domínio, são interfaces genéricas que representam itens que estarão presentes nos vários artefatos da família de software criados (MARIANO, GIRARDI, et al.. 2008).

A interface de Login (Figura 16) permite a identificação do usuário, a partir da qual é possível o registro da navegação e adaptação dos links, de forma a personalizar o site. No contexto deste trabalho não existirá um usuário anônimo, uma vez que apenas usuários cadastrados poderão utilizar o site, cujo perfil é preenchido no momento do cadastramento do usuário.

Figura 16 Interface de Login

A Figura 17 mostra a interface utilizada para que o usuário especifique a sua identificação, com seu nome, e-mail, “login” e senha, permitindo que se consulte e altere os valores informados.

Figura 17 Interface de Identificação

A interface de usuário, aba “Categoria de Interesse”, Figura 18, permite especificar os interesses, os quais serão utilizados para gerar as recomendações.

Figura 18 Interface de Categorias de Interesse

Na interface “Recomendação Personalizada”, Figura 19, são listados os itens recomendados para o usuário com a respectiva similaridade.

Personalized recommendations	
Item	Similarity
Information Item 1	75.00% FH

Figura 19 Interface de Recomendação

Na Interface “Avaliação das recomendações” (Figura 20) o usuário realiza um feedback explícito, através do preenchimento da coluna avaliação, conforme

julgue que tenha sido menos ou mais relevante a recomendação. A partir de tais dados será possível melhorar as próximas personalizações.

Recommendations Evaluation		
Item	Evaluation	Similarity
Information Item 1	0	75.00%

Figura 20 Interface de Avaliação da Recomendação

4.2 Projeto de Domínio

Terminada a fase de Análise de Domínio, a qual teve por finalidade identificar e esclarecer o problema, tem-se a fase de Projeto de Domínio, a qual tem por objetivo a identificação de uma solução computacional aos requisitos especificados no modelo de domínio. Portanto nas seções seguintes irá se detalhar as tarefas e produtos da fase de Projeto de Domínio.

4.2.1 Modelagem do Conhecimento da Sociedade Multiagente

O propósito desta tarefa é especificar o conhecimento utilizado na comunicação dos agentes (GIRARDI, LINDOSO, 2005). O Modelo do Conhecimento da Sociedade Multiagente do ONTOSERS (Figura 21) informa que as recomendações personalizadas são baseadas no modelo do usuário e nos itens filtrados. Os itens filtrados são espécies dos itens de informação. O modelo de usuário representa o perfil do usuário, que é composto pela identificação do usuário e pelos seus interesses.

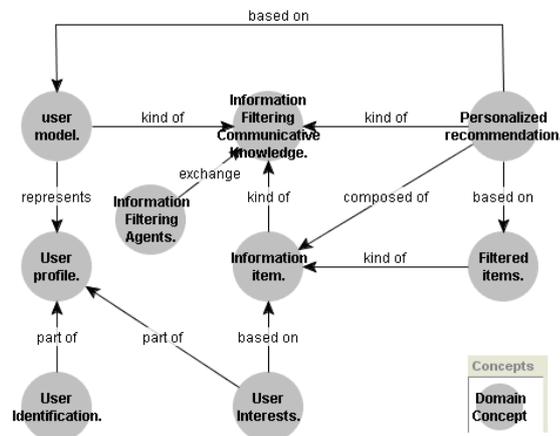


Figura 21 Modelo de Conhecimento da Sociedade Multiagente

4.2.2 Projeto Arquitetural

O propósito desta tarefa é desenvolver uma solução para o problema especificado no modelo de domínio, do ponto de vista arquitetural (GIRARDI, LINDOSO, 2005). Esta etapa é composta das tarefas: “Modelagem da Sociedade Multiagente”, “Modelagem das Interações dos Agentes”, “Modelagem das Atividades” e “Modelagem dos Mecanismos de Coordenação e Cooperação”.

4.2.2.1 Modelagem da Sociedade Multiagente

No Modelo da Sociedade Multiagente os papéis são agrupados em agentes, de acordo com os critérios de coesão funcional. Nesta tarefa, para cada responsabilidade foram definidas as destrezas associadas. Nos casos em que a responsabilidade executada pelo agente é associada a mais de uma destreza foi colocado um ponto de variabilidade na responsabilidade e cada destreza em um grupo, onde cada grupo é uma variante alternativa, de forma que uma delas terá de ser executada, a fim de que a responsabilidade seja realizada.

4.2.2.1.1 O Agente “Interface do Usuário” na Modelagem da Sociedade Multiagente

Verificou-se que os papéis “Monitor do Usuário”, “Interface de Entrada”, “Modelador do Usuário”, “Seletor” e “Interface de Saída”, da fase Análise de Domínio, relacionavam-se com o usuário sendo agrupados no agente “Interface de Usuário” (Figura 22).

Na identificação das destrezas relacionadas à cada responsabilidade foram utilizados os trabalhos de (PIERRAKOS, PALIOURAS et al., 2003), o qual apresenta:

- Na responsabilidade “Aquisição Explícita do perfil do usuário” foi identificada a destreza preenchimento de um formulário web, no qual o usuário expressa os seus interesses.
- Na responsabilidade “Aquisição Implícita do perfil do usuário” foram identificadas as destrezas:

- Uso de um Applet Java – que realiza o monitoramento da aplicação, por meio de pequenos programas embutidos nas páginas do site (SHAHABI, BANAEI-KASHANI, FARUQUE, 2001).
- Técnicas de armazenamento e análise de sessão – que armazenam dados da sessão do usuário e depois realizam a análise desta para identificar padrões (PIERRAKOS, PALIOURAS et al., 2003).
- Técnicas de armazenamento e análise de cookies – que armazenam os cookies (pequenos pedaços de informações) na máquina do usuário e depois realizam a análise desta para identificar padrões (PIERRAKOS, PALIOURAS et Al 2003).
- Técnicas de Aprendizagem de Máquina – que usam as técnicas de aprendizagem supervisionada para criar os agrupamentos (PETASIS, VICHOT, WOLINSKI 2001).
- Técnicas de Mineração de Dados – que usam as diversas técnicas de mineração, como as de uso, conteúdo e de link (FACCA, LANZI 2005) (PIERRAKOS et Al 2003).
- Técnicas do browser Mosaic – Usam as técnicas em que o Browser grava as páginas Web que foram visitadas por um usuário e envia elas juntamente com outras informações como tempo de acesso e tempo de retorno da resposta para o servidor (TAUSCHER, GREENBERG, 1997).
- Na responsabilidade “Criação e Atualização do Modelo de Usuário representado Ontologicamente” foram identificadas as destrezas:
 - Regras de Associação (FACCA, LANZI 2005) (VASCONCELOS, CARVALHO, 2004).
 - Classificação (FACCA, LANZI 2005).
 - Representação Ontológica do Perfil.
 - Inferência Ontológica do Perfil.
 - Técnicas de Agrupamento – Técnicas que dividem um conjunto em grupos que são distintos entre si e cujos

membros são similares (SHAHABI, BANAEI-KASHANI, 2003). Estas técnicas são técnicas de aprendizagem não supervisionada

- Na responsabilidade “Produção de Recomendação Personalizada” foi identificada a destreza “Técnicas de Seleção”, a qual se baseia em adaptar o item de informação ao meio que será utilizado para a entrega. Uma referência sumária da informação da existência de um item recomendado poderá ser encaminhada para o e-mail, o celular, etc.
- Na responsabilidade “Entrega de Recomendação Personalizada” foi identificada a destreza “Interface de Saída para entrega dos itens de informações filtrados”. Nesta responsabilidade é definido o momento da entrega, o qual pode ser no primeiro acesso ao site, semanalmente, etc.

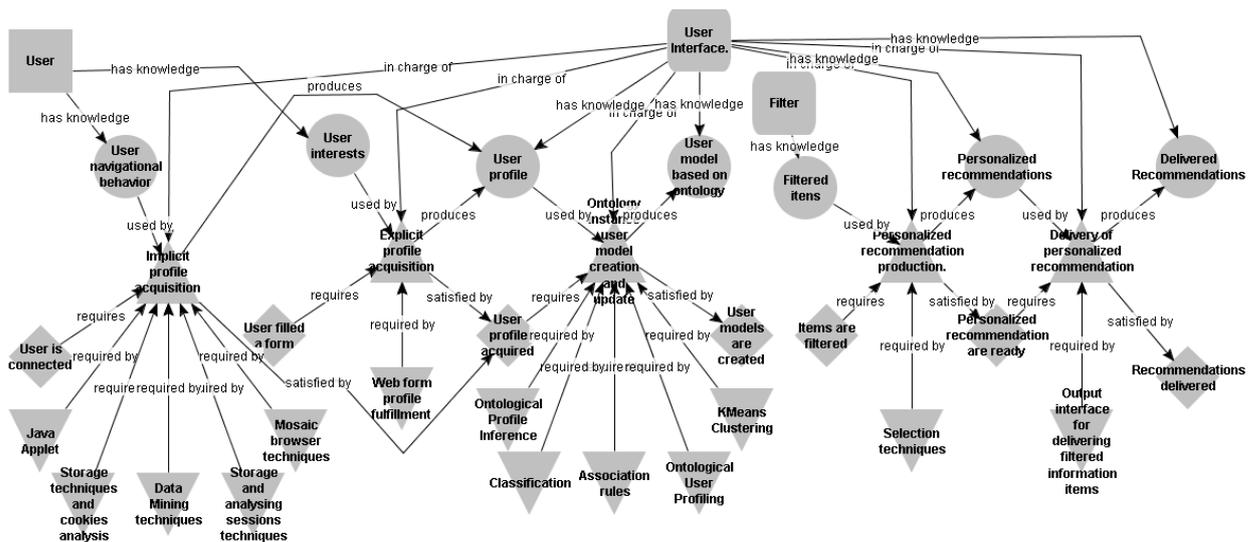


Figura 22 Modelo da Sociedade Multiagente relacionadas ao agente “Interface do Usuário”

Na Figura 23 é mostrado o ponto de variabilidade colocado no agente “Interface do Usuário”, o qual tem as variantes alternativas: “Aquisição Explícita do Perfil”, “Aquisição Implícita do Perfil” e ambas.

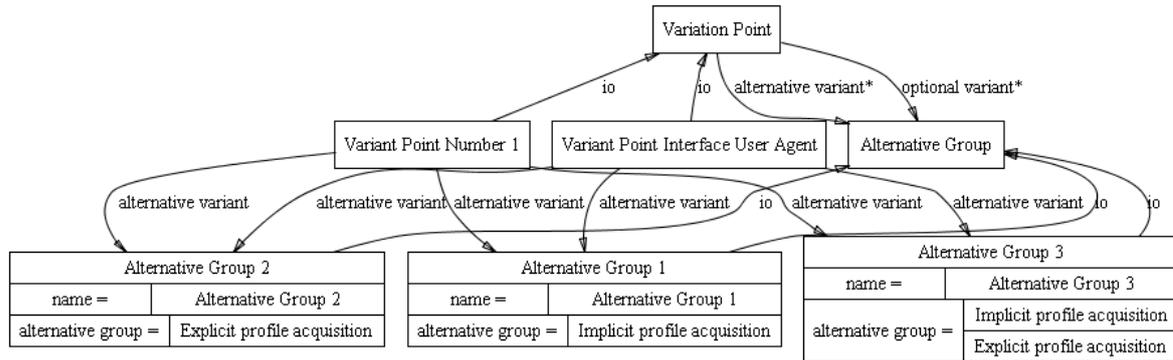


Figura 23 Variabilidade relacionada ao Agente “Interface do Usuário”

4.2.2.1.2 O Agente “Filtrador” na Modelagem da Sociedade Multiagente

As responsabilidades “Agrupamento dos modelos de Usuários” (papel “Minerador”), “Representações dos itens de Informações baseados em Instâncias de Ontologia” (papel “Modelador de Itens”) e “Análise de Similaridade” (papel “Filtrador”) serão realizadas pelo agente “Filtrador” (Figura 24). Este agente terá a função de formar os agrupamentos dos diversos perfis do usuário, de forma a identificar usuários com características comuns. Na responsabilidade “Agrupamento dos modelos do usuário” será utilizado o conhecimento “Modelo de usuário baseado em ontologia” e as destrezas:

- Técnicas de “Agrupamento (Clustering) e Mineração (Mining)” – onde Clustering é o processo de agrupamento dinâmico de um conjunto de dados em classes similares chamadas “Clusters” e que são diferentes dos objetos dos outros “Clusters” (BREESE, HECKERMAN, KADIE, 1998).
- Técnicas de “Identificação de Comunidade de Práticos Baseados em Ontologias” – Visa identificar comunidades (grupos de usuários com interesses com um interesse particular), cujas características sejam transversais ao domínio representado na Ontologia. Por exemplo, nos agrupamentos de usuários jurídicos, tem-se juízes, procuradores, advogados, promotores, todos têm características específicas, mas pode existir uma comunidade de usuários que agregue determinados juízes, procuradores, advogados e promotores que tenham interesses nos instrumentos normativos

4.2.2.2 Modelagem das Interações dos Agentes

Definido o que cada agente faz na Sociedade Multiagente tem-se de especificar como interagem para atingir os seus objetivos. As mensagens são especificadas de acordo com o protocolo FIPA/ACL (GREENWOOD, 2004).

No Modelo de Interações entre Agentes da Figura 25, o agente "Interface do Usuário" realiza as interações com o usuário, através da captura do comportamento navegacional e das necessidades de informação de longo prazo e envia para o usuário as recomendações personalizadas.

Quando o agente "Filtrador" tem conhecimento de um novo item de informação, ele irá propor ao agente Interface de usuário através de uma diretiva PROPOSE, a análise de similaridade do novo item de informação com o modelo do usuário, o qual aceitando retornará com a diretiva ACCEPT. Sendo entregue através da performativa INFORM_REF os itens filtrados, os quais serão repassados para o usuário, pelo agente Interface do usuário.

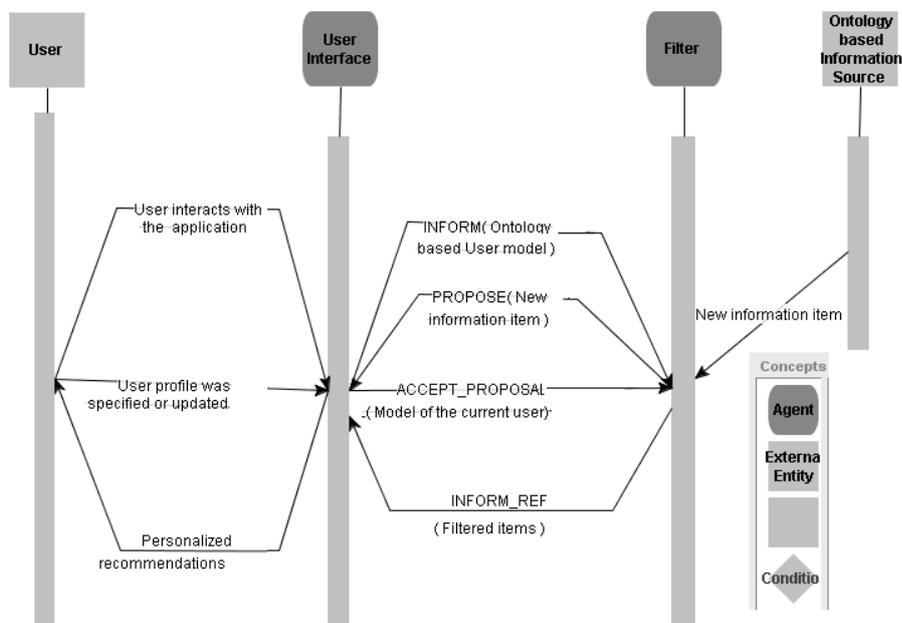


Figura 25 Modelo de Interação dos Agentes

4.2.2.3 Modelagem das Atividades

Esta modelagem tem o detalhamento de como os agentes exercem as responsabilidades e como elas se correlacionam, em termos das atividades, o que é mostrado na Figura 26.

De forma geral, o agente "Interface do Usuário", no exercício da responsabilidade "Aquisição Explícita do Perfil", necessita executar a atividade "Extrair os dados do formulário preenchido pelo usuário", a fim de criar o perfil do usuário, o qual será utilizado pela responsabilidade "Criação e atualização do modelo de usuário ontologicamente", na atividade "Criação e atualização do modelo de usuário". Esta atividade também é executada, quando do exercício da atividade "Aquisição do perfil de usuário a partir do comportamento", cuja atividade anterior foi a "Monitorar o comportamento do Usuário", na responsabilidade "Aquisição Implícita do Perfil".

O agente Filtrador executa a atividade "Agrupamento dos modelos de usuário", quando o modelo de usuário estiver disponível, pela execução da atividade "Criação e atualização do modelo de usuário ontologicamente", do agente "Interface do Usuário".

O agente Filtrador, na responsabilidade e atividade "Análise de Similaridade colaborativa", gera os itens filtrados, os quais são encaminhados ao agente "Interface de usuário", para execução das responsabilidades de "Produção da Personalização da Recomendação", na qual são executadas as atividades "Definição do Quando Entregar", "Definição do Como Entregar", para finalmente executar a atividade e responsabilidade "Entrega da Recomendação Personalizada". Outras fontes de recomendação estão baseadas nas filtragens baseadas no conteúdo e híbrida, para o que são necessário que sejam executadas as atividades "Monitoramento da Fonte de Informação" e "Representação do novo item de informação, como instância da ontologia", pelo agente "Interface do Usuário", bem como a atividade e responsabilidade "Análise de Similaridade baseada no conteúdo" e "Análise de Similaridade híbrida", pelo agente Filtrador.

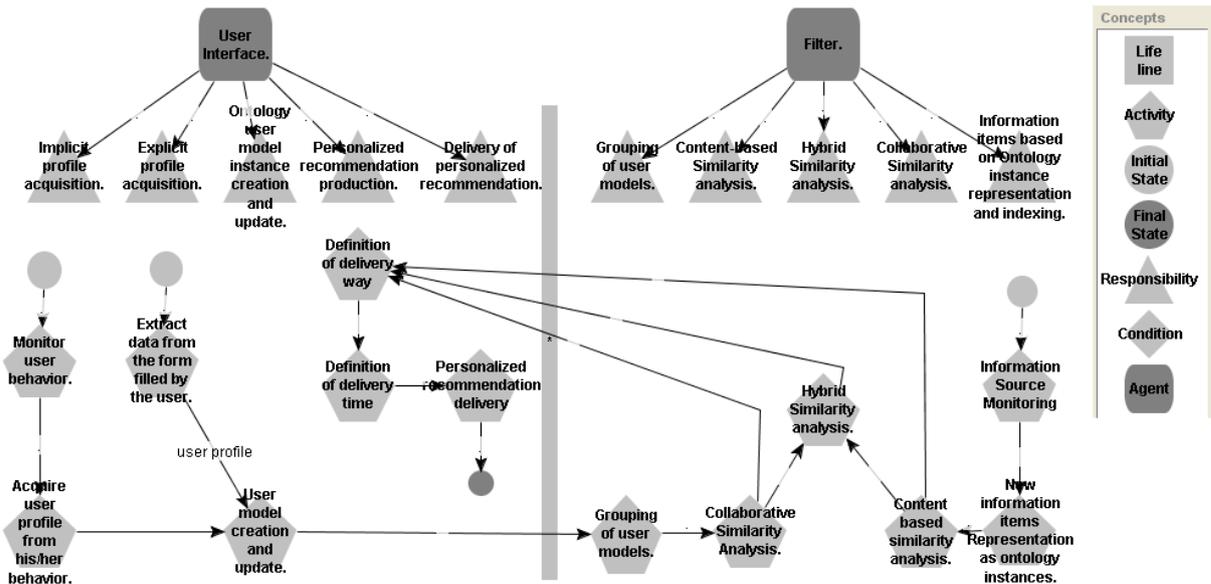


Figura 26 Modelo de Atividade

4.2.2.4 Modelagem dos Mecanismos de Coordenação e Cooperação

Não é foco deste trabalho a discussão do impacto das configurações do mecanismo de coordenação e cooperação, no desempenho das recomendações. Escolheu-se o modelo de 2 camadas (Figura 27).

Para uma noção sobre as diferentes configurações há os trabalhos de (SILVA JR., 2003) e (OLIVEIRA, 2004), onde se identificam diversos contextos associados a cada configuração.

A escolha pelo modelo de 2 camadas é em função deste modelo permitir isolar o desenvolvimento do conteúdo de cada camada, a qual precisa saber apenas: os agentes externos, com os quais irá se relacionar; as interfaces das camadas imediatamente superior e inferior e os respectivos agentes que compõem a camada.

Para o ONTOSERS-DD foi definida a camada “Processamento do Usuário”, a qual faz o front-end com o usuário e que tem o agente “Interface do Usuário” e a camada “Processamento da Informação”, a qual possui o agente “Filtrador” e que monitora a “Fonte de informação baseada em Ontologia”.

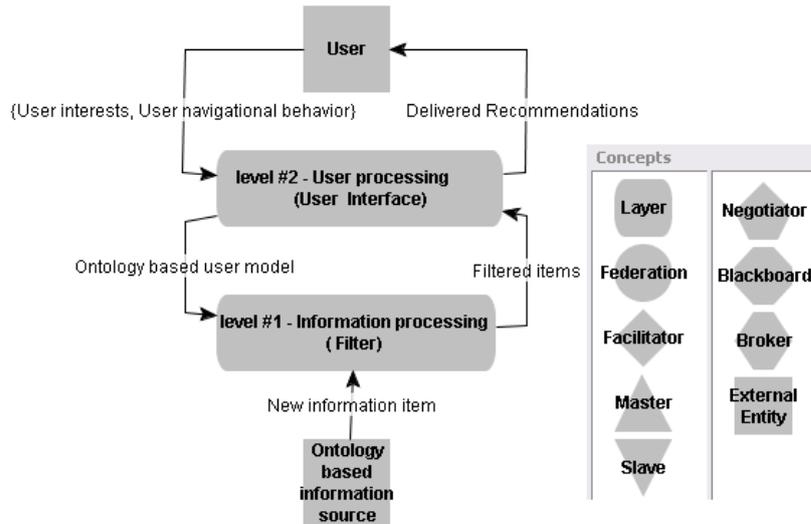


Figura 27 Modelo do Mecanismo de Coordenação e Cooperação

4.2.3 Projeto dos Agentes

Um agente de software é uma entidade computacional com as características de: autonomia, sociabilidade, percepção e raciocínio. Para isto é necessário definir o conhecimento de cada agente e os estados deles, o que é formalizado de forma genérica, a fim de se terminar a fase de Projeto de Domínio, do ONTOSERS.

4.2.3.1 Modelagem do Conhecimento do Agente "Interface do Usuário"

Na Figura 28 o agente "Interface do Usuário" tem o conhecimento "Perfil do Usuário", que possui o conhecimento "Comportamento navegacional do usuário" e "Interesses do Usuário". O "Perfil do Usuário" é representado como um "Modelo de Usuário baseada em Ontologia". As recomendações personalizadas satisfazem ao perfil do usuário (Figura 28).

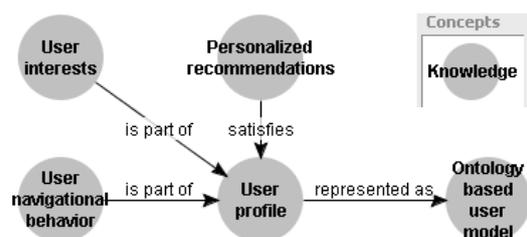


Figura 28 Modelo de Conhecimento do Agente "Interface do Usuário"

A variabilidade no Modelo de Conhecimento do agente “Interface do Usuário” é representada pelo ponto de variabilidade colocado sobre o conhecimento “Perfil do Usuário”, que tem como variantes os conhecimentos “Interesses do Usuário” e “Comportamento Navegacional do Usuário” (Figura 29).

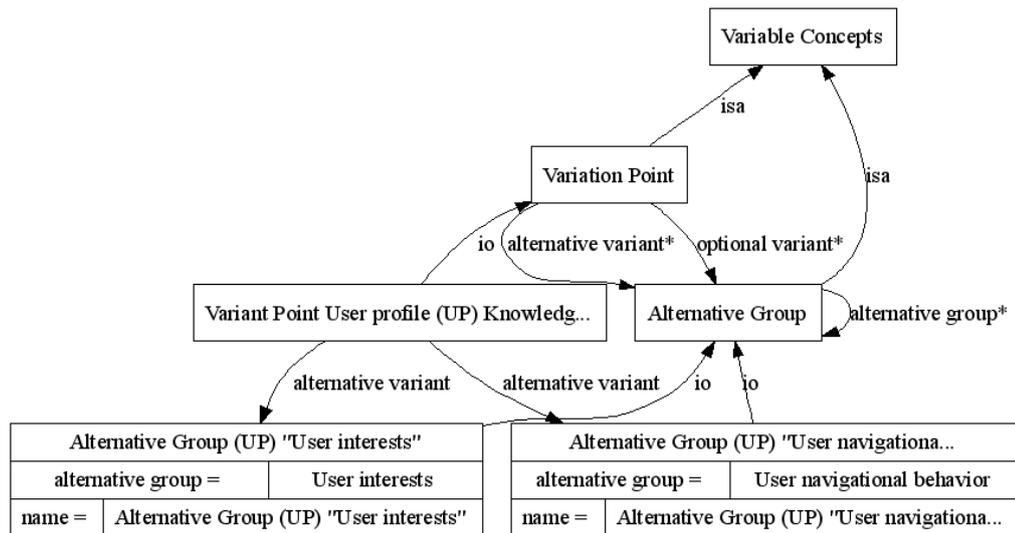


Figura 29 Variabilidade do Modelo de Conhecimento do Agente “Interface do Usuário”

4.2.3.2 Modelagem de Conhecimento do Agente “Filtrador”

O Modelo de Conhecimento do agente Filtrador utiliza o conhecimento “Itens Filtrados”, o qual é baseado no “Modelo de Item de Informação” e nos “Grupos de modelos de usuário”. Este é um tipo de “Novo item de informação”, o qual é uma espécie de item filtrado (Figura 30).

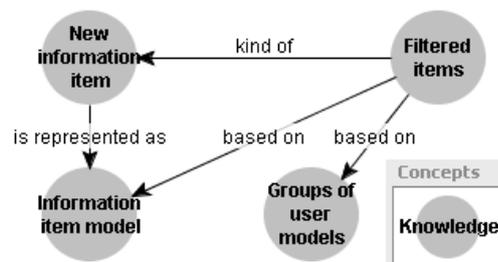


Figura 30 Modelo de Conhecimento do Agente “Filtrador”

4.2.3.3 Modelagem do Estado do Agente “Interface do Usuário”

A Modelagem do Estado do Agente tem por finalidade representar os estados e as transições de estados pelos quais um agente passa durante o seu ciclo de vida no Sistema Multiagente (COSTA, 2006).

O Modelo de Estado do agente “Interface do Usuário” é mostrado na Figura 31. Quando o aplicativo é iniciado o agente entra no estado “Aguardando interação do usuário com o sistema ou por itens filtrados”. Depois que o usuário estiver conectado ao sistema, o agente começa a fiscalizar o seu comportamento. Quando o usuário preenche um formulário indicando os seus interesses, o agente vai para o estado “Criação ou atualização do modelo do usuário”. Se o agente “Interface de Usuário” receber itens filtrados, irá para o estado “Produção e entrega de recomendações personalizadas”. Quando o aplicativo é encerrado o agente atinge o seu estado final.

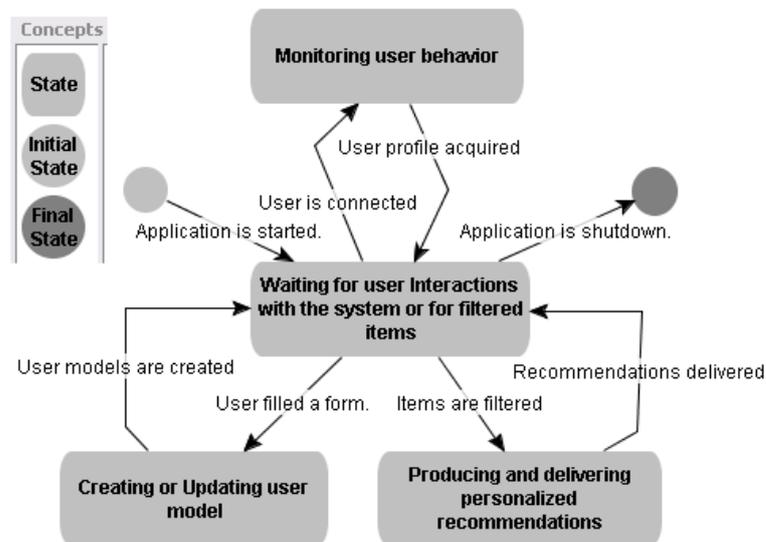


Figura 31 Modelo de Estado do Agente “Interface do Usuário”

Um ponto de variação foi identificado no Modelo de Estado do agente “Interface do Usuário”. Este ponto está relacionado com os estados “Criação ou atualização dos modelos de usuário” e “Monitoramento do comportamento do usuário”, mostrados na Figura 32.

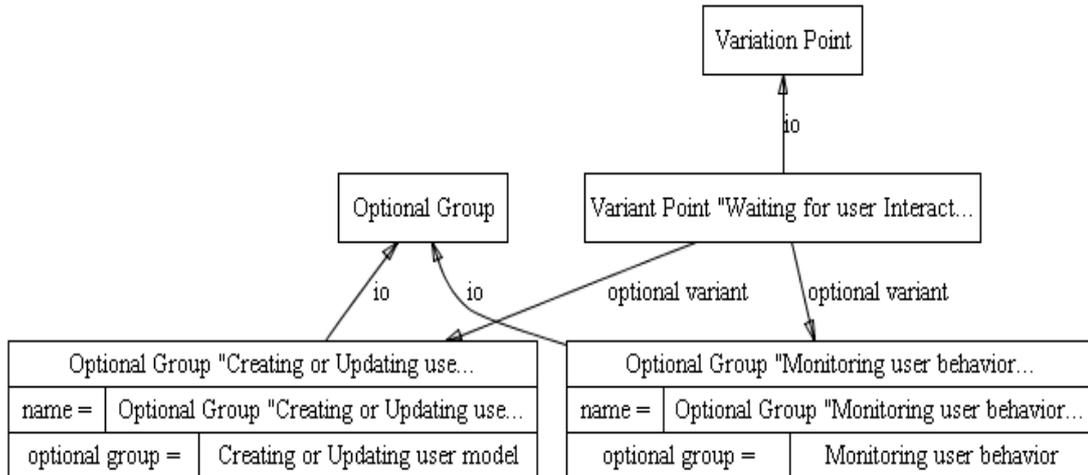


Figura 32 Variabilidade do Modelo de Estado do Agente “Interface do Usuário”

4.2.3.4 Modelagem do Estado do Agente “Filtrador”

O agente “Filtrador” começa a sua execução com o estado “Aguardando novos itens de informações ou Modelos de Usuários”. Quando há um novo item de informação disponível o agente “Filtrador” passa para o estado “Criando o modelo do novo item de informação”, onde se cria a sua representação interna e a seguir passa-se para o estado “Filtrando itens de informação”, no qual realiza a análise de similaridade. O aparecimento de novos modelos de usuários leva o agente para o estado “Classificando ou Agrupamento do novo modelo de usuário”, em que ele classifica o modelo novo usuário em um grupo existente ou se reagrupam os modelos existentes, no qual se executa o estado “Filtrando itens de informação”.

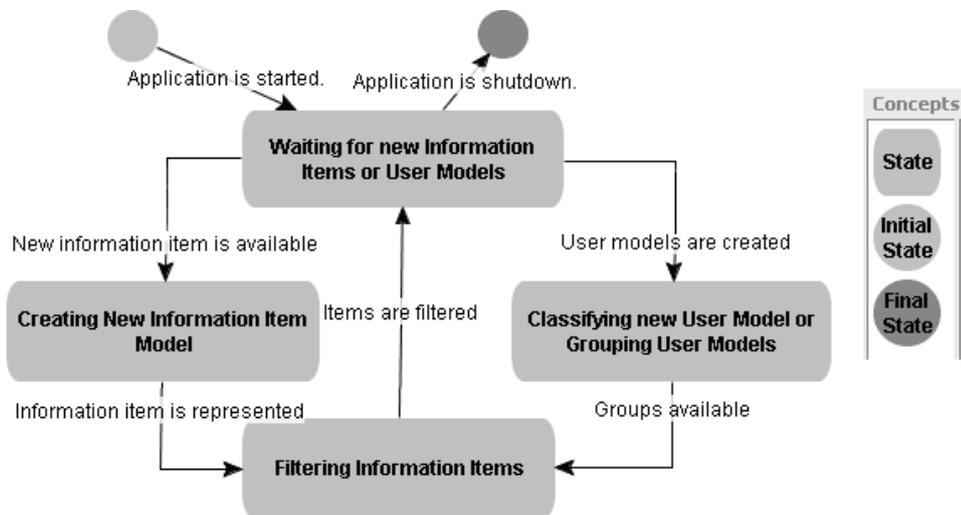


Figura 33 Modelo de Estado do Agente “Filtrador”

No Modelo de Estado, do agente “Filtrador”, tem-se o ponto de variabilidade relacionado com o estado “Aguardando novos itens de informação ou modelos de usuários”, o qual tem o grupo opcional “Criação de um novo modelo de usuário ou de um grupo de modelo de usuário” e o “Criação de uma nova informação”, Figura 34. Um ponto de variação no estado “Itens de informações filtrados” também é possível e teria os mesmos grupos opcionais.

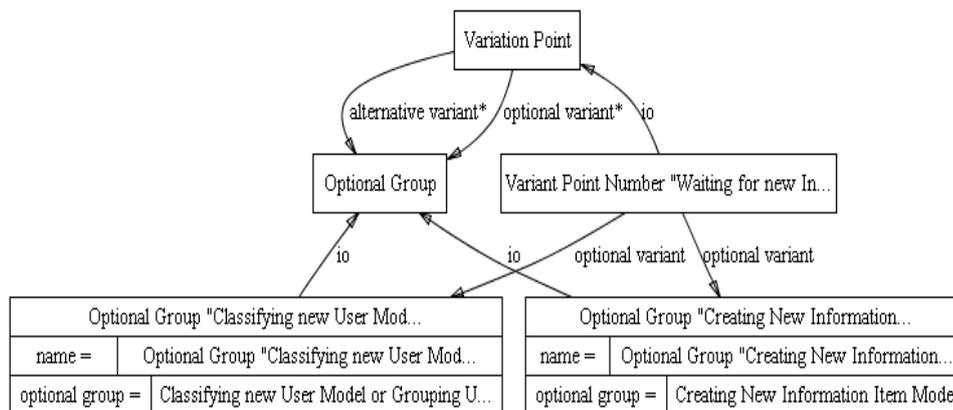


Figura 34 Ponto de Variação do Modelo de Estado do Agente “Filtrador”

4.3 Implementação do Domínio

A fase Implementação de Domínio utiliza como insumos os produtos da fase Projeto de Domínio para a criação de modelos a serem implementados. Esta fase é composta de 3 tarefas: Modelagem de comportamento, Modelagem de Atos de Comunicação e Implementação dos Agentes.

Os pontos de variação foram identificados no ONTOSERS - DI nos seguintes itens:

- Modelo de Agentes e Comportamentos na variabilidade do JADE Agent em relação a Jade Behavior e do Jade Comportamento em relação ao Método Jade Behavior
- Modelo de Atos de Comunicação dos Agentes, na variabilidade do JADE Agente em relação a Jade Behavior e do Jade Comportamento em relação ao Método Jade Behavior.

4.3.1 Mapeamento do Projeto para a Implementação dos Agentes e dos Comportamentos

Na Modelagem de Comportamento é analisado o Modelo de Atividades, a fim de definir como os agentes atuarão. Nesta fase os agentes são mapeados para seus comportamentos, em JADE.

No ONTOSERS-DI foi desenvolvido o Modelo de Agentes e Comportamentos do agente “Interface do Usuário” apresentado na Figura 35.

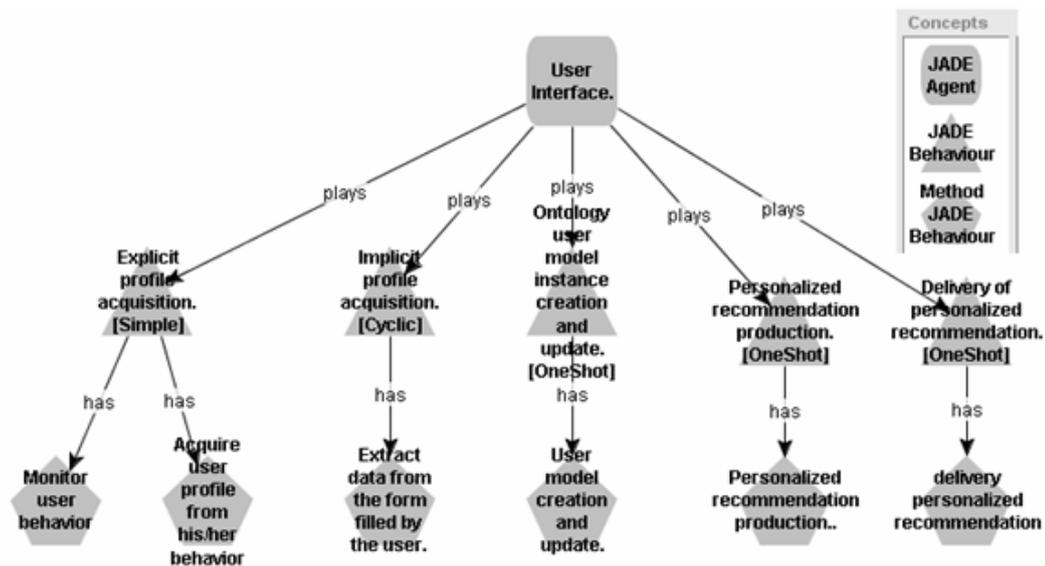


Figura 35 Modelo de Agentes e Comportamentos do agente “Interface de Usuário”

Na Figura 36 é exibido o modelo de agente e comportamento relacionado ao agente “Filtrador”.

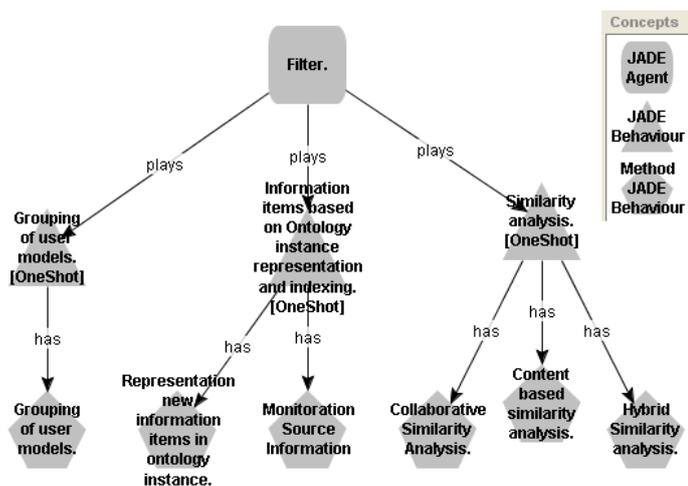


Figura 36 Modelo de Agentes e Comportamentos do agente “Filtrador”

Cada agente de projeto foi mapeado para um agente JADE. Cada responsabilidade foi mapeada para uma classe JADE Behaviour mais um arquivo Jess e cada atividade para um método da classe JADE Behavior respectiva.

O agente “Filtrador” tem um ponto de variação representado pela Figura 37, onde aparece os métodos JADE Behavior que são executados em cada contexto.

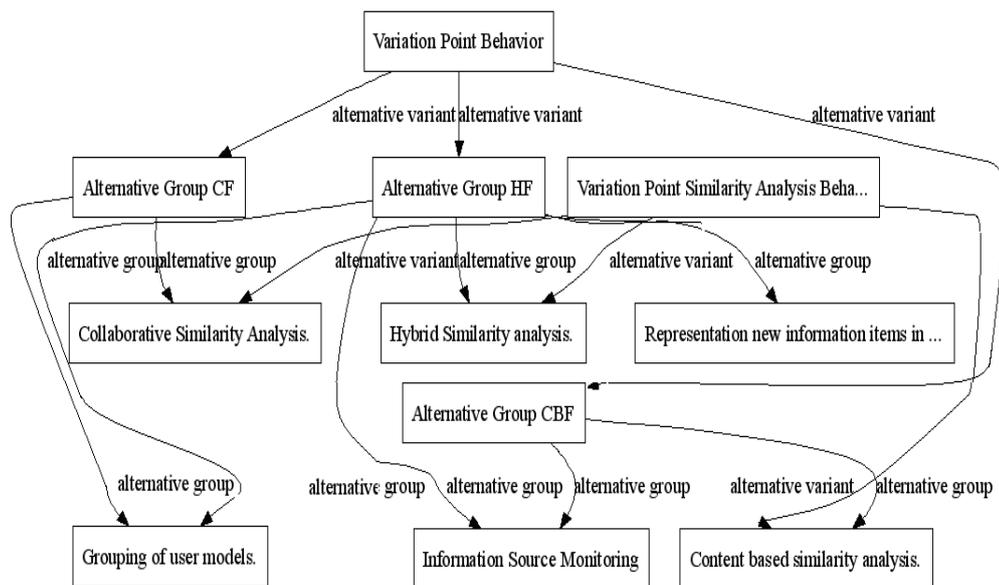


Figura 37 Ponto de Variação do Behavior do Agente “Filtrador”

4.3.2 Mapeamento das Interações entre Agentes para Atos de Comunicação

Nesta parte são especificadas as diretivas de comunicação trocadas entre os agentes, como mostrado na Figura 38. Estes atos comunicativos foram mapeados do Modelo de Interação do Agente. O Modelo de Atos de Comunicação determina as mensagens enviadas. Por exemplo, a mensagem PROPOSE (Novo item de informação) é gerada no método “Representação de novos itens informações como ontologia” e é recebida pelo método “Avaliar as novas itens de informações proposto”.

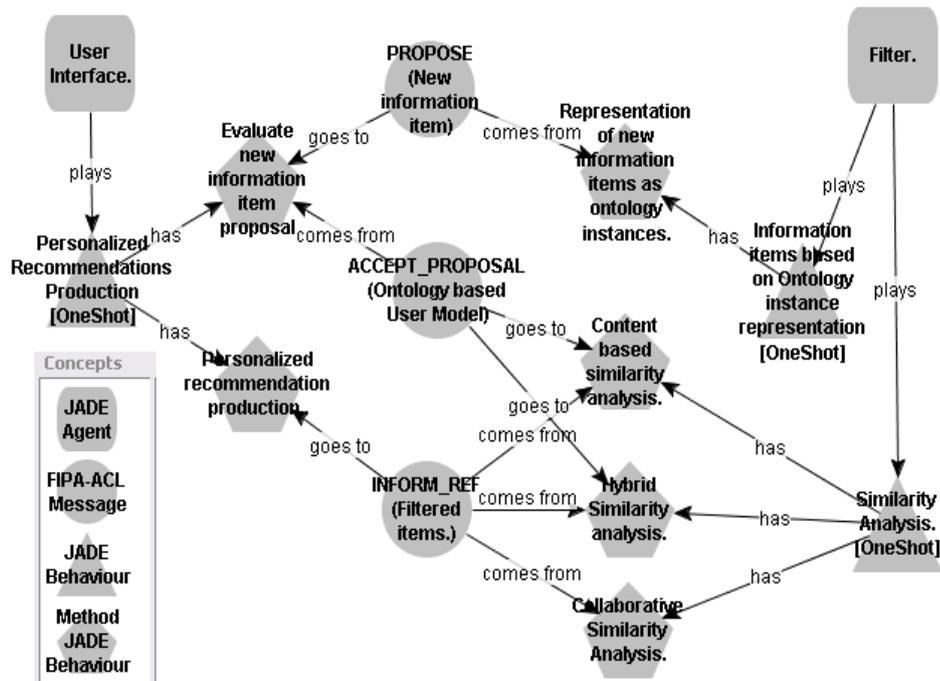


Figura 38 Modelo de Atos de Comunicação

4.3.3 Implementação dos Agentes

A família ONTOSERS é formada por dois agentes (o agente “Interface do Usuário” e o “Filtrador”), os quais foram gerados de forma a representarem um modelo genérico, capaz de ser reusado.

Ambos foram desenvolvidos utilizando JADE e o Jess, a fim de que os agentes tivessem capacidade deliberativa. A solução computacional tem no pacote ONTOSERS (Figura 39), diversos diretórios. No diretório jessInJADE foram colocados os códigos fonte do JessAgent e do JessBehaviour, a fim de fazer reuso dos stubs e skeletons dos agentes deliberativos desenvolvidos por (MASCARDI, MARTELLI, et al.. 2008). No diretório tuPinJADE foram tratadas as mensagens de erros, conforme o modelo de agentes deliberativos desenvolvidos em (MASCARDI, MARTELLI, et al.. 2008). Apesar deste diretório, no modelo utilizado, permitir o envio de mensagens também para agentes definidos em Prolog, neste trabalho, não foi usada essa capacidade. No diretório userProcessingLayer foi definida a camada de processamento do usuário, composta dos agentes de interface do usuário UserInterfaceAgent e ManagerAgent. O ManagerAgent é o responsável em localizar os diversos UserInterfaceAgent, a partir do seu registro no Yellow Pagers, do Jade.

No diretório `informationProcessingLayer` foi definida a camada de processamento da informação, composta dos agentes de filtragem `XFAgent` e o programa java que executa a análise de similaridade. No diretório `util` foram colocadas as funções relacionadas à conexão e às operações com o Banco. Outros diretórios necessários ao projeto no NetBeans 6.0.1 são o pacote de testes, as bibliotecas e as bibliotecas de testes.

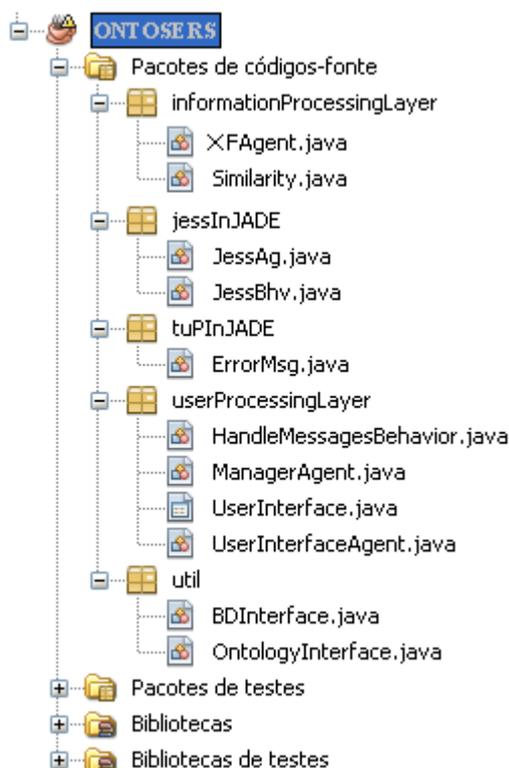


Figura 39 Estrutura do Código-fonte da família ONTOSERS

4.3.3.1 O Modelo Genérico do Agente de interface do Usuário

No ONTOSERS foram desenvolvidos stubs, skeletons e outros artefatos de código capazes de serem utilizadas em unidades funcionais da família de software. Neste sentido o “UserInterfaceAgent” foi desenvolvido a partir do trabalho de (MASCARDI, MARTELLI, et al.. 2008), no qual é dado o stub e skeleton básico para o desenvolvimento dos agentes jess no Jade. Além destes programas em java é necessário um arquivo CLP, no qual terá o jess necessário à deliberação dos agentes. Os códigos-fonte de todos os agentes desenvolvidos poderão ser consultados no site (GESEC 2001).

No interface.clp foram definidos os: Deftemplate user_model; Defglobal login; Deffunction createReply(?element) e createRequestReply(); um defrule para cada estado. No interface.clp foram colocados os estados: start-up, waiting-state, recommending-state, msg-received e evaluate-proposal.

No interface.java foi especificado que este: carregaria uma ontologia; teria o Protocolo definido como "interface", com setType("user-interfacing") e setName("Infotrib-user-modeling"); na variável jessFile seria informado interface.clp e que existiriam os métodos: acquireNewProfile, readOntologyFromFile, saveUserModel, deliverRecommendations, takeDown e initTables.

4.3.3.2 O Modelo Genérico do Agente Filtrador

A classe XFilter.java estende a jessInJADE.JessAg, adiciona o protocolo "xfilter"; prepara os comandos jess que utilizará e carrega o jessFile "xfilter.clp" e registra os candidatos à recomendação. No arquivo xfilter.clp define-se os templates do usuário e do item de informação e os estados que dispararão as ações, como: startup, msg-receiver, sending-filtered-item.

4.4 Extração e Representação de Padrões

A análise do ONTOSERS, para identificação dos padrões não é objetivo deste trabalho, dado que um padrão é uma solução reiterada e este é o primeiro ensaio sobre a família de sistema de recomendações.

4.5 Considerações Finais

O desenvolvimento da família ONTOSERS utilizou o seguinte conjunto de ferramentas: a ONTORMAS, para a modelagem e representação dos modelos; o Jade⁸, por ser um framework de desenvolvimento de agente java open source; o

⁸ JADE - Disponível em <http://jade.tilab.com/>

Jess⁹, por possibilitar incorporar a deliberação nos agentes JADE; o jessInJADE Package¹⁰, por ter um pacote para criação de agentes com Jess e o DCaseLP¹¹ e um ambiente de prototipagem para sistemas multiagente.

A contribuição do ONTOSERS foi a de gerar diversos produtos, com base na MADEM, representados no Modelo de Domínio, no Framework Multiagente, no Modelo de Implementação da Sociedade Multiagente e nos agentes de software executáveis relacionados aos Sistemas de recomendações Baseados na Tecnologia da Web Semântica. Estes produtos permitem o desenvolvimento de aplicações multiagente a partir do reuso dos mesmos. Tais artefatos foram utilizados no desenvolvimento do INFOTRIB.

⁹ Jess – Disponível em <http://herzberg.ca.sandia.gov/>

¹⁰ jessInJADE – Disponível em <http://www.disi.unige.it/person/MascardiV/Software/jessInJADE-Manual.pdf>

¹¹ DCaseLP - Disponível em [http://lia.deis.unibo.it/confs/lads/papers/2.1%20paper_10%20\(gungui\).pdf](http://lia.deis.unibo.it/confs/lads/papers/2.1%20paper_10%20(gungui).pdf)

5. AVALIAÇÃO DOS AGENTES DE FILTRAGEM

Para a análise da efetividade dos diferentes algoritmos de filtragem vários experimentos foram conduzidos (Figura 40).

Nas fases I e II foram avaliadas a filtragem baseada em conteúdo e a colaborativa, de forma isolada. Na fase III as modalidades de filtragens foram combinadas na filtragem híbrida, utilizando as técnicas ponderada e alternada, sendo que a ponderada foi realizada por Intersecção e por pesos.

Todos os experimentos realizados utilizaram modelagem explícita dos usuários, a fim de que o perfil adquirido não tivesse impacto na efetividade da filtragem.

Para validar a aplicação desenvolvida foram executados experimentos off-line (para verificar a capacidade de recomendar corretamente) e on-line (para validar a satisfação do usuário, quanto às recomendações).

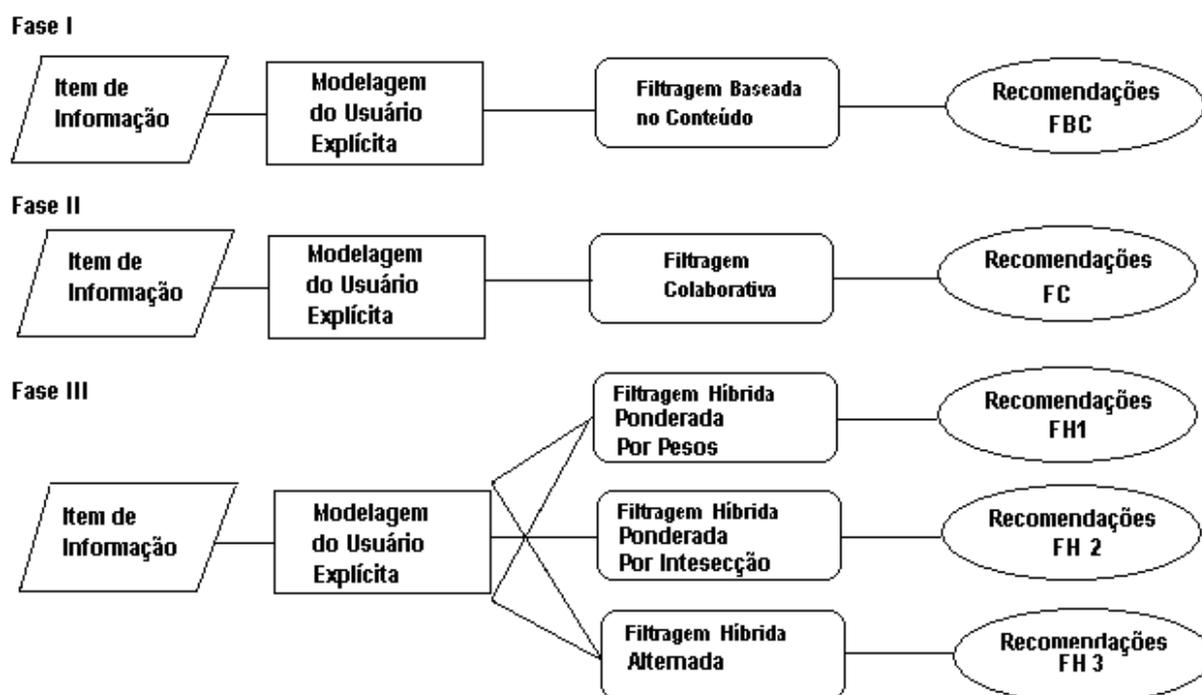


Figura 40 Os diferentes Experimentos Realizados

A medida de similaridade utilizada nos experimentos foi a medida semântica apresentada em (DRUMOND, GIRARDI 2008).

A fonte de informação utilizada foi as instâncias da ontologia ONTOTRIB.

5.1 A Ontologia Tributária ONTOTRIB

A Ontologia Tributária, ONTOTRIB, tem a função de representar um domínio específico do Direito Brasileiro, o Direito Tributário, sobre o qual serão realizadas recomendações, a partir da versão atualizada da ONTOJURIS (DRUMOND, GIRARDI, 2008), da qual é uma extensão. Para o contexto desta dissertação, será utilizada a ONTOTRIB restrita à recomendação de instrumentos normativos tributários.

Os principais conceitos de interesse, para o escopo deste trabalho, na ONTOTRIB, são os instrumentos normativos tributários e os elementos da relação tributária. Não serão analisadas todas as leis tributárias brasileiras, mas apenas algumas, relativas aos Estados do CE, MA, BA, ao Brasil e ao município de São Luís.

No seu atual estágio, a ONTOTRIB não pretende representar todo o conhecimento tributário, mas aquele necessário para recomendar leis tributárias. Logo, focou-se em responder perguntas do tipo:

- Quais são as novas leis tributárias, compatíveis com os interesses de seu usuário, das quais ele ainda não possui conhecimento?
- Quais as leis tributárias similares àquelas as que já foram consultadas?
- Quais são as leis tributárias recomendadas para pessoas com o perfil similar ao do consultor?

5.1.1 Estrutura das classes principais da ONTOTRIB

Da Tabela 11 até a Tabela 14 são apresentadas as principais classes e atributos da ONTOTRIB. A classe “Instrumento Normativo Tributário” é uma especialização (subclasse) da classe “Instrumento Normativo” da ONTOJURIS e suas instâncias representam normas do Direito Tributário.

Tabela 11 Classe “Instrumento Normativo”

Atributos	Descrição
Tipo	Espécie da norma: Lei (Ordinária, Complementar, Delegada), Constituição, Medida Provisória, Decreto, Resolução, etc.
Número	Número dado à norma. Por exemplo, LC 87 (Lei Complementar do ICMS, chamada Lei Kandir).
Data	Quando foi aprovada.
Sumário	Resumo da Lei
Dispositivos	Quais dispositivos compõem a norma. Suas instâncias são membros da classe “Dispositivo”.
Autoridades	Quem assinou a norma.
Ramo Legal	A qual ramo do Direito pertence a norma.
Sanção	Quando a norma foi sancionada, nos casos em que é necessário sanção.
Promulgação	Data em que ocorreu a promulgação.
URL	Local na web, de onde se retirou a norma.

Tabela 12 Subclasse “Instrumento Normativo Tributário” sem os slots herdados

Atributos	Descrição
Aplicação Judicial	Escopo de aplicação da norma tributária (Federal – Brasil, Estado ou Município específico)
Tributo	Espécie tributária a que se relaciona.
Conceitos tributários	Sobre qual(is) conceito(s) a norma trata.

A classe “Usuário Legal Tributário” representa uma especialização da classe usuário, da ONTOJURIS, na qual se agrupam os usuários que tenham interesses jurídicos tributários comuns.

Tabela 13 Classe “Usuário Legal”

Atributos	Descrição
Nome	Nome do usuário
E-mail	e-mail para envio das recomendações
Ramos legais	Ramos do Direito pelo qual se tem interesse
Tipos	Espécie de norma de interesse: Lei (Ordinária, Complementar, Delegada), Constituição, Medida Provisória, Decreto, Resolução, etc.

Tabela 14 Subclasse “Usuário Legal Tributário”, sem os slots herdados

Atributos	Descrição
Tributos	Espécie tributária a que se relaciona.
Aplicação tributária	Escopo de aplicação da norma tributária (Federal – Brasil, Estado ou Município específico)
Conceitos tributários	Sobre qual(is) conceito(s) a norma trata.

A Figura 41 ilustra as subclasses herdadas da ONTOJURIS e as classes especificamente definidas na ONTOTRIB.

O atributo “tributo”, da classe “Instrumento Normativo Tributário” pode ter zero ou mais tributos associados. Em função do Direito Tributário não ser formado apenas por leis que regulam tributos. Leis, como a dos “Crimes contra a Ordem Financeira”, não se refere a um tributo, mas é uma Lei Tributária. Já a Lei 7.799, do Maranhão, que institui o Sistema Tributário Estadual, define todos os impostos, taxas e contribuições de melhoria estaduais. Logo, referem-se aos N tributos estaduais, do Maranhão.

Na classe “Tributo” (Figura 43), utilizou-se a Teoria Pentapartida dos Tributos, do STF (CAPEZ, MALTINTI, 2006). Por esta teoria, a qual é aceita pelo STF, o tributo seria gênero, cujas espécies seriam: impostos, taxas, contribuições de melhoria, contribuições especiais e empréstimo compulsório.

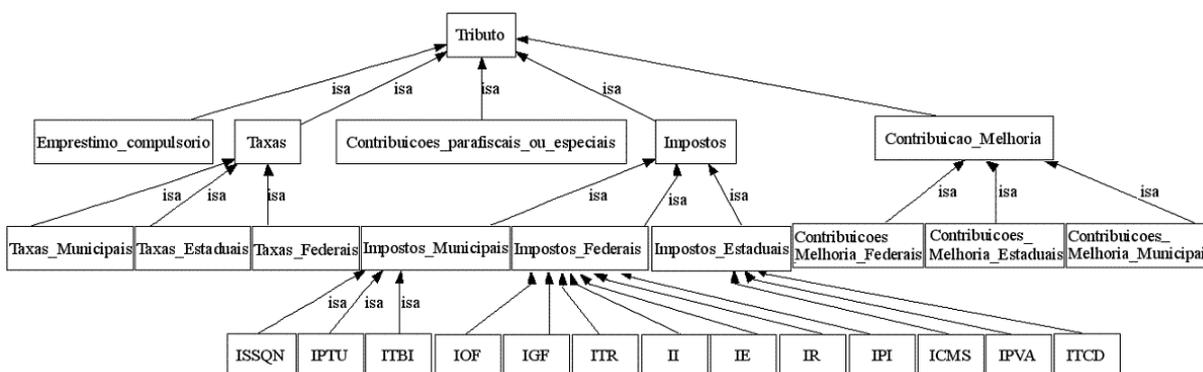


Figura 43 Classificação dos Tributos

Para a ONTOTRIB foi necessário especializar o usuário legal, cujas propriedades foram herdadas da ONTOJURIS, o que foi feito através da classe “Usuário Legal Tributário”, a qual tem vários atributos para atender o domínio tributário, tanto de registro das necessidades de informação tributária, quanto de entrega das recomendações (Figura 44). Os interesses tributários foram especificados em termos de tributos, modalidades e aplicações. Incluiu-se o conceito de confiança, de MASSA e AVESANI (2007), no qual os algoritmos colaborativos não são aplicados para todo o conjunto de usuários, mas só para aqueles com os quais se possua relação de confiança.

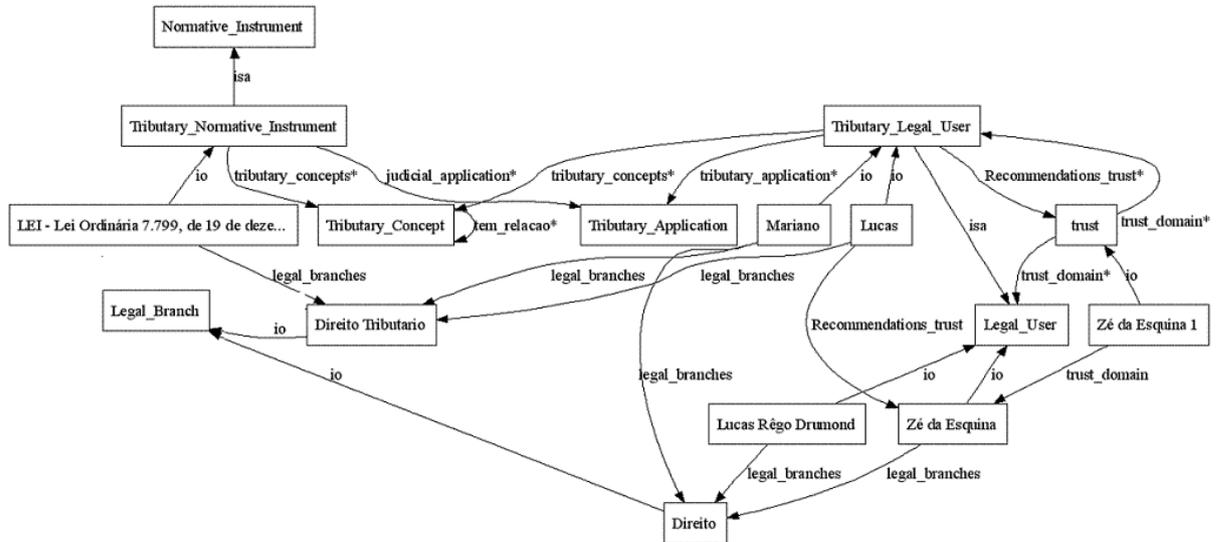


Figura 44 Instâncias da Classe “Usuário Legal Tributário” da ONTOTRIB

A Figura 45 e Figura 46 apresentam as principais classes, relacionamentos e instâncias da ontologia ONTOTRIB.

Na classe “Conceitos Tributários” tem-se a definição dos conceitos, seus sinônimos, as generalizações, as especializações e a fonte da informação. Na Figura 45 é apresentada o relacionamento entre a classe “Conceitos Tributários” e as instâncias da classe “Elementos da relação tributárias”.

Na Figura 46 há um detalhamento das classes e instâncias que compõem os elementos da relação tributária. Por estas duas figuras têm-se uma visão geral do Direito Tributário. Os tipos de sujeição: ativa ou passiva. Os tipos de obrigação: principal ou acessória. As modalidades de obrigação tributária: fazer, não fazer ou tolerar. Os aspectos do fato gerador: temporal, espacial, quantitativo, etc.

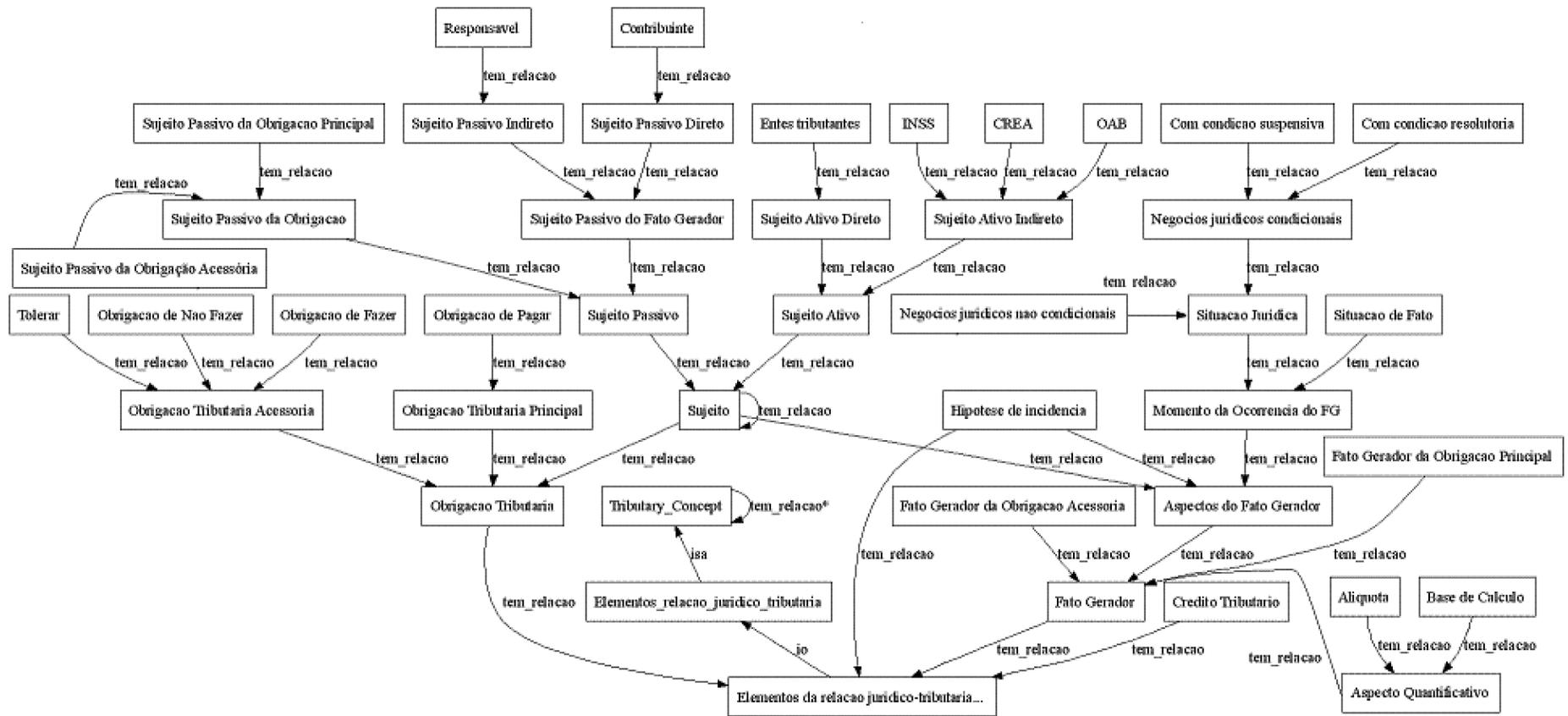


Figura 46 Detalhamentos dos relacionamentos entre instâncias tributárias da ONTOTRIB.

5.2 A Medida Semântica de Similaridade

As etapas deste Modelo de Análise de Similaridade Baseado em Casos Semânticos estão mostrados na Figura 47 (DRUMOND 2007). Um caso semântico é uma característica do elemento de informação através da qual os interesses do usuário podem ser especificados. A um caso semântico é associada uma hierarquia de termos, que são os possíveis valores que um caso semântico pode assumir em um modelo de item de informação. Cada caso semântico em um modelo de usuário possui um peso. Este peso representa a relevância do caso semântico para o usuário e pode ser atribuído explicitamente pelo mesmo ou inferido a partir de suas avaliações ou de seu comportamento. De forma análoga, um item de informação pode ser descrito por um conjunto de casos semânticos (Figura 47).

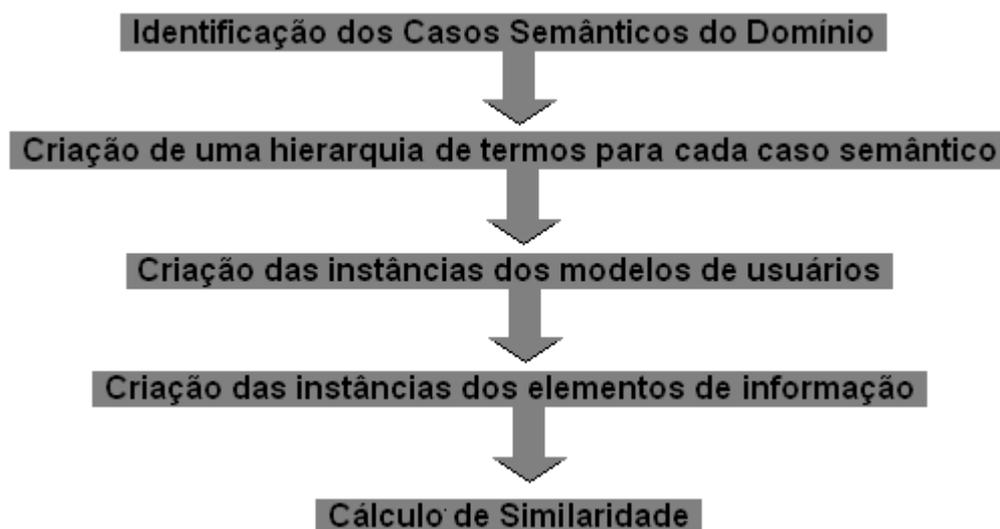


Figura 47 Resumo do Modelo de Análise de Similaridade Baseado em Casos Semânticos

Um conjunto de casos semânticos pode ser utilizado para representar um modelo de usuário em um sistema de filtragem de informação, indicando os seus interesses. Os casos semânticos presentes nas páginas pelas quais ele navegou ou buscou as páginas pelas quais ele voltou. Um conjunto de casos semânticos pode também representar os conceitos presentes em uma consulta a um sistema de recuperação de informação ou descrever um item de informação. Deste modo, a similaridade entre um conjunto de casos semânticos pode ser utilizada por qualquer aplicação.

Formalmente, pode se representar um caso semântico como um conceito C em uma T-Box T . Os termos do caso semântico C são todos os conceitos de T , que são especializações de C , ou seja, um conceito D é um termo do caso semântico C se, e somente se, $C \sqsupseteq D$.

O Modelo de Usuário e os Itens de Informações são compostos por um conjunto de casos semânticos. Os casos semânticos em um conjunto podem assumir vários valores. A similaridade entre um conjunto de casos semânticos pode ser utilizada por qualquer aplicação desde que esta disponha de uma ontologia e possa associar um conjunto de conceitos desta ontologia aos elementos que precisam ser comparados.

No modelo de análise de similaridade de (DRUMOND 2007), um modelo de usuário, denotado por u pode ser definido como um conjunto de casos semânticos: $u = \{i / i \text{ é um caso semântico de interesse do usuário } u\}$

Cada caso semântico possui um peso, denotado por w_i , e pode assumir vários valores dentro do modelo de usuário u :

$$u_i = \{j / j \text{ é um termo do caso semântico } i\}$$

Uma representação de um elemento de informação será denotada por ε :

$$\varepsilon = \{i / i \text{ é um caso semântico utilizado para descrever o elemento } \varepsilon\}$$

Analogamente ao modelo de usuário, cada caso semântico que descreve o elemento de informação está associado a um conjunto de termos:

$$\varepsilon_i = \{j / j \text{ é um termo do caso semântico } i\}$$

A similaridade entre casos semânticos é dada por:

$$\text{similaridade}(u, \varepsilon) = \sum_{\forall i \in \varepsilon} w_i \bullet \max_{\forall j \in \varepsilon_i} \text{sim}(u_{ij}, \varepsilon_{ik})$$

O que será verdade deste que:

$$\sum_{\forall i \in \varepsilon} w_i = 1$$

Partindo-se destes conceitos foi definido \mathcal{L} como sendo o conjunto de todos os conceitos em uma hierarquia, cuja similaridade entre dois conceitos desta hierarquia é uma função $\text{sim}: \mathcal{L} \times \mathcal{L} \rightarrow [0,1]$ tal que:

$$sim(C, D) = \frac{2 |C_{\Xi} \cap D_{\Xi}|}{|C_{\Xi}| + |D_{\Xi}|}$$

Transportando para o caso da ONTOTRIB, exemplificando os cálculos da similaridade semântica, utilizando parte da hierarquia associada, com base na Figura 48 tem-se que:

$(C)_{\Xi}$ para (ICMS MA) Ξ {ICMS MA, ICMS, Estaduais, Impostos }

$(D)_{\Xi}$ para (IPVA) Ξ {IPVA, Estaduais, Impostos }

$C_{\Xi} \cap D_{\Xi}$ para (ICMS MA) \cap (IPVA) = { Estaduais, Impostos }

Portanto,

$$sim(C, D) = \frac{2 |C_{\Xi} \cap D_{\Xi}|}{|C_{\Xi}| + |D_{\Xi}|} = \frac{2 \times 2}{4+3} = 0,5714$$

$(C)_{\Xi}$ para (ICMS MA) Ξ {ICMS MA, ICMS, Estaduais, Impostos }

$(C)_{\Xi}$ para (ICMS CE) Ξ {ICMS CE, ICMS, Estaduais, Impostos }

$C_{\Xi} \cap D_{\Xi}$ para (ICMS MA) \cap (IPVA) = { ICMS, Estaduais, Impostos }

Logo,

$$sim(C, D) = \frac{2 |C_{\Xi} \cap D_{\Xi}|}{|C_{\Xi}| + |D_{\Xi}|} = \frac{2 \times 3}{4+4} = 0,75$$

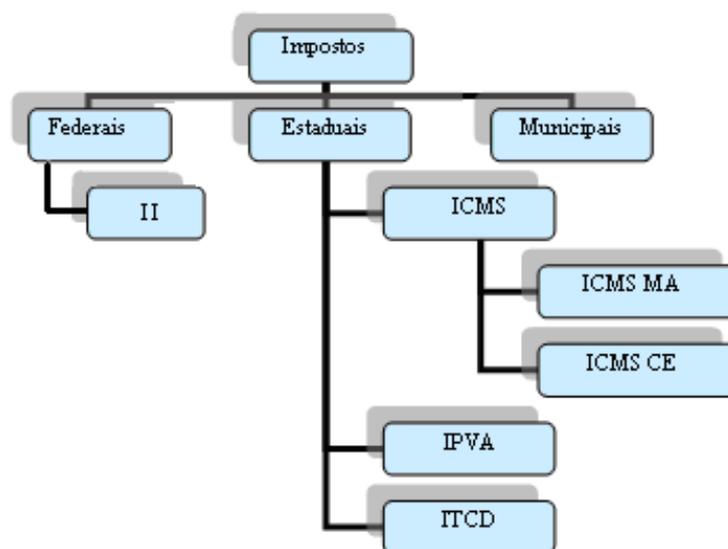


Figura 48 Parte da Hierarquia da ONTOTRIB

Para o foco da ONTOTRIB, os casos semânticos serão a hierarquia dos tributos brasileiros, os instrumentos normativos tributários, os elementos da relação tributária, seus conceitos e o usuário tributário.

A hierarquia tributária, em termos de tributos, é aquela definida nos artigos 145, 147, 148, 149, 149-A, 153, 154, 155, 156, 194, 195, 239, 240 da Constituição Federal de 1988 e demais artigos associados.

Quanto aos instrumentos normativos tributários estes serão todas as normas que se relacionem com o Direito Tributário, quer tratem, ou não, de tributos, como a Lei Dos Crimes Contra a Ordem Tributária.

Os elementos da relação tributária e seus conceitos serão aqueles definidos pela Doutrina e Jurisprudência, apesar de se que trabalhará apenas alguns nos testes a serem realizados.

O usuário tributário será uma especialização do usuário legal.

5.3 A Medida Semântica de Similaridade e o KNN

O algoritmo do Vizinho mais Próximo (KNN), de forma geral, tem duas partes. Uma de cálculo da proximidade, onde se pode utiliza o cosseno, o coeficiente de correlação de Pearson, etc. e a segunda que utiliza uma inferência do valor que o usuário daria para um determinado item de informação, o qual ele ainda não avaliou, a esta segunda parte a literatura chama de predição (CRESPO,1997).

Cada forma de cálculo da proximidade tem seus prós e contra. No coeficiente de correlação de Pearson a distribuição tem de ser linear (CRESPO,1997). Ou seja, há a possibilidade de um esforço computacional inútil, o que não ocorre com a medida semântica de similaridade. Portanto utilizaremos a medida semântica de similaridade, para o cálculo da proximidade, e a predição para informar o valor que determinado item teria se fosse avaliado, com base em seus vizinhos.

A predição é a média ponderada das avaliações dos vizinhos, somada à média das avaliações dadas pelo usuário-alvo (CHAVES, 2005). Este procedimento corresponde à equação a seguir:

	<p>RES – Resolução,</p> <ul style="list-style-type: none"> - Aplicação Tributária: Brasil, CE, PI, MA - Tributos.....: ICMS, IPI, IOF,ISSQN, IPVA - Conceitos Tributários:Base de Cálculo, Fato Gerador, Incidência, Sujeito, Alíquota - Perfil Profissional: Empresário, Juiz, Advogado - Formas de Entrega: On site, e-mail - Freqüência na entrega: imediata. 	<p>ICMS, IPI, IOF,ISSQN, IPVA }, { BC, FG, Inc, Suj., Aliq.}</p> <p>}</p>
US4	<p>Usuário Jurídicos Tributário:</p> <ul style="list-style-type: none"> - Nome:jose - E-mail: uiratan@gmail.com Senha:uiratan - Ramos Jurídicos.....: Direito Civil, Direito Tributário, Direito Previdenciário, Direito Comercial - Tipos Normativos: CON – Constituição, LEI – Lei Ordinária, LCP – Lei Complementar RES – Resolução, - Aplicação Tributária: Brasil, CE, PI, MA - Tributos.....: ICMS, IPI, IOF,ISSQN, IPVA - Conceitos Tributários:Base de Cálculo, Fato Gerador, Incidência, Sujeito, Alíquota - Perfil Profissional: Empresário, Juiz, Advogado - Formas de Entrega: On site, e-mail - Freqüência na entrega: imediata. 	<p>1 { Usuário Jurídicos Tributário , Direito Tributário, { CON – Constituição, LEI – Lei Ordinária, LCP – Lei Complementar, RES – Resolução }, { Brasil, CE, PI, MA}, { ICMS, IPI, IOF,ISSQN, IPVA }, { BC, FG, Inc, Suj., Aliq.}</p> <p>}</p>
US5	<p>Usuário Jurídicos Tributário:</p> <ul style="list-style-type: none"> - Nome:X - E-mail: x@gmail.com Senha:x - Ramos Jurídicos.....: - - Tipos Normativos: - - Aplicação Tributária: - - Tributos.....: - Conceitos Tributários: - Perfil Profissional: - Formas de Entrega: On site, e-mail - Freqüência na entrega: 	<p>1 { Usuário Jurídicos Tributário}</p>
US6	<p>Usuário Jurídicos Tributário:</p> <ul style="list-style-type: none"> - Nome:Y - E-mail: y@gmail.com Senha:y - Ramos Jurídicos.....: Direito Público, Direito Tributário - Tipos Normativos: CON – Constituição, LEI – Lei Ordinária, LCP – Lei Complementar RES – Resolução, - Aplicação Tributária: São Luís - Tributos.....: ISSQN - Conceitos Tributários: - - Perfil Profissional: - Formas de Entrega: On site - Freqüência na entrega:semanal. 	<p>1 { Usuário Jurídicos Tributário , Direito Tributário, { CON – Constituição, LEI – Lei Ordinária, LCP – Lei Complementar }, {São Luís}, { ISSQN}</p> <p>}</p>
US7	<p>Usuário Jurídicos:</p> <ul style="list-style-type: none"> - Nome: W E-mail: W@gmail.com - Ramos Jurídicos.....: Direito Penal, Direito Tributário - Tipos Normativos: CON – Constituição, LEI – Lei Ordinária, LCP – Lei Complementar 	<p>1 {UJ, Direito Tributário, {CON, LEI, LCP} }</p>
US8	<p>Usuário Jurídicos Tributário:</p> <ul style="list-style-type: none"> - Nome:A E-mail: a@gmail.com Senha:A - Ramos Jurídicos.....: Direito Tributário - Tipos Normativos: CON – Constituição, EC – Emenda Constitucional, LEI – Lei Ord. - Aplicação Tributária: Brasil - Tributos.....: ITCD - Conceitos Tributários: - - Perfil Profissional: - - Formas de Entrega: On site - Freqüência na entrega:mensal. 	<p>1 {UJT, Direito Tributário, {LEI, EC, CON} {ITCD} }</p>
US9	<p>Usuário Jurídicos Tributário:</p> <ul style="list-style-type: none"> - Nome:B - E-mail: b@gmail.com Senha:B - Ramos Jurídicos.....: Direito Público, Direito Tributário 	<p>1 { Usuário Jurídicos Tributário , Direito Tributário, {</p>

	<ul style="list-style-type: none"> - Tipos Normativos: CON – Constituição, EC – Emenda Constitucional - Aplicação Tributária: Brasil - Tributos.....: IR PF, IR PJ - Conceitos Tributários: Alíquota, Sujeito Passivo - Perfil Profissional: - Formas de Entrega: On site - Frequência na entrega:mensal. 	CON – Constituição, EC – Emenda Constitucional}, { Brasil }, { IR PF, IR PJ } { Aliq., Suj. Passivo}
US10	Usuário Jurídicos Tributário: <ul style="list-style-type: none"> - Nome:Y2 - E-mail: y@gmail.com Senha:y - Ramos Jurídicos.....: Direito Agrário, Direto Tributário - Tipos Normativos: LEI – Lei Ordinária, LCP – Lei Complementar - Aplicação Tributária: - Tributos.....: ITR, IPTU, ITCD. - Conceitos Tributários: Incidência - Perfil Profissional: - Formas de Entrega: On site - Frequência na entrega:semanal. 	1 { Usuário Jurídicos Tributário , Direito Tributário, { LEI – Lei Ordinária, LCP – Lei Complementar }, { ITR, IPTU, ITCD }, { Incidência}

Na Tabela 16 estão as especificações dos instrumentos normativos utilizados nos experimentos e neles são especificadas as condições em que há interesse pelos instrumentos normativos . Será zero quando não for do Direito Tributário e um quando for do Direito Tributário.

Tabela 16 Instâncias de Instrumentos Normativos de Teste

Id.	Instâncias	Valores
IN1	Instrumento Normativo: <ul style="list-style-type: none"> - Tipo: Lei Ordinária - Data: 28/02/2008 - Ramo legal: - Tributos: - Número: 0001 - Sumário: Lei de teste 1 - Aplicação Judicial: - Conceitos 	0 (Lei)
IN2	Instrumento Normativo: <ul style="list-style-type: none"> - Tipo: Lei Complementar - Data: 28/02/2008 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: Brasil - Tributos: IPVA - Número: 0002 - Sumário: Lei de teste 2 - Conceitos Tributários: 	1 { LC – Lei Complementar, Direito Tributário, Brasil, IPVA}
IN3	Instrumento Normativo: <ul style="list-style-type: none"> - Tipo: Lei Ordinária - Número: 0003 - Sumário: Lei de teste 3 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: MA - Tributos: ITCD - Conceitos Tributários:Fato Gerador, Sujeito Ativo, Sujeito Passivo - Data: 28/02/2008 	1 { LEI – Lei Ordinária, Direito Tributário, MA, ITCD, {FG, SA, SP}}
IN4	Instrumento Normativo: <ul style="list-style-type: none"> - Tipo: Lei Ordinária - Data: 28/02/2008 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: MA - Conceitos Tributários:Fato Gerador, Sujeito Ativo, Sujeito Passivo - Número: 0004 - Sumário: Lei de teste 4 - Tributos: ICMS 	1 { LEI – Lei Ordinária, Direito Tributário, {MA}, ICMS, {FG, SA, SP}}

IN5	Instrumento Normativo: - Tipo: Lei Ordinária - Número: 0005 - Data: 28/02/2008 - Sumário: Lei de teste 5 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: MA - Tributos: IPVA - Conceitos Tributários: Aliq., BC	1 { LEI – Lei Ordinária, Direito Tributário, { MA}, IPVA, { Aliq., BC} }
IN6	Instrumento Normativo: - Tipo: Lei Ordinária - Número: 0006 - Data: 28/02/2008 - Sumário: Lei de teste 6 - Ramo Jurídicos: Direito Civil	0 (Lei, Direito Civil)
IN7	Instrumento Normativo: - Tipo: Lei Ordinária - Número: 0007 - Data: 28/02/2008 - Sumário: Lei de teste 2 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: Brasil - Tributos: IR PJ - Conceitos Tributários:Fato Gerador, Sujeito Ativo, Sujeito Passivo	1 { LEI – Lei Ordinária, Direito Tributário, { Brasil }, IR, {FG, SA, SP}}
IN8	Instrumento Normativo: - Tipo: Lei Ordinária - Número: 0008 - Data: 28/02/2008 - Sumário: Lei de teste 2 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: MA - Tributos: ICMS - Conceitos Tributários: Sujeito Passivo	1 { LEI – Lei Ordinária, Direito Tributário, { MA}, ICMS, {SP}}
IN9	Instrumento Normativo: - Tipo: Lei Complementar - Número: 0009 - Data: 28/02/2008 - Sumário: Lei de teste 8 - Ramo Jurídicos: Direito Tributário - Aplicação Judicial: Brasil, MA - Tributos: ITR - Conceitos Tributários:Fato Gerador, Sujeito Ativo, Sujeito Passivo, Alíquota, Isenção	1 { LEI – Lei Ordinária, Direito Tributário, {BRASIL}, ITR, {FG, SA,SP, Aliq., Isenção}}
IN10	Instrumento Normativo: - Tipo: Emenda Constitucional - Número: 00010 - Data: 28/02/2008 - Sumário: IR PF - Ramo Jurídicos: Direito Constitucional, Direito Tributário - Aplicação Judicial: Brasil - Tributos: Imposto - Conceitos Tributários:Fato Gerador, Sujeito Ativo, Sujeito Passivo, incidência	1 { EC – Emenda Constitucional, Direito Tributário, {BRASIL}, IR, {FG, SA,SP, Incidência}}

Da análise entre, o que o instrumento normativo é, e, o que o usuário deseja, criou-se a Matriz de Resultados Esperados, representada por E(us,in) (Tabela 17), a fim de se poder mensurar a validade das recomendações nas

diferentes técnicas, confrontando-se a matriz de resultados esperados com a matriz de resultados obtidos em cada técnica.

Tabela 17 Matriz de Resultados Esperados $E(us,in)$

	IN1	IN2	IN3	IN4	IN5	IN6	IN7	IN8	IN9	IN10
US1	0	1	1	1	1	0	1	1	1	1
US2	0	1	1	1	1	0	1	1	1	1
US3	0	1	0	1	1	0	0	1	0	0
US4	0	1	0	1	1	0	0	1	0	0
US5	0	1	1	1	1	0	1	1	1	1
US6	0	0	0	0	0	0	0	0	0	0
US7	0	1	1	1	1	0	1	1	1	1
US8	0	0	1	0	0	0	0	0	0	0
US9	0	0	0	0	0	0	0	0	0	1
US10	0	0	1	0	0	0	1	0	1	0

Na Tabela 17 as linhas representam os usuários da Tabela 15 e as colunas, os instrumentos tributários da Tabela 16. O um (1) em uma célula da Matriz de Resultados Esperados representa que o instrumento normativo deve ser recomendado ao usuário e o 0 significa que não deve. Os dados da Tabela 17 são o padrão estipulado para conferência com os resultados das filtragens.

5.4.1 Medidas de Avaliação

Para avaliar os resultados obtidos, da matriz “Resultados Esperados ($E_{(us,in)}$)”, em relação à matriz “Resultados obtidos ($O_{us,in}$)” foram definidos os seguintes medidas de avaliação dos agentes de filtragem na execução dos algoritmos de filtragem FBC, FC e FH:

- Índice de Acertos por Instrumento Normativo (IA_{IN});
- Índice de Acertos por usuário (IA_{US}),
- Índice Médio de Acertos por usuário (IMA_{US}),
- Recall e
- Precisão.

Para a definição das medidas foi especificado acerto em uma célula (i,j), a partir do comparativo das matrizes “Resultados Obtidos” e “Resultados Esperados”, cujo valor seria um (1) se aparece 1 em ambas as matrizes e zero (0), nos demais casos.

$$\text{Acerto}(i,j) = \begin{cases} 0, & \text{se } E_{i,j} \neq O_{i,j} \\ 1, & \text{se } E_{i,j} = O_{i,j} \text{ e } E_{i,j} = 1 \end{cases}$$

Da definição do que é Acerto(i,j) é possível especificar os demais medidas de avaliação como sendo:

- Índice de Acertos por Instrumento Normativo (IA_{IN}) – indica a proporção de acertos de recomendação de determinado instrumento normativo por usuário, expressa por:

$$IA_{IN}(j) = \frac{\sum_{i=0}^{us} \text{acerto}(i, j)}{us} \quad \text{(I)}$$

- Índice Médio de Acertos por Instrumento Normativo (IMA_{IN}) – indica a média de acertos de recomendação de determinado instrumento normativo, expressa por:

$$IMA_{IN}(j) = \frac{\sum_{j=0}^{in} IA_{in}(i)}{\sum in} \quad \text{(II)}$$

- Índice de Acertos por usuário (IA_{US}) - indica a proporção de acertos de recomendação de instrumentos normativos para determinado usuário, expressa por:

$$IA_{US}(i) = \frac{\sum_{j=0}^{in} \text{acerto}(i, j)}{in} \quad \text{(III)}$$

- Índice Médio de Acertos por usuário (IMA_{US}) - indica a média de acertos de recomendação de instrumentos normativos para os usuários, expressa por:

$$IMA_{US}(i) = \frac{\sum_{j=0}^{in} IA_{us}(i)}{\sum in} \quad \text{(IV)}$$

- Recall - indica a proporção total de sucesso obtidos em relação ao total de sucessos esperados, expressa por:

$$\text{Recall} = \frac{\sum_{i=0}^{us} \sum_{j=0}^{in} \text{acerto}_{obtidos}(i, j)}{\sum_{j=0}^n \sum_{i=0}^{us} \text{existentes} 1} \quad (\text{V})$$

- Precisão - indica a proporção total de sucesso obtidos em relação ao total de elementos recuperados da amostra, expressa por:

$$\text{Precisão} = \frac{\sum_{i=0}^{us} \sum_{j=0}^{in} \text{acerto}_{obtidos}(i, j)}{\sum_{j=0}^n \sum_{i=0}^{us} \text{recuperados} 1} \quad (\text{VI})$$

No anexo I há um exemplo de cálculo utilizando as medidas de avaliação especificadas acima.

5.4.2 Resultados

Como a medida de similaridade possui um valor entre 0 e 1 decidiu-se recomendar apenas os itens cujo valor de similaridade esteja acima do valor limiar especificado. Os experimentos foram repetidos com valores limiares variando de 0 a 100%, com o objetivo de se montar o gráfico do limiar segundo o recall e a precisão.

Calculou-se todos os índices definidos, para cada filtragem. Montou-se o gráfico com os resultados obtidos e apresentou-se a análise realizada sobre os dados.

5.4.2.1 Os Resultados na FBC

Os resultados dos testes de filtragem baseado no conteúdo (FBC) são mostrados na Tabela 18. Estes resultados foram obtidos a partir da ativação dos agentes de filtragem com a ontologia sem instrumentos normativos, mas com todos os usuários cadastrados. A seguir houve a inclusão das instâncias de instrumentos normativos de teste off-line e foi calculada a similaridade do usuário em relação a todos os dez instrumentos normativos.

A Tabela 18 mostra os resultados da análise de similaridade e não das recomendações.

Tabela 18 Matriz de Similaridade Obtida na FBC

	1	2	3	4	5	6	7	8	9	10
US1	0	100	100	100	100	0	100	100	100	100
US2	0	72	72	72	72	0	100	72	86	100
US3	0	70	86	100	100	0	86	100	86	86
US4	0	86	86	100	86	0	86	100	86	86
US5	0	72	72	72	72	0	86	72	100	86
US6	0	72	72	72	72	0	72	72	72	72
US7	0	72	72	72	72	0	86	69	86	86
US8	0	86	100	86	86	0	72	86	72	72
US9	0	72	72	72	72	0	100	72	86	100
US10	0	73	90	76	76	0	76	73	90	86

Da Matriz de Similaridade chegou-se à Matriz de Resultados Obtidos, pela definição de valores limiares, onde se o valor de uma célula da Matriz de Similaridade fosse maior ou igual do que o valor limiar o valor da correspondente célula na matriz de resultados obtidos seria um, caso contrário seria zero. Por exemplo, para um limiar de 25%, obteve-se a Tabela 19, onde cada célula será um (1) se o valor da célula correspondente na Tabela 18 for maior ou igual a 25 e zero (0) se for menor.

Tabela 19 Matriz de Resultados Obtidos O(us,in), na FBC

	1	2	3	4	5	6	7	8	9	10
US1	0	1	1	1	1	0	1	1	1	1
US2	0	1	1	1	1	0	1	1	1	1
US3	0	1	1	1	1	0	1	1	1	1
US4	0	1	1	1	1	0	1	1	1	1
US5	0	1	1	1	1	0	1	1	1	1
US6	0	1	1	1	1	0	1	1	1	1
US7	0	1	1	1	1	0	1	1	1	1
US8	0	1	1	1	1	0	1	1	1	1
US9	0	1	1	1	1	0	1	1	1	1
US10	0	1	1	1	1	0	1	1	1	1

Da matriz de resultados obtidos, com limiar de 25%, foram registrados 22 ocorrências de recomendações (1), das 22 possíveis da matriz de Resultados Esperados. Logo, o recall foi:

$$\text{Recall} = \frac{\text{Total de Sucesso Obtidos}}{\text{Total de sucessos esperados}} = \frac{22}{22} = 1 = 100\%$$

Dos 80 itens recomendados com o limiar de 25% apenas 45 correspondem aos resultados esperados, de acordo com a Tabela 17. Com isso a precisão é a seguinte:

$$\text{Precisão} = \frac{\text{Total sucesso obtidos}}{\text{Total de elementos recuperados}} = 45/80 = 56\%$$

A Figura 49 mostra o gráfico do recall e da precisão para níveis de limiar variando de 10 em 10% até 100%, nos quais se verifica que até 60% (0,6) é possível recuperar todos os instrumentos normativos desejados, mas com a precisão de 0,56 (56%). Para valores limiares maiores que 0,7 ocorre a diminuição do recall, o qual atinge o valor de 0,4, para um valor limiar igual a 1. Em relação à precisão, até o valor limiar de 0,7 estava em 0,56, passando a subir para valores maiores que 0,7, atingindo o pico de 0,95, para o valor limiar igual a 1. Portanto, se o objetivo for maximizar o recall e minimizar a precisão, os valores de valor limiar deverão ser inferiores a 0,6. Se for maximizar a precisão deverá ser o valor limiar igual a 1.

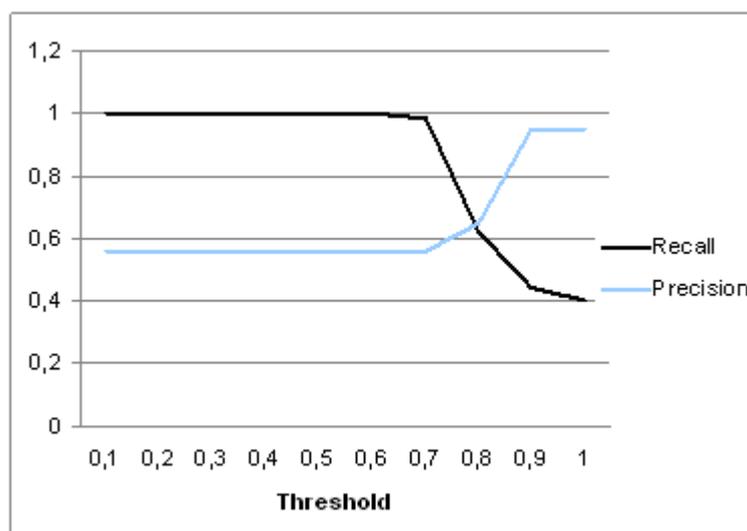


Figura 49 Gráfico do recall e da precisão da FBC em função do limiar utilizado

No índice de acertos por instrumento normativo (IA_{in}), Tabela 20, com exceção dos IA_{in1} e IA_{in6} , que não foram recomendados para nenhum usuário, os demais sofreram influência do valor limiar para valores maiores do que 0,7, mostrando que houve diminuição do acerto.

Tabela 20 Índice de Acertos por Instrumento Normativo (IAin) (Definição I)

Valor limiar	IAin1	IAin2	IAin3	IAin4	IAin5	IAin6	IAin7	IAin8	IAin9	IAin10	IAinm
0,1		1	1	1	1		1	1	1	1	1
0,2		1	1	1	1		1	1	1	1	1
0,3		1	1	1	1		1	1	1	1	1
0,4		1	1	1	1		1	1	1	1	1
0,5		1	1	1	1		1	1	1	1	1
0,6		1	1	1	1		1	1	1	1	1
0,7		1	1	1	1		1	0,83	1	1	0,98
0,8		0,33	0,5	0,5	0,5		0,8	0,5	1	1	0,64
0,9		0,17	0,5	0,5	0,33		0,4	0,5	0,6	0,6	0,45
1,0		0,17	0,3	0,5	0,33		0,4	0,5	0,4	0,6	0,40

No índice de acertos por usuário (IA_{us}) (Tabela 21), os usuários IAus1, IAus8 e IAus9 não sofreram influência independentemente do valor limiar utilizado comprovando a capacidade de gerar recomendações pelo sistema de recomendação e destas serem aceitas por estes usuários.

Tabela 21 Índice de Acertos por Usuário (IAus) (Definição III)

Valor limiar	IAus1	IAus2	IAus3	IAus4	IAus5	IAus6	IAus7	IAus8	IAus9	IAus10	lausm
0,1	1	1	1	1	1	0	1	1	1	1	0,90
0,2	1	1	1	1	1	1	1	1	1	1	1
0,3	1	1	1	1	1	1	1	1	1	1	1
0,4	1	1	1	1	1	1	1	1	1	1	1
0,5	1	1	1	1	1	1	1	1	1	1	1
0,6	1	1	1	1	1	0	1	1	1	1	0,90
0,7	1	1	1	1	1	0	0,88	1	1	1	0,89
0,8	1	0,38	0,75	1	0,38	0	0,38	1	1	0,67	0,65
0,9	1	0,25	0,75	0,5	0,13	0	0	1	1	0,67	0,53
1,0	1	0,25	0,75	0,5	0,13	0	0	1	1	0	0,46

5.4.2.2 Os Resultados na FC

Uma vez que o algoritmo de filtragem colaborativa utilizado é baseado no KNN, um algoritmo de agrupamento, onde a predição para um dado usuário é realizada com base nos seus K vizinhos mais próximos, foram realizados experimentos dos diversos valores de K nos resultados da filtragem.

Os resultados em termos de recall estão representados na Figura 50, dos quais se conclui que não há diferença em função da quantidade de vizinhos em

relação ao recall. Uma vez que o recall é o mesmo para o k variando de 1 a 10, em cada valor de limiar calculado.

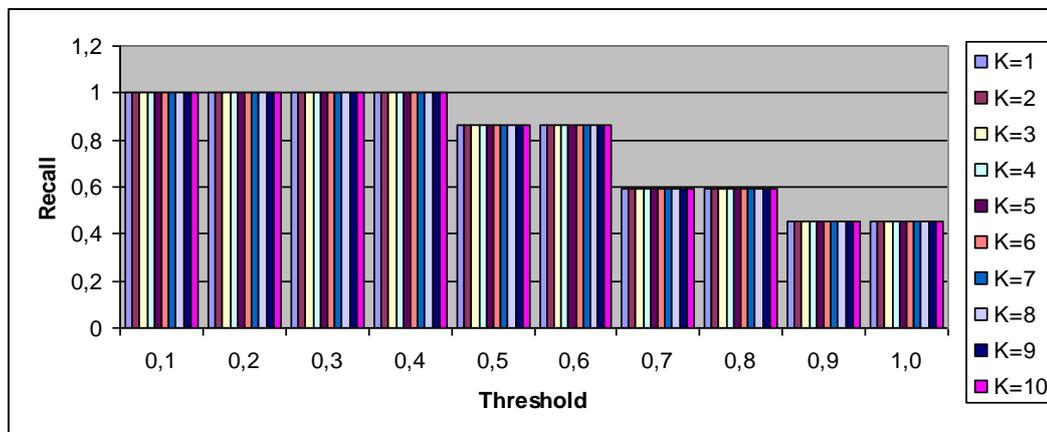


Figura 50 Influência do Parâmetro K, no recall, para os limiares de 10% a 100%

Os resultados em termos de precisão estão representados na Figura 51, os quais refletem a existência de uma pequena variação.

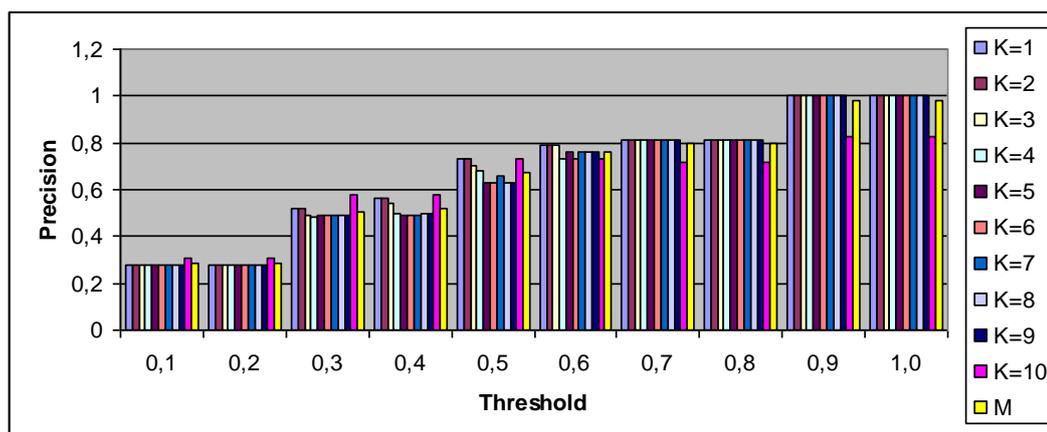


Figura 51 Influência do Parâmetro K, na precisão, para os limiares de 10% a 100%

Na FC nem todos os itens de informação (instrumentos normativos) no ONTOTRIB foram avaliados (Tabela 22), a fim de representar a realidade dos sistemas de recomendações colaborativos, os quais tendem a terem itens não avaliados. As avaliações variam de zero (0) a cinco (5) ocorreram em função que estes eram os valores apresentados ao usuários-especialistas que realizaram as respectivas avaliações.

Tabela 22 Avaliações Realizadas pelos Usuários

	1	2	3	4	5	6	7	8	9	10
US1		4	3	3	2		3	3	2	3
US2		3	3	2	4		2	4	4	3
US3		5	1	4	5		1	5	1	1
US4		1	1	5	1		1	5	1	1
US5		1	1	2	1		1	2	5	1
US6		1	1	2	2		1	1	2	2
US7		1	2	1			1		1	
US8		1	5				1			2
US9		1	1	1			1			5
US10		1	5	1			4	1	5	1

A Tabela 22 das avaliações dos itens de informação preenchida pelos usuários resultou na Tabela 23, quando do cálculo da similaridade do KNN, com uso da similaridade semântica e cálculo da predição.

Tabela 23 Matriz de Similaridade Obtida na FC, com K-3

	1	2	3	4	5	6	7	8	9	10
US1		32,30	85,63	72,06	37,78		52,30	72,06	72,06	38,97
US2		58,76	51,77	75,82	43,84		51,77	91,59	43,84	79,74
US3		55,00	48,33	81,67	48,33		48,33	75,00	48,33	55,00
US4		58,33	25,00	51,67	51,67		25,00	51,67	25,00	31,67
US5		25,86	46,84	42,93	16,34		39,85	42,93	39,85	18,87
US6		45,00	11,67	58,33	31,67		11,67	65,00	5,00	11,67
US7		41,47	8,82	34,12	26,76		8,82	40,44	26,76	8,82
US8		56,05	52,97	43,04	59,30		45,98	49,06	45,98	25,00
US9		42,89	41,91	28,90	38,00		28,90	48,00	35,90	38,00
US10		44,05	64,05	52,86	32,86		37,38	52,86	72,86	44,05

A Figura 52 mostra o gráfico do recall e da precisão, do algoritmo de filtragem colaborativa para os valores limiares variando entre 10% e 100%. Sendo que de 0,3 até 0,7 há um decaimento do recall com um aumento da precisão. De 0,7 a 0,9 o recall melhora e piora, com valores maiores do que 0,9 sendo nulo. A precisão de 0,7 a 0,9 tem sempre uma diminuição, o qual não tem sempre a mesma constância e de 0,9 em diante é nulo. O resultado ruim, para valores limiares maiores do que 0,3, foi resultado do baixo número de usuários, o qual afeta a filtragem colaborativa.

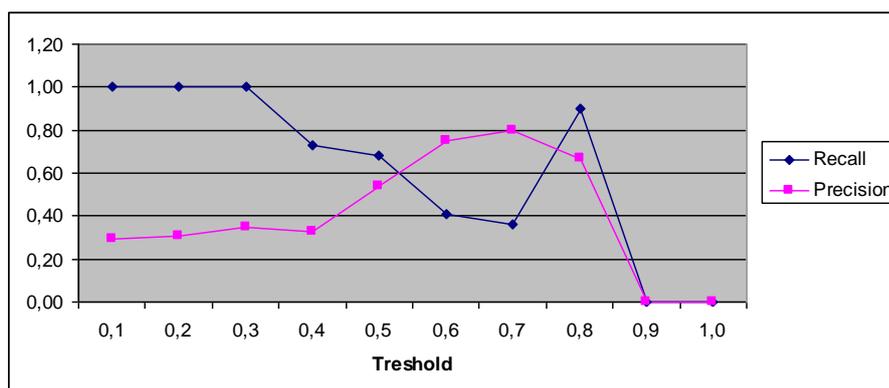


Figura 52 Gráfico do recall e da precisão da FC, em função do limiar utilizado, para $k = 3$

A causa da instabilidade para valores limiares maiores do que 0,3 é a baixa quantidade de usuários, mas como não é foco deste trabalho a análise além do necessário dos demais itens associados à filtragem, será utilizado na filtragem híbrida as filtragens baseada no conteúdo e colaborativa para valores limiares iguais a 0,3, os quais não sofreram esta influência.

5.4.2.3 Os Resultados na FH alternada

A Tabela 24 mostra os resultados dos experimentos da filtragem híbrida, com uso da técnica alternada, na qual se tem uma função de escolha entre o tipo de filtragem a ser utilizada. No caso deste experimento a escolha entre as filtragens colaborativa e baseada no conteúdo foi realizada com base no maior valor entre as duas.

Tabela 24 Matriz Similaridade Obtidos na FH alternada

	1	2	3	4	5	6	7	8	9	10
US1	0	100	100	100	100	0	100	100	100	100
US2	0	72	72	72	80	0	100	80	86	100
US3	0	100	86	100	100	0	86	100	86	86
US4	0	86	86	100	86	0	86	100	86	86
US5	0	72	72	72	72	0	86	72	100	86
US6	0	72	72	72	72	0	72	72	72	72
US7	0	72	72	72	72	0	86	86	86	86
US8	0	86	100	86	86	0	72	86	72	72
US9	0	72	72	72	72	0	100	72	86	100
US10	0	73	100	76	76	0	80	73	100	86

Na FH, com uso da técnica alternada, o recall permaneceu estável até 0,8, quando muda de 1 para 0,95, em cujo valor permanece constante. Diferentemente, a precisão permanece estável apenas até 0,7, após o que sobe, ficando estável em 0,9, quando passa a ter o mesmo valor do recall.

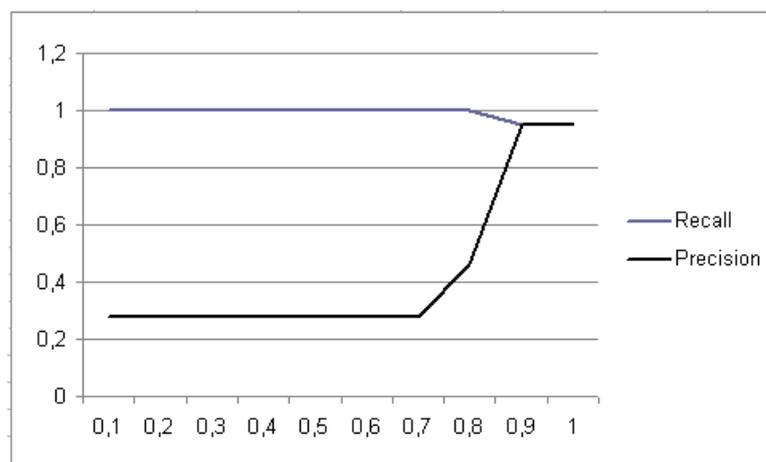


Figura 53 Gráfico do recall e da precisão na FH alternada, em função do limiar utilizado

Na filtragem híbrida, com uso da técnica alternada, alcançou-se alta precisão e recall, para valores limiares maiores do que 0,8.

5.4.2.4 Os Resultados na FH, com a técnica ponderada, por Intersecção

Na filtragem híbrida, com uso da técnica ponderada, por intersecção, são selecionadas apenas as recomendações dos itens de informação comuns à filtragem colaborativa e baseada no conteúdo. Nesta modalidade obteve-se os resultados mostrados na Tabela 25.

Tabela 25 Matriz de Similaridade Obtida na FH Intersecção

	1	2	3	4	5	6	7	8	9	10
US1	0	90	80	80	70	0	80	80	70	80
US2	0	66	66	56	76	0	70	76	83	80
US3	0	100	53	90	100	0	53	100	53	53
US4	0	53	53	100	53	0	53	100	53	53
US5	0	46	46	56	46	0	53	56	100	53
US6	0	46	46	56	56	0	46	46	56	56
US7	0	46	56	46	38.92	0	53	61.12	53	51.3
US8	0	53	100	55.5	53	0	46	55.5	63.5	56
US9	0	46	46	46	55	0	60	60	62	100
US10	0	46.57	95	48	53.71	0	78	46.57	95	53

Dos dados da Tabela 25, obteve-se a Figura 54, na qual até 0,7 o recall ficou estável em 1 e a partir daí decaiu até o valor de 0,36. Já a precisão ficou estável até 0,4, alcançando o ápice em 0,7 e decaindo a partir daí. Ou seja, a FH ponderada por intersecção conseguiu melhorar o recall, em relação ao apresentado pela FC, mas não a precisão.

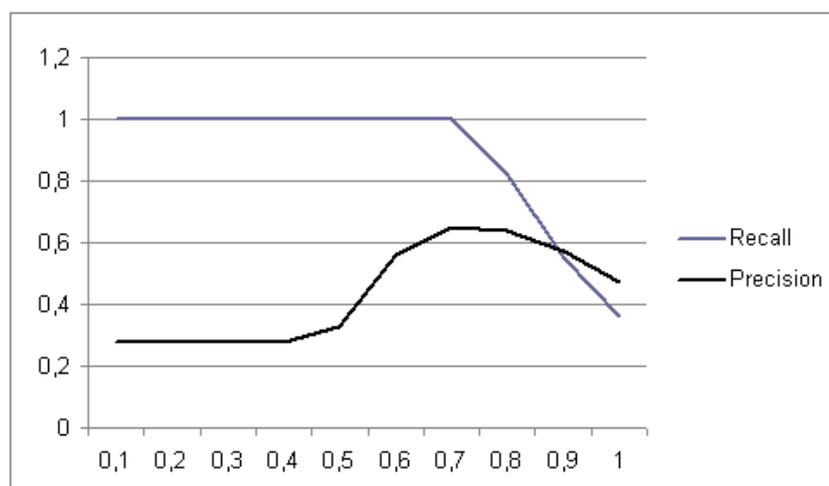


Figura 54 Gráfico do recall e da precisão da FH ponderada por intersecção, em função do limiar utilizado

5.4.2.5 Os Resultados na FH ponderada por pesos

Na filtragem híbrida, com uso da técnica ponderada, por pesos, há uma função com pesos para a FBC e para a FC. O resultado híbrido será o somatório das parcelas do produto dos pesos pelas respectivas filtragens. Na Tabela 26 são apresentados os resultados dos experimentos com os pesos da FBC igual a 0,4 e da FC igual a 0,6.

Tabela 26 Matriz de Similaridade Obtida na FH ponderada

	1	2	3	4	5	6	7	8	9	10
US1	0	90	80	80	70	0	80	80	70	80
US2	0	66	66	56	76	0	70	76	83	80
US3	0	100	53	90	100	0	53	100	53	53
US4	0	53	53	100	53	0	53	100	53	53
US5	0	46	46	56	46	0	53	56	100	53
US6	0	46	46	56	56	0	46	46	56	56
US7	0	46	56	46	38.92	0	53	61.12	53	51.3
US8	0	53	100	55.5	53	0	46	55.5	63.5	56
US9	0	46	46	46	55	0	60	60	62	100
US10	0	46.57	95	48	53.71	0	78	46.57	95	53

Dos dados da Tabela 26, obteve-se a Figura 55, cujo recall manteve-se estável em 0.64 até o valor limiar de 0.6, após o que decresceu para 0.5, 0.23, 0.14 e 0.5. A precisão, por sua vez, ficou em 0.19, até o valor limiar de 0.5, após o que decaiu até atingir o valor de 0.02, no valor limiar de 1. Ou seja, na FH ponderada, por pesos, teve-se um recall menor do que o da FH ponderada, por intersecção, e o decaimento a partir de 0,7 foi de forma mais rápida. A precisão foi inferior a registrada na FH ponderada, por intersecção. Logo, para os pesos de FBC=0.4 e FC=0,6 a FH ponderada, por pesos foi mais sensível aos problemas apresentados na FC.

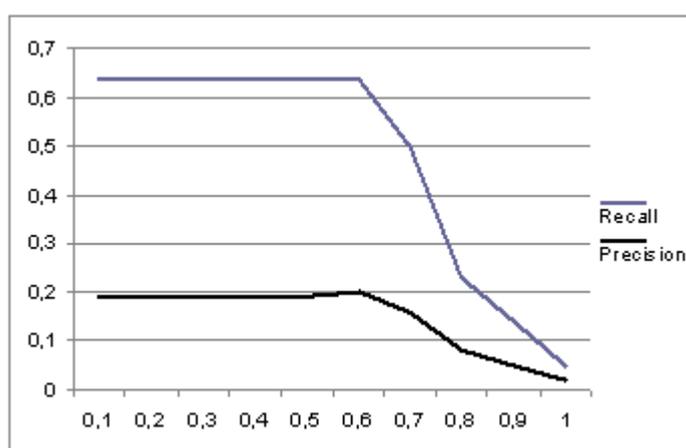


Figura 55 Gráfico do recall e da precisão da FH ponderada, por pesos, com o peso do FBC = 0.4 e FC = 0.6, em função do limiar utilizado

Na FH ponderada, por pesos, (FH PP) é possível a calibração da recomendação por meio dos pesos. Da Figura 56 a Figura 64 estão representados nos gráficos o estudo realizado mudança do recall e da precisão, em função da variação dos pesos da FBC de 0.1 a 0.9 e a da FC de 0.9 a 0.1. Logo, na Figura 55, FH PP, FBC 0.1 e FC 0.9 apresenta o gráfico da filtragem híbrida, na técnica ponderada por pesos, com o peso da FBC sendo 0.1 e o peso da FC sendo 0.9.

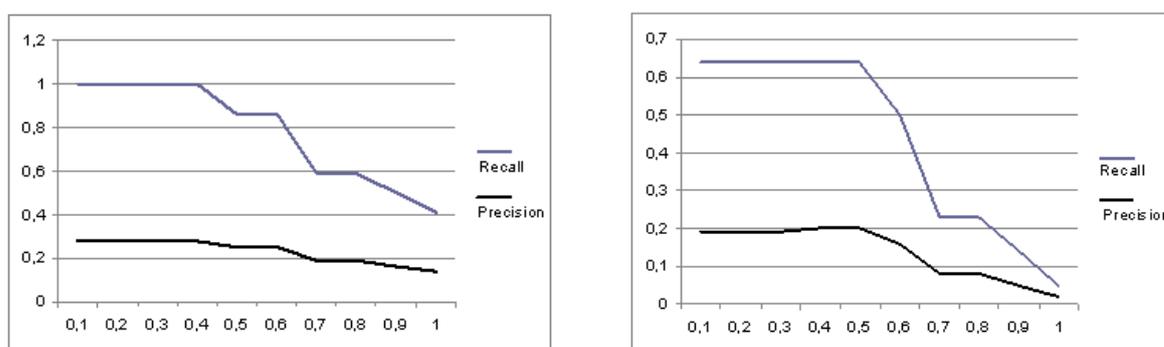


Figura 56 FH PP, FBC 0.1 e FC 0.9

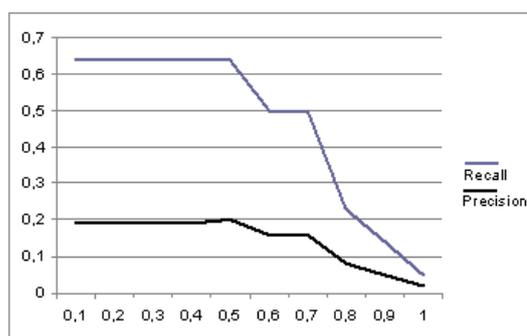


Figura 58 FH PP, FBC 0.3 e FC 0.7

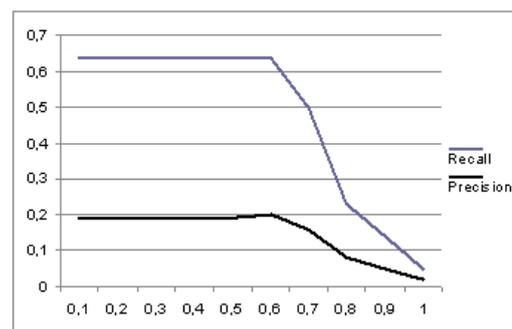


Figura 59 FH PP, FBC 0.4 e FC 0.6

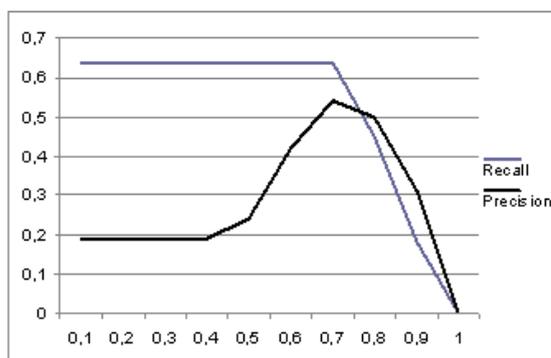


Figura 60 FH PP, FBC 0.5 e FC 0.5

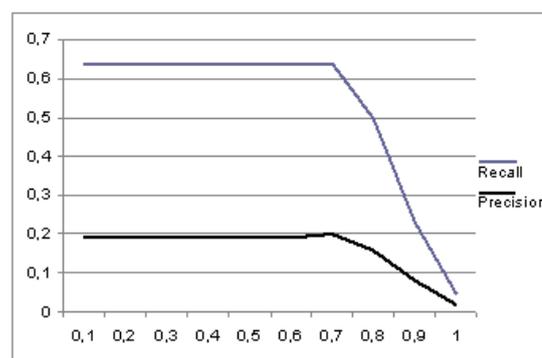


Figura 61 FH PP, FBC 0.6 e FC 0.4

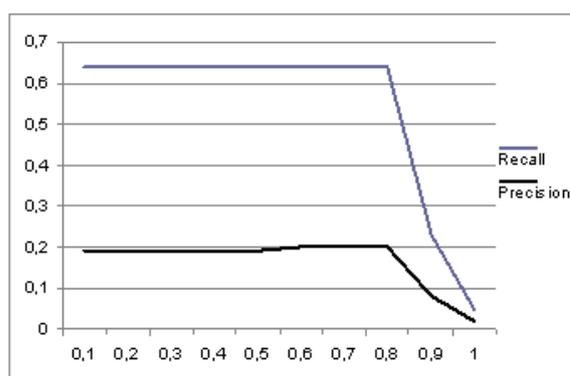


Figura 62 FH PP, FBC 0.7 e FC 0.3

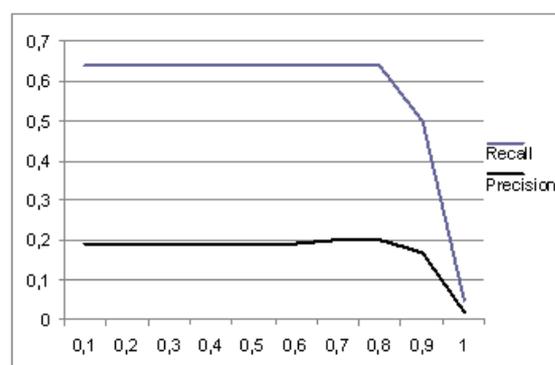


Figura 63 FH PP, FBC 0.8 e FC 0.2

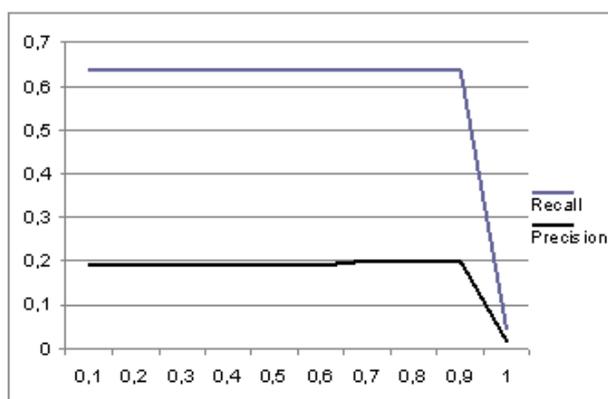


Figura 64 FH PP, FBC 0.9 e FC 0.1

Da Figura 56 a Figura 64 observa-se que quanto maior o peso utilizado para a FBC, mais estável é o gráfico, com alto recall. Apenas quando os pesos foram iguais a 0.5 é que a precisão elevou-se para 0.55. Logo, na FH ponderada, por pesos, quando os pesos da FC e da FBC são iguais a 0,5 tem-se o mesmo gráfico da FH ponderada, por intersecção, e quanto maior o peso da filtragem mais estável maior (FBC) e menor o da instável (FC) mais estável será a FH ponderada, por pesos.

5.5 Análises dos resultado das experiências off-line

Nas filtragens colaborativa e baseada no conteúdo separadas, a FBC apresentou recall máximo 1 e baixa precisão, para valor limiar até 0,7. Para valores limiares maiores do que 0,9 foram observados valores de recall e precisão maiores do que 0,8. Na FC se teve recall máximo 1, para limiar até 0,3. Para valores maiores do que 0,3 os experimentos se mostraram instáveis terminando por ficarem em 0, para valores limiares maiores que 0,9.

Na FH, com uso da técnica alternada, obteve-se, entre os experimentos realizados o gráfico mais estável, minimizando a instabilidade da FC, com alta precisão e recall, para valores limiares maiores do que 0,8.

Na FH, com uso das técnicas ponderada, por Intersecção apresentou o mesmo problema da FH, com uso das técnicas ponderada, por pesos, para FBC=0.5 e FC=0.5, cujo melhor recall e precisão ocorreram para valor de limiar 0.8, no qual se obteve o valor de recall e precisão igual a 0.52.

Na FH, com uso das técnicas ponderada, por pesos, realizou-se a calibração dos pesos da FBC e da FC, os quais estão consolidados na Figura 65 e Figura 66, as quais apontam discrepância da FH ponderada, com peso de FBC=0.1 e peso de FC=0.9. Até o valor limiar de 0,5, os demais experimentos com o exceção do anterior apresentaram o mesmo recall. Na Figura 65 foram colocadas em um único gráfico todas os recall obtidos com os pesos da FBC variando de 0.1 a 0.9 e os pesos da FC variando de 0.9 a 0.1. Na Figura 66 foram colocadas em um único gráfico todas as Precisões obtidas com os pesos da FBC variando de 0.1 a 0.9 e os

pesos da FC variando de 0.9 a 0.1. Em ambas as figuras na linha R(X,Y) ou P(X,Y) o X representa o peso dado à FBC e o Y o peso dado à FC.

Na precisão além da discrepância da FH ponderada, com peso de FBC=0.1 e peso de FC=0.9, detectou-se a da FH ponderada, com peso de FBC=0.6 e peso de FC=0.4. E os demais permaneceram estáveis para valores limiares menores ou igual a 0.5.

Os experimentos realizados mostraram que o uso da filtragem híbrida com as técnicas alternada e ponderada, apresenta melhores resultados, do que o uso isolado das filtragens colaborativa e baseada no conteúdo.

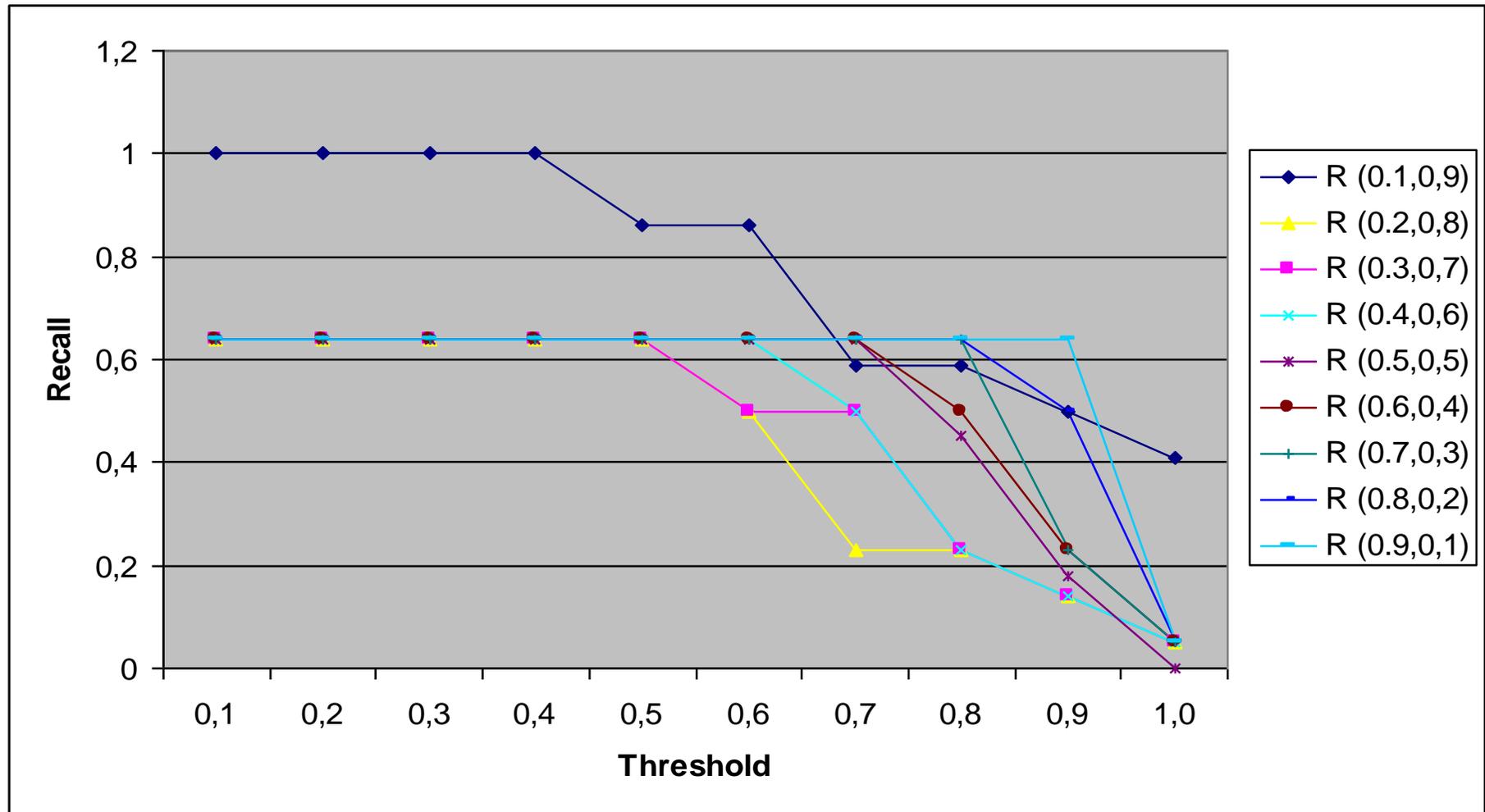


Figura 65 Gráfico do recall da FH ponderada, por pesos, com o peso de FBC = 0.1 e FC = 0.9 a FBC = 0.9 e FC = 0.1, em função do limiar utilizado

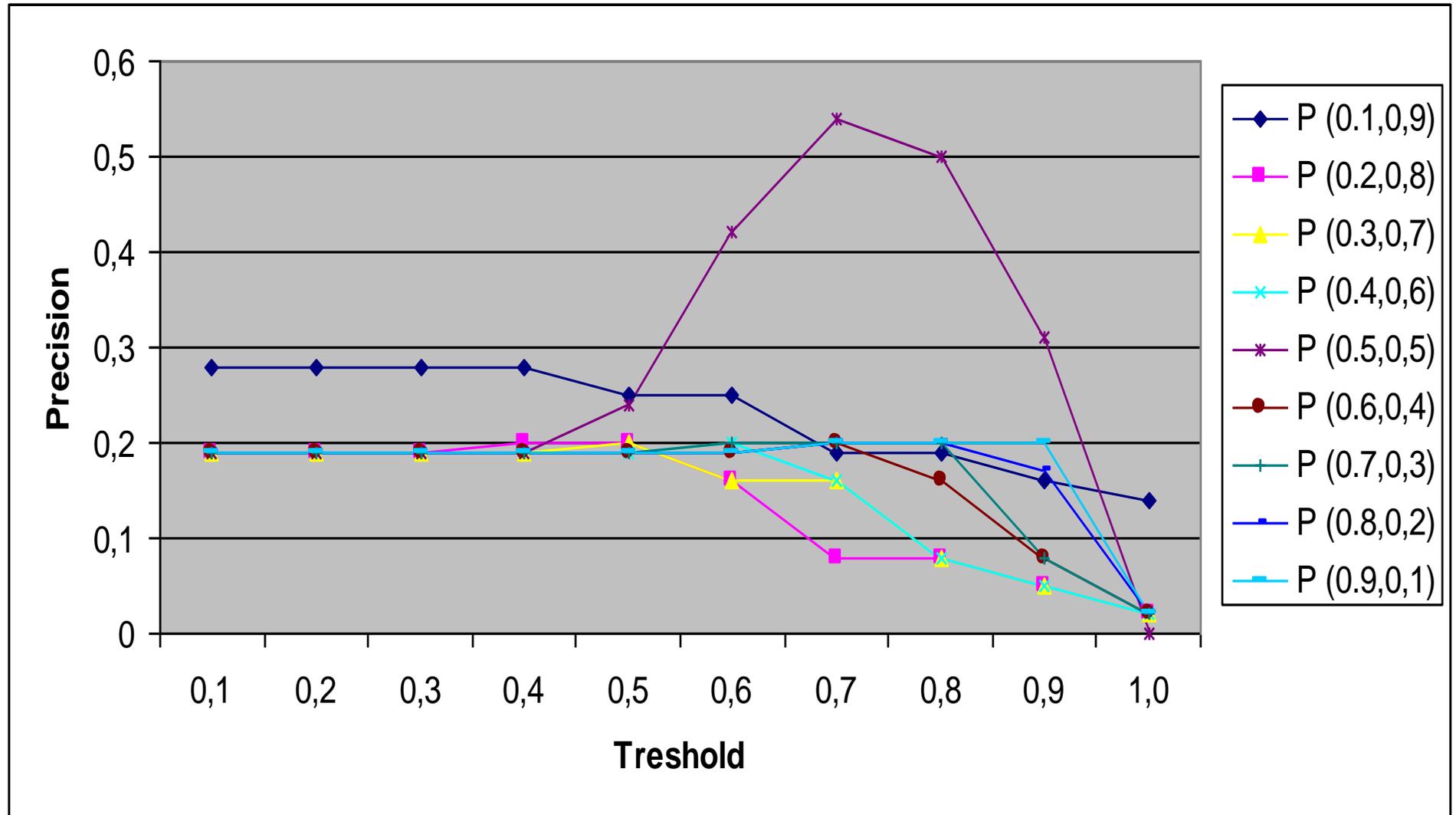


Figura 66 Gráfico da precisão da FH ponderada, por pesos, com o peso de FBC = 0.1 e FC = 0.9 a FBC = 0.9 e FC = 0.1, em função do limiar utilizado

Na Figura 67 foi colocado o recall das diversas filtragens, onde se observa que para valor limiar até 0,3 todas as filtragens apresentam alto recall. De 0,4 a 0,7 a filtragem colaborativa (FC) é a única cujo valor diminui. Em 0,8 somente a filtragem híbrida alternada (FHS) permanece estável em 1.0, sendo a mais efetiva. Em 0,8 as filtragens baseada no conteúdo (FBC) e colaborativa apresentam melhor recall do que a filtragem híbrida ponderada por intersecção (FHI) e por pesos (FHW). Em 0,9 a FHS apresenta a melhor efetividade, seguida pela FBC, FHI e FHW, com as últimas com o mesmo valor.

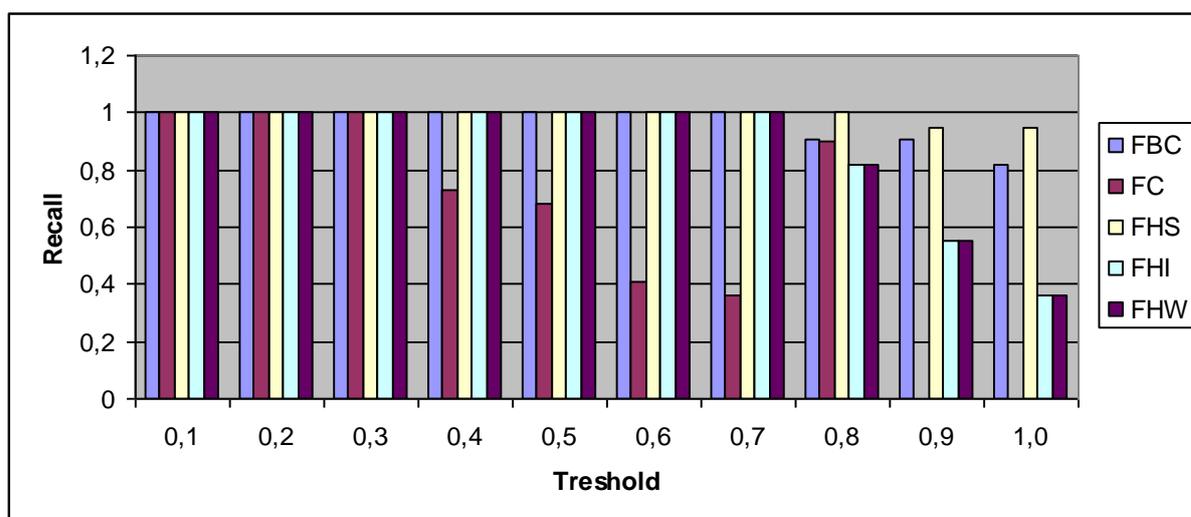


Figura 67 Recall das filtragens FBC, FC, FHC, FHI e FHW

Na Figura 68, foi colocada a precisão obtida nas diversas filtragens. De 0.1 a 0.8 observa-se uma melhor efetividade da filtragem colaborativa. Até 0.4 as demais apresentam a mesma efetividade. De 0.5 a 0.8 as filtragens híbridas ponderadas por intersecção e por peso apresentam efetividade menor do que a colaborativa mais superior às demais. Em 0.9 a precisão da filtragem baseada no conteúdo e híbrida alternada são iguais e as maiores. Em 1.0 a filtragem híbrida alternada passa a ser maior do que as demais.

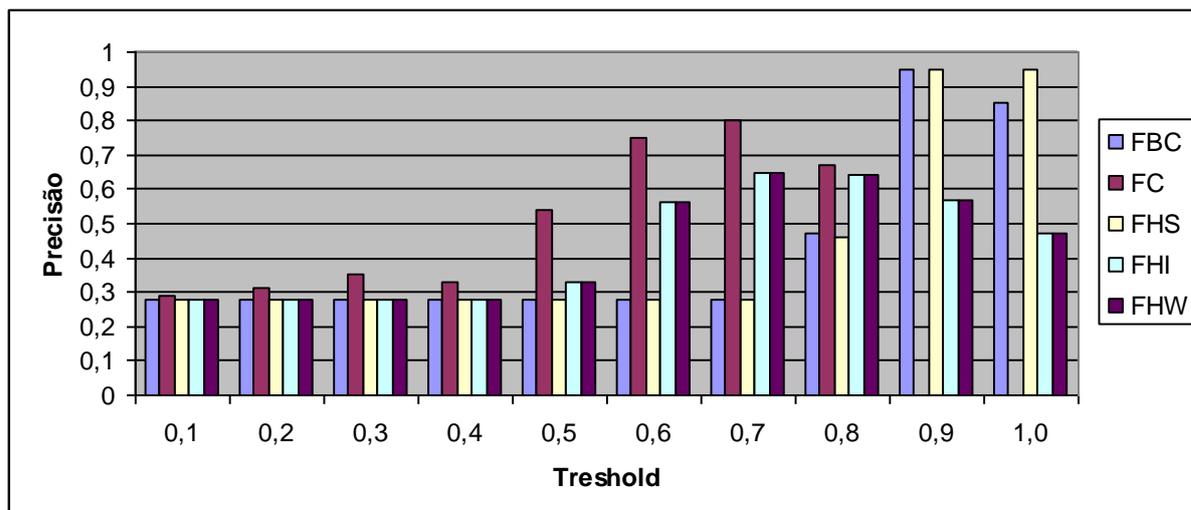


Figura 68 Precisão das filtragens FBC, FC, FHC, FHI e FHW

5.6 Experiências On-line

Nos testes on-line foi utilizado o feedback explícito, no qual o usuário atribui à recomendação uma nota de 1 a 5, significando a importância da recomendação realizada, onde 1 era pouco relevante e 5 muito relevante. Posteriormente se realizou para cada usuário a tabulação, foi verificada a correlação, em relação a cada técnica de recomendação.

5.6.1 O Feedback dos Usuários

A partir da tela de feedback os usuários expressaram os seus interesses, de um (1) a cinco (5), nos itens recomendados, o qual está expresso na Tabela 27, a qual apresentou valores próximos a da FC, em muitos casos, mas não os mesmos, no seu geral.

As células da tabela sem avaliação expressam o desinteresse no instrumento normativo, independentemente do motivo: desconhecimento, inadequação do conteúdo, etc.

Tabela 27 Matriz Feedback Explícito dos Usuários

	1	2	3	4	5	6	7	8	9	10
US1	0	4	3	3	2		3	3	2	3
US2	0	3	4	2	4		2	4	4	4
US3	0	4	1	4	5		1	5	1	2
US4	0	1	1	5	1		1	4	1	1
US5	0	2	1	2	1		1	2	4	1
US6	0	1	1	2	2		1	1	2	2
US7	0	1	2	1			1		1	
US8	0	1	4				1			2
US9	0	1	2	2			1			5
US10	0	1	4	1			4	1	5	1

Os valores obtidos no feedback explícito foram multiplicados por 20, a fim de se poder comparar os resultados obtido nas diversas matrizes de similaridade, uma vez que nas diversas matrizes de similaridade os valores variam de 0 a 100.

Apesar de que os resultados da FH, com uso da técnica alternada foram mais estáveis e efetivos, nos testes off-line, foi confrontado o feedback do usuário com as recomendações em cada tipo de filtragem, para se identificar com qual o usuário se aproximava mais.

Os resultados foram avaliados em termos das suas correlações em relação ao feedback realizado, Figura 69.

Na Figura 69, FBC significa filtragem baseada no conteúdo; FC, filtragem colaborativa; FHS, filtragem híbrida alternada; FHI, filtragem híbrida ponderada por Interseção e FHW, filtragem híbrida ponderada por Peso.

A filtragem híbrida alternada recomendou com o mesmo valor da FBC, para 1, 4, 5, 6,8 e 9 usuários. Ela apresentou resultados mais efetivos do que a FBC, com 2, 3, 7 e 10 usuários. A filtragem híbrida alternada apresentou resultados inferiores às filtragem híbrida ponderada por Interseção e filtragem híbrida ponderada por peso.

A filtragem híbrida ponderada por Interseção (FHI) recomendou sempre com melhor correlação em relação ao feedback dos usuários do que a FBC e a FC. A FHI recomendou com o mesmo valor da filtragem híbrida ponderada por Peso, com 1, 2,3,4,5,6 e 9 usuários, superou esta com 7, 8 e 10 usuários.

A filtragem híbrida ponderada por peso (FHW), apesar de não ter sido mais efetiva do que a FHI, conseguiu em alguns casos igualar e sempre foi superior do que a FBC e a FC.

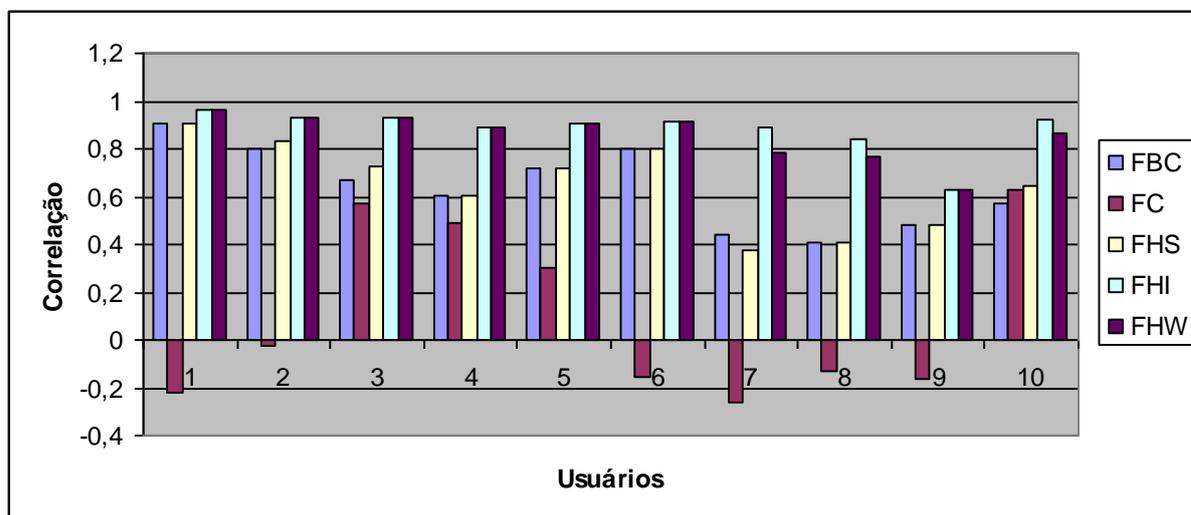


Figura 69 Correlação dos resultados das filtragens com o feedback dos usuários

No anexo II há um detalhamento do processo de análise das filtragens com o feedback.

5.7 Comparativo com os resultados de outros autores

No trabalho de (TORRES, 2004) são apresentados os algoritmos pure-CBF, o qual corresponde ao FBC isolado e pure-CF, o qual corresponde ao FC isolado deste trabalho. Apesar de (TORRES, 2004) ter experimentado com diferentes técnicas de filtragem ele não chegou a fazer um comparativo das filtragens isoladas com as híbridas. Ele apresenta comparações só entre a FBC pura, a combinada e separada e entre a FC pura e densa. Em relação à filtragem híbrida, ele analisou a filtragem híbrida com características aumentadas e a mista, mas em ambos os casos da análise foi baseada nas variações de configurações possíveis, de cada tipo de filtragem híbrida.

Em relação aos testes on-line, o resultado final de (TORRES, 2004) resumiu-se a apontar se o usuário agradava-se ou desagradava-se mais com uma, ou outra recomendação de alguma filtragem. Nos resultados apresentados por

(TORRES, 2004) o usuário ficou mais satisfeito com a FBC, seguido pela híbrida mista, com uso da fusão, seguido pela filtragem híbrida com características aumentadas, em que o algoritmo FBC foi combinado com o FC. Em terceiro e quarto lugar ficaram os algoritmos FC puro e FC-FBC separado. Observa-se que em (TORRES, 2004) os resultados dos experimentos com as técnicas híbridas não chegaram a ser os melhores para os usuários. A comunidade científica tende a ter pouca socialização, trabalhando-se menos de forma colaborativa, do que de forma baseada no conteúdo. Tanto que normalmente, a partir de um artigo científico, procura-se ler os demais artigos referenciados, a fim de se ter o mesmo conhecimento do autor do artigo inicial, em vez de se tentar interagir com o autor.

Já a comunidade jurídica, tanto se baseia na norma (conteúdo), quanto na doutrina (domínio das lições, ensinamentos e descrições explicativas do direito posto, elaboradas pelos mestres e pelos juristas especializados) e na jurisprudência (decisões reiteradas). O estudo da norma jurídica reflete a filtragem baseada no conteúdo, a doutrina reflete a filtragem colaborativa. Por isso que neste trabalho, os usuários estiveram mais satisfeitos com as filtrações híbridas do que os de (TORRES, 2004), pois no Direito há um maior uso de ambas as filtrações, o que é o foco da filtragem híbrida.

5.8 Considerações Finais

Para o sistema de recomendação INFOTRIB, o gráfico off-line de recall (Figura 67) que apontavam uma preponderância da filtragem híbrida alternada com uma significativa efetividade da filtragem baseada no conteúdo, bem como o gráfico off-line da precisão (Figura 68), na qual a filtragem colaborativa foi preponderante sendo depois superada pela filtragem alternada e colaborativa, não foram confirmados nos testes on-line.

Nos testes on-line, a filtragem híbrida ponderada por intersecção (FHI), conforme (Figura 69), apresenta os valores mais correlacionados com o feedback do usuário, para o domínio tributário e instâncias definidas no ONTOTRIB. Ou seja, dentre as diversas filtrações utilizadas, a FHI foi o mais efetivo para os usuários. A filtragem baseada no conteúdo sempre se correlacionou positivamente,

independentemente da quantidade de usuários existentes, comprovando a característica desta filtragem, da similaridade ser do item de informação com outros anteriormente consultados pelo usuário e não em relação a outros usuários. A filtragem colaborativa apresentou recomendações mais e menos correlacionadas, conforme os grupos de usuários utilizados fossem mais coeso, disperso ou contraditório (tivesse ovelhas-negras). Interessante observar que o fato de um determinado subconjunto de usuários ter tido correlação negativa não implicou que o seguinte obrigatoriamente tivesse, possibilitando recomendações inesperadamente boas, que é uma característica desta filtragem.

A ONTOTRIB mostrou ser compatível com o trabalho de MARTINS (2006), onde ambos tratam do desenvolvimento de uma ontologia jurídica no ramo do Direito Tributário e consistentes com os elementos do Direito Tributário (SABBAG, 2005).

6. ESTUDO DE CASO: O SISTEMA INFOTRIB

6.1 Introdução

No capítulo 4 foram mostrados os modelos genéricos do ONTOSERS, nos quais se tem os objetivos específicos: “Modelar usuário”, “Filtrar Informação” e “Entregar Informação”. Neste capítulo os modelos genéricos do ONTOSERS e seus artefatos de software, desenvolvidos na MADEM, serão selecionados e adaptados para uso na aplicação específica INFOTRIB, desenvolvida conforme as diretrizes da metodologia MAAEM.

O INFOTRIB é um sistema de recomendações na área do Direito Tributário, que será utilizado para realizar a avaliação da família de sistemas de recomendações ONTOSERS.

A análise, projeto e implementação dos agentes de filtragem híbrida terá por foco o desenvolvimento de agente deliberativos, utilizando JADE com Jess (BELLIFEMINE, CAIRE et ali. 2003).

A filtragem híbrida fará uso da filtragem baseada no conteúdo e da filtragem colaborativa. As técnicas de filtragem híbrida utilizadas serão a ponderada e a alternada. A filtragem híbrida, com método ponderado será utilizado nas modalidades por peso e por intersecção. A filtragem híbrida, com uso do método ponderado, por pesos, é aquela em que as recomendações resultam da ponderação, segundo pesos predefinidos das respostas da FBC e FC. A filtragem híbrida, com uso do método ponderado, por intersecção, é a aquela em que as recomendações resultam na seleção apenas das recomendações comuns à FBC e a FC e a filtragem híbrida, com uso da técnica alternada, é aquela onde é utilizado um critério de seleção entre as recomendações (BURKE 2002) (BURKE 2007).

6.2 Engenharia de Requisitos do INFOTRIB

6.2.1 Modelo de Conceitos

A Figura 70 apresenta o modelo de conceitos da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, apareceram conceitos específicos do Direito Tributário, tais como: tributos, elementos normativos tributários e elementos tributários.

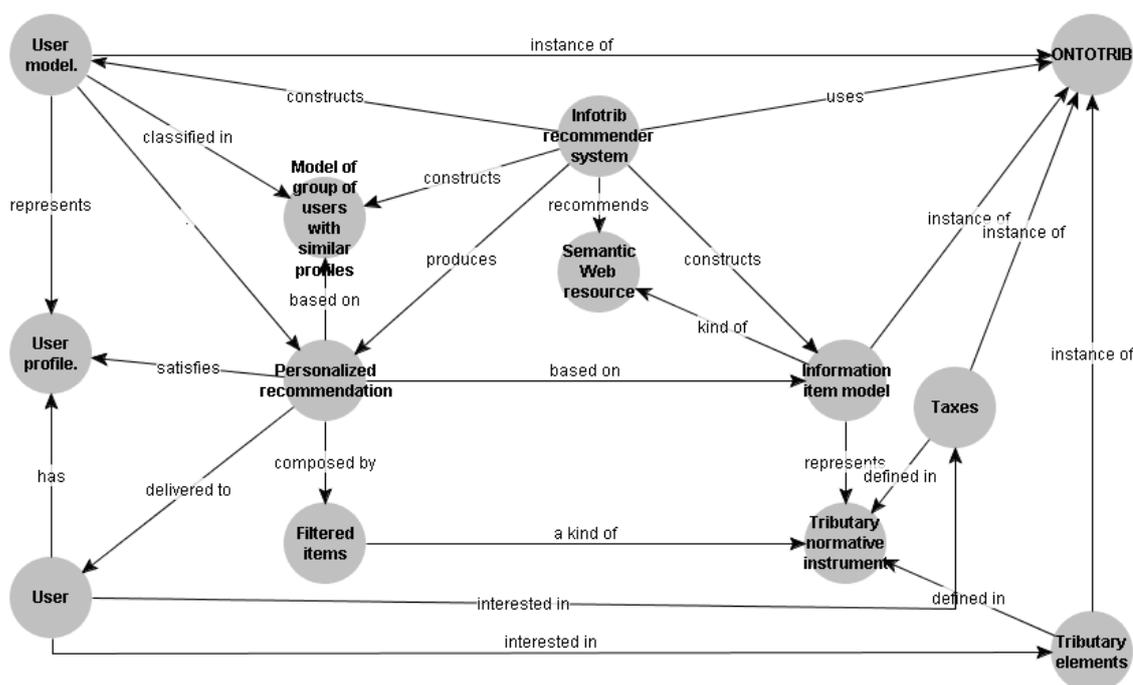


Figura 70 Modelo de Conceitos

6.2.2 Modelo de Objetivos

A Figura 71 apresenta o modelo de objetivos da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, apareceram todas as responsabilidades, onde o que não é utilizado pela FBC é utilizado na FC. O usuário passou a ser o usuário tributário e a fonte de informação passou a ser a ONTOTRIB. O objetivo geral é o de “Prover recomendações no domínio tributário”

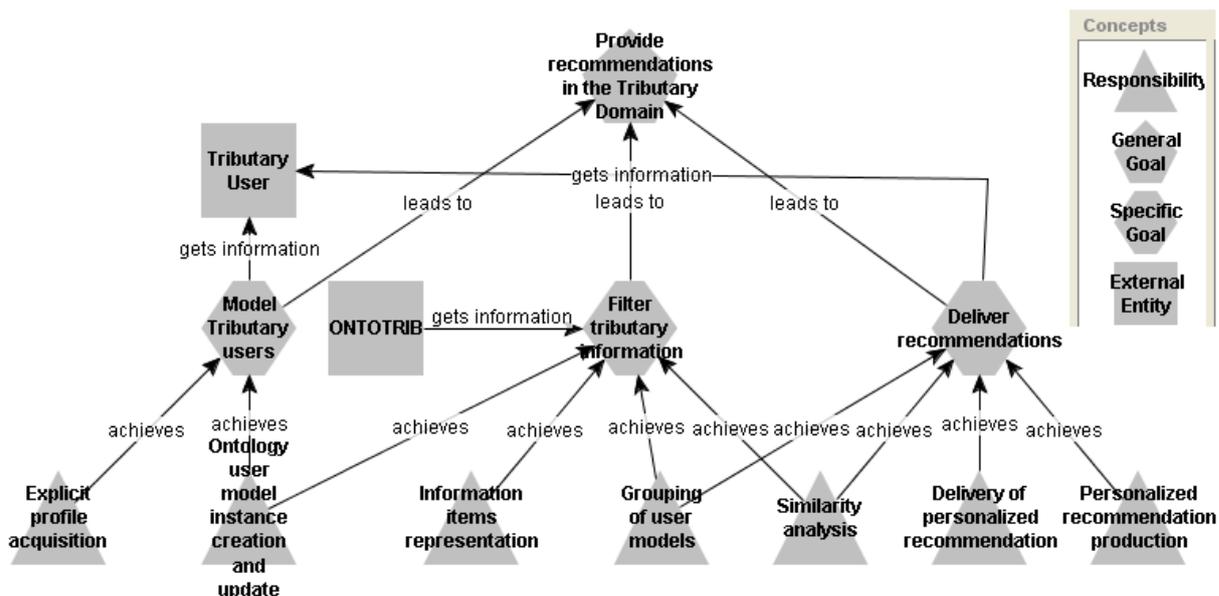


Figura 71 Modelo de Objetivos

6.2.3 Modelo de Papéis do INFOTRIB, do Objetivo Específico “Modelar Usuário”

A Figura 72 apresenta o modelo de papéis, do objetivo específico “Modelar Usuários”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, foi alterada a entidade externa, que agora é o usuário tributário.

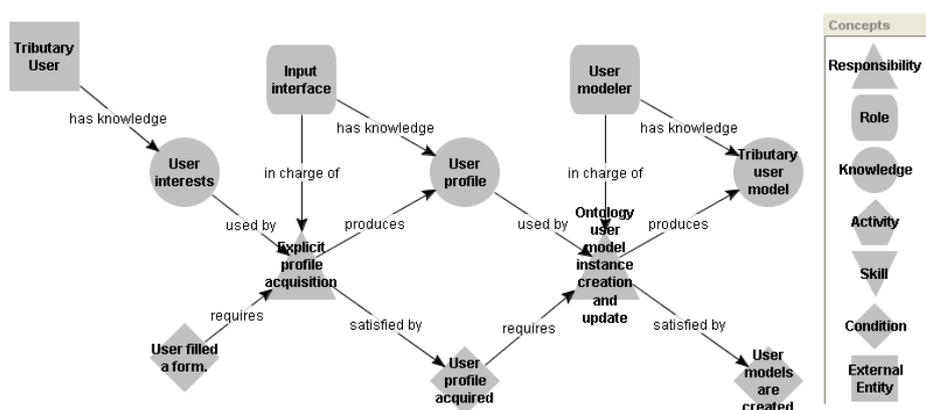


Figura 72 Modelo de Papéis do INFOTRIB, do Objetivo Específico “Modelar Usuário”

6.2.4 Modelo de Papéis do Objetivo Específico “Filtrar Informação”

A Figura 73 apresenta o modelo de papéis, do objetivo específico “Filtrar Informação”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, a fonte de informação passou a ser a ONTOTRIB.

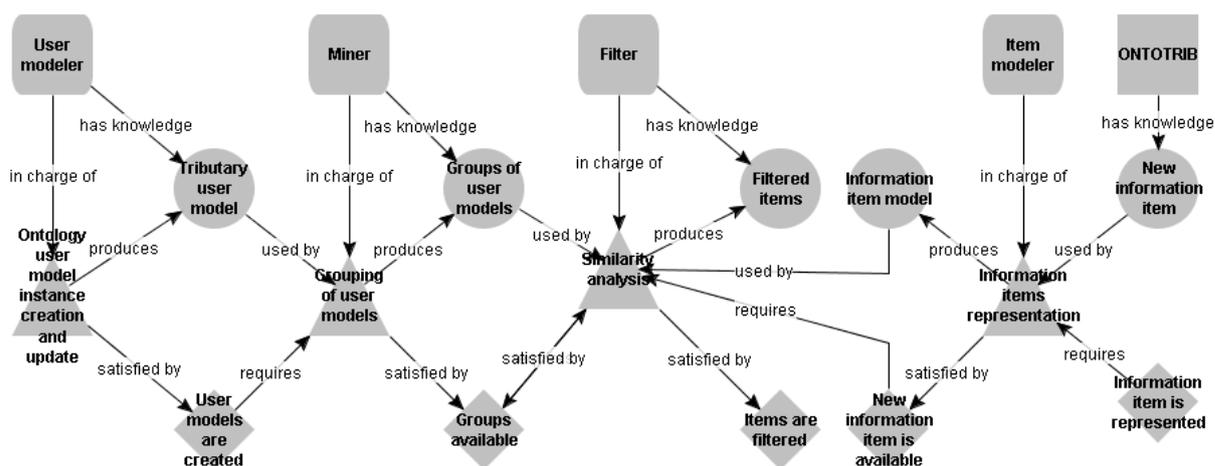


Figura 73 Modelo de Papéis do Objetivo Específico “Filtrar Informação”

6.2.5 Modelo de Papéis do Objetivo Específico “Entregar Informação”

Este modelo não sofreu modificações, em relação ao que fora apresentado na ONTOSERS (Figura 74).

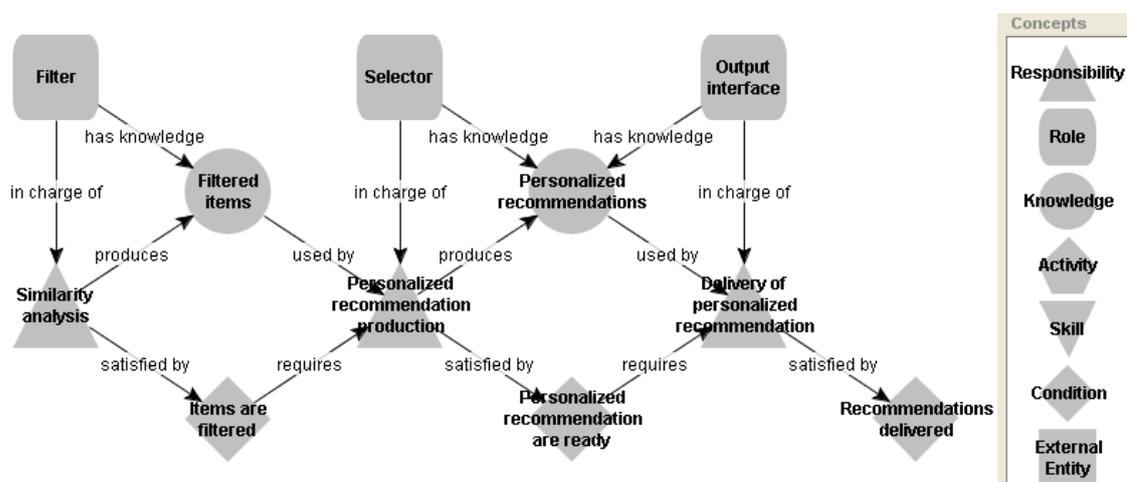


Figura 74 Modelo de Papéis do Objetivo Específico “Entregar Informação”

6.2.6 Modelo de Interações entre Papéis do Objetivo Específico “Modelar Usuário”

A Figura 75 apresenta o modelo de interações entre papéis, do objetivo específico “Modelar Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, o usuário passou a ser o usuário tributário.

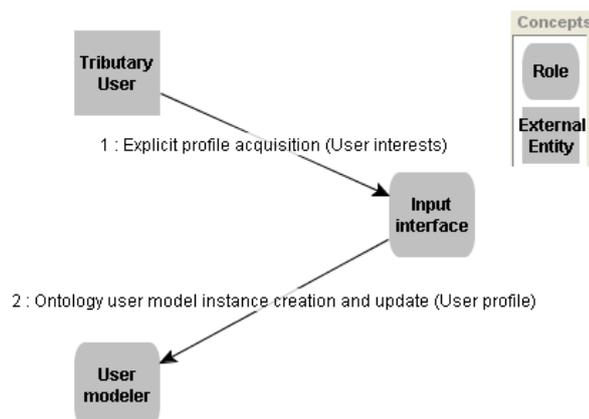


Figura 75 Modelo de Interação entre Papéis do Objetivo Específico “Modelar Usuário”

6.2.7 Modelo de Interações entre Papéis do Objetivo Específico “Filtrar Informação”

A Figura 76 apresenta o modelo de interações entre papéis, do objetivo específico “Filtrar Informação”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, a fonte de informação é a ONTOTRIB e o usuário é o usuário tributário.

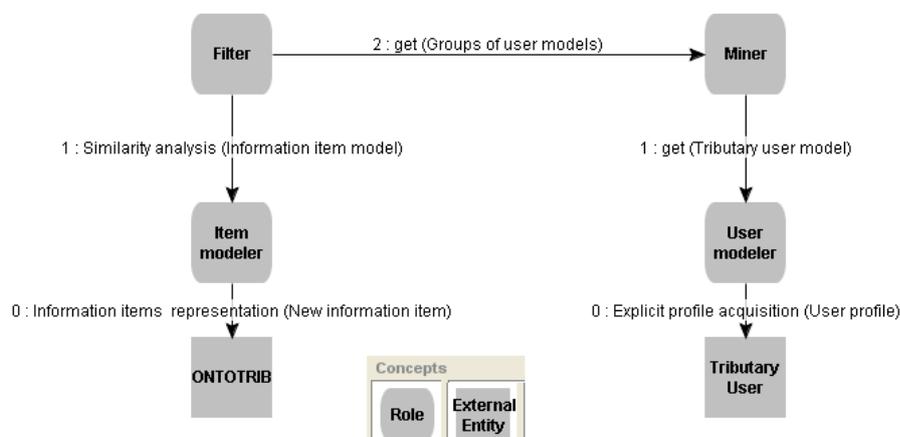


Figura 76 Modelo de Interação entre Papéis do Objetivo Específico “Filtrar Informação”

6.2.8 Modelo de Interações entre Papéis do Objetivo Específico “Entregar Recomendações”

A Figura 77 apresenta o modelo de interações entre papéis, do objetivo específico “Entregar Recomendações”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, o usuário é o usuário tributário.

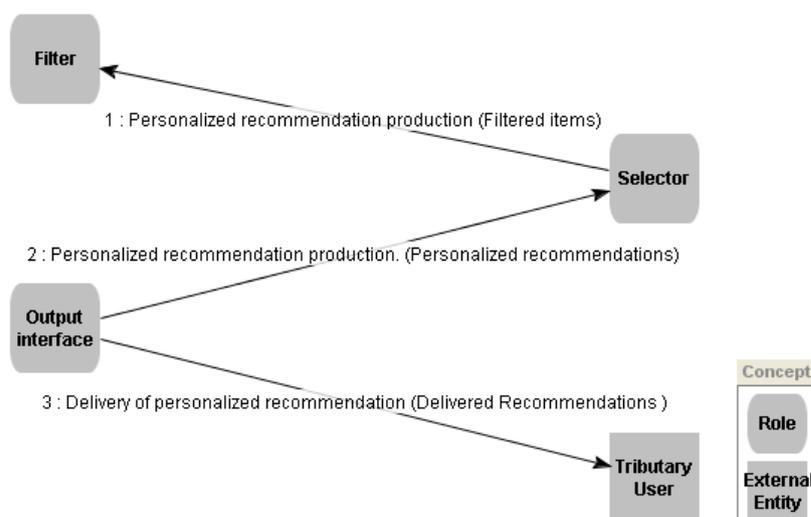


Figura 77 Modelo de Interação entre Papéis do Objetivo Específico “Entregar Recomendação”

6.2.9 Protótipo das Interfaces

As telas não são mais genéricas. Representando exatamente as telas que serão implementadas.



Figura 78 Tela Apresentação

A tela de Apresentação tem as informações sobre o usuário e o login para acesso ao INFONORMA (Figura 78).

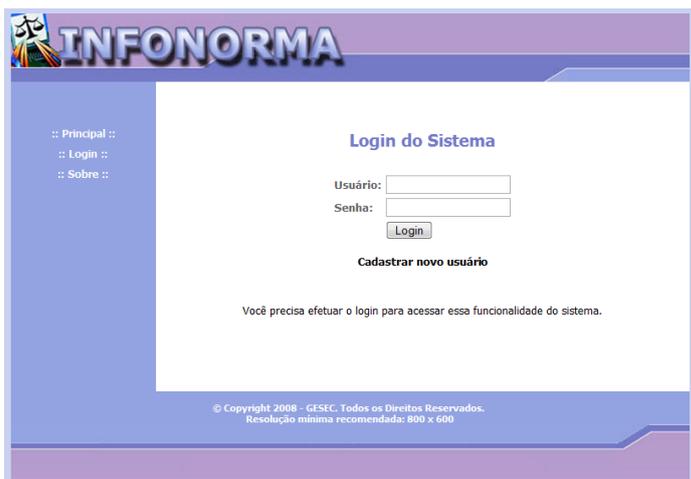


Figura 79 Tela Login



Figura 80 Tela Opções



Figura 81 Tela Perfil do Usuário

A tela de Login serve para o usuário se cadastrar ou acessar o INFONORMA (Figura 79).

Após o login o usuário tem acesso ao menu principal, no qual pode atualizar seu perfil ou ver as recomendações que o INFONORMA tem para ele (Figura 80).

Para atualizar o perfil o usuário tem-se as opções: Identificação, Tipos Normativos e Ramos Jurídicos (Figura 81).

Figura 82 Tela Identificação

Na tela de identificação o usuário há os campos: nome; e-mail, login e senha. O e-mail é para as entregas das recomendações off-line e o site para as on-line (no site) (Figura 82).

Figura 83 Tela Tipos Normativos

A tela Tipos Normativos permite que se escolha os diversos tipos normativos sobre os quais o usuário tem interesse (Figura 83).

Figura 84 Tela Ramos Jurídicos

A tela Ramos Jurídicos permite que se escolha os diversos ramos do Direitos sobre os quais o usuário tem interesse (Figura 84).



Figura 85 Tela Opções Extendida

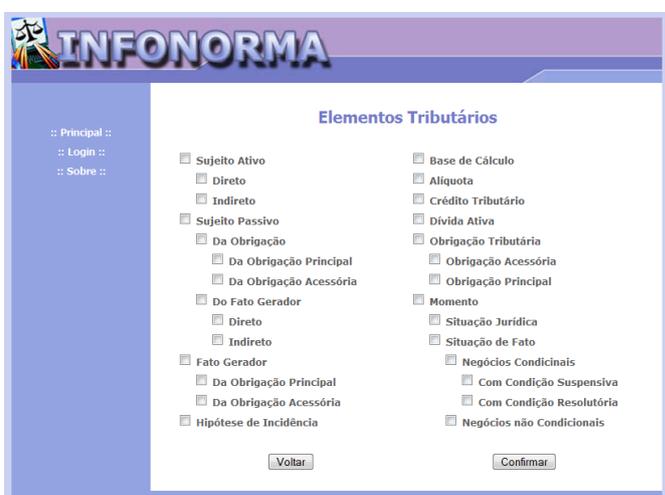


Figura 86 Tela Elementos Tributários



Figura 87 Tela Tributos

A tela opções estendida aparece para os usuários que tenham assinalado o ramo jurídico Direito Tributário, para o qual será necessário identificar os elementos e tributos sobre os quais se tem interesse (Figura 85).

Na tela Elementos Tributários são apresentados todos os elementos da relação tributária (Figura 86).

Na tela Tributos é possível especificar quais tributos das diversas esferas se tem interesse (Figura 87).



Figura 88 Tela Tributos Federais

A aba de Tributos Federais exibe as diversas modalidades nesta esfera administrativa, para que o usuário escolha as de seu interesse e passe a receber as respectivas recomendações relacionadas (Figura 88).

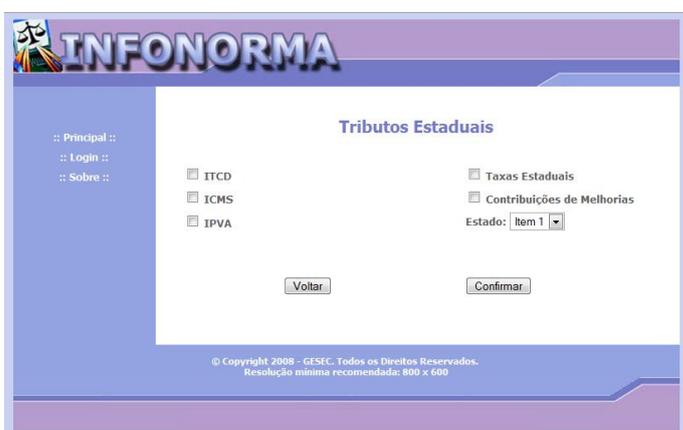


Figura 89 Tela Tributos Estaduais

Idem à anterior, com relação aos tributos estaduais, sendo que se tem de especificar de qual Estado (Figura 89).



Figura 90 Tela de Tributos Municipais

Idem à anterior, com relação aos tributos municipais, sendo que se tem de especificar de qual município (Figura 90).

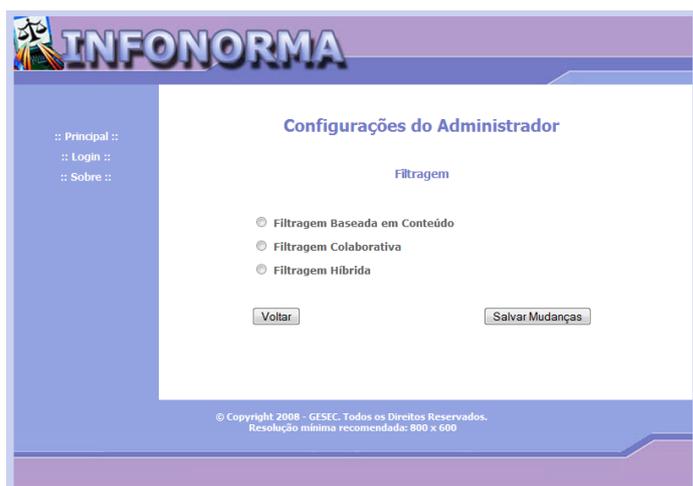


Figura 91 Configurações do Administrador

O INFONORMA tem o módulo de configurações do Administrador, o qual possibilita que ele escolha a filtragem a ser utilizada para as recomendações (Figura 91).

6.3 Projeto de Aplicação

6.3.1 Modelo de Conhecimentos da Sociedade Multiagente

A Figura 92 apresenta o modelo de conhecimento da sociedade multiagente, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, os conhecimentos foram especializados para o domínio tributário.

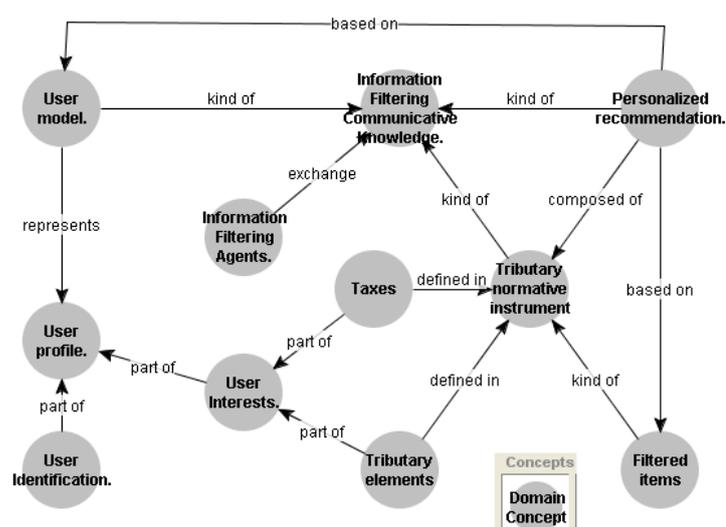


Figura 92 Modelo de Conhecimento da Sociedade Multiagente

6.3.2 Modelo da Sociedade Multiagente do Agente “Interface do Usuário”

A Figura 93 apresenta o modelo da sociedade multiagente, do agente “Interface do Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, a entidade externa não é o usuário e sim o usuário tributário

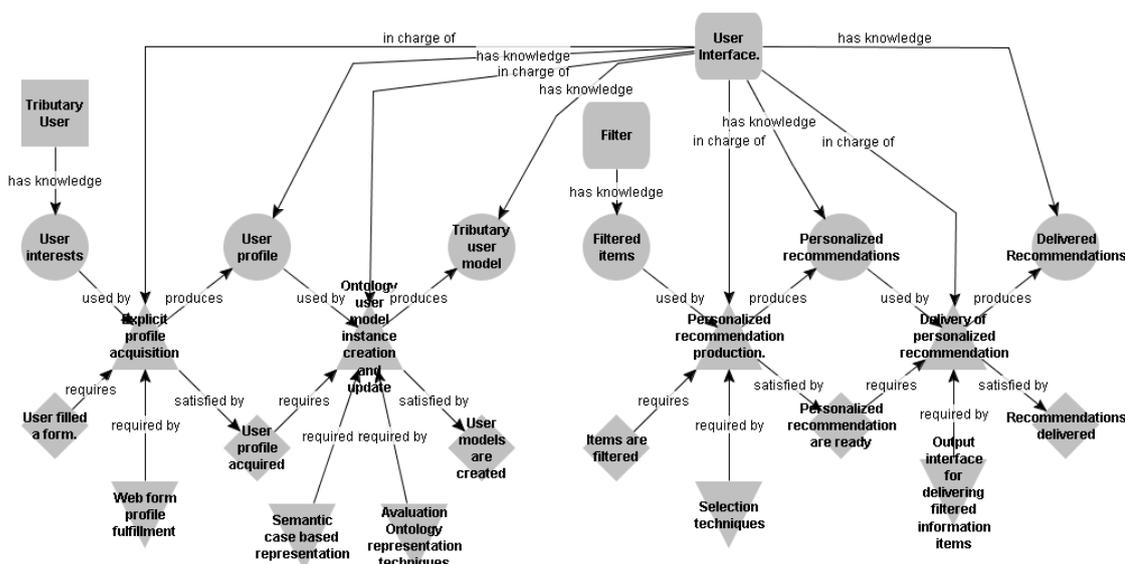


Figura 93 Modelo da Sociedade Multiagente do Agente “Interface do Usuário”

6.3.3 Modelo da Sociedade Multiagente do Agente “Filtrador”

A Figura 94 apresenta o modelo da sociedade multiagente, do agente “Filtrador”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, o agente “Filtrador”, realiza a filtragem colaborativa e a filtragem baseada no conteúdo. Para a filtragem baseada no conteúdo são utilizados os conceitos “Modelo do usuário tributário” e “Modelo do item de informação”. Para a filtragem colaborativa são utilizados os conceitos “Modelo do grupo de usuários”, “Modelo de usuários” e “Modelo do item de informação”. Os métodos de filtragens híbrida utilizados serão os métodos ponderado e o alternado. A fonte de informação será a ONTOTRIB. As demais responsabilidades necessárias serão: a medida semântica; o KNN, quanto à predição; as filtragens colaborativa e baseada no conteúdo.

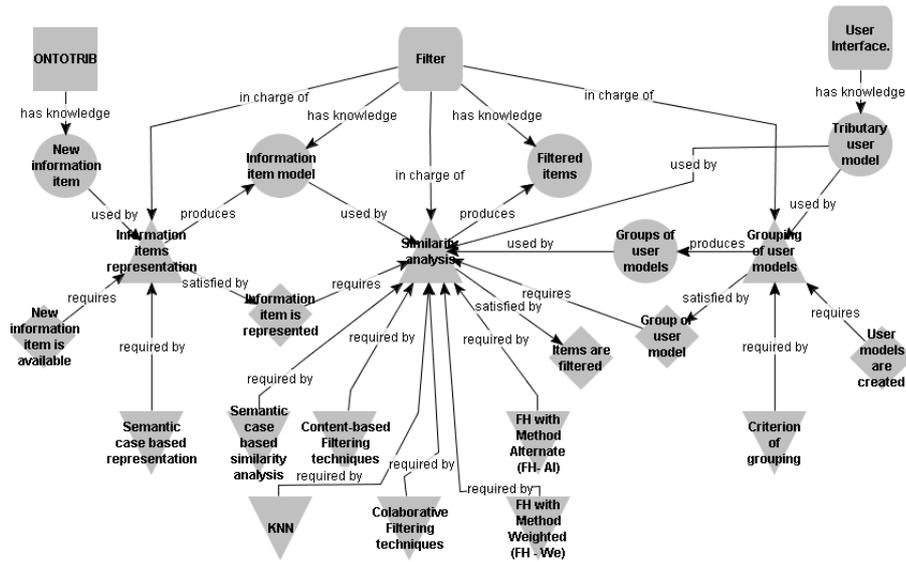


Figura 94 Modelo da Sociedade da Multiagente do Agente “Filtrador”

6.3.4 Modelo de Interação entre os Agentes da INFOTRIB

A Figura 95 apresenta o modelo de interações entre os agentes da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS, observa-se que neste, as fontes externas passaram a ser o usuário tributário e a ONTOTRIB.

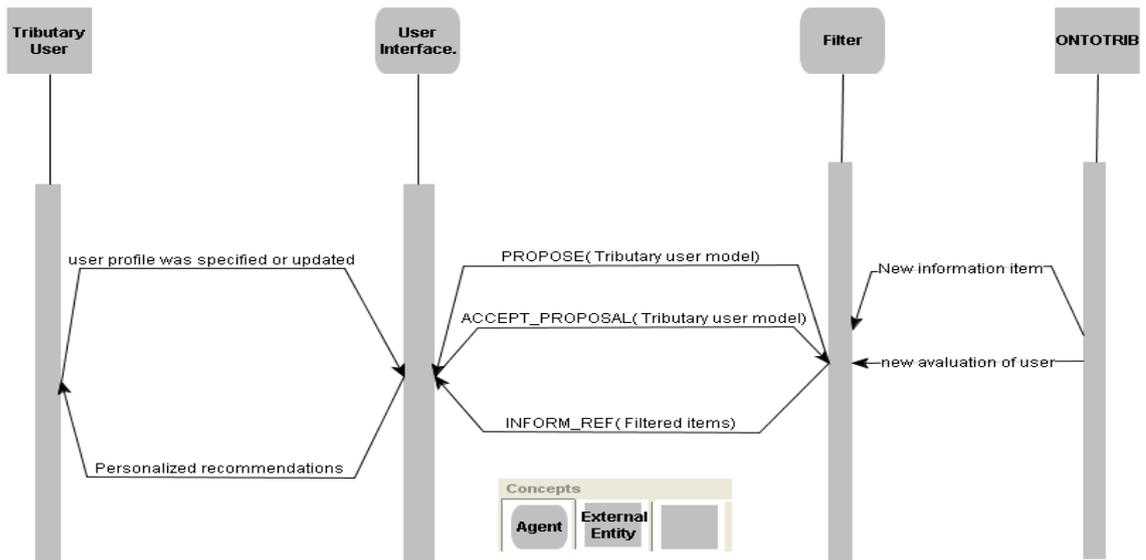


Figura 95 Modelo de Interação entre Agentes

6.3.5 Modelo de Atividade do INFOTRIB na FH ponderada

A Figura 96 apresenta o modelo de atividade, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, as filtragens baseada no conteúdo e a colaborativa necessárias para gerar a filtragem híbrida seguem este modelo.

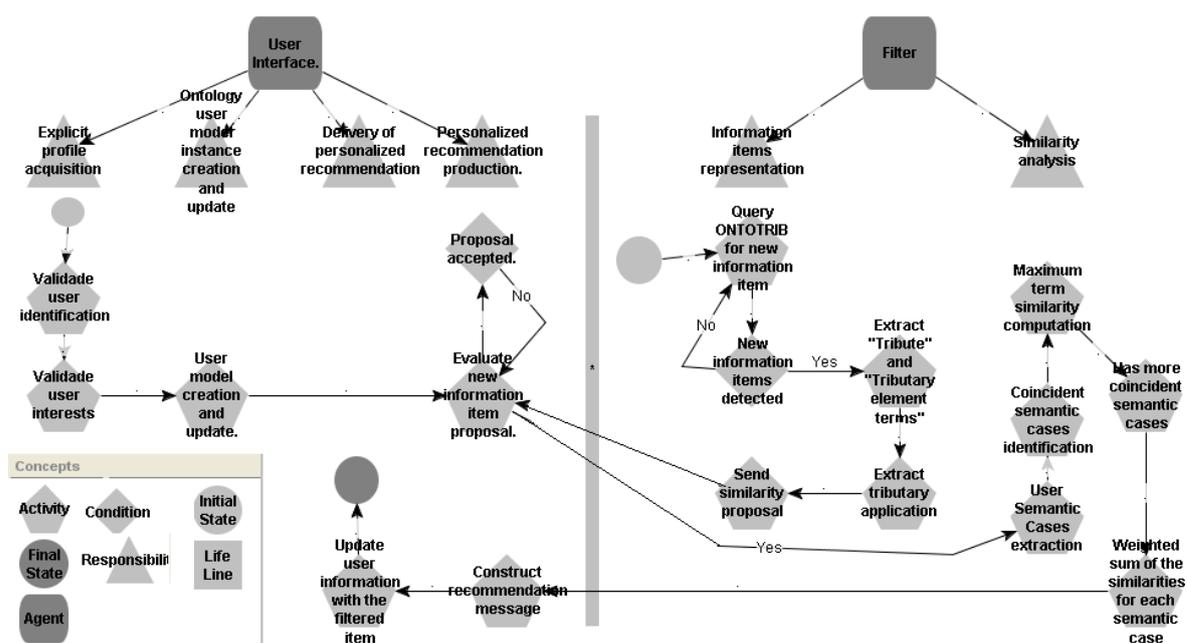


Figura 96 Modelo de Atividade da Sociedade Multiagente

A filtragem híbrida ponderada (BURKE 2007), Figura 97, na primeira etapa gera candidatos. A lista de candidatos é encaminhada para a filtragem colaborativa e as suas preferências por itens de informação, para a filtragem baseada no conteúdo. Cada filtragem gera recomendações, que são as predições, representadas como recomendações 1 e 2. Tais recomendações são ordenadas numa lista de candidatos. Na segunda etapa são combinados as notas (“scores”) das duas filtragens, levando-se em conta os pesos das recomendações de cada filtragem, por meio de uma função de ponderação, a fim de se obter a nota combinada. Terminando com a apresentação das notas combinados ordenadas.

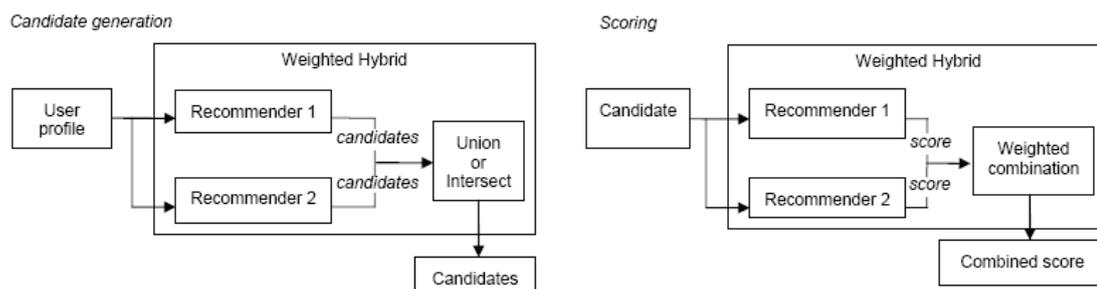


Figura 97 Filtragem híbrida ponderada (BURKE 2007)

Exemplificando: Na abordagem da filtragem híbrida ponderada tem-se o perfil do usuário X, que na filtragem colaborativa, apresentou as seguintes recomendações T (FC 0,5), Y (FC 0,8) e Z (FC 0,95). Na filtragem baseada no conteúdo, os itens de informação de preferência de X foram similares aos de A (FBC 0,5), B (FBC 0,8) e T (FBC 0,95) e Y (FBC 0,4). Portanto os candidatos seriam: A, B, T, Y e Z. Supondo que o peso da FC é 0,8 e a FBC é 0,2 e que se utilize uma combinação linear dos pesos, da forma:

$$FH(P) = \text{Peso FC} \times FC(P) + \text{Peso FBC} \times FCB(P)$$

Com os seguintes valores:

$$FH(A) = 0,2 \times 0,5 = 0,1$$

$$FH(B) = 0,2 \times 0,8 = 0,16$$

$$FH(T) = 0,8 \times 0,5 + 0,2 \times 0,5 = 0,40 + 0,10 = 0,5$$

$$FH(Y) = 0,8 \times 0,8 + 0,2 \times 0,8 = 0,48 + 0,16 = 0,64$$

$$FH(Z) = 0,8 \times 0,95 = 0,49$$

Logo, a recomendação dos três primeiros seria: YFH=0,64, TFH=0,5 e Z FH=0,49.

Outra forma de se realizar a filtragem híbrida ponderada é pela intersecção das listas de cada recomendação. Ou seja, serão selecionados os itens que aparecerem em ambas as filtragens.

Pela abordagem da intersecção, considerando o exemplo anterior se teria T e Y em ambas as filtragens. Portanto estes seriam o universo de itens e se teria YFH=0,8, TFH=0,5.

6.3.5.1.1 Modelo de Atividade, na FH, com método ponderado

Com base nas duas formas de se fazer filtragem híbrida ponderada foi gerada a modelagem de atividades da Figura 98, a qual aplica as filtragens colaborativa e baseada no conteúdo gerando a lista de candidatos a serem recomendados. Esta lista de candidatos é utilizada para gerar a lista de ponderações por intersecção e por pesos.

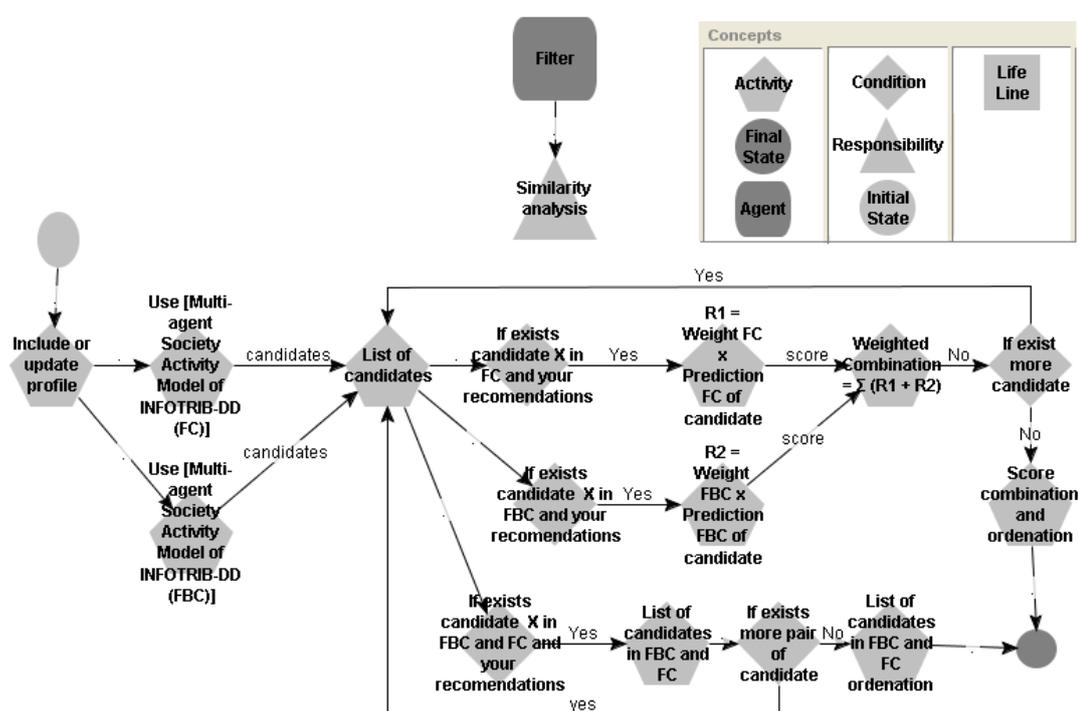


Figura 98 Modelo de Atividades da filtragem híbrida ponderada

6.3.5.1.2 Modelo de Atividade, na FH, com método Alternado

Na filtragem híbrida, com método Alternado, o sistema utiliza um critério para determinar a técnica que gera a recomendação, conforme Figura 99.

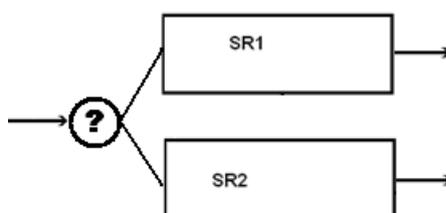


Figura 99 Formas de se fazer filtragem híbrida alternada

Na modelagem realizada, apresentada na Figura 100, utilizou-se o critério de selecionar a lista contendo as recomendações das técnicas que tivesse o maior valor de recomendação.

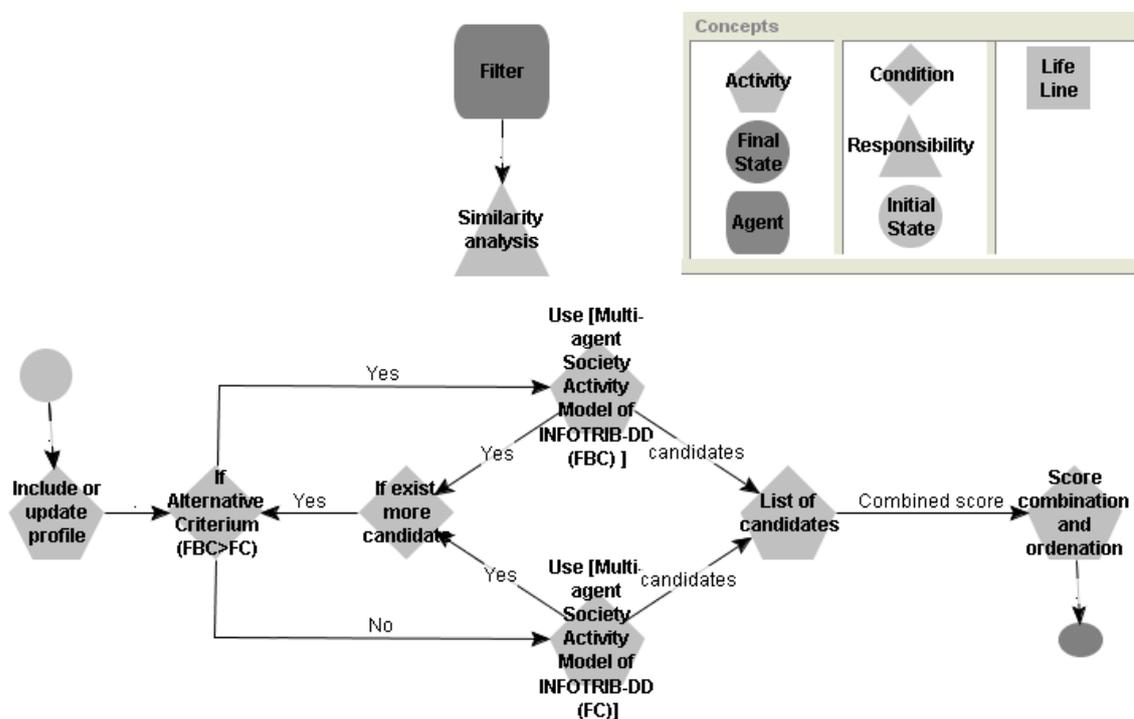


Figura 100 Modelo de Atividade da filtragem híbrida alternada

6.3.6 Projeto do Agente

6.3.6.1 Modelo do Mecanismo de Coordenação e Cooperação

A Figura 101 apresenta o modelo do mecanismo de coordenação e cooperação, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS, observa-se que neste, o usuário é o usuário tributário e que a ontologia é a ONTOTRIB.

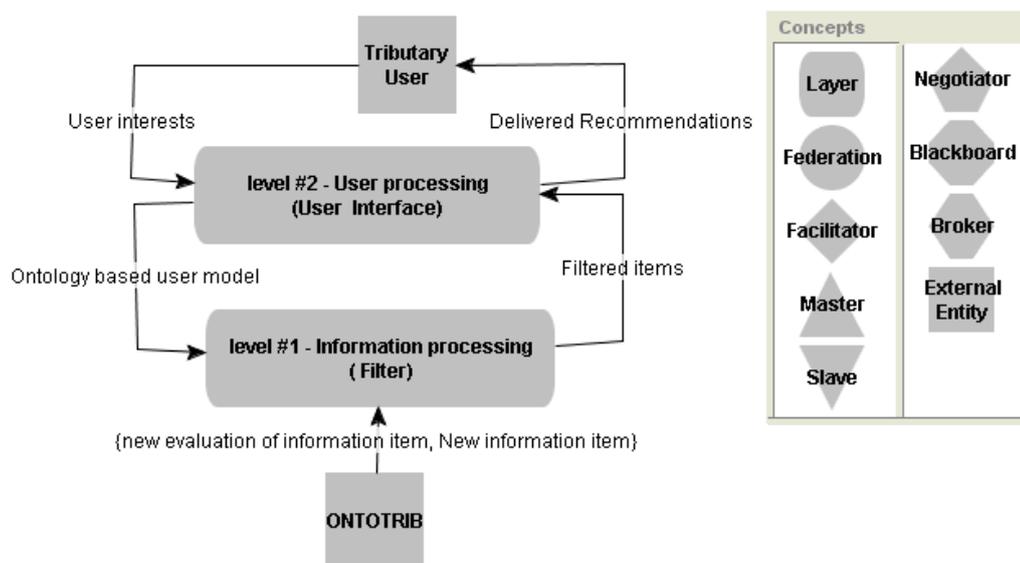


Figura 101 Modelo do Mecanismos de Coordenação e de Cooperação

6.3.6.2 Modelo de Conhecimento do Agente “Interface do Usuário”

A Figura 102 apresenta o modelo de conhecimento do agente “Interface do Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve a exclusão do conhecimento “comportamento navegacional do usuário”, o qual não foi definido neste trabalho.

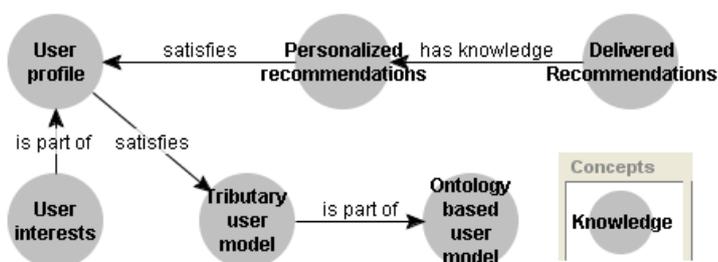


Figura 102 Modelo de Conhecimento do Agente “Interface do Usuário”

6.3.6.3 Modelo de Conhecimento do INFOTRIB do Agente “Filtrador”

A Figura 103 apresenta o modelo de conhecimento do agente “Filtrador Híbrido”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve a inclusão de uma série de conceitos relativos ao domínio tributário.

Agente Filtrador Híbrido tem o conhecimento necessário à filtragem colaborativa e a baseada no conteúdo. Da primeira ele tem o conhecimento “Modelo

do Usuário Corrente” e o dos “Grupos de Modelos do Usuário Corrente”, os quais são baseados no conhecimento “Modelo do Usuário Tributário”. Da segunda o agente Filtrador tem o conhecimento “Novos itens de informação tributária” e do “Modelo do Usuário Corrente”, os quais são mapeados como “Representações de Casos Semânticos”.

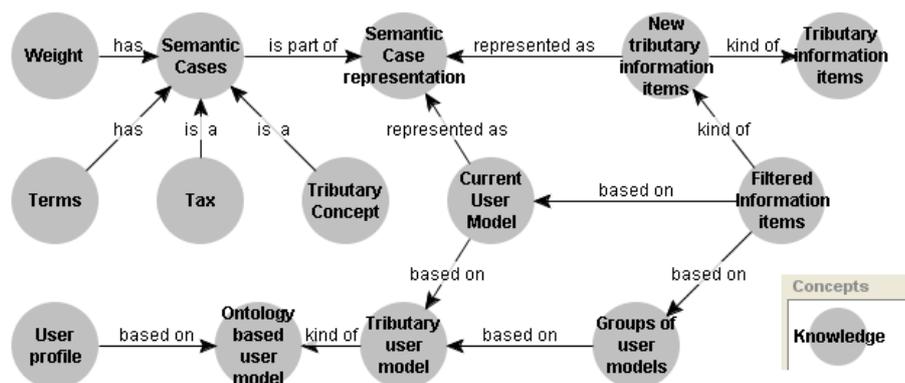


Figura 103 Modelo de Conhecimento do Agente “Filtrador”

6.3.6.4 Modelo de Estado do INFOTRIB do Agente “Interface do Usuário”

A Figura 104 apresenta o modelo de estado do agente “Interface do Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve o completo reuso.

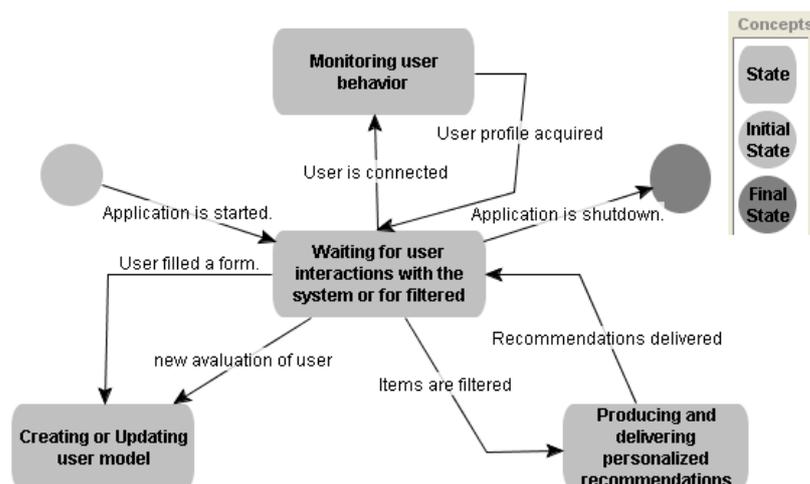


Figura 104 Modelo de Estado do Agente “Interface do Usuário”

6.3.6.4.1 Modelo de Estado do INFOTRIB do Agente Filtrador

A Figura 105 apresenta o modelo de estado do agente “Filtrador”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve o completo reuso.

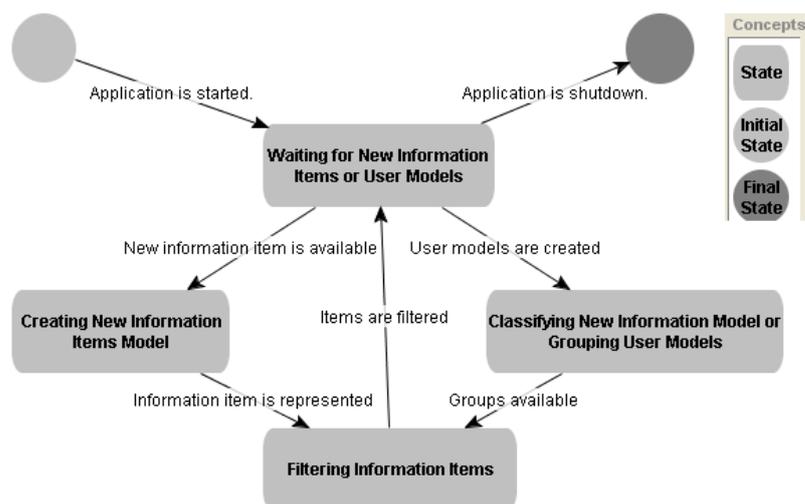


Figura 105 Modelo de Estado do Agente “Filtrador”

6.3.6.5 Implementação da Aplicação INFOTRIB na FH

6.3.6.5.1 Modelo de Comportamento do Agente Interface do Usuário no INFOTRIB FH

A Figura 106 apresenta o modelo de comportamento do agente “Interface do Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve a exclusão do JADE Behaviour relativo a aquisição implícita do perfil.

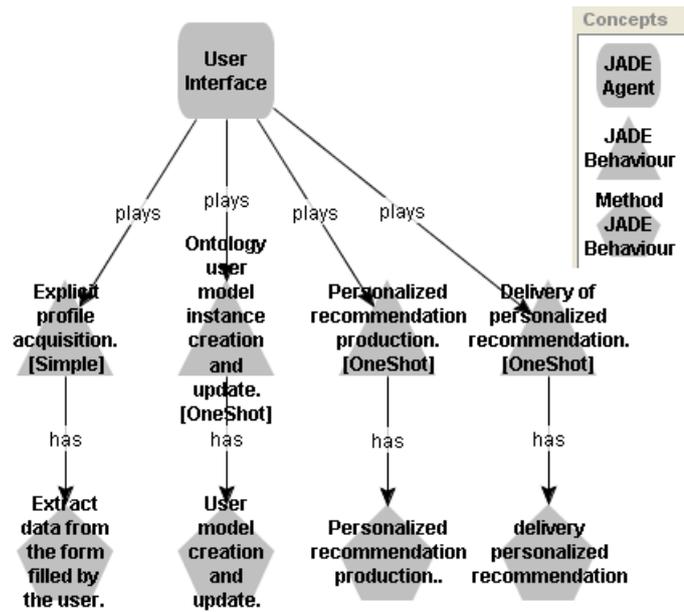


Figura 106 Modelo de Comportamento do Agente “Interface do Usuário”

6.3.6.5.2 Modelo de Comportamento do Agente Filtrador no INFOTRIB FH

A Figura 107 apresenta o modelo de comportamento do agente “Interface do Usuário”, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve o reuso completo.

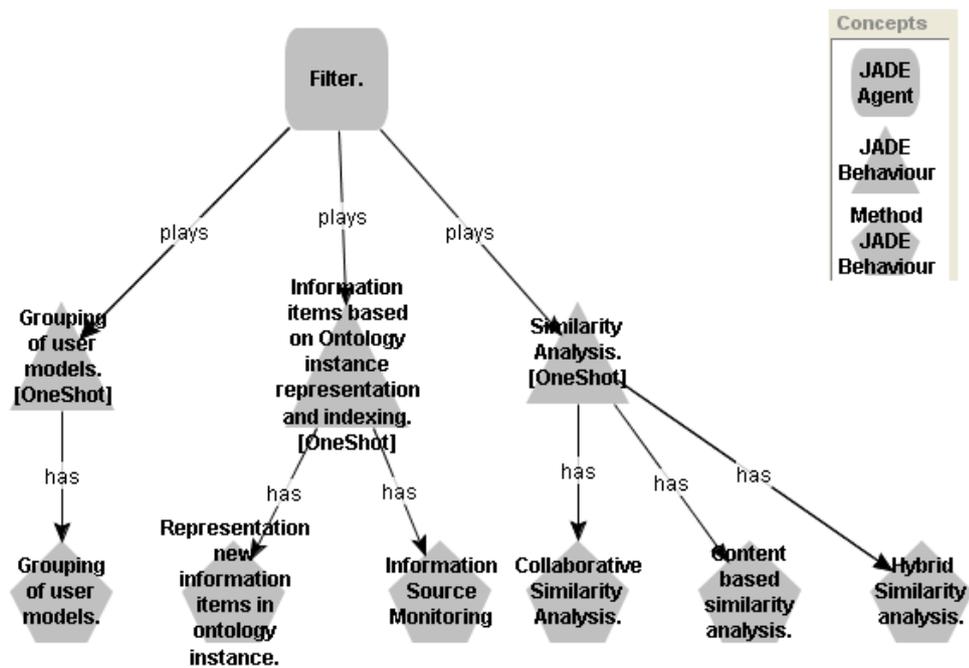


Figura 107 Modelo de Comportamento do Agente “Filtrador”

6.3.6.5.3 Modelo de Atos de Comunicação dos Agentes no INFOTRIB FH

A Figura 108 apresenta o modelo de atos de comunicação dos agentes, da INFOTRIB. Em relação ao correspondente modelo da ONTOSERS observa-se que neste, houve o reuso tendo ocorrido um rearranjo dos elementos da figura.

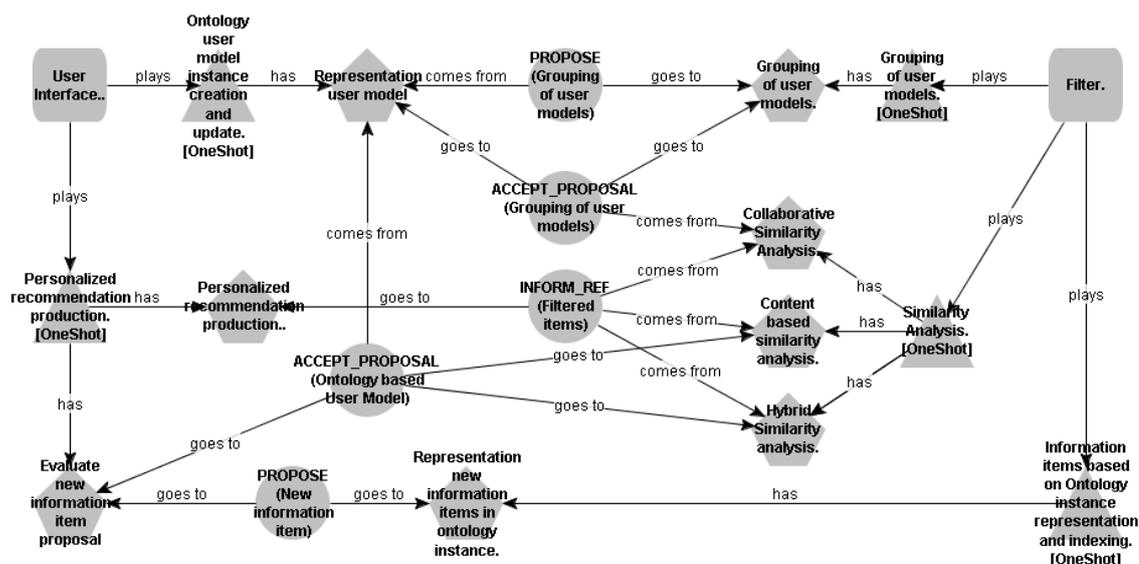


Figura 108 Modelo de Atos de Comunicação dos Agentes no INFOTRIB FH

6.4 Considerações Finais

O desenvolvimento do sistema de recomendação INFOTRIB permitiu uma primeira avaliação da família de sistemas ONTOSERS. Vários modelos sofreram apenas pequenas adaptações no INFOTRIB. Outros foram utilizados da forma que foram modelados.

À medida que novos algoritmos de filtragem sejam implementados e novas técnicas incorporadas será transparente a construção de sistemas de recomendações da família do ONTOSERS, ou do seu representante INFOTRIB, nos quais se possa fazer análise comparativa com outros anteriormente implementado. Possibilitando que se escolha o mais adequado para cada domínio e instâncias respectivas.

7. CONCLUSÕES

Este trabalho contribui com a análise, projeto, implementação e avaliação de agentes de filtragem híbrida, com uso da tecnologia da Web Semântica.

Os agentes foram desenvolvidos de acordo com as diretrizes da metodologia MADEM e integrados no ONTOSERS, uma família de sistemas multiagentes para o desenvolvimento baseado no reuso de sistemas de recomendações.

A família ONTOSERS, por sua vez, foi avaliada através do desenvolvimento de uma aplicação multiagente específica, o sistema INFOTRIB, de acordo com as diretrizes da metodologia MAAEM.

O INFOTRIB é um sistema multiagente de recomendações de instrumentos normativos jurídico-tributários.

7.1 Resultados e contribuições da pesquisa

As principais contribuições desta pesquisa foram:

- Desenvolvimento da família ONTOSERS. A família ONTOSERS define uma solução computacional multiagente reutilizável na construção de sistemas de recomendações baseados em diferentes técnicas de filtragem de informação: baseada em conteúdo, colaborativa e híbrida. A família integra diferentes agentes de software para filtragem que foram avaliados através de diversas experiências que mostraram a capacidade da filtragem híbrida superar os problemas comumente encontrados nos algoritmos de filtragens baseado em conteúdo e colaborativo usados isoladamente. Assim, a ONTOSERS fornece um framework apropriado para a experimentação e análise dos diversos algoritmos de filtragem de informação.
- Desenvolvimento do sistema de recomendação INFOTRIB, no domínio tributário brasileiro. A família ONTOSERS foi reutilizada na construção do sistema INFOTRIB, que fornece aos usuários com interesse no

Direito Tributário, recomendações de instrumentos jurídico-normativos nesse domínio.

- Avaliação das metodologias MADEM e MAAEM. O trabalho realizado permitiu uma avaliação das metodologias MADEM e MAAEM. A metodologia MADEM foi utilizada no desenvolvimento e implementação da família de sistemas ONTOSERS. Suas técnicas mostraram-se apropriadas para a identificação e representação das partes fixas e variáveis das abstrações de software da família, tornando assim, possível, o seu reuso e aplicabilidade no desenvolvimento de aplicações específicas. A metodologia MAAEM, utilizada no desenvolvimento do sistemas de recomendações INFOTRIB forneceu um adequado suporte para a sua construção através do reuso da família ONTOSERS.

7.2 Trabalhos futuros

Pela sua complexidade, o escopo desta pesquisa foi limitado em vários aspectos que deverão ser abordados em futuros trabalhos, entre eles:

- Enriquecimento da variabilidade da família ONTOSERS com a incorporação de novas técnicas de filtragem.
- Reuso da ONTOSERS na construção de novos sistemas de recomendações em outros domínios de problema de interesse.
- Extensão da ontologia ONTOTRIB com novos conceitos e relacionamentos do domínio tributário.
- Utilização da ontologia ONTOTRIB no desenvolvimento de sistemas baseados em conhecimento que atendam outros requisitos funcionais do domínio tributário.

REFERÊNCIAS

ADOMAVICIUS, Gediminas. TUZHILIN, Alexander. **Toward the Next Generation of Recommender Systems: A Survey of the State-of-Art and Possible Extensions** – IEEE Transactions on Knowledge and Data Engineering, v. 17, n 6. pp. 734-749. 2005.

ALANI, H., DASMAHAPATRA, S., O'HARA, K. and SHADBOLT, N. **Identifying Communities of Practice through Ontology Network Analysis**. IEEE Intelligent Systems, 18 (2). pp. 18-25. 2003.

ANTONIOU, G. and HARMELEN, F. V. **A Semantic Web Primer**. MIT Press, 2004.

ARAÚJO, Isabel. DRUMOND, Lucas. MARIANO, Roberval. GIRARDI, Rosario. **ONTOJURIS e ONTOTRIB: ontologias para a modelagem do conhecimento jurídico**. SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL UFF - IACS - Departamento de Ciência da Informação – Niterói 08/2008. Disponível em <<http://www.uff.br/ontologia/artigos/314.pdf>>. Acessado em 27/10/2008.

BAEZA-YATES, R. RIBEIRO-NETO, B. **Modern Information Retrieval**, Addison-Wesley, 1998.

BELKIN, N.J. CROFT, W. B. **Information Retrieval and Filtering: Two Sides of the Same Coin?** Communications of the ACM, v. 35, n.12, December 1992.

BELLIFEMINE, F., CAIRE, G., POGGI, A. and RIMASSA, G.: **JADE A White Paper**. Exp v. 3 n. 3, Sept 2003. <http://jade.tilab.com/>.

BERTOLDI, Anderson. CHISHMAN, Rove. ALVES, ISA. **A semântica adjetival e sua representação em uma ontologia de domínio jurídico**. Calidoscópio. Vol. 4. Num. 3. pp. 185-195. Set./Dez. 2006.

BREESE, J.S., HECKERMAN D., KADIE, C. **Empirical Analysis of Predictive Algorithms for Collaborative Filtering**. Proc. 14th Conf. Uncertainty in Artificial Intelligence, July 1998.

BURKE, R.: **Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-adapted Interaction**. [S.l.], v. 12, n. 4, pp. 331-370, November 2002.

BURKE, Robin. **Hybrid Web Recommender Systems**. in The Adaptive Web, LNCS 4321, pp. 377 – 408, 2007. Springer-Verlag Berlin Heidelberg 2007.

CAPEZ, Fernando. MALTINTI, Eliana R. **Direito Tributário (Perguntas e Respostas)**. Pág. 48. Editora Saraiva, 2006.

CASTELLS, Pablo. FERNÁNDEZ, Miriam. and VALLET, David. **An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval**. IEEE Transactions on Knowledge and Data Engineering 19(2), special issue on Knowledge and Data Engineering in the Semantic Web Era, February 2007, pp. 261-272. Disponível em <http://nets.ii.uam.es/~search/publications/tkde07.pdf>.

CF/88 - **Constituição da República Federativa do Brasil**, de 1988 e emendas.

CHAVES, Geysa H. **Abordagem colaborativa para Sistemas de Recomendação: Método dos K Vizinhos Mais Próximos**. UFMA. 2005.

CORDEIRO, João P. **Extracção de Elementos Relevantes em Texto/Páginas da World Wide Web**. Dissertação (Mestrado). Faculdade de Ciências do Porto. Porto (PORTUGAL). 2003.

COSTA, Adriana L. **Especificação de um Processo para a Engenharia de Domínio Multiagente**. Monografia de Especialização. UFMA. Maranhão. Brasil. 2006.

CRESPO, Antônio A. **Estatística Fácil**. Editora Saraiva. São Paulo, 1997.

CTN - **Código Tributário Nacional** - LEI Nº 5.172, DE 25 DE OUTUBRO DE 1966.

CUNNINGHAM, A. **Australian Government Locator Service: enabling seamless online access to government**. 1998. Disponível em: <<http://www.naa.gov.au/records-management/publications/agls-element.aspx>>. Acesso em 25/10/2008.

CZARNECKI, Krzysztof. EISENECKER, Ulrich. **Generative Programming: Methods, Tools, and Applications**. ACM Press/Addison-Wesley Publishing Co., New York, NY, 2000.

DE SÁ, Ana Maria Lopes. **Bases do Direito Tributário**. Edipro. Rio de Janeiro, Brasil. 1981.

DIAS, Tatiane D. SANTOS, Neide. **Web Semântica: Conceitos Básicos e Tecnologias Associadas**. Cadernos do IME. Série Informática. Volume 14. 2003.

DRUMOND, Lucas. **Análise de Similaridade para a Filtragem Baseada em Conteúdo Utilizando a Tecnologia da Web Semântica**. Monografia de Graduação. Universidade Federal do Maranhão. São Luís, Brasil. 2007.

DRUMOND, Lucas Rego. GIRARDI, Rosário. **Uma Análise das Camadas de Ontologia e Lógica da Web Semântica e de suas Implicações para o Acesso à Informação**. Jornada de Informática do Maranhão JIM 2006 São Luís. 29 de novembro de 2006.

DRUMOND, Lucas. GIRARDI, Rosário. **Uma Análise das Técnicas e Ferramentas para o Desenvolvimento de Aplicações para a Web Semântica**. Journal REIC - Revista Eletrônica de Iniciação Científica. Volume 6. Number 1. 2006.

DRUMOND, Lucas. GIRARDI, Rosario. **A multi-agent legal recommender system.** Journal Artificial Intelligence and Law. Publisher Springer Netherlands Volume 16, Number 2 / June, 2008. p. 175-207.

DRUMOND, Lucas Rego ; GIRARDI, Rosario ; LEITE, Adriana . **Architectural Design of a Multi-Agent Recommender System for the Legal Domain.** In: Eleventh International Conference on ARTIFICIAL INTELLIGENCE and LAW, 2007, Palo Alto. Proceedings of the Eleventh International Conference on ARTIFICIAL INTELLIGENCE and LAW. New York : ACM Press, 2007. p. 183-188.

DRUMOND, Lucas Rego. GIRARDI, Rosário. LEITE, Adriana. **A Case Study on the Application of the MAAEM Methodology for the Specification Modeling of Recommender Systems in the Legal Domain.** Ninth International Conference on Enterprise Information Systems (ICEIS 2007), Volume SAIC, Funchal, Madeira, Portugal, June 12-16, 2007. pp. 155-160.

DRUMOND, Lucas Rego. GIRARDI, Rosário. LINDOSO A., and MARINHO, Leandro. **A Semantic Web Based Recommender System for the Legal Domain.** Proceedings of the European Conference on Artificial Intelligence (ECAI 2006) Workshop on Recommender Systems, Ed. IOS Press, - Riva del Garda, Itália. August 2006.

DRUMOND, Lucas Rego. LINDOSO A., and GIRARDI, Rosário. **InfoNorma: Um Sistema de Recomendação baseado em Tecnologias da Web Semântica.** Journal of Computer Science INFOCOMP. Volume 6, N. 3, September. 2007. p.93-100.

EIRINAKI, Magdalini. **Web Mining: A RoadMap.** Athens University. Chapter 2. Technical Report, DB-NET 2004, available at <http://www.soe.ucsc.edu/~eirinaki/papers/NEMIS.pdf>.

FACCA, Federico. LANZI, Pier. **Mining interesting knowledge from weblogs: a survey.** Data & Knowledge Engineering 53 (2005). Elsevier. Pp. 225-241.

FANIZZI, Claudia.. ESPÓSITO, Floriana. **A Semantic Similarity Measure for Expressive Description Logics**. Universita di Bari. Italy. 2005.

GESEC. **Grupo de Pesquisa em Engenharia de Software e Engenharia do Conhecimento**. 2001. Disponível em <<http://maae.deinf.ufma.br/produtos/index2.php?id=5>>. Acesso em 16/11/2008 .

GIRARDI, Rosario. MARINHO, Leandro. **A Domain Model of Web Recommender Systems based on Usage Mining and Collaborative Filtering**, Requirements Engineering Journal, vol.12 n. 1, Ed. Springer-Verlag, pp. 23-40. London (on line first). 2007.

GIRARDI, Rosario. LINDOSO, Alisson. **An Ontology-based Methodology for Multi-agent Domain Engineering**. In: 3rd Workshop on Multi-Agent Systems: Theory and Applications (MASTA 2005) at 12th Portuguese Conference on Artificial Intelligence (EPIA 2005), 2005, Covilhã. Proceedings MASTA/EPIA 2005. Los Alamitos : IEEE CS Press, 2005. p. 321-324.

GIRARDI, Rosario. LINDOSO, Alisson Neres. **DDEMAS: A Domain Design Technique for Multi-agent Domain Engineering**. In: The Seventh International Bi-conference Workshop on Agent-Oriented Information Systems (AOIS-2005) at The 24th International Conference on Conceptual Modeling (ER 2005). 2005. Klagenfurt. Proceedings of ER Workshops, LNCS, vol. 3770. Springer. Berlin Heidelberg New York, pp 141-150.

GIRARDI, Rosario. LINDOSO, Alisson Neres. **An Ontology-based Methodology and Tool for Multi-agent Domain Engineering**. 2006. Disponível em citeseerx.ist.psu.edu/showciting.jsessionid=463E049912D3B48460BEE99087010AF3?cid=4971201.

GIRARDI, Rosario. **Classification and Retrieval of Software through their Descriptions in Natural Language**, Ed. Imprimerie de l'Université de Geneve, Geneva, Switzerland. 1995.

GREENWOOD, Dominic. **FIPA - The Foundation for Intelligent, Physical Agents**. Whitestein Technologies AG. 2004. Disponível em <http://jade.tilab.com/papers/JADETutorialIEEE/JADETutorial_FIPA.pdf>. Acessado em 28/10/2008.

HAMATI, Cecília M. **Curso de Direito Tributário**. Editora Malheiros. São Paulo (Brasil). 1996.

LINDOSO, Alisson Neres. **Uma Metodologia baseada em Ontologia para a Engenharia de Aplicações Multiagente**. Tese de Mestrado. São Luís. 2006.

MARIANO, Roberval. GIRARDI, Rosário. LEITE, Adriana, DRUMOND, Lucas, MARANHÃO, Djefferson. **A Case Study on Domain Analysis of Semantic Web Multi-Agent Recommender Systems**. Porto, Portugal. Published in Third International Conference on Software and Data Technologies. ICSoft 2008. July 5-8, 2008. pp. 160-167.

MARTINS, Maria Cleci Coti. **Ontologia Legal – Estudo Sobre a Modelagem do Conhecimento Legal no Contexto do Direito Tributário**. ESAF. 2006. Disponível em http://www.esaf.fazenda.gov.br/esafsite/premios/schontag/Monografias_premiadas_arquivos/monografia/monografias6/MENCAO-HONRO_SA.pdf. Acessado em 05/01/2008.

MASCARDI, Viviane. MARTELLI, Maurizio. and GUNGUI, Ivana. **DCaseLP: a Prototyping Environment for Multi-Language Agent Systems**. In Proceedings of the First Workshop on Languages, methodologies and Development tools for multi-agent systems, LADS'07 Post-proceedings, M. Dastani, A. El Fallah Seghrouchni, J. Leite, P. Torroni eds., pages 139-155, LNCS 5118, Springer-Verlag, 2008.

MASSA, Paolo. AVESANI, Paolo. **Trust-aware Recommender Systems**. ACM Recommender Systems Conference, Minneapolis, Minnesota, USA. 2007. Disponível em http://gnuband.org/files/papers/trustaware_recommender_systems_paolo_massa.pdf. Acessado em 17/01/2008.

MENCZER, Fillipo. **Combining Link and Content Analysis to Estimate Semantic Similarity**. WWW2004. May 17-22, 2004. ACM 1-581113-912-8. New York. USA.

MENCZER, Fillipo. **Finding semantic needles in haystacks of Web text and links**. IEEE Internet Computing. May/June 2005. Disponível em www.informatics.indiana.edu/fil/Papers/maps.pdf.

MIDDLETON, Stuart E. ALANI, Harith. DE ROURE, David C. de. **Exploiting synergy between ontologies and recommender systems**. In Proceedings of the WWW International Workshop on the Semantic Web, volume 55 of CEUR Workshop Proceedings, Maui, HI, USA, May 2002.

MOBASHER, Bamshad. COOLEY, Robert. SRIVASTAVA, Jaideep. **Automatic Personalization Based on Web Usage Mining**. Communication of ACM, Volume 43, Issue 8, August, 2000.

MORAIS, E. ; AMBROSIO, A. P. L. . **Associação Semântica entre Ontologias e Documentos**. In: Simpósio Brasileiro de Engenharia de Software (SBES), 2006, Florianópolis/SC. Proceedings of the First Workshop on Ontologies and Metamodeling Software and Data Engineering. Vitória/ES. : Giancarlo Guizzardi-UFES., 2006. p. 83-92.

OLIVEIRA, Ismênia R. de, **Um Sistema de Padrões Baseados em Agentes para a Modelagem de Usuários e Adaptação de Sistemas**. 2004. Dissertação (Mestrado em Curso de Pós Graduação Em Eng de Eletricidade) - Universidade Federal do Maranhão, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Maria del Rosario Girardi.

PAPATHEODOROU C. **Machine Learning in User Modeling. Machine Learning and Applications.** Lecture Notes in Artificial Intelligence. Springer Verlag, 2001.

PETASIS, Georgios. VICHOT, Frantz. WOLINSKI, Francis. **Using Machine Learning to Maintain Rule-based Named-Entity Recognition and Classification Systems.** Annual Meeting of the Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. Toulouse, França, 2001. p. 426 – 433.

PIERRAKOS, D. PALIOURAS, G. PAPATHEODOROU, C. and SPYROPOULOS, C. **Web Usage Mining as a Tool for Personalization: A Survey. User Modeling and User-Adapted Interaction** v. 13, pp. 311-372, 2003.

POHL, Klaus· BÖCKLE, Günter. LINDEN, Frank. **Software Product Line Engineering - Foundations, Principles, and Techniques With 150 Figures and 10 Tables.** Springer. 1998.

POKORNÝ, Jaroslav. **Web Searching and Information Retrieval.** Web Engineering. IEEE 2004. July/August 2004. pp 43-48.

RUSSEL, Stuart. NORVIG, Peter. **Inteligência Artificial.** Editora Campos. pp. 629-656, 690-737. 2004.

SABBAG, Eduardo de M. **Elementos do Direito – Direito Tributário.** 7ª Edição. Editoria Premier Máxima.São Paulo. 2005.

SALTON, Gerard. MCGILL, Michael. **Introduction to Modern Information Retrieval.** McGraw-Hill, Inc. New York, NY, USA. 1986.

SCHAFER, B. **MetaLens: A Framework for Multi-Source Recommendations.** Thesis in Computer Science. University of Minnesota. Minneapolis, 2001. EUA. pp. 37-54. Disponível em http://cns2.uni.edu/~schafer/publications/schafer_thesis.pdf. Acessado em 13/01/2008.

SHAHABI, C., BANAEI-KASHANI, F., “**Efficient and Anonymous Web Usage Mining for Web Personalization**”, *INFORMS Journal on Computing - Special Issue on Data Mining*, Vol.15, No.2, Spring. 2003.

SHAHABI, C., BANAEI-KASHANI, F. and FARUQUE, J.: **A Reliable, Efficient, and Scalable System for Web Usage Data Acquisition**. In: *WebKDD'01Workshop in conjunction with the ACM SIGKDD 2001*, San Francisco, CA, August 2001.

SILVA JR, Geovane B. **Padrões Arquiteturais para o Desenvolvimento de Aplicações Multiagente**. 2003. Dissertação (Mestrado em Curso de Pós Graduação Em Eng de Eletricidade) - Universidade Federal do Maranhão, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: María del Rosario Girardi.

TAUSCHER, L. and GREENBERG, S.: **How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems**. *International Journal of Human Computer Studies*. Special issue on World Wide Web Usability, 47(1), 97-138, 1997.

TORRES, Roberto. **Personalização na Internet**. Editora Novatec. São Paulo. 2004.

TORRES, Roberto. **Combining Collaborative and Content-based Filtering to Recommend Research Papers**. Tese de Mestrado. Porto Alegre (RS). UFRS. 2004.

VASCONCELOS, Arnaldo. **Teoria da Norma Jurídica**. 6ª Edição. Editora Malheiros. 2006.

VASCONCELOS, Livia. CARVALHO, Cedric. **Aplicação de Regras de Associação para Mineração de Dados na Web**. Technical Report RT-INF_004-04. UFF. Goiás. Brasil. 2004.

WEBB, G. PAZZANI, M. BILLSUS, D. **Machine Learning for User Modeling. User Modeling and User-Adapted Interaction**. p. 19–29. Kluwer Academic Publishers Hingham. Volume 11 , Issue 1-2. 2001. MA, USA.

ZIEGLER, Cai-Nicolas. SCHIMITDT-THIEME, Lars. LAUSEN, George. **Exploiting Semantic Product Descriptions for Recommender Systems.** SWIR 04. ACM SIGIR Semantic Web and Information Retrieval Wokshop. July, 29, 2004. Sheffield, UK.

ANEXO I – EXEMPLO DE CÁLCULO DAS MEDIDAS DE AVALIAÇÃO UTILIZADAS NOS EXPERIMENTOS OFF-LINE

A fim de demonstrar o cálculo dos índices serão utilizadas a Tabela 28 e a Tabela 29, as quais representam nas colunas os instrumentos normativos (IN) e nas linhas os usuários (US). As células representam a avaliação dada por determinado usuário a um instrumento normativo, no caso da matriz resultados obtidos e serão os resultados que se espera obter na matriz resultados esperados.

Tabela 28 Matriz Resultados Esperados E(us,in)

	IN1	IN2	IN3	IN4
US1	0	1	0	1
US2	1	0	1	0
US3	1	1	1	0
US4	0	0	0	1

Tabela 29 Matriz Resultados Obtidos O(us,in)

	IN1	IN2	IN3	IN4
US1	0	1	1	1
US2	1	1	0	1
US3	1	1	0	1
US4	0	0	1	1

Considerando

$$\text{Acerto } (i,j) = \begin{cases} 0, & \text{se } E_{i,j} \neq O_{i,j} \\ 1, & \text{se } E_{i,j} = O_{i,j} \text{ e } E_{i,j} = 1 \end{cases}$$

Tem-se os seguintes valores:

Índice de Acertos por Instrumento Normativo (IA_{IN})

$$IA_{IN}(j) = \frac{\sum_{i=0}^{us} \text{acerto}(i, j)}{us}$$

$$IA_{in}(\text{IN1}) = \frac{0+1+1+0}{2} = 1 \text{ (100\%)}$$

$$IA_{in}(\text{IN2}) = \frac{1+0+1+0}{2} = 1 \text{ (100\%)}$$

$$IA_{in}(\text{IN3}) = \frac{0+0+0+0}{2} = 0 \text{ (0\%)}$$

$$IA_{in}(\text{IN4}) = \frac{1+0+0+1}{2} = 1 \text{ (100\%)}$$

Com base nos resultados obtidos para este exemplo, representados na Figura 109, conclui-se que, para esta filtragem tem-se bons resultados em relação ao instrumento normativo 1, 2 e 4 (100% acerto). Mas em relação ao instrumento normativo 3 (0% acerto), esta filtragem não é adequada.

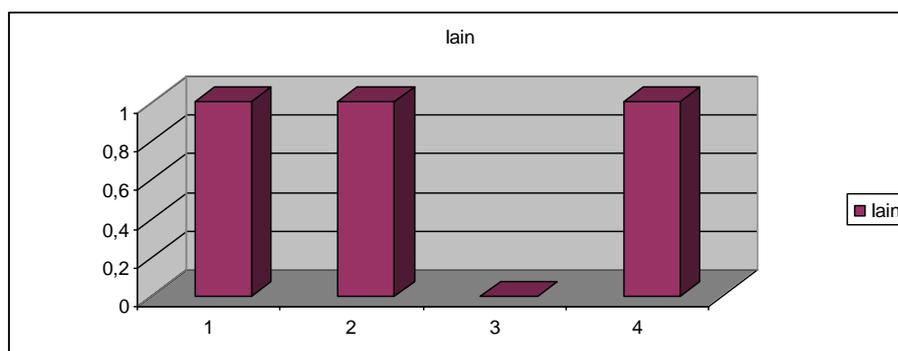


Figura 109 Índice de Acertos por Instrumento Normativo (IA_{in})

Índice Médio de Acertos por Instrumento Normativo (IMA_{IN})

$$IMA_{IN}(j) = \frac{\sum_{j=0}^{in} IA_{IN}(j)}{in}$$

$$IMA_{IN}(j) = \frac{1+1+0+1}{4} = \frac{3}{4} = 0,75 \text{ (75\%)}$$

Índice de Acertos por usuário (IA_{us})

$$IA_{us}(i) = \frac{\sum_{j=0}^{in} acerto(i, j)}{in}$$

$$IA_{us}(US1) = \frac{0+1+0+1}{2} = 1 \text{ (100\%)} \quad IA_{us}(US2) = \frac{1+0+0+0}{2} = 0,5 \text{ (50\%)}$$

$$IA_{us}(US3) = \frac{1+1+0+0}{3} = 0,25 \text{ (25\%)} \quad IA_{us}(US4) = \frac{0+0+0+1}{1} = 1 \text{ (100\%)}$$

Índice Médio de Acertos por usuário (IMA_{us})

$$IMA_{us}(i) = \frac{\sum_{j=0}^{in} IA_{us}(i)}{\sum in}$$

$$IMA_{us}(i) = \frac{1+0,5+0,25+1}{4} = \frac{2,75}{4} = 0,6875 \text{ (68,75\%)}$$

$$\text{Precisão} = \frac{\sum_{i=0}^{us} \sum_{j=0}^{in} \text{acerto}(i, j)}{\sum_{j=0}^n \sum_{i=0}^{us} 1} = \frac{8}{11} = 0,72 \text{ (72\%)}$$

$$\text{Recall} = \frac{\sum_{i=0}^{us} \sum_{j=0}^{in} \text{acerto}(i, j)}{\sum_{j=0}^n \sum_{i=0}^{us} 1} = \frac{6}{8} = 0,75 \text{ (75\%)}$$

Com base nos resultados obtidos para este exemplo, representados na Figura 110, conclui-se que para esta filtragem tem-se bons resultados em relação ao instrumento normativo 1 e 4 (100% acerto).

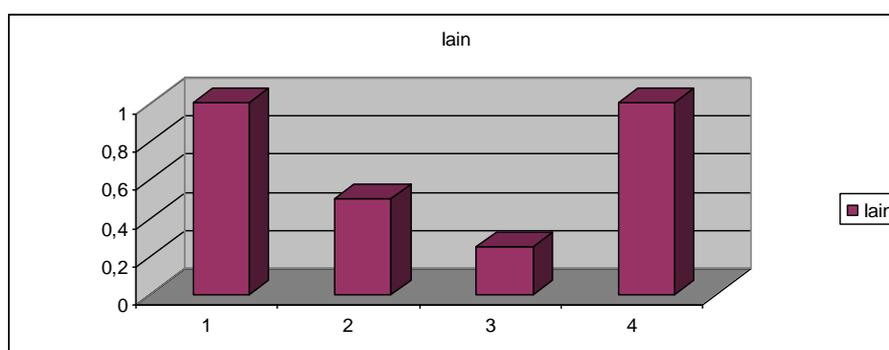


Figura 110 Índice de Acertos por Usuário (IAus)

Na Figura 111 observa-se que para estes dados o recall foi maior do que o precisão.

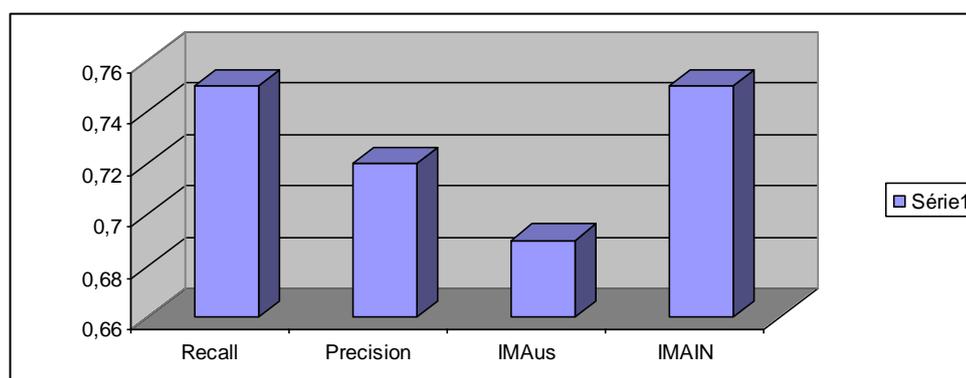


Figura 111 Comparativo entre os diversos índices

ANEXO II – EXEMPLO DE CÁLCULO DAS MEDIDAS DE AVALIAÇÃO UTILIZADAS NOS EXPERIMENTOS ON-LINE

De nada adianta os sistemas de recomendações proporem instrumentos normativos que não sejam do interesse do usuário, neste aspecto é fundamental o feedback dos usuários, a fim de se realizar as calibrações, com base em quem usa o sistema de recomendações.

Neste trabalho a análise do feedback foi tratado a partir da sua correlação com cada uma das filtragens e a seguir foi gerado um gráfico, que demonstrou visualmente os resultados. Quanto mais próximo de um (1) mais correlacionadas.

Para exemplificar o processo suponha a Tabela 30, a qual possui as colunas: instrumentos normativos, filtragem baseada no conteúdo (FBC), filtragem colaborativa (FC), filtragem híbrida (FH) e de feedback.

Tabela 30 Valores Obtidos das filtragens e do Feedback

Instrumento Normativo	FBC	FC	FH	Feedback
Lei X (CTN MA)	70%	80%	80%	77%
Lei Y (Lei tributária AC)	90%	10%	50%	37%
Lei Z (Nfe no MA)	60%	90%	40%	80%
Lei W (ICMS Simples)	10%	90%	50%	63%

A partir dos valores da Tabela 30 foi calculada a correlação de Pearson, o que resultou na Tabela 31, a qual foi convertida no gráfico da Figura 112. A fórmula do coeficiente de correlação de Pearson é:

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Onde:

\bar{r}_a - é a média das avaliações que o usuário "a" deu.

$r_{u,i}$ - é a avaliação que o usuário "u" deu para o item "i".

\bar{r}_u - é a média das avaliações que o usuário "u" deu.

$w_{a,u}$ - é o coeficiente de correlação de Pearson do usuário a, em relação a u.

No caso do exemplo apresentado o feedback esteve mais próximo da filtragem colaborativa.

Utilizou-se o coeficiente de correlação de Pearson, porque não se estava analisando os itens de informação na ontologia ONTOTRIB, mas sim os dados obtidos nas diversas filtragens representados na Tabela 30.

Tabela 31 Valores Obtidos das filtragens e do Feedback

Relação com o Feedback	Correlações
FBC	-0,313250909
FC	0,902906253
FH	0,220707595

Os valores de correlação resultaram na Tabela 31, a qual foi representada na Figura 112.

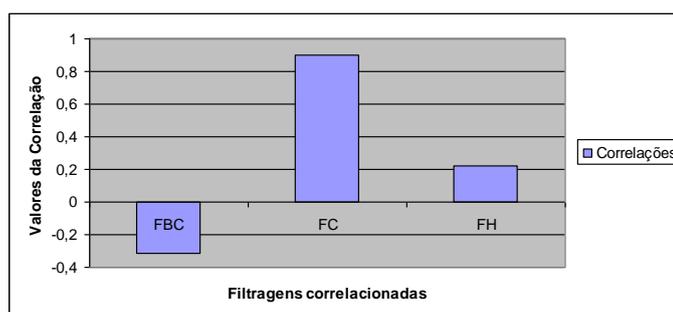


Figura 112 Correlação d as filtragens com o feedback do usuário