

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

SÍLVIO CÉSAR CAZELLA

**Aplicando a Relevância da Opinião de  
Usuários em Sistema de Recomendação  
para Pesquisadores**

Tese apresentada como requisito parcial para a  
obtenção do grau de Doutor em Ciência da  
Computação

Prof. Dr. Luis Otávio Campos Alvares  
Orientador

Porto Alegre, agosto de 2006.

## CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Cazella, Sílvio César

Aplicando a Relevância da Opinião de Usuários em Sistema de Recomendação para Pesquisadores / Sílvio César Cazella – Porto Alegre: Programa de Pós-Graduação em Ciência da Computação da UFRGS, 2006.

180 f.:il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Ciência da Computação, Porto Alegre, BR – RS, 2006. Orientador: Luis Otávio Campos Alvares.

1. Sistemas de Recomendação. 2. Modelagem de Perfil de Usuário. 3. Sistemas Multiagentes. 4. Mineração de dados. 5. Comunidades Virtuais. I. Alvares, Luis Otávio Campos. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Vice-reitor: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora de Pós-Graduação: Profa Valquíria Linck Bassani

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenador do PPGC: Prof. Carlos Alberto Heuser

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“Deixei os bosques por uma razão tão boa quanto a que me levou para lá. Talvez por ter me parecido que eu tinha várias vidas para viver, e não podia desperdiçar mais tempo com aquela. É impressionante a facilidade com que insensivelmente caímos numa determinada rotina e fazemos pra nós uma trilha batida.”*

*Henry David Thoreau (1817-1862)*

**Walden**

*“Por isso, mais uma vez, somos responsáveis, também por nós. Somos no mínimo co-responsáveis pelo que fazemos com a bagagem que nos deram para esse trajeto entre nascer e morrer. Carregamos muito peso inútil. Largamos no caminho objetos que poderiam ser preciosos e recolhemos inutilidades. Corremos sem parar até que aquele fim temido, raramente nos sentamos para olhar em torno, avaliar o caminho, e modificar ou manter nosso projeto pessoal.”*

*Lia Luft (2003)*

**Perdas e Ganhos**

## TABACARIA

(ÁLVARO DE CAMPOS)

*"Fiz de mim o que não soube,  
E o que podia fazer de mim não o fiz.  
O dominó que vesti era errado.  
Conheceram-me logo por quem não era e  
não desmenti, e perdi-me.  
Quando quis tirar a máscara,  
Estava pegada à cara.  
**QUANDO A TIREI E ME VI AO  
ESPELHO,  
JÁ TINHA ENVELHECIDO."***

## A VIDA

(MÁRIO QUINTANA)

*A vida são deveres que trouxemos para  
fazer em casa.  
Quando se vê, já são seis horas!  
Quando se vê, já é sexta-feira...  
Quando se vê, já é Natal...  
Quando se vê, já terminou o ano...  
Quando se vê, perdemos o amor da nossa  
vida...  
Quando se vê, passaram-se 50 anos!  
Agora, é tarde demais para ser reprovado...  
**Se me fosse dado, um dia, outra  
oportunidade, eu nem olhava o relógio.**  
*Seguiria sempre em frente e iria jogando,  
pelo caminho, a casca dourada e inútil das  
horas.*  
**SEGURARIA O MEU AMOR, QUE ESTÁ  
HÁ MUITO À MINHA FRENTE, E DIRIA  
EU TE AMO.**  
*Dessa forma, eu digo: não deixe de fazer  
algo que gosta devido à falta de tempo.  
Não deixe de ter alguém ao seu lado por puro  
medo de ser feliz.  
A única falta que terá, será desse tempo que  
infelizmente...  
Não voltará...."**

## AOS MOÇOS

(CORA CORALINA)

*Eu sou aquela mulher  
a que o tempo muito ensinou.  
Ensinou a amar a vida.*  
**Não desistir da luta.  
Recomeçar na derrota.  
Renunciar a palavras e pensamentos negativos.  
Acreditar nos valores humanos.  
Ser Otimista!**  
*Creio numa força imanente  
que vai ligando a família humana  
numa corrente luminosa  
de fraternidade Universal.*  
*Creio na solidariedade Humana  
Creio na superação dos erros  
e angústias do presente.*  
*Acredito nos moços.  
Exalto sua confiança,  
generosidade e idealismo.  
Creio nos milagres da ciência  
e na descoberta de uma profilaxia futura  
dos erros e violências do presente.*

**APRENDI QUE MAIS VALE LUTAR  
DO QUE RECOLHER DINHEIRO FÁCIL.  
ANTES ACREDITAR DO QUE DUVIDAR.**

## AGRADECIMENTOS

Quando terminamos uma tarefa sempre temos que agradecer alguém. Quanto maior o desafio que esta tarefa nos impõe, mais pessoas nós temos que agradecer. Quando este desafio constitui-se em um projeto tão longo como uma tese de doutorado, não podemos deixar de agradecer a nós mesmos pela perseverança de levar uma idéia até o fim, sendo assim, obrigado Silvio!

Gostaria de agradecer a pessoas distintas de vários grupos que fazem parte da minha vida, tornando-a muito mais interessante. Agradeço primeiro a família pelo apoio e incentivo, e por terem acreditado mais nessa idéia do que eu mesmo acreditava, obrigado Sérgio (pai), Sígla (mãe), Silviane (irmã), Sergio Ery (irmão) e Márcia (cunhada).

Pelos incentivos e ajuda nas horas difíceis e fáceis, pela demonstração do que realmente é amizade, eu quero agradecer aos amigos (as) do período inesquecível do Canadá: Camila e Tibério, Moises e Ana, Oricélia e Alexandre, Delane e Tristan, Stanley e Dayse, Nelson e Scott, Mário, Dabo, Ayzea, Sunet, Luca, Reza e Leila, Luciano, Magali e Felipe e os esquilos; do período da vida: Yanin, Marla, Marcelo, André, Natália, Mauren, Beto, Joyce, Débora, Anelise, Maria do Carmo, Carla, Alessandra, Stela, Paulo, Alexandre, Cláudia, e Marileide; da turma dos focas (eternos e saudosos focas): Bia, Bida, Denise, Gabriela, Augusto e Andréa (a esta amiga mais do que especial meu muito obrigado...valeu o apoio guria).

Do meio acadêmico gostaria de agradecer aos profissionais que me auxiliaram nesta empreitada e que dignificam, ao meu ver, o qualificador “doutor”: Julita Vassileva, Gord McCalla (*University of Saskatchewan*, Canadá), Osmar Zaiane, Nelson Amaral, Mário Nascimento, Renée Elio (*University of Alberta*, Canadá), Fernando Santos Osório, Marinho Barcellos, Renata Vieira, Soraia Musse, Mara Abel, Eliseo Reategui, Rejane Frozza, Carine Weber, Marcelo Walter, Juliana Herbert, Valter Roesler, Luciano Gaspary, Lisandro Granvile e em especial ao Luis Otavio Campos Alvares (meu orientador) pela sua paciência com a minha pessoa.

Da Unisinos não poderia deixar de agradecer as pessoas que me apoiaram de alguma forma nesta empreitada, meu muito obrigado a Silvia Dutra, Denise Bandeira, Amarolinda Saccol, Marie Macadar, Guilherme Liberali, Carlo Belini, Celso Pires Machado, Viviane Todt, Izabel Santos, Mirian Pires (e sua trupe no preparo para o Toefl) e toda a equipe da secretaria por entender os momentos de sufoco.

Agradeço aos meus orientandos por terem acreditado na minha idéia e aceitarem o desafio de trabalhar junto comigo nesta pesquisa: Gabriela Veit Volkveis, Fábio Bressler, Wagner Dandas da Silva Filho, Juliana Gonçalves de Freitas, Tiago Masiero,

Fábio Milani, Patrícia Marmitt, Fabiano Ferreira da Rosa, Paulo Ricardo Guglieri Gillmeister, Marcos de Oliveira Machado e Sérgio Cristiano Cruz.

Gostaria de agradecer, também, a toda a equipe do Instituto de Informática da UFRGS, equipe da pós-graduação, equipe da biblioteca, a equipe de professores, ao grupo de colegas (principalmente Vânia Bogorny), a equipe da segurança e a todos os CACHORROS do campus dos quais tive de correr. Obrigado senhor por conseguir correr!!!

Por fim, agradeço a um grande amigo e médico que me ajudou nos momentos difíceis, mostrando-se um grande ser humano, valeu Daniel Cirelli Guedes!

Agradeço, também, a oportunidade de ter parte dos impostos pagos por mim, revertidos em bolsa para minha qualificação. Obrigado a CAPES pela bolsa **(BEX1357/03-4)**.

Por fim, abro um espaço para uma pessoa muito especial desta grande família na qual me insiro, que do alto dos seus quase 90 anos me deu o maior incentivo possível para terminar esta tarefa. Quando eu já estava cansado de recomendar, minerar, colaborar, e estava a fim de seguir outros caminhos, com toda a ternura e com a sabedoria de quem transforma quedas em saltos ornamentais, frustrações em incentivos pra seguir em frente, tristezas em lágrimas que se confundem com a emoção da palavra e tudo isso em ternura. Este trabalho eu dedico a você **Dona Ida Borsatto Ferreira!**Pela honra de ser seu neto.

Agradeço a você também, que teve paciência para chegar até aqui. Espero que a leitura seja proveitosa!

# SUMÁRIO

<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>10</b>
<b>LISTA DE FIGURAS .....</b>	<b>12</b>
<b>LISTA DE TABELAS .....</b>	<b>15</b>
<b>RESUMO.....</b>	<b>17</b>
<b>ABSTRACT.....</b>	<b>18</b>
<b>1 INTRODUÇÃO .....</b>	<b>19</b>
1.1 Motivação .....	21
1.2 O que é a tese.....	22
1.3 Objetivos.....	22
1.4 Contribuições da tese .....	22
1.5 Organização do Texto .....	23
<b>2 SISTEMAS DE RECOMENDAÇÃO .....</b>	<b>25</b>
2.1 Recuperação de Informação .....	26
2.2 Filtragem de Informação X Recuperação de Informação .....	27
2.3 Sistemas de Recomendação.....	27
2.4 Técnicas de Filtragem aplicadas a Sistemas de Recomendação .....	29
2.4.1 Filtragem Baseada em Conteúdo .....	29
2.4.2 Filtragem Colaborativa.....	30
2.4.3 Filtragem Híbrida.....	34
2.5 Exemplos de Sistemas de Recomendação.....	35
2.5.1 Projetos Acadêmicos .....	37
2.5.2 Sites Comerciais .....	45
2.6 Limitações dos Sistemas de Recomendação .....	48
2.6.1 Problema do primeiro avaliador.....	48
2.6.2 Similaridade .....	48
2.6.3 Novo item .....	49
2.6.4 Novo usuário.....	49
2.6.5 Escalabilidade .....	49
2.6.6 Esparsialidade .....	49

2.6.7	Superespecialização.....	50
2.7	<b>Privacidade em Sistemas de Recomendação .....</b>	<b>50</b>
<b>3</b>	<b>A RELEVÂNCIA DA OPINIÃO DO RECOMENDADOR .....</b>	<b>53</b>
3.1	<b>O modelo Mo-DROP .....</b>	<b>55</b>
3.1.1	A métrica para o Ranqueamento do Recomendador.....	57
3.1.2	Equações.....	57
3.1.3	Exemplo de Aplicação.....	59
3.2	<b>Aplicando o Mo-DROP .....</b>	<b>60</b>
3.3	<b>Utilizando a métrica em recomendação colaborativa .....</b>	<b>62</b>
<b>4</b>	<b>PREDIZENDO INTERESSES DOS USUÁRIOS .....</b>	<b>64</b>
4.1	<b>Regras de Associação .....</b>	<b>66</b>
4.2	<b>Um framework para prever o interesse dos usuários (F-UIP) .....</b>	<b>68</b>
4.2.1	Fase I – Mineração de dados .....	68
4.2.2	Fase II – Filtrando regras de associação a partir da relevância da opinião do usuário .....	70
4.2.3	Fase III – Selecionando usuários para aplicar o conhecimento novo .....	71
4.3	<b>Algoritmo para predição de interesses .....</b>	<b>71</b>
<b>5</b>	<b>O PROTÓTIPO W-RECMAS .....</b>	<b>74</b>
5.1	<b>CV-Lattes e Qualis .....</b>	<b>76</b>
5.2	<b>Agentes de software.....</b>	<b>80</b>
5.2.1	Agente Buscador .....	80
5.2.2	Agente Analista.....	81
5.2.3	Agente Recomendador .....	83
5.2.4	Agente Pessoal .....	84
5.2.5	Agente Comunitário .....	85
5.3	<b>Arquitetura do protótipo W-RECMAS.....</b>	<b>88</b>
5.3.1	Implementação do W-RECMAS.....	90
5.3.2	Interface do protótipo .....	91
5.4	<b>Limitações e Dificuldades.....</b>	<b>99</b>
<b>6</b>	<b>EXPERIMENTOS E RESULTADOS .....</b>	<b>100</b>
6.1	<b>Descrição dos experimentos realizados .....</b>	<b>100</b>
6.1.1	Primeiro Experimento: Ponderação de atributos do CV-Lattes.....	100
6.1.2	Segundo Experimento: Aplicando o modelo Mo-DROP .....	105
6.1.3	Terceiro Experimento: Interfaces de Recomendação.....	111
6.1.4	Quarto Experimento: Avaliação humana x agente de software.....	116
6.1.5	Quinto Experimento: Aplicando o <i>framework</i> F-UIP .....	119
6.1.6	Sexto Experimento: Recomendando itens.....	122
6.2	<b>Conclusões .....</b>	<b>126</b>
6.2.1	Limitações dos experimentos.....	127
<b>7</b>	<b>CONCLUSÕES .....</b>	<b>128</b>
7.1	<b>Trabalhos Futuros.....</b>	<b>130</b>



<b>7.2</b>	<b>Limitações.....</b>	<b>132</b>
<b>7.3</b>	<b>Considerações Finais.....</b>	<b>132</b>
	<b>REFERÊNCIAS.....</b>	<b>133</b>
	<b>ANEXO A ANÁLISE DOS SISTEMAS DE RECOMENDAÇÃO E W-RECMAS .....</b>	<b>144</b>
	<b>ANEXO B TEXTO DO PRIMEIRO EXPERIMENTO .....</b>	<b>150</b>
	<b>ANEXO C TEXTO DO TERCEIRO EXPERIMENTO .....</b>	<b>154</b>
	<b>ANEXO D TEXTO DO QUARTO EXPERIMENTO .....</b>	<b>157</b>
	<b>ANEXO E TEXTO DO SEXTO EXPERIMENTO.....</b>	<b>159</b>
	<b>ANEXO F BANCO DE DADOS .....</b>	<b>163</b>
	<b>ANEXO G COMUNIDADES VIRTUAIS .....</b>	<b>176</b>

## LISTA DE ABREVIATURAS E SIGLAS

AAE	Avaliação de um agente de software na área especialista
ACL	<i>Agent Communication Language</i>
AHE	Avaliação humana na área especialista
AHNE	Avaliação humana na área não especialista
AI	Área de Interesse
ANE	Avaliação de um agente de software na área não especialista
APP	Artigos publicados em periódicos
AT	Demais tipos de produção técnica (apresentação de trabalhos)
CoCoA	<i>Compilation Compiler Advisor</i>
CoFE	<i>Collaborative Filtering Engine</i>
CORA	<i>Collaborative Recommender Agent</i>
DAE	Desvio na área especialista
DCBD	Descoberta de conhecimento em base de dados
DM	Dissertação de mestrado
DNE	Desvio na área não especialista
DTP	Demais tipos de produção bibliográfica
DTR	Demais trabalhos relevantes
EA	Especialização/Aperfeiçoamento
FBC	Filtragem baseada em conteúdo
FC	Filtragem colaborativa
FConf	Fator de confiança
FE	<i>Feedback</i> explícito
FH	Filtragem híbrida
FIPA	<i>Foundation for Intelligent Physical Agents</i>
FO	Formação
FP	<i>Feedback</i> pontuação
FSup	Fator de suporte
F-UIP	<i>Framework</i> para predizer o interesse dos usuários
GFF	<i>Gradual forgetting function</i>
IA	Inteligência Artificial
IAD	Inteligência Artificial Distribuída
IC	Iniciação científica
J2SE	<i>Java 2 Standard Edition</i>
JADE	<i>Java Agent Development</i>

LA	Livros Acessados
LCL	Livros ou Capítulos de livro
LHS	<i>Left Hand Side</i>
LL	Livros Lidos
Mo-DROP	Modelo para Determinação da Relevância da Opinião do usuário
OC	Número de orientações concluídas
OA	Orientações em andamento
P2P	<i>Peer-to-Peer</i>
PBCJ	Participação em banca de comissões julgadoras
PBTC	Participação em banca de trabalho de conclusão
PE	Participação em eventos
PP3	<i>Platform for privacy preferences</i>
PT	Produtos Tecnológicos
RCF	Recomendações com <i>feedback</i>
RF	Recomendações futuras
RHS	<i>Right Hand Side</i>
RR	Ranqueamento do Recomendador
RRarea	Ranqueamento do Recomendador na área de interesse do usuário
RRtotal	Ranqueamento total do Recomendador
RSF	Recomendações sem <i>feedback</i>
SGBD	Sistema Gerenciador de Banco de Dados
SW	<i>Software</i>
TAE	Trabalhos em anais de eventos
TC	Trabalho de conclusão
TD	Tese de doutorado
TJR	Textos em jornais ou revistas
TT	Trabalhos técnicos
W3C	<i>World wide web consortium</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
W-RECMAS	<i>Web-RECommender system based on Multi-Agent System for academic paper recommendation</i>
XML	<i>Extensible Markup Language</i>

## LISTA DE FIGURAS

Figura 2.1: Filtragem Híbrida.....	35
Figura 2.2: Visão geral da arquitetura do Fab [BAL 97] .....	40
Figura 2.3: Interface da ferramenta CORA [LUE 98] .....	41
Figura 2.4: Interface de recomendações do MovieLens .....	42
Figura 2.5: Criação de compilações no CoCoA .....	43
Figura 2.6: Cadastramento para construção de perfil no CoCoA.....	44
Figura 2.7: Página pessoal com compilações no CoCoA .....	44
Figura 2.8: Interface do Drugstore.com .....	48
Figura 2.9: Política de privacidade da loja Amazon.com <sup>TM</sup> .....	51
Figura 2.10: Selo da empresa TRUSTe no <i>web site</i> da empresa NetRaker .....	52
Figura 4.1: Visão macro da descoberta de relações entre áreas .....	66
Figura 4.2: <i>Framework</i> para predizer o interesse dos usuários .....	68
Figura 4.3: Algoritmo de predição de possíveis interesses dos usuários .....	73
Figura 5.1: Visão macro da arquitetura do protótipo W-RECMAS .....	75
Figura 5.2: Indicadores de produção no CV-Lattes do pesquisador.....	77
Figura 5.3: Classificação Qualis .....	78
Figura 5.4: Classificação Qualis - Ciência da Computação .....	78
Figura 5.5: Esquema do cálculo do RR.....	79

Figura 5.6: Informações para o cálculo do RR dos usuários.....	79
Figura 5.7: Arquitetura interna e interfaces do Agente Buscador .....	81
Figura 5.8: Arquitetura interna e interfaces do Agente Analista.....	83
Figura 5.9: Arquitetura interna e interfaces do Agente Recomendador .....	84
Figura 5.10: Arquitetura interna e interfaces do Agente Pessoal .....	84
Figura 5.11: Aplicando a tecnologia <i>push</i> para recomendar.....	85
Figura 5.12: Arquitetura interna e interfaces do Agente Comunitário .....	86
Figura 5.13: Estabelecendo uma comunidade .....	87
Figura 5.14: Visão geral da arquitetura do protótipo W-RECMAS .....	88
Figura 5.15: Comunicação para estabelecer uma comunidade no W-RECMAS.....	90
Figura 5.16: Interface de <i>login</i> do protótipo W-RECMAS.....	92
Figura 5.17: Interface de cadastramento no protótipo W-RECMAS .....	93
Figura 5.18: Interface de busca e posterior inclusão de artigos no protótipo W-RECMAS .....	94
Figura 5.19: Interface de recomendação de artigos no protótipo W-RECMAS .....	95
Figura 5.20: (a) Legenda da relevância dos recomendadores; (b) Legenda da predição calculada para recomendação.....	95
Figura 5.21: Interface de feedback da recomendação no protótipo W-RECMAS .....	96
Figura 5.22: Interface de Recomendação de Usuários.....	97
Figura 5.23: Interface de Mensagens .....	97
Figura 5.24: Interface do Perfil e Comunidade .....	98
Figura 6.1: Acadêmicos e atuação com análise prévia de CV-Lattes.....	102
Figura 6.2: Acadêmicos e tempo de doutoramento .....	102
Figura 6.3: Acadêmicos e áreas de atuação.....	103

Figura 6.4: Interface 1 do terceiro experimento .....	112
Figura 6.5: Interface 2 do terceiro experimento .....	113
Figura 6.6: Distribuição dos acadêmicos e profissionais do terceiro experimento .....	114
Figura 6.7: Avaliação humana x avaliação agente software .....	118
Figura 6.8: Distribuição das formações na base de dados sintética.....	120
Figura 6.9: Áreas de interesse e usuários .....	120
Figura 6.10: Distribuição dos valores do Ranqueamento do Recomendador .....	121
Figura A.1: Banco de dados do protótipo .....	163

## LISTA DE TABELAS

Tabela 2.1: Matriz de avaliações usuário x item .....	31
Tabela 2.2: Itens X Usuários .....	33
Tabela 2.3: Similaridade através do coeficiente de Pearson .....	34
Tabela 2.4: Classificação de Sistemas de Recomendação .....	36
Tabela 3.1: Atributos qualitativos dos usuários .....	59
Tabela 3.2: Atributos quantitativos dos usuários.....	60
Tabela 3.3: Valores não normalizados dos atributos .....	61
Tabela 3.4: Atributos normalizados para o cálculo do RR .....	61
Tabela 3.5: Resultados do Ranqueamento do Recomendador .....	62
Tabela 4.1: Usuário e áreas de interesse .....	69
Tabela 4.2: Regras e Ranqueamento do Recomendador.....	70
Tabela 6.2: Atributos quantitativos do CV-Lattes e seus pesos .....	106
Tabela 6.3: Peso das categorias de publicações .....	107
Tabela 6.4: Atributos não normalizados para o cálculo do RR.....	109
Tabela 6.5: Atributos normalizados para o cálculo do RR .....	110
Tabela 6.6: Resultados do Ranqueamento do Recomendador .....	111
Tabela 6.7: Respostas do terceiro experimento.....	115
Tabela 6.8: Resultados das avaliações .....	117

Tabela 6.9: Regras Geradas.....	122
Tabela A.1. Análise dos Sistemas de Recomendação.....	145
Tabela G.1: Número de membros em comunidades virtuais .....	177
Tabela G.2: Relação de usuários e áreas de interesse comum .....	179



## RESUMO

As pessoas têm acesso a uma vasta gama de informações devido a grande oferta e aos recursos da Internet, porém despendem muito tempo na busca do que realmente é interessante ou útil para elas. A dificuldade de encontrar a informação correta é aumentada quando a informação disputa a atenção de uma pessoa com uma série de outras informações não tão relevantes.

Procurando minimizar esta dificuldade e auxiliar no acesso a informação interessante, são aplicados desde sistemas de recuperação de informação até sistemas de filtragem de informação. Os sistemas de recuperação são amplamente difundidos na Internet através dos motores de busca (por exemplo, google.com, av.com, citeseer.ist.psu.edu), porém um problema neste tipo de aplicação constitui-se na necessidade do usuário apresentar os termos (palavras-chave) que são relevantes para a consulta.

A filtragem de informação, tendo como representante os Sistemas de Recomendação, surge como uma nova abordagem que procura liberar o usuário da exigência de criar consultas com palavras-chave, ou seja, a filtragem baseada em conteúdo procura casar o perfil do usuário e o conteúdo dos itens a serem recomendados, e então, oferecer alguns destes itens aos usuários. Por fim, surgiram sistemas que não se baseavam na análise do conteúdo dos itens, mas sim na reputação de um item junto aos outros usuários, ou seja, o usuário recebe a recomendação de um item que pode ser do seu interesse frente à colaboração de outros usuários que avaliaram o item. A questão maior nesta abordagem está no quanto a opinião de um usuário que avaliou um item é relevante para servir como colaboração na elaboração da recomendação para outro usuário.

Esta tese constitui-se em uma proposta para modelar e incluir a relevância da opinião do usuário no processo de recomendação colaborativa, ou seja, apresenta uma abordagem de Sistemas de Recomendação para recomendar itens baseando-se em informação adicional – definida como relevância da opinião do usuário – além das típicas informações utilizadas na grande maioria dos Sistemas de Recomendação. Esta inclusão da relevância da opinião constitui-se em uma alternativa para que o usuário alvo da recomendação consiga identificar qual a importância de um determinado item recomendado frente à relevância de opinião dos recomendadores. A idéia apresentada é a de que pessoas com maior relevância de opinião poderiam melhor avaliar e recomendar itens.

**Palavras-Chave:** Sistemas de Recomendação, Modelagem de Perfil de Usuário, Sistemas Multiagentes, Mineração de Dados e Comunidades Virtuais.

## **“Applying user’s opinion relevance in a Recommender System to Researchers.”**

### **ABSTRACT**

Nowadays, people have access to a huge amount of information due to the Internet's resources. However they spend too much time searching for interesting, adequate or useful information. The difficulty to find worthwhile information increases when interesting things dispute the user's attention.

Information retrieval and information filtering systems are applicable in order to minimize search difficulties, aiming to aid the user in the search for worthwhile information. Information retrieval systems are widely spread in the Internet through search engines (e.g., google.com, av.com, citeseer.ist.psu.edu). However there is a problem in this kind of application, which consists in compelling the user to know the terms (keywords) that are relevant for the search.

Recommender Systems are an information filtering solution. They present a different approach that frees the user from creating queries with keywords. It means that the system tries to match the user's profile (historical interests) with the content of items to be recommended, and then offers these items to the user (recommendee). In parallel, an alternative approach to item recommendation was proposed, this one based on the offering of items based on other users’ opinion, i.e. the user receives an item recommendation based on the evaluation of other users (collaborative filtering or social filtering). However, a different question is raised here – how much the opinion of a user who evaluated an item is relevant to be employed in the recommendation process applying a collaborative method?

This thesis presents a new approach to model and include in the collaborative recommendation process additional information named Recommender's Rank, which represents the relevance of the user's opinion and complements the typical information used in the large majority of Recommender Systems. This approach is an alternative to aid the user to identify the importance of a recommended item based on other users' opinions, as people with higher relevance of opinion are more likely to better evaluate and recommend items.

**Keywords:** Recommender Systems, User Modeling, Multiagent Systems, Data Mining and Virtual Web Communities.

# 1 INTRODUÇÃO

Todas as pessoas dependem de informações para realizar de maneira adequada suas atividades diárias, sejam estas de caráter profissional ou pessoal. A grande questão refere-se a onde concentrar esforços para buscá-las. Caso seja analisada a quantidade de informação disponível atualmente, poder-se-ia concluir que uma pessoa está exposta a uma quantidade de informação muito maior do que conseguiria lidar.

As pessoas, normalmente, despendem muito tempo na busca de uma informação que realmente seja útil ou interessante, pois esta informação muitas vezes disputa a atenção de uma pessoa com uma série de outras informações, o que pode prejudicar a sua tomada de decisão ou o foco da sua análise.

Zaiane afirma que atualmente a produção de informação é maior do que um indivíduo poderia lidar durante todo o seu tempo de existência [ZAI 2000]. Verifica-se que este aumento considerável na quantidade de informação disponível deve-se em muito ao rápido desenvolvimento da Internet — uma das fontes mais democráticas de informações da atualidade. Nesta rede, qualquer pessoa pode incluir informações que passarão a fazer parte deste grande repositório tipicamente distribuído. Este aumento considerável somado a democratização no uso (onde um indivíduo apenas necessita de um computador e um provedor de acesso para fazer parte desta grande rede) e rápido aumento no número de usuários ao redor do mundo, têm permitido a formação de uma sociedade de informação [MUR 2004] [HAG 97] de aspecto global (também referida como comunidades virtuais ou *Internet-mediated Communities*) [WEL 2002] [BEC 2005].

Esta sociedade apresenta um aumento considerável em número de indivíduos e uma produção jamais vista, o que resulta em uma super oferta de informação. Por um lado esta super oferta e democratização da informação apresenta pontos positivos tais como a possibilidade de que todos tenham acesso ao seu conteúdo, porém por outro lado apresenta alguns pontos negativos, como a dificuldade de encontrar exatamente o que se deseja, ou seja, aquela informação que efetivamente irá auxiliar na atividade desempenhada [MON 2003].

Com esta grande oferta ocorre uma sobrecarga de informação (*information overload*), onde o usuário pode vir a sentir-se perdido, sem estímulo ou ajuda apropriada para conseguir encontrar entre tanta informação aquela que realmente poderá atendê-lo [BEL 2000]. A questão que pode ser levantada neste momento refere-se a –

qual recurso o usuário poderia utilizar para diminuir esta sensação de sobrecarga? Uma possibilidade são os vários motores de busca oferecidos pela Internet. Como exemplo de motores de busca, amplamente divulgados na Internet, pode-se citar: AltaVista ([www.av.com](http://www.av.com)), Google ([www.google.com](http://www.google.com)), Google Scholar ([scholar.google.com](http://scholar.google.com)), MetaCrawler ([www.metacrawler.com](http://www.metacrawler.com)), HotBot ([www.hotbot.com](http://www.hotbot.com)), Lycos ([www.lycos.com](http://www.lycos.com)) e um dos precursores, o Yahoo ([www.yahoo.com](http://www.yahoo.com)). Estes motores de busca são muito úteis na chamada recuperação de informações, ou seja, na aplicação de técnicas que procuram apresentar toda a informação que “casar” com as palavras-chave de uma consulta realizada pelo usuário.

Um problema deste tipo de ferramenta é que o usuário obriga-se a explicitar o seu interesse, ou seja, deve saber apresentar as palavras-chave que são relevantes para a consulta. Estas ferramentas apresentam-se como excelentes soluções, porém retornam uma quantidade muito grande de *links*, sendo que a grande maioria destes *links* não apresenta a qualidade que o usuário espera. A grande questão hoje seria o quanto estas ferramentas estão sendo realmente úteis para o usuário encontrar a informação desejada [MON 2003].

Procurando auxiliar na solução do desafio de encontrar informações úteis, surgem novas propostas de sistemas que buscam focar não somente na recuperação de informação pelo casamento de consulta e conteúdo, mas procuram focar na filtragem de informação e possível recomendação das mesmas. Este esforço na busca por informações úteis poderia ser altamente amenizado caso os usuários da Internet fossem encorajados a disseminar suas informações e dividir suas opiniões com outros indivíduos, estabelecendo uma cultura de compartilhar informações (*information sharing*) [LUE 98].

No mundo real a recomendação de produtos, serviços ou qualquer tipo de item constitui-se em uma prática muito antiga, podendo ser feita através do que Pattie Maes e Paul Resnick denominam de recomendação boca-a-boca (*word-of-mouth*) [RES 97][SHA 95]. Surge desta forma o que foi denominado pelos autores de Sistema de Recomendação. Este tipo de sistema deve aprender sobre o usuário do sistema para então poder recomendar produtos, serviços ou qualquer outro tipo de item [RES 97], apresentando algumas limitações técnicas como, por exemplo, o problema de superespecialização, o qual refere-se ao fato do sistema só conseguir recomendar itens muito semelhantes àqueles que o usuário já avaliou, tornando-se, com o passar do tempo, um sistema pouco atraente ao usuário.

Segundo Miquel Montaner [MON 2003] as pesquisas na área de Inteligência Artificial têm empregado vários recursos na solução do problema de busca e recomendação da correta informação. Estes recursos aplicam uma extensão dos motores de busca, utilizando motores personalizáveis, Sistemas de Recomendação, agentes de software (assistentes inteligentes), mineração de dados na Web, tudo para facilitar este processo de busca de informações e recomendação acertada. Segundo John Riedl [RIE 98], o paradigma de agentes pode auxiliar na realização desta atividade por incluir indivíduos computacionais e autônomos. Em um de seus artigos Pattie Maes [MAE 94] apresenta preocupação com esta sobrecarga de informações a que o usuário está exposto e afirma que a combinação de modelagem de preferências do usuário, construção de

modelos e modelagem de padrões sociais com agentes inteligentes, poderia auxiliar a suprir a necessidade de informações.

Assistentes inteligentes podem auxiliar na realização da tarefa de decidir a forma mais acertada do que deve ser recomendado ao usuário e como. Os assistentes inteligentes ou agentes de software são entidades modulares e autônomas [BRW 98], e encontram seu embasamento teórico na Inteligência Artificial (IA) e na Inteligência Artificial Distribuída (IAD). Estes se constituem em especialistas no trabalho que realizam e não trabalham isoladamente, mas de forma cooperativa tentam resolver um problema da melhor maneira possível, caracterizando uma área da Inteligência Artificial Distribuída denominada de Sistemas Multiagentes [WEI 99] [FRE 2002] [SIN 99] [YOS 99].

Com o crescimento dos recursos oferecidos pela Internet, a possibilidade de substituir a presença física pela eletrônica neste tipo de atividade tomou forma [MAE 94]. Muitas das atividades desempenhadas pelos usuários poderiam ser delegadas a assistentes inteligentes [BRA 97] [ETZ 95], os quais teriam a finalidade de auxiliá-los na busca de informações interessantes [BOL 98], deixando tempo disponível para que o usuário realize outras atividades. Uma aplicação interessante resultante da união da abordagem de Sistemas Multiagentes e filtragem de informação, é apresentada por Moukas [MOU98] no sistema denominado Amalthea.

## 1.1 Motivação

A Internet, sem sombra de dúvidas, tornou-se a maior fonte de informações para os usuários. O problema está na dificuldade de um usuário encontrar informações realmente relevantes e confiáveis para a realização de seu trabalho. Existem muitos *sites* e geralmente as informações são disponibilizadas por outros usuários, sendo assim a questão seria qual fonte consultar. Possivelmente a melhor fonte será oferecida pelo usuário que tiver maior conhecimento ou relevância de opinião na área em que este disponibiliza informações.

Cinco pontos podem ser destacados como sendo motivadores para esta tese de doutorado:

- 1º) a crescente sobrecarga de informação a que os usuários têm sido submetidos;
- 2º) o desafio de encontrar o que lhes interessa, constituindo informação útil e interessante;
- 3º) a dificuldade de identificar outros usuários que partilhem de interesses e gostos comuns;
- 4º) a falta de uma métrica que represente a relevância da opinião de usuários em Sistemas de Recomendação;
- 5º) a necessidade de lidar com o problema da superespecialização.

## 1.2 O que é a tese

Esta tese constitui-se em uma proposta para modelar e incluir a relevância da opinião do usuário no processo de recomendação colaborativa, ou seja, apresenta uma abordagem de Sistemas de Recomendação para recomendar itens baseando-se em informação adicional definida como relevância da opinião do usuário, além das típicas informações utilizadas na grande maioria dos Sistemas de Recomendação.

A tese busca modelar a relevância da opinião dos usuários nas suas respectivas áreas de interesse, e aplicar esta relevância ao recomendar itens aos usuários de Sistemas de Recomendação.

Cabe destacar nesta seção que esta tese concentra seu foco em modelar a relevância da opinião de usuários (autoridade) para aplicação em Sistemas de Recomendação, e não em relações de confiança (*trust*) entre usuários e nem reputação (*reputation*) de usuários de Sistemas de Recomendação.

## 1.3 Objetivos

O objetivo principal desta tese de doutorado é permitir modelar a relevância da opinião de um usuário em seu domínio de atuação e desta forma contribuir para a área de Sistemas de Recomendação a partir da inclusão de uma métrica que representa a relevância da opinião.

Objetivos Específicos:

1º) criação de um modelo genérico que permita a identificação da relevância da opinião do usuário em um Sistema de Recomendação;

2º) aplicar o modelo desenvolvido no processo de recomendação colaborativa;

3º) buscar reduzir a problemática da superespecialização.

## 1.4 Contribuições da tese

As principais contribuições desta tese para a área de Sistemas de Recomendação estão listadas a seguir:

- 1) Uma grande questão em Sistemas de Recomendação refere-se à definição da relevância da opinião de quem participa de um Sistema de Recomendação, uma vez que em filtragem colaborativa, a opinião de um usuário sobre dado item é fundamental. Nesta tese, é apresentado um modelo completo que permite representar em Sistemas de Recomendação a relevância da opinião de um usuário para sua área de interesse.
- 2) Uma outra questão de relevância para a comunidade que pesquisa os Sistemas de Recomendação recai sobre a dificuldade do sistema criar o perfil atualizado dos usuários, uma vez que na grande maioria dos sistemas o

usuário tende a possuir um perfil estático, ou seja, não progride com o passar do tempo, causando desinteresse dos usuários. Nesta tese, é apresentado um *framework*, que procura auxiliar na preservação dos usuários, pois através da aplicação do mesmo, pode-se predizer novos interesses dos usuários, baseando-se para tanto na opinião do grupo, e “surpreender” o usuário com novas ofertas de produtos (itens) em áreas de interesse distintas.

- 3) A apresentação e especificação de uma arquitetura completa de Sistema de Recomendação baseada em agentes de software e mineração de dados constitui-se em uma outra contribuição deste trabalho.
- 4) Uma nova abordagem no cálculo da recomendação de um item ao usuário do sistema de recomendação. A inclusão da relevância da opinião do usuário que avalia um item no cálculo de predição de um item para um usuário alvo constitui-se em contribuição para a área de Sistemas de Recomendação.

## 1.5 Organização do Texto

O texto desta tese encontra-se organizado em oito capítulos, incluindo os capítulos da Introdução, Conclusão e Bibliografia.

No Capítulo 2 são apresentados os conceitos de Sistemas de Recomendação, filtragem de informação, recuperação de informação, técnicas para recomendação e suas limitações. O conteúdo deste capítulo foi publicado parcialmente em [CAZ 2005d].

No terceiro capítulo é descrito em detalhes e exemplificado o Modelo para Determinação da Relevância da Opinião do Recomendador (Mo-DROP), modelo publicado em sua totalidade em [CAZ 2005b], e parcialmente em [MIL 2005]. Também é descrita a nova abordagem para o cálculo da recomendação utilizando a relevância da opinião como um dos fatores [SIL 2005]. O modelo proposto atende mais especificamente a primeira contribuição descrita na Seção 1.4 e a abordagem de como calcular a recomendação atende a quarta contribuição descrita na mesma seção.

O Capítulo 4 descreve em detalhes e exemplifica o *framework* para predição de interesse do usuário (F-UIP) baseado em tarefa de mineração de dados, para predizer interesses dos usuários em Sistemas de Recomendação. Este *framework* foi publicado em sua totalidade em [CAZ 2005c], as propostas iniciais do mesmo foram publicadas em [CAZ 2003], [CAZ 2004] e [CAZ 2005a]. Este *framework* apresenta uma proposta de solução para minimizar a problemática da superespecialização em Sistemas de Recomendação descrita no objetivo da tese, e atende mais especificamente a segunda contribuição descrita na Seção 1.4.

No Capítulo 5 é apresentado em detalhes o protótipo desenvolvido bem como a arquitetura multiagentes proposta. A descrição do protótipo foi publicada de acordo com seu progresso em [CAZ 2003], [CAZ 2004], [CAZ 2005a], [BRE 2005], [MIL 2005], [SIL 2005] e [CAZ 2006]. Este Capítulo descreve cada um dos agentes bem como as suas interações, tarefas e arquitetura interna. Esta arquitetura constitui-se na terceira contribuição descrita na Seção 1.4.

O sexto capítulo apresenta os experimentos realizados (e seus resultados) com a utilização de base de dados real e com a participação de acadêmicos (neste caso buscando verificar a percepção dos mesmos quanto ao ganho com as soluções propostas por esta tese).



## 2 SISTEMAS DE RECOMENDAÇÃO

A quantidade de informação produzida e disponibilizada pela Internet pode ocasionar uma sobrecarga sobre o usuário final. Devido a este problema várias tecnologias têm surgido visando apoiar a seleção, recuperação e filtragem da informação desejada ou de interesse do usuário. Esta recuperação na maioria das vezes é realizada por uma apresentação explícita das necessidades do usuário, ou seja, uso de palavras-chave.

A demanda por tecnologias de filtragem de informação não é algo novo, como afirma Loeb [LOE 92]. A preocupação com a filtragem de informação já vem de longa data. Peter Denning já escrevia no volume 25 da Communications of ACM do ano de 1982 [DEN 82] sobre a preocupação no que se refere à quantidade de informação que estava sendo gerada pelos diversos tipos de sistemas e recebida pelos usuários (o que o autor denominou de *eletronic junk*). Ele destacava que toda a atenção estava concentrada na geração da informação para suprir as necessidades do usuário, e destacava que já era tempo de começar a se preocupar com o recebimento da informação, ou seja, o processo de controlar, recuperar e filtrar a informação para que esta alcançasse a pessoa que deveria realmente utilizá-la.

A questão maior é como deve ser entendida a expressão filtragem de informação. Para Belvin [BEJ 92] a filtragem de informação é o nome utilizado para descrever uma variedade de processos que envolvem a entrega de informação para as pessoas que realmente necessitam delas. Segundo o autor este termo tornou-se muito popular, e artigos técnicos descrevem aplicações focadas em filtragem aplicada a e-mail, documentos eletrônicos em escritórios, mas não fica clara a distinção entre filtragem e processos relacionados com recuperação de informações somente.

Um sistema de filtragem de informação geralmente apresenta as seguintes características [BEJ 92]:

1. Constitui-se em um sistema projetado para suportar e trabalhar com dados não estruturados ou semi-estruturados. Isto contrasta bastante com as aplicações típicas de banco de dados, as quais envolvem dados altamente estruturados.
2. Sistemas de filtragem de informação lidam basicamente com informação textual, incluindo também dados de imagem, voz, e vídeo.

3. Sistemas de filtragem envolvem grande quantidade de dados. Aplicações típicas devem lidar com *gigabytes* de texto ou com grande quantidade de dados de outras mídias.

4. Filtragem também tem sido usada para descrever a maneira de acessar e recuperar informações a partir de base de dados remotas, neste caso os dados são um resultado de pesquisa na base de dados. Este cenário é bastante utilizado por desenvolvedores para gerar “agentes inteligentes” para buscas.

5. Filtragem baseia-se na descrição das preferências (perfis) de indivíduos ou grupos. Segundo o autor os perfis representam os interesses de longa duração.

## 2.1 Recuperação de Informação

Segundo Belvin [BEJ 92] existem muitas semelhanças entre os termos filtragem e recuperação de informação, mas mesmo assim, é possível realizar uma separação entre estas terminologias. Uma definição apresentada por Robertson (Robertson apud [BEJ 92]) diz que um sistema de recuperação é assim considerado se o mesmo tiver a funcionalidade de conduzir o usuário para aqueles documentos que melhor irão habilitá-lo a satisfazer as suas necessidades de informação.

Segundo Foltz e Herlocker [FOL 92] [HER 2000] as aplicações de recuperação de informação envolvem tecnologias de consulta e indexação baseadas na análise automática do conteúdo do item, essencialmente para documentos textuais. Indexar pode ser descrito como o processo de examinar coleções de conteúdos e criar uma estrutura de dados que contenha descrições dos itens avaliados. Uma das questões-chaves neste processo está na identificação do que é realmente importante e do que é irrelevante, sendo que retirar o que é irrelevante é fundamental. A identificação de termos semelhantes também é importante para que não ocorra perda de conteúdo dito relevante, ou seja, caso seja solicitada a recuperação de informações relacionadas com o assunto “agentes cognitivos”, seria importante recuperar conteúdos, por exemplo, referentes a “modelagem de conhecimento”.

Os sistemas de recuperação são uma ótima solução para os usuários que sabem realmente o que desejam buscar, ou seja, os que sabem definir com alguma exatidão os termos de sua consulta, mas muito pode ser perdido pelo fato do usuário não ter conhecimento de outros conteúdos que também seriam interessantes.

Os motores de busca de informações disponíveis na Internet são os maiores exemplos de ferramentas para recuperação de informações. Estas ferramentas são muito úteis na aplicação de técnicas que procuram apresentar toda a informação que casar com uma consulta realizada pelo usuário. Um problema deste tipo de aplicação é que o usuário é obrigado a explicitar o seu interesse, ou seja, o mesmo deve saber apresentar exatamente os termos que são relevantes para a busca (palavras-chave). Neste tipo de ferramenta o usuário deve estar disposto a despende um tempo significativo nas buscas através da aplicação de esforço próprio [LUE 98] [HER 2000]. Até um certo ponto estas ferramentas apresentam-se como soluções excelentes no processo de recomendação,

retornando uma grande quantidade de *links*, porém a grande maioria dos *links* não apresenta a qualidade que o usuário esperava.

## 2.2 Filtragem de Informação X Recuperação de Informação

As expressões “recuperação de informação” e “filtragem de informação” descrevem soluções que procuram auxiliar na resolução de problemas referentes à busca de informações. A expressão “recuperação de informação” tradicionalmente envolve armazenamento, índices, e tecnologia para recuperação de documentos textuais. Para que a recuperação realmente ocorra o usuário deve descrever sua necessidade de informação para que uma consulta seja realizada. A partir deste momento, o sistema busca casar a necessidade explicitada com os documentos armazenados. Este tipo de abordagem tende a manter uma base com característica mais estática no armazenamento das informações, sendo que a interação com o sistema acontece por iniciativa do usuário.

Diferentemente da recuperação [FOL 92] a filtragem propõe uma abordagem distinta. Esta abordagem geralmente mantém um perfil dos interesses do usuário. Neste tipo de sistema o grande diferencial está na existência da especificação clara de um perfil. A idéia é que esta abordagem tenha como característica um maior tempo de duração no que tange ao casamento de interesses, ou seja, a filtragem não se refere ao momento, mas sim às preferências, enquanto que a recuperação baseia-se na percepção de uma necessidade do momento pelo usuário. Esta filtragem deve ser aplicada a cada novo item adicionado na base de dados procurando verificar se este atende ao usuário.

## 2.3 Sistemas de Recomendação

Com a quantidade de informações e com a disponibilidade facilitada das mesmas pelo uso da Internet, as pessoas se deparam com uma diversificação muito grande de opções. Muitas vezes uma pessoa não possui nenhuma ou quase nenhuma experiência pessoal para realizar escolhas entre as várias alternativas que lhe são apresentadas durante uma busca por informações.

A questão relevante neste momento refere-se a como proceder nestes casos. Para minimizar as dúvidas e necessidades que temos frente à escolha entre alternativas, geralmente confiamos nas recomendações que são passadas por outras pessoas, as quais podem chegar de forma direta (*word of mouth*) [SHA 95], cartas de recomendação, opiniões de revisores de filmes e livros, impressos de jornais, entre outras.

Os sistemas de recomendação já não são novidade na aplicação em empresas e na Internet [AGG 99] [SLY 2002]. Tradicionalmente, os Sistemas de Recomendação lidam com aplicações que juntam dois tipos de entidades: usuários e itens. Resnick e Varian [RES 97] definem Sistemas de Recomendação como sistemas que utilizam as opiniões de uma comunidade de usuários para auxiliar indivíduos desta mesma comunidade a identificarem conteúdos de interesse em um conjunto de opções que poderiam caracterizar uma sobrecarga.

Segundo Schafer [SCH 99] [SCH 2001] os Sistemas de Recomendação são usados pelos *sites* de comércio eletrônico para sugerir produtos para seus clientes e fornecer informações que procuram auxiliar os clientes sobre qual produto deve ser adquirido.

Os Sistemas de Recomendação auxiliam no aumento da capacidade e da eficácia deste processo de indicação já bastante conhecido na relação social entre seres humanos [RES 97]. Em um sistema típico as pessoas fornecem avaliações de itens como entrada e o sistema agrega e direciona estas recomendações para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios deste tipo de sistema é realizar o correto casamento entre os itens que estão sendo avaliados e os usuários que estão recebendo a referida recomendação, ou seja, definir e descobrir este relacionamento de interesses.

O processo de recomendação inicia com a especificação de um conjunto inicial de pontuações que são fornecidas pelos usuários de forma explícita, ou de forma implícita inferida pelo Sistema de Recomendação. Uma vez que esta pontuação é fornecida, o Sistema procura estimar a pontuação pela função  $R$  para o par (usuário, item) nos itens ainda não pontuados pelo usuário alvo [ADO 2005].

$$R: \text{Usuários} \times \text{Itens} \Rightarrow \text{Pontuações} \quad (1)$$

Conceitualmente, uma vez que a função  $R$  é estimada para todo domínio de Usuários  $\times$  Itens, um Sistema de Recomendação pode selecionar o item  $i'_u$  com a maior pontuação (ou um conjunto dos  $K$  itens com as maiores pontuações) para o usuário  $u$  e recomendar estes itens para o usuário alvo [ADO 2005].

$$\forall u \in \text{Usuários}, i'_u = \underset{i \in \text{Itens}}{\text{arg Max}} R(u, i) \quad (2)$$

Os proponentes do primeiro Sistema de Recomendação denominado Tapestry [GOD 92], criaram a expressão “filtragem colaborativa”, visando designar um tipo de sistema específico no qual a filtragem de informação era realizada com o auxílio humano, ou seja, pela colaboração entre os grupos de interessados. Resnick [RES 97] prefere utilizar a expressão Sistemas de Recomendação, por ser um termo genérico e defende este posicionamento por dois motivos: primeiro porque os recomendadores podem não explicitar colaboração com os que as recebem, pois um pode não conhecer o outro, e por último os recomendadores podem sugerir itens de interesse particular, incluindo aqueles que poderiam ser desconsiderados. A questão aqui é como verificar o real interesse em colaborar, pois na realidade as recomendações esboçadas pelos recomendadores estão concentradas em seus interesses próprios, na verdade o grupo a ser formado para receber as recomendações deveria pelo menos apresentar indivíduos com interesses comuns (explícitos) ou comportamento comum (implícito).

Sistemas de filtragem colaborativa e sistemas de filtragem baseada em conteúdo são tipos de Sistemas de Recomendação que aplicam abordagens distintas, mas possuem como finalidade única à recomendação de itens para os usuários. Montaner [MON 2003] destaca que existe um terceiro tipo de filtragem de informação denominada de filtragem demográfica. A filtragem demográfica, por sua vez, utiliza a descrição de um

indivíduo para aprender o relacionamento entre um item em particular e o tipo de indivíduo que poderia vir a se interessar. Este tipo de abordagem utiliza as descrições das pessoas para conseguir aprender o relacionamento entre um item e o tipo de pessoa que gostaria deste item [PAZ 99].

Existem pesquisas como as descritas em [SAR 2000] [PRE 99] [TER 2001] [HAH 2002] [MON 2003] [ADO 2005a] que abordam o assunto Sistemas de Recomendação em detalhes. Estes artigos apresentam a análise de alguns Sistemas de Recomendação e de diferentes tecnologias aplicadas a estes sistemas.

## **2.4 Técnicas de Filtragem aplicadas a Sistemas de Recomendação**

A seguir são descritas três técnicas aplicadas a Sistemas de Recomendação, segundo os autores Herlocker [HER 2000] e Ansari [ANS 2000].

### **2.4.1 Filtragem Baseada em Conteúdo**

Segundo Herlocker [HER 2000], por muitos anos os cientistas têm direcionado seus esforços para aliviar o problema ocasionado com a sobrecarga de informações através de projetos que integram tecnologias que automaticamente reconhecem e categorizam as informações. Os sistemas baseados nesta técnica têm como objetivo gerar de forma automática, descrições dos conteúdos dos itens e comparar esta descrição de cada item com a descrição dos interesses dos usuários ou com o histórico de consumo do mesmo [BAL 97], visando verificar se o item é ou não relevante para o usuário em questão.

A descrição de interesses do usuário é obtida através de informações fornecidas pelo próprio usuário, através de uma consulta, ou aprendendo com os itens que o usuário consome (aqueles que o usuário gostou). Esta técnica é chamada de filtragem baseada em conteúdo (FBC) [HER 2000] [ANS 2000] porque o sistema realiza filtragem baseada na análise de conteúdo do item e no perfil do usuário.

Muitas ferramentas que aplicam esta forma de filtragem aplicam a técnica de indexação de frequência de termos (Salton apud [HER 2000]). Neste tipo de indexação, informações dos documentos e necessidades dos usuários são descritas por vetores com uma dimensão para cada palavra que ocorre na base de dados. Cada componente do vetor corresponde à frequência que uma respectiva palavra ocorre em um documento ou na consulta do usuário. Os vetores dos documentos que estão próximos aos vetores da consulta do usuário são considerados os mais relevantes para o usuário. Outros exemplos de tecnologias que são aplicadas para filtragem baseada em conteúdo são índices de busca booleana, onde a consulta constitui-se em um conjunto de palavras chaves unidas por operadores booleanos (Cleverdon apud [HER 2000]); sistemas de filtragem probabilística, onde o raciocínio probabilístico é aplicado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário (Fuhr, Robertson, Wong apud [HER 2000]).

Uma maneira de trabalhar com a filtragem baseada em conteúdo é através de uma solicitação de análise de itens feita ao próprio usuário, onde este deve avaliar

alguns itens indicando se estes são de interesse ou não. Uma vez realizada a avaliação, o sistema busca itens que estão casando em conteúdo com o que foi classificado como de interesse e desconsidera os que estão casando em conteúdo com o que foi classificado de não interesse.

Em recomendação baseada em conteúdo a pontuação  $R(u,i)$  do item  $i$  pelo usuário  $u$  é estimada baseando-se nas pontuações  $R(u,i')$  fornecidas pelo mesmo usuário  $u$  para outros itens  $i' \in Itens$  que são similares ao conteúdo do item  $i$ . Por exemplo, em uma aplicação clássica de recomendação de filmes para que o sistema recomende algum filme ao usuário  $u$ , este procura entender as preferências do usuário através da análise do conteúdo de um item pontuado com valor alto no passado e que case com o conteúdo (similar) de um filme ainda não assistido pelo usuário em questão. Desta forma somente os itens que apresentem um alto nível de similaridade com este filme já pontuado deverão ser recomendados [ADO 2005].

Este tipo de filtragem pode ser mais formalmente definida da seguinte forma, sendo  $Conteúdo(i)$  um conjunto de atributos que caracterizam o item  $i$ . Este conteúdo é computado, usualmente, pela extração de um conjunto de características do item  $i$  (conteúdo). Vários sistemas de recomendação baseados em conteúdo (exemplos no Anexo A) são aplicados na recomendação de itens de texto (artigos científicos, por exemplo), geralmente, o conteúdo destes itens é descrito por palavras-chave. A importância destes termos em relação ao texto pode ser medida pela frequência em que o termo ocorre ao longo do texto.

### 2.4.2 Filtragem Colaborativa

A abordagem da filtragem colaborativa (FC) foi proposta para atender pontos que estavam em aberto na filtragem baseada em conteúdo [HER 2000] [ANS 2000]. A Filtragem Colaborativa não se constitui em algo novo e apresenta uma proposta diferenciada dos sistemas de filtragem uma vez que não requerem compreensão ou conhecimento de conteúdo dos itens.

Nos sistemas colaborativos a essência está na troca de experiências entre as pessoas que possuem interesses comuns. Nestes sistemas os itens são filtrados baseado na avaliação feita pelos usuários daqueles itens, ao invés do conteúdo dos itens.

O primeiro sistema criado com esta abordagem foi o Tapestry [GOD 92] que permitia ao usuário especificar uma consulta como: “mostre-me todos os memorandos que uma determinada pessoa considera como importante”. A idéia neste tipo de abordagem é a de que membros de uma dada comunidade podem ser beneficiados pela experiência de outros antes de decidir por consumir uma dada informação.

Sistemas de Recomendação Colaborativa procuram prever a pontuação de um item para um consumidor em particular baseado em como outros usuários com gostos semelhantes ao usuário alvo previamente pontuaram o mesmo item. Mais formalmente, a pontuação  $R(u,i)$  do item  $i$  para o usuário  $u$  é estimada baseada na pontuação  $R(u',i)$  atribuída ao mesmo item  $i$  pelos usuários  $u'$  que são similares ao usuário em questão. O valor de uma pontuação desconhecida  $r_{u,i}$  para o usuário  $u$  e item  $i$  é normalmente

computado como uma agregação de pontuações de outros usuários (por exemplo, os  $N$  mais similares) para o mesmo item  $i$  [ADO 2005]:

$$r_{u,i} = \mathit{aggr}_{u' \in \hat{U}} r_{u',i} \quad (3)$$

Onde  $\hat{U}$  denota o conjunto de  $N$  usuários  $u'$  que são mais similares ao usuário  $u$  e que pontuaram o item  $i$ .

Uma questão importante em sistemas colaborativos refere-se a como coletar as preferências dos usuários, como definir no que o mesmo encontra-se interessado [GLO 2001]. Basicamente existem duas abordagens propostas: uma refere-se à obtenção destas preferências de forma explícita, onde o usuário apresenta explicitamente o que lhe é importante; outra propõe que se aprenda de forma implícita sobre o usuário. A idéia fundamental é aprender sobre o usuário dentro do comportamento [BEA 2002] que o mesmo apresenta (ações que este realiza).

A FC reforça o conceito de comunidade, na qual cada usuário contribui com suas avaliações para o desempenho geral do sistema. A entrada tradicional de um algoritmo de FC é uma matriz na qual as linhas representam usuários e as colunas os itens. Cada dado na matriz corresponde à avaliação (*rating*) dada por um indivíduo para determinado item. A tabela 2.1 exemplifica tal matriz, onde os itens referem-se a artigos científicos:

Tabela 2.1: Matriz de avaliações usuário x item

Usuários	Itens			
	(A1) Fab: Content-Based, Collaborative Recommendation	(A2) Using collaborative filtering to weave an information Tapestry	(A3) Social information filtering: Algorithms for automating "word of mouth"	(A4) GroupLens: Applying Collaborative Filtering to Usenet News
João	4	4	5	2
Maria	-	3	-	5
José	4	5	5	?
Rosa	-	2	3	5

Na tabela 2.1 pode-se observar uma interrogação onde se pretende especificar qual a avaliação que o usuário José daria para o artigo A4 caso tivesse acesso ao mesmo. Por análise verifica-se que João é o melhor vizinho de José, pois ambos possuem avaliações semelhantes aos artigos que leram no passado e avaliaram. A opinião de João para o artigo A4 influenciará fortemente na predição para José. Por outro lado, Maria e Rosa não concordam com José em certos artigos, logo não são considerados bons vizinhos e influenciarão menos do que João no cálculo da predição.

A técnica de Filtragem Colaborativa pode ser descrita em três fases [HER 99] [HER 2000], explicada do ponto de vista das recomendações realizadas para um usuário, também chamado de usuário alvo (*target*):

1. Calcular o peso de cada usuário do sistema em relação à similaridade com o usuário alvo (métrica de similaridade).
2. Selecionar um subconjunto de usuários com maiores similaridades (vizinhos) para considerar na predição.
3. Normalizar as avaliações e computar as predições ponderando as avaliações dos vizinhos com seus pesos.

Esta técnica também é chamada de “*k-nearest-neighbor*” ou “*user-based*” [HER 2000]. No primeiro passo, a definição da similaridade pode ser realizada através de diversos coeficientes, sendo mais comumente aplicado o coeficiente de correlação de Pearson [HER 2000] [TOR 2004].

$$W_{a,u} = \frac{\sum_{i=1}^m [(r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)]}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 * \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}} \quad (4)$$

Na equação (4),  $w_{a,u}$  é a correlação do usuário ativo  $a$  com um determinado usuário  $u$ ,  $r_{a,i}$  é a avaliação que o usuário ativo deu para o item  $i$ , e  $\bar{r}_a$  é a média de todas as avaliações do usuário ativo  $a$ , e  $r_{u,i}$  é o conjunto de avaliações do vizinho. Observa-se a necessidade de mais de uma avaliação em comum para que a correlação seja viável, e os resultados variam entre 1 para similaridade total, e -1 para total dissimilaridade.

No passo três, o cálculo da predição pode ser efetuado através da equação (5).

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n |w_{a,u}|} \quad (5)$$

O valor da predição  $p_{a,i}$  do item  $i$  para o usuário ativo  $a$  é a média ponderada das avaliações dadas ao item  $i$  pelos  $n$  vizinhos  $u$  do usuário ativo  $a$ . A quantidade  $n$  de melhores vizinhos (com maiores correlações) é uma escolha de cada sistema que utiliza a FC [TOR 2004].



Por exemplo, de acordo com a tabela 2.2 o usuário U2 avaliou os itens de A até J com as seguintes pontuações [5;1;2;2;1;7;1;-;6;5] (em uma escala de avaliação de [1-7] referindo-se a itens menos interessantes (1) a itens mais interessantes (7), conforme sua avaliação). Como se pode verificar, o item H não foi pontuado (avaliado) por este usuário.

Tabela 2.2: Itens X Usuários

Usuários	Itens									
	A	B	C	D	E	F	G	H	I	J
U1	1	-	-	2	1	7	-	-	2	3
U2	5	1	2	2	1	7	1	-	6	5
U3	1	5	6	4	3	1	2	-	-	2
U4	6	5	4	1	2	-	-	5	-	3
U5	1	-	2	5	6	7	-	1	-	4
U6	-	2	-	3	-	1	-	2	-	-
U7	4	2	2	-	-	6	1	6	-	6
U8	-	1	-	3	7	6	-	3	-	1
U9	6	-	-	6	6	2	-	-	5	6

De acordo com a matriz apresentada, o usuário U2 não gostou do item B, avaliando-o com o valor 1. Por outro lado, este usuário gostou do item I, avaliando-o com valor 6. Pode-se observar também que o usuário U2 tende a concordar com U7 e discordar de U3 de acordo com o valor de pontuação apresentado.

Portanto, a filtragem colaborativa recomenda itens de acordo com a similaridade de avaliações fornecidas pelos usuários para os mesmos itens. Para medir essa similaridade, pode-se utilizar o coeficiente de Pearson [TOR 2004].

Tendo novamente como exemplo os usuários U2 e U7 da tabela 2.2, as avaliações dos itens em comum seriam as pontuações [5;1;2;7;1;5] e [4;2;2;6;1;6] respectivamente, gerando uma média de 3,5 para ambos usuários. Aplicando a equação (4), obtêm-se o valor de 0,94.

Como se pode observar, os usuários U2 e U7 são bastante similares, tendo em vista que eles concordam em suas avaliações. Tendo como usuário alvo U2 e aplicando o algoritmo do coeficiente de Pearson em todos usuários, têm-se os seguintes resultados apresentados na tabela 2.3:

Tabela 2.3: Similaridade através do coeficiente de Pearson

Usuário	Pearson (U2)
U1	0,63
U3	-0,66
U4	0,34
U5	0,09
U6	-0,78
U7	0,94
U8	0,11
U9	-0,66

De acordo com os resultados da tabela 2.3 observa-se que o usuário U2 é bastante similar a U7, relativamente similar a U4 e não é similar a U3, U6 e U9 em gostos. Uma vez, descobertos quais são os usuários similares ao usuário alvo, pode-se então gerar a predição de novos itens para este usuário.

Aplicando-se a equação (5) para predizer a nota que o usuário U2 atribuiria ao item H, levando em consideração todos os itens que os usuários vizinhos, com limiar superior a 0,7 (limite assumido), pontuaram em comum com o usuário alvo, tem-se o resultado de 5,22. Este resultado significa que caso o usuário U2 consumisse o item H este daria como feedback uma nota igual a 5,22 (predição) para este item (tendo como base a avaliação dos vizinhos mais próximos).

Herlocker apresenta uma revisão completa dos objetivos, base de dados, e algoritmos referentes a filtragem colaborativa em [HER 2004].

### 2.4.3 Filtragem Híbrida

A abordagem da filtragem híbrida (FH) procura combinar filtragem colaborativa e filtragem baseada em conteúdo visando criar um sistema que possa melhor atender às necessidades do usuário [HER 2000] [ANS 2000]. Um exemplo de sistema que aplica esta abordagem é o Fab [BAL 97], o Recommendz [GAR 2005] e o Rascal [MCC 2005].

Essa abordagem é constituída de vantagens proporcionadas pela filtragem baseada em conteúdo e pela filtragem colaborativa, unindo o melhor das duas técnicas e eliminando as fraquezas de cada uma, conforme apresentado pela Figura 2.1.

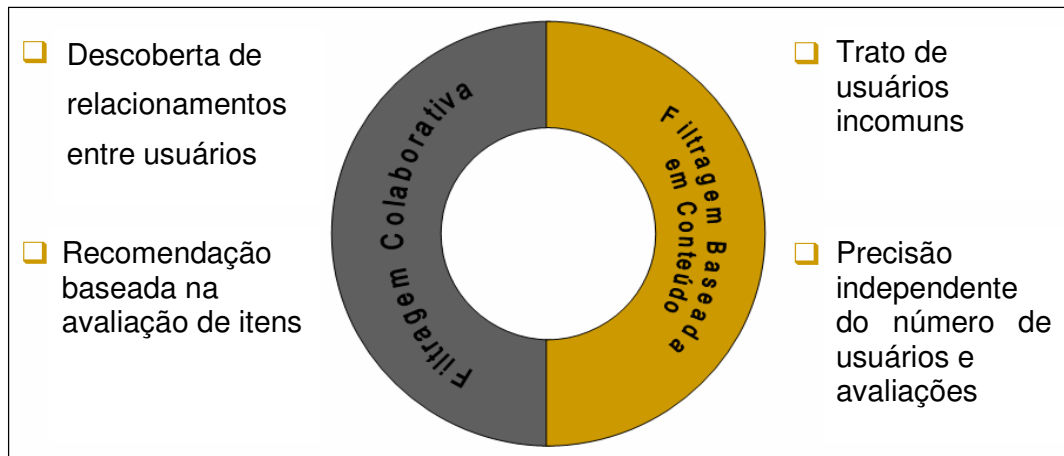


Figura 2.1: Filtragem Híbrida

## 2.5 Exemplos de Sistemas de Recomendação

Nesta seção serão descritos alguns exemplos de Sistemas de Recomendação, tanto acadêmicos como comerciais. Deve-se destacar que nenhum dos sistemas descritos a seguir propõe soluções para incluir a relevância da opinião dos usuários nas recomendações geradas, sendo que esta tese de doutorado apresenta preocupação com esta questão e propõe uma maneira de incluir a mesma em Sistemas de Recomendação.

A seguir é apresentada uma tabela de classificação de técnicas aplicadas a Sistemas de Recomendação proposta por Adomavicius [ADO 2005a]. Nesta tabela o autor cita uma série de pesquisas que fazem uso das técnicas apresentadas na Seção 2.4, classificando-as por técnica de recomendação e outras técnicas aplicadas. Esta tabela procura ilustrar o número de pesquisas que têm sido realizadas na área de Sistemas de Recomendação, destacando a aplicação de técnicas variadas.

Deve-se observar que o autor apresenta duas grandes classes de técnicas aplicadas a Sistemas de Recomendação (Breese apud [ADO 2005a]): baseada em Heurística (baseado em memória) e baseada em Modelo. O autor define técnicas baseadas em Heurística como sendo heurísticas que fazem previsões de avaliações baseando-se em toda a coleção de avaliações feitas pelos usuários aos itens, enquanto técnicas baseadas em Modelo utilizam a coleção de avaliações para aprender o modelo, o qual será utilizado para realizar previsões de avaliações.

Tabela 2.4: Classificação de Sistemas de Recomendação

Técnicas de Recomendação	Técnicas Aplicadas	
	Baseada em Heurística	Baseada em Modelo
Baseada em Conteúdo	<p>Técnicas mais utilizadas: TF-IDF (recuperação de informação) Agrupamento</p> <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> <li>• Lang, 1995</li> <li>• Balabanovic &amp; Shoham 1997</li> <li>• Pazzani &amp; Billsus 1997</li> </ul>	<p>Técnicas mais utilizadas: Classificadores Bayesianos Agrupamento Árvores de Decisão Redes Neurais</p> <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> <li>• Pazzani &amp; Billsus 1997</li> <li>• Billsus &amp; Pazzani 1999, 2000</li> <li>• Mooney et al. 1998</li> <li>• Mooney &amp; Roy 1999</li> <li>• Zhang et al. 2002</li> </ul>
Collaborativa	<p>Técnicas mais utilizadas: Vizinhança mais próxima (coseno, correlação) Teoria dos Gráficos</p> <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> <li>• Resnick et al. 1994</li> <li>• Hill et al. 1995</li> <li>• Shardanand &amp; Maes 1995</li> <li>• Breese et al. 1998</li> <li>• Nakamura &amp; Abe 1998</li> <li>• Aggarwal et al. 1999</li> <li>• Delgado &amp; Ishii 1999</li> <li>• Pennock &amp; Horwitz 1999</li> <li>• Sarwar et al. 2001</li> </ul>	<p>Técnicas mais utilizadas: Redes Bayesianas Agrupamento Redes Neurais Regressão Linear Modelos Probabilísticos</p> <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> <li>• Billsus &amp; Pazzani 1998</li> <li>• Breese et al. 1998</li> <li>• Ungar &amp; Foster 1998</li> <li>• Chien &amp; George 1999</li> <li>• Getoor &amp; Sahami 1999</li> <li>• Pennock &amp; Horwitz 1999</li> <li>• Goldberg et al. 2001</li> <li>• Kumar et al. 2001</li> <li>• Pavlov &amp; Pennock 2002</li> <li>• Shani et al. 2002</li> <li>• Yu et al. 2002, 2004</li> <li>• Hofmann 2003, 2004</li> <li>• Marlin 2003</li> <li>• Si &amp; Jin 2003</li> </ul>

Técnicas de Recomendação	Técnicas Aplicadas	
	Baseada em Heurística	Baseada em Modelo
Híbrida	Combinando componentes baseados em conteúdo e colaborativos: Combinação linear de avaliação previstas Esquemas variados de votação Incorporando um componente como parte da heurística de outro Exemplos de Pesquisas: <ul style="list-style-type: none"> <li>• Balabanovic &amp; Shoham 1997</li> <li>• Claypool et al. 1999</li> <li>• Good et al. 1999</li> <li>• Pazzani 1999</li> <li>• Billsus &amp; Pazzani 2000</li> <li>• Tran &amp; Cohen 2000</li> <li>• Melville et al. 2002</li> </ul>	Combinando componentes baseados em conteúdo e colaborativos: Incorporando um componente como parte de um modelo em outro Construindo um modelo unificado Exemplos de Pesquisas: <ul style="list-style-type: none"> <li>• Basu et al. 1998</li> <li>• Condliff et al. 1999</li> <li>• Soboroff &amp; Nicholas 1999</li> <li>• Ansari et al. 2000</li> <li>• Popescul et al. 2001</li> <li>• Schein et al. 2002</li> </ul>

Fonte: [ADO 2005a], p. 742.

A seguir são descritos alguns Sistemas de Recomendação. Uma listagem completa de Sistemas de Recomendação e comparação entre os mesmos pode ser encontrada em [MON 2003] e seu conteúdo apresenta-se compilado no Anexo A desta tese.

### 2.5.1 Projetos Acadêmicos

Abaixo são descritos alguns Sistemas de Recomendação desenvolvidos em projetos acadêmicos:

#### 2.5.1.1 Referral Web

Este projeto procurava possibilitar que através de um sistema fosse possível encontrar e visualizar redes sociais, que de acordo com o autor seriam grupos de pessoas ligadas por atividades profissionais [KAU 97].

Para criar esta rede o sistema utiliza-se de textos já existentes na Web. Uma vez que o usuário entra com seus dados é feita uma varredura na Web buscando documentos que mencionem o usuário (aplica um motor de busca e recuperação de informações). A partir dos documentos recuperados e relacionados com o usuário, são extraídos nomes de outros indivíduos citados.

No final é construída uma rede global de citações representada na forma de um grafo. Neste sistema, a formação da rede está baseada somente na relação entre usuário,

documento e citações, a importância e relevâncias dos trabalhos ou indivíduos não são trabalhadas.

### 2.5.1.2 RINGO

O RINGO é um sistema desenvolvido para recomendação personalizada de música e foi desenvolvido no *Massachusetts Institute of Technology* [SHA 95].

Este trabalho explorava similaridades entre os gostos dos diferentes usuários para recomendar itens. Isto se baseia no fato de que os gostos das pessoas apresentam tendências gerais e padrões entre os gostos das pessoas e entre os grupos de pessoas.

Neste sistema os usuários descreviam suas preferências musicais para o sistema através da avaliação de algumas músicas. Estas avaliações constituíam o perfil dos usuários. O sistema usava estes perfis para gerar as recomendações para usuários individuais.

O Ringo comparava os perfis dos usuários para determinar quais os usuários que apresentavam gostos similares (gostavam dos mesmos álbuns e/ou não gostavam dos mesmos, por exemplo).

Para o seu funcionamento, primeiramente usuários similares eram identificados, a partir desta identificação e comparação de perfis, o sistema podia prever o quanto o usuário gostaria de um álbum / artista que ainda não foi avaliado pelo mesmo.

Quando o usuário acessava pela primeira vez o Ringo era apresentado a uma lista de 125 artistas. O usuário avaliava os artistas de acordo com o quanto gostava de ouvi-los. Caso o usuário não estivesse familiarizado com o artista ou não possuísse uma forte opinião sobre o mesmo, este era solicitado a não avaliar para não ocorrerem distorções. Os usuários eram aconselhados a pontuar os artistas única e exclusivamente pelo fato de gostarem ou não dos mesmos. Para esta pontuação era aplicada uma escala de avaliação de 7 pontos, sendo o valor 1 (não gosta), 4 (indiferente) e 7 (adora).

A lista de artistas enviada para os usuários era selecionada em duas partes. Uma parte da lista era gerada pelos artistas mais pontuados (Ranqueamento), isto assegurava que um novo usuário tivesse a oportunidade de pontuar artistas que outros já tivessem pontuado, então existiria algo em comum entre os perfis das pessoas. A outra parte da lista era gerada através de uma seleção randômica a partir de um banco de dados sobre os artistas.

Uma vez que o perfil inicial do usuário tivesse sido submetido, o usuário podia pedir ao Ringo por predições, ou seja, uma pessoa podia pedir ao Ringo para:

- 1) sugerir novos artistas / álbuns que o usuário gostaria de obter ou ouvir;
- 2) listar artistas / álbuns que o usuário não gostaria;
- 3) realizar uma predição sobre um artista / álbum específico.

O retorno dado pelo Ringo aos usuários não incluía nenhuma informação em particular sobre a identidade dos outros usuários que contribuíram com as recomendações. Deve-se destacar esta última informação, em sistemas colaborativos ou de informação social, a identidade de quem avaliou deve ser mantida em segredo.

O sistema Ringo também permitia a escrita de comentários (como *feedback* explícito) sobre o produto recomendado. Os próprios usuários podiam incluir novos artistas e álbuns no banco de dados. O usuário podia receber uma mensagem sobre os novos artigos, bem como, sobre as novidades do sistema.

#### 2.5.1.3 *GroupLens*

O sistema GroupLens [KON 97] foi um projeto implantado usando Filtragem Colaborativa para recomendar leituras na Usenet News<sup>1</sup>. Desenvolvido por pesquisadores da Universidade de Minnesota, considera que o grande volume de informações produzidas nesse meio, somadas às diferenças existentes de perfis de usuários, formam um candidato ideal para uso da FC.

O Sistema coleta a avaliação dos usuários referente aos artigos que estes leram (escala de [1-5] pontos) e utilizava esta avaliação para identificar os vizinhos mais próximos com avaliações semelhantes e predizer se os usuários iriam gostar dos artigos novos baseando-se nas avaliações dos usuários vizinhos.

#### 2.5.1.4 *Fab*

O Fab constitui-se em um sistema que procura combinar sistemas de filtragem colaborativa e filtragem baseada em conteúdo (abordagem híbrida), procurando eliminar muitas das fraquezas encontradas em cada uma das abordagens. Foi desenvolvido pela Universidade de Stanford que procurava recomendar páginas Web (documentos) para usuários [BAL 97].

Objetivando a construção de uma arquitetura híbrida, são mantidos perfis de usuários baseado na análise de conteúdo, e estes perfis são comparados diretamente para determinar usuários similares para recomendação colaborativa. O sistema constitui-se em uma implementação distribuída e foi construído como parte do projeto da biblioteca digital da Universidade de Stanford.

O sistema implementava dois agentes, um responsável pela coleta dos documentos na Web (*Collection Agent*) e outro responsável pela seleção de páginas (*Selection Agent*) que seriam apresentadas aos usuários como recomendação, conforme a Figura 2.2. Após a recomendação o usuário poderia avaliar o item aplicando uma escala de avaliação de [1-7] pontos.

---

<sup>1</sup> *Usenet News* constitui-se em uma idéia inicial de grupos de discussão na Internet. Os tópicos são separados em categorias de grupos, por exemplo, “*alt.rec.humor*” para assuntos de humor e “*comp.lang.java*” com relação à linguagem de programação Java.

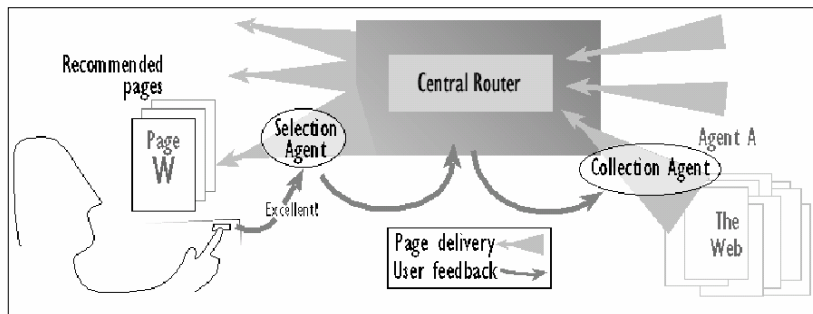


Figura 2.2: Visão geral da arquitetura do Fab [BAL 97]

Buscando criar uma arquitetura híbrida, os perfis dos usuários eram criados usando a análise do conteúdo dos documentos que seriam aplicados para descobrir quais são os usuários semelhantes. Uma vez que os usuários semelhantes fossem obtidos aplicava-se a filtragem colaborativa para recomendar. Sendo assim o usuário recebe recomendação de algo que feche com o seu perfil e com o perfil de seus usuários similares. Com esta arquitetura híbrida o sistema minimizava os problemas existentes nas duas abordagens e conseguia as seguintes vantagens:

- aplicando a abordagem de filtragem colaborativa as experiências dos usuários são levadas em consideração;

- aplicando a abordagem de filtragem baseada em conteúdo torna-se possível levar consideração itens não vistos por outros usuários e caso não existam usuários com gostos semelhantes ao usuário, mesmo assim é possível fazer recomendações.

#### 2.5.1.5 *Agente de recomendação colaborativo*

O Agente de recomendação colaborativo - CORA (Christoph apud [LUE 98]) constitui-se em um sistema distribuído assíncrono para filtragem colaborativa ativa de documentos da Web, desenvolvido na Universidade de Zurique. O CORA habilitava os usuários a recomendar URLs através de uma simples clicagem de mouse.

Os usuários eram providos com um agente pessoal de recomendação que monitorava o comportamento do usuário no uso do *browser*. O agente reconhecia a página Web que o usuário estava visualizando e apresentava a mesma em uma janela de recomendação especial. Caso o usuário decidisse recomendar a URL para um grupo de usuários pré-definido, a URL poderia ser enviada pela clicagem de um ícone específico. Assim como o agente monitorava o envio de URLs este também monitorava a chegada de URLs enviadas por outros usuários.



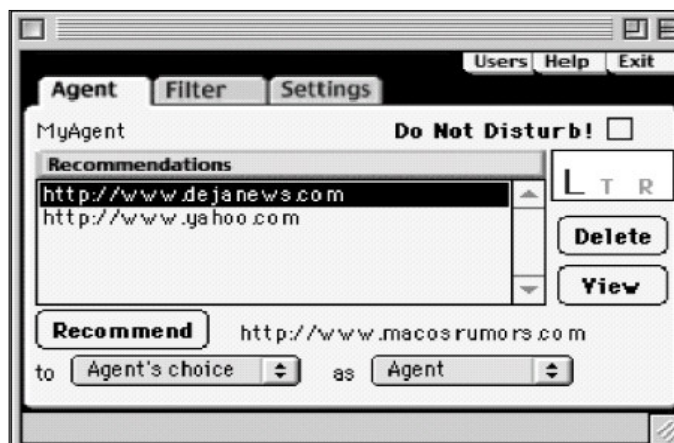


Figura 2.3: Interface da ferramenta CORA [LUE 98]

Na interface apresentada na Figura 2.3 pode-se observar três letras L,T e R que de acordo com o seu tamanho procuravam simbolizar o montante de *Links* seguidos tendo as páginas recomendadas como ponto de partida, o Tempo utilizado no acesso da página, e um fator especial de Recomendação, o qual é computado com base no uso passado das recomendações. Lueg [LUE 98] destaca que sempre que alguém está querendo buscar uma informação de valor, este possui um tempo determinado para aplicar nesta busca, mas este esforço apresenta um limite claro. Pesquisas indicam que o esforço empreendido para ter a informação desejada (a recomendação desejada) em um sistema colaborativo deve ser o menor possível.

#### 2.5.1.6 *MovieLens*

O projeto *MovieLens*<sup>2</sup> foi desenvolvido por pesquisadores da Universidade de Minnesota constituindo-se em um projeto que visa realizar recomendações de filmes. O projeto é aberto ao público, que pode se cadastrar, efetuar algumas avaliações para formar um perfil inicial para receber as recomendações [RIE 99].

<sup>2</sup> <http://movielens.umn.edu>

Os usuários podem solicitar recomendações geradas por predições, por exemplo, quando se quer saber qual seria a avaliação do usuário para um determinado filme ainda não visto pelo mesmo. A Figura 2.4 apresenta um exemplo de recomendações realizadas pelo MovieLens: nota-se que o filme “Fugitive, The” não foi ainda avaliado, e portanto uma predição está sendo realizada de quatro e meio estrelas.

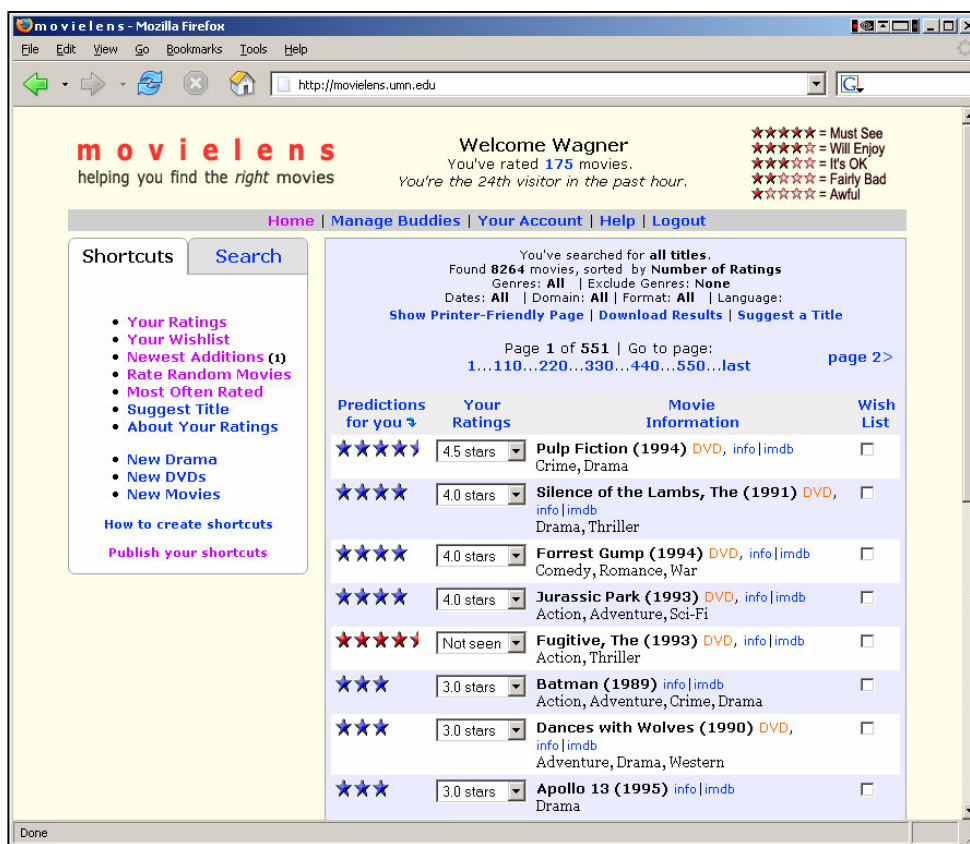


Figura 2.4: Interface de recomendações do MovieLens

Uma das grandes contribuições deste projeto para a comunidade de Sistemas de Recomendação é disponibilizar gratuitamente uma base de dados<sup>3</sup> com cerca de 100 mil avaliações dadas por seus usuários. Esta base de dados tem sido utilizada por diversos projetos no momento da validação e experimentação de suas propostas.

<sup>3</sup> <http://www.grouplens.org>

### 2.5.1.7 *Compilation Compiler Advisor*

O sistema CoCoA<sup>4</sup> (*Compilation Compiler Advisor*) [AGU 2002] constitui-se em um Sistema de Recomendação que sugere músicas clássicas com base nas compilações criadas pelo usuário no passado de uso do sistema.

Este sistema baseia-se na base do Karadar, o que significa um total de 11.000 músicas clássicas em formato MP3 sem a ocorrência de problemas de copyright, biografias de compositores, 5.000 textos de músicas clássicas e 1.100 arquivos midi.

Atualmente o sistema conta com um total de 1.400 usuários conectados todo o dia, realizando suas compilações e contribuindo como sistema de recomendação sobre músicas clássicas.

Este sistema é utilizado na tese de doutorado de Paolo Massa da Universidade de Trento na Itália. As figuras 2.5, 2.6 e 2.7 ilustram a compilação de músicas clássicas pelo usuário, o cadastramento do usuário para fins de criação de perfil e por fim uma página pessoal, onde são apresentados os *download* do usuário e compilações particulares, respectivamente.



Figura 2.5: Criação de compilações no CoCoA

<sup>4</sup> <http://cocoa.itc.it>



Figura 2.6: Cadastramento para construção de perfil no CoCoA



Figura 2.7: Página pessoal com compilações no CoCoA

## 2.5.2 Sites Comerciais

Abaixo são apresentados alguns *sites* comerciais que aplicam os conceitos referentes a Sistemas de Recomendação.

### 2.5.2.1 Amazon.com

O *site* mundialmente conhecido Amazon (www.amazon.com) é um portal de vendas on-line com diversas categorias de produtos. Este realiza vendas de livros entre outros itens. Os exemplos abaixo são focados na Seção de compras de livros [SCH 99] [SCH 2001].

- **Cientes que compraram:** A Amazon.com utiliza a técnica “Clientes que compraram” que consiste em recomendar livros freqüentemente comprados por clientes que compraram esse mesmo livro em questão. A outra lista de recomendações que o *site* utiliza é recomendar autores cujos livros são freqüentemente comprados por clientes que compraram trabalhos do autor do livro em questão.
- **Amazon.com entregas:** Os clientes selecionam a partir de listas de categorias os seus interesses pessoais para que periodicamente lhes sejam enviados por e-mail as últimas recomendações e novidades nas categorias previamente selecionadas.
- **Classificação de livros:** Essa técnica permite que os clientes façam suas classificações aos livros comprados. Eles podem classificá-los em uma escala que vai do “gostar” ao “detestar”. Aqui se tem um retorno direto do cliente sobre os produtos negociados. Depois de classificar um produto o cliente pode solicitar outras recomendações que serão trazidas conforme o gosto do cliente.
- **Comentários dos clientes:** Essa característica permite que os clientes recebam recomendações textuais sobre o produto pesquisado, baseado na opinião de pessoas que já utilizaram o produto. Na página de cada produto encontra-se a classificação dele em uma escala de avaliação de [1-5] e os comentários escritos dos clientes que compraram. Neste caso valeria ressaltar que não existe nenhuma referência à relevância das opiniões emitidas pelos clientes que avaliaram um produto.

### 2.5.2.2 eBay

Assim como no site da Amazon.com<sup>TM</sup>, o *site* de leilões *on-line* eBay.com<sup>TM</sup>, acessível no endereço www.ebay.com, possui mais de uma variação de estratégias de recomendação, também apresentado por Konstan [KON 97]. Segue a explicação de cada uma destas estratégias [SCH 99] [SCH 2001]:

- **Direito de resposta:** Permite aos compradores e vendedores avaliar o seu parceiro de negócio de acordo com o grau de satisfação da transação de compra. Isto gera uma pontuação que demonstra a confiabilidade de cada negociante (reputação). Quanto mais positiva a pontuação, mais confiável é o negociante, e quanto mais negativa a pontuação, menos confiável é o negociante.

- **Comprador pessoal:** Permite aos clientes indicar os itens que eles têm interesse em comprar, para que, em uma periodicidade definida, o *site* possa enviar os resultados da busca por esses itens.

### 2.5.2.3 Rádio Terra

O Rádio Terra (radio.terra.com.br) constitui-se em um *site* de rádios *on-line* pertencente ao portal do provedor Terra Networks Ltda. Contém rádios com mais de 100 estilos, além de serviços que possibilitam ao usuário criar Rádios Pessoais, realizar buscas no acervo, comprar cds, etc.

Abaixo estão listados alguns métodos de recomendação que o *site* utiliza:

- **Recomendação de Álbuns:** A Rádio Terra disponibiliza uma página com informações sobre cada álbum, e nesta página indica até 15 outros álbuns relacionados ao artista do álbum em questão.
- **Recomendação de Artistas:** Informações sobre o artista. Cada artista possui uma página com informações do mesmo, bem como todas as músicas próprias ou que participa. Existe também uma listagem de indicações de até 10 artistas relacionados.
- **TOP 10:** A listagem das 10 músicas mais ouvidas na Rádio Terra é indicada semanalmente aos usuários. Os usuários, podem acessar a página do álbum, do artista, ou simplesmente ouvir a música.
- **OneByOne:** O serviço *OneByOne*, possui uma lista pré-selecionada de 40 músicas, onde os usuários escolhem quais desejam ouvir. Nessa lista entram diversas músicas, lançamentos, acervo, *singles* e etc.
- **Freeway:** Mensalmente é gerada uma lista de 30 artistas que tiveram dentre suas músicas as mais escutadas do mês. Dentro desta mesma seção existe um artista que fica em destaque durante uma semana. O artista em destaque é exibido junto à listagem de todos os seus álbuns.

### 2.5.2.4 CD NOW

O site CD NOW (www.cdnow.com) é um portal de venda *on-line* de cd (*compact discs*), vídeos (VHS) e dvd (*digital video disc*). O portal possui vários serviços de recomendação listados abaixo [SCH 99] [SCH 2001]:

- **Recomendador de Álbuns:** O recomendador de álbuns chamado *Album Advisor*, do *site* CDNOW™ (www.cdnow.com) trabalha de dois modos diferentes: no modo de visualização por álbum, o *site* disponibiliza uma página com informações de cada álbum e recomenda também 10 outros álbuns que podem despertar interesse do usuário e na visualização por múltiplos artistas, os consumidores podem selecionar até três artistas e o sistema recomenda 10 álbuns relacionados aos artistas em questão.
- **My CDNOW:** A Seção My CDNOW permite aos consumidores criar sua própria “biblioteca” de músicas, baseada em álbuns e artistas de sua preferência. Os clientes indicam quais são os álbuns e artistas favoritos. Ao comprar do *site*, os artistas e álbuns são incluídos automaticamente na sua lista de adquiridos. Ao

entrar nessa lista, o cliente pode, classificar essas informações como “adquirido e preferido” ou “adquirido, porém não preferido”.

Quando os clientes solicitam recomendações, o sistema faz predições de 6 álbuns que o consumidor poderia gostar, baseado na lista de adquiridos e preferidos. A lista de sugestões contém campos para que o usuário preencha e com isso registre um feedback sobre a qualidade da recomendação. Cada álbum sugerido pode ser classificado em “Eu compro”, “Adicionar a lista de desejados” ou “Não é para mim”.

#### 2.5.2.5 *Reel.com*

O Reel.com ([www.reel.com](http://www.reel.com)) constitui-se em um *site* destinado a oferecer a seus usuários sinopses e outras informações de filmes. Existem diversas seções, como lançamentos, filmes em cartaz, acervo e etc. O portal disponibiliza uma busca de filmes e alguns serviços de recomendação descritos a seguir [SCH 99] [SCH 2001].

- **Movie Matches:** Consiste em oferecer recomendações nas páginas de cada filme. Essas recomendações estão divididas em “*close matches*” e “*creative matches*” que trazem listas de filmes sugeridos. As recomendações são baseadas em filmes similares que correspondem ao mesmo assunto, história, tipo de audiência, etc.
- **Movie Map:** Essa característica do reel.com recomenda filmes conforme características sintáticas. Os usuários entram com informações tipo gênero, formato, preço e restringem a busca entre “*sleepers*” e “*best of this genre*”. O *site* traz então uma lista de recomendações dos editores para filmes que se encaixam nas especificações solicitadas pelo usuário.

#### 2.5.2.6 *Drugstore.com*

É uma farmácia *on-line* onde os clientes indicam suas preferências quando compram um produto de determinada categoria. Cada usuário preenche um formulário informando idade e os sintomas da doença. Por meio dessas informações o sistema retorna uma lista de produtos recomendados que atendam às necessidades do cliente. A figura 2.8 apresenta a Interface do Drugstore.com.

O *site* também oferece um *test drive* para novos produtos, onde um grupo de clientes voluntários é formado para avaliar os novos produtos. Tais clientes fornecem uma pontuação e comentários sobre a avaliação realizada.

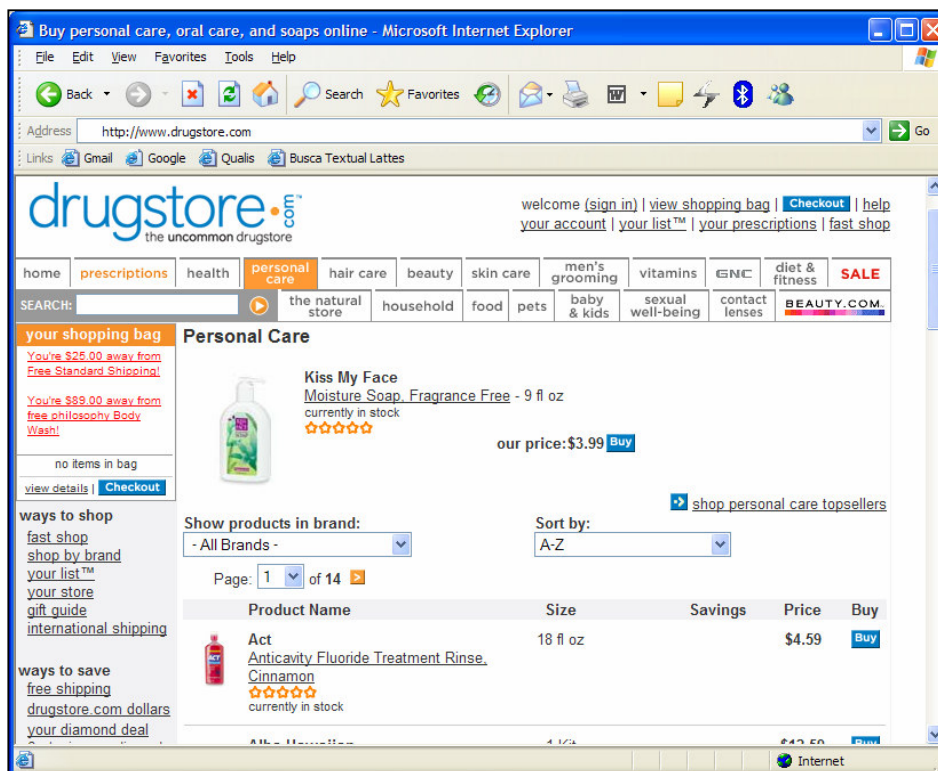


Figura 2.8: Interface do Drugstore.com

## 2.6 Limitações dos Sistemas de Recomendação

Conforme já discutido, a técnica FC resolve alguns dos principais problemas da abordagem baseada em conteúdo. Entretanto, também possui limitações [SAR 2000] [HER 2004] [MAS 2004a], e estas serão apresentadas a seguir.

### 2.6.1 Problema do primeiro avaliador

Não existe maneira de um determinado item ser recomendado para o usuário por filtragem colaborativa até que mais informações sobre o item sejam obtidas através de outro usuário. Sendo assim é necessário realizar a avaliação inicial do item. A questão está em como obter esta primeira avaliação. Quem deveria avaliar e qual estratégia deveria ser aplicada?

### 2.6.2 Similaridade

O funcionamento de um algoritmo de FC (clássico) está baseado em encontrar pessoas com gostos similares e a partir daí proceder com a busca de uma recomendação, baseando-se nas avaliações de pessoas com gostos similares. Sendo assim, caso um usuário tenha gostos que variam do normal, o Sistema de Recomendação terá dificuldades para encontrar outros usuários com gostos similares, sendo assim as recomendações podem tornar-se pobres ou inviáveis.



### 2.6.3 Novo item

Na filtragem colaborativa, as recomendações são feitas baseada no perfil do usuário e de seus vizinhos mais próximos. Itens bem avaliados pelos vizinhos são recomendados ao usuário. Entretanto, quando um novo item aparece no sistema e não faz parte do perfil de nenhum usuário, não é possível realizar recomendações ou predições sobre ele.

Devido ao cálculo da similaridade de Pearson, um item deve ser avaliado por pelo menos dois usuários, caso contrário, ele não entrará no cálculo da similaridade e será desconsiderado do perfil do usuário. A este problema é dado o nome de “Problema do Novo Item”.

No sistema Fab, é utilizada a abordagem híbrida para resolver este problema, recomendando o novo item baseado no seu conteúdo. No MovieLens existe uma opção “Novas Adições”, onde são apresentados para o usuário itens recentemente cadastrados para que possam ser avaliados.

### 2.6.4 Novo usuário

Quando um usuário é cadastrado no sistema, ainda não existem avaliações realizadas por ele, portanto, seu perfil de avaliações está vazio. Como já visto, para as recomendações serem realizadas é necessário encontrar os vizinhos mais próximos do usuário, e como ele ainda não realizou avaliações, não existirão vizinhos. A este problema é dado o nome de “Problema do Novo Usuário”. Também é referenciado na literatura como “*Cold Start User*”.

No MovieLens, é solicitado ao usuário que avalie uma certa quantidade de filmes durante a etapa de cadastramento no sistema, desta forma, quando iniciar o uso do sistema provavelmente terá vizinhos e estará apto a receber recomendações.

### 2.6.5 Escalabilidade

Um dos problemas mais críticos da FC é a escalabilidade. Quando o volume de usuários, itens e avaliações são muito grandes, os sistemas que fazem o cálculo da vizinhança (“*k-nearest-neighbor*”) de forma *on-line* (a cada cálculo de predição) podem chegar a um tempo de resposta inaceitável.

### 2.6.6 Esparsialidade

Outra importante limitação de sistemas baseados em FC ocorre na medida que o número de itens na base de dados vai aumentando. Isso reduz as chances dos usuários possuírem itens em comum (informação necessária para o funcionamento dos Sistemas de Recomendação de FC), resultando na redução do tamanho médio da vizinhança dos usuários [ADO 2005a].

O objetivo dos sistemas de filtragem colaborativa é ajudar pessoas, focando em documentos lidos ou itens adquiridos. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema existe um grande risco das pontuações tornarem-se muito esparsas. Por conseguinte, o sistema terá menos vizinhos para

realizar as predições, causando um impacto negativo na confiabilidade das recomendações.

### 2.6.7 Superespecialização

Este problema refere-se ao fato do sistema só conseguir recomendar itens muito semelhantes àqueles que o usuário já avaliou. Por exemplo, um usuário do MovieLens que tem o perfil formado basicamente de filmes de guerra, receberá recomendações, em grande parte, de outros filmes de guerra. Isto pode tornar-se um problema, uma vez que os interesses dos usuários tendem a apresentar mudanças com o passar do tempo [BAL 97] [ADO 2005a].

A superespecialização apresenta-se presente mais comumente em Sistemas de Recomendação baseados em conteúdo, porém a sua presença indesejada também é percebida em Sistemas de Recomendação de Filtragem Colaborativa. Para que um Sistema de Recomendação apresente o sucesso pretendido este necessita “cativar” o usuário, e a busca de conteúdo inesperado que venha a surpreender o usuário pode ser uma estratégia que venha a fidelizá-lo.

## 2.7 Privacidade em Sistemas de Recomendação

Para que um Sistema de Recomendação possa recomendar itens apropriadamente para seus usuários, o sistema deve constantemente coletar dados sobre os usuários. Estes dados referem-se aos comportamentos (por exemplo, navegação e compra) e possivelmente os demográficos. Esta coleta é feita muitas vezes de forma implícita, isto é, sem que o usuário necessariamente perceba que informações sobre ele estão sendo armazenadas na medida em que utiliza o sistema. Esta prática traz algumas questões relacionadas à privacidade [CAZ 2005d].

Pesquisas mostram que a maior parte dos usuários está disposta a fornecer informações suas para que possam receber ofertas personalizadas. No entanto, o censo americano mostra que 75% dos usuários daquele país se preocupam com a possível divulgação de dados que fornecem às empresas [TOR 2004]. Os usuários buscam sempre conhecer os objetivos da coleta de dados e se estes dados serão fornecidos a terceiros.

Portanto, a utilização de Sistemas de Recomendação não teria problema algum se as empresas que os empregam tivessem uma coleta de dados associada a uma política de privacidade adequada. Entretanto, não é incomum encontrar empresas, que além de coletar dados para personalizar o relacionamento com seus clientes, vendam estes dados para outras empresas, uma prática que alimenta o aumento de *spam*.

Atualmente não existem leis de privacidade *online* muito claras, sendo importante que as políticas de privacidade das empresas sejam precisas. A Figura 2.9 mostra um trecho da página contendo a política de privacidade da loja Amazon.com<sup>TM</sup>.

**Amazon.com Privacy Notice**

Last updated: April 3, 2003

Amazon.com knows that you care how information about you is used and shared, and we appreciate your trust that we will do so carefully and sensibly. This notice describes our privacy policy. **By visiting Amazon.com, you are accepting the practices described in this Privacy Notice.**

**What Personal Information About Customers Does Amazon.com Gather?**

The information we learn from customers helps us personalize and continually improve your shopping experience at Amazon.com. Here are the types of information we gather.

- **Information You Give Us:** We receive and store any information you enter on our Web site or give us in any other way. [Click here](#) to see examples of what we collect. You can choose not to provide certain information, but then you might not be able to take advantage of many of our features. We use the information that you provide for such purposes as responding to your requests, customizing future shopping for you, improving our stores, and communicating with you.
- **Automatic Information:** We receive and store certain types of information whenever you interact with us. For example, like many Web sites, we use "cookies," and we obtain certain types of information when your Web browser accesses Amazon.com. [Click here](#) to see examples of the information we receive. A number of [companies](#) offer utilities designed to help you visit Web sites anonymously. Although we will not be able to provide you with a personalized experience at Amazon.com if we cannot recognize you, we want you to be aware that these tools exist.

Figura 2.9: Política de privacidade da loja Amazon.com<sup>TM</sup>

Uma política de privacidade não é apenas uma exigência legal, mas também a única forma de garantir que uma empresa agirá com honestidade e protegerá as informações dos seus clientes. No momento em que existir um padrão e uma forma simples de identificação dos clientes, as políticas de privacidade serão mais eficientes.

Uma nova alternativa está sendo desenvolvida pela *World Wide Web Consortium (W3C)* e chamada de *Platform for Privacy Preferences (PP3)*. A PP3 é um protocolo que tem como objetivo principal implementar métodos para definir políticas de privacidade padronizadas, as quais podem ser compreendidas e processadas por computadores. Quando um cliente visita um *website* a PP3 é executada e pergunta ao cliente quais as informações ele deseja compartilhar com o *site*, criando assim a política de privacidade. Por exemplo, um cliente talvez deseje informar o seu e-mail e número do telefone com a promessa de que o *site* somente utilizará essas informações em suas iterações e com a devida permissão.

Algumas empresas estão propondo selos que garantem que a política de privacidade de um *website* é adequada e cumprida com rigor. A Figura 2.10, apresenta um destes selos, fornecido pela organização TRUSTe<sup>5</sup>.

<sup>5</sup> <http://www.truste.org/>

The screenshot shows the NetRaker website's 'Company' page. At the top, the NetRaker logo is displayed with the tagline 'Perfecting the Customer Experience'. A navigation bar includes links for Home, Solutions, Products, Research, Partners, Support, and Company. Below the navigation, a breadcrumb trail reads 'Home > Company >'. A sidebar menu on the left lists 'Company' (selected), Overview, Management, Newsroom, Customer List, and Contact Us. The main content area features the title 'NetRaker-Privacy Policy' and the statement 'NetRaker is committed to respecting your privacy.' To the left of the text is the TRUSTe seal, which includes the text 'CLICK TO REVIEWED BY TRUSTe SITE privacy statement VERIFY'. The main text explains that NetRaker is a licensee of the TRUSTe Privacy Program and details its commitment to privacy, including information collection, use, and security procedures.

Figura 2.10: Selo da empresa TRUSTe no *web site* da empresa NetRaker

## RESUMO DO CAPÍTULO:

Neste capítulo apresentou-se o significado de Sistemas de Recomendação, bem como, a diferença conceitual entre este e os sistemas de recuperação de informação. Buscou-se definir filtragem baseada em conteúdo, filtragem colaborativa e filtragem híbrida e apresentar alguns exemplos de sistemas que utilizam estas técnicas. Questões referentes à privacidade em Sistemas de Recomendação também foram apresentadas. Parte do conteúdo deste capítulo encontra-se publicado em [CAZ 2003] e [CAZ 2005d].

### 3 A RELEVÂNCIA DA OPINIÃO DO RECOMENDADOR

A Internet através de dispositivos de busca oferece acesso fácil a uma grande quantidade de conteúdo e informações. O problema está na dificuldade de um usuário encontrar informações relevantes e confiáveis para a realização de seu trabalho.

Existem muitos *sites* e geralmente as informações são disponibilizadas por outros usuários, sendo assim a questão seria qual fonte consultar. Possivelmente, a melhor fonte de consulta será oferecida pelo usuário que tiver maior conhecimento ou relevância de opinião na área em que este disponibiliza informações. Como exemplo pode-se citar o caso da enciclopédia eletrônica Wikipedia<sup>6</sup> (a enciclopédia livre da Internet), a qual se propõe a ser uma enciclopédia aberta onde os usuários podem contribuir com seus conhecimentos para criar uma grande base de informações (verbetes) para auxiliar nas consultas de outros usuários. A idéia é que qualquer verbete possa ser atualizado por qualquer pessoa que se disponha a contribuir.

O problema neste caso é que o conteúdo a ser ofertado aos usuários torna-se pouco confiável, pois os especialistas que contribuem acabam sendo iguais aos amadores/curiosos que também contribuem. Neste caso, considera-se que a relevância de opinião ou nível de conhecimento de quem auxilia no abastecimento da base de informações torna-se fundamental. Esta enciclopédia trabalha com recuperação de informação e não com recomendação, mas caso fosse implementada uma funcionalidade referente a Sistemas de Recomendação, a preocupação com a relevância da opinião deveria ser redobrada.

Sistemas de Recomendação são amplamente aplicados em *sites* de comércio eletrônico para auxiliar os clientes em suas decisões sobre compras. Conforme [SAR 2000] [HER 2004], itens podem ser recomendados baseando-se nos itens mais vendidos pelo *site*, em dados demográficos referentes ao cliente, ou analisando-se o comportamento passado de compra do cliente. As informações sobre o comportamento podem ser utilizadas na predição de futuras ofertas e possíveis opiniões de um cliente em relação ao item ainda não consumido por este.

Muitos *sites* de comércio eletrônico contêm um sistema que fornece recomendações tipo “quem comprou o produto X também comprou o produto Y ” [HER

---

<sup>6</sup> <http://pt.wikipedia.org>

2004] [CAZ 2005d]. Entretanto, uma questão muito importante permanece: qual a relevância da opinião desta pessoa que comprou X e depois comprou Y? O quanto esta pessoa conhece destes produtos para poder recomendá-los ou avaliá-los? Qual o peso da opinião do mesmo no Sistema de Recomendação? Herlocker [HER 2004] apresenta preocupação semelhante ao comentar que encontrar recomendadores confiáveis (nos quais o recomendado possa depositar confiança) passa a ser uma necessidade em questão de Sistemas de Recomendação.

Quando esta preocupação sobre a relevância de opiniões é trazida para o comércio eletrônico de equipamentos de uso mais profissional (por exemplo, máquinas fotográficas digitais) observa-se que a relevância da opinião do usuário em uma área específica é muito importante, pois este pode auxiliar um indivíduo a tomar decisão de qual equipamento deveria ou não adquirir. Deve-se observar que a relevância não apresenta relação com o quanto o usuário é bem visto ou não pela comunidade em geral.

O site “*Digital Camera Product Reviews / Previews*”<sup>7</sup> constitui-se em um exemplo deste tipo de problema sobre relevância de opinião, pois este apresenta opiniões de usuários sobre diversos modelos de câmeras fotográficas digitais. Estas opiniões possuem como propósito auxiliar o consumidor no momento de tomar a decisão sobre a compra de uma câmera ou não. O problema neste *site* é que as opiniões não são acompanhadas pela relevância da opinião de cada usuário que comentou um produto, então em qual comentário confiar?

Uma pessoa iniciando na profissão e desejando adquirir um equipamento necessita encontrar informação relevante; a questão é — como? Aplicando a abordagem proposta por esta tese, a solução surge da combinação de Sistema de Recomendação e a relevância da opinião do usuário. Nesta tese assume-se que usuários com maior peso de relevância de opinião em uma determinada área de conhecimento teriam melhor condições de avaliar um item de uma determinada área.

Cabe destacar que existem várias pesquisas que procuram trabalhar com conceitos ligados à confiança (*trust*) e reputação (*reputation*). Estas pesquisas concentram seus esforços na definição de confiança e reputação do indivíduo frente a percepção dos demais membros de uma comunidade de usuários, ou seja, a reputação ou o nível de confiança de um usuário não se refere a algo próprio do indivíduo mas sim a algo que lhe é atribuído pela comunidade.

Por exemplo, o trabalho realizado por Wang [WAN 2004] em análise de redes P2P (*peer-to-peer*), no qual o problema refere-se a disponibilizar mais recursos do sistema aos usuários com maior reputação e mais confiáveis, bem como diminuir a ação dos usuários com baixa reputação e confiança, é trabalhado frente a opinião (avaliação) dos demais membros da comunidade em relação a um usuário fonte (*source*), ou seja, a reputação e confiança depende da visão dos demais em relação a este usuário. Outro exemplo amplamente conhecido refere-se ao *site* de compras e-bay<sup>8</sup>, onde o usuário tem

---

<sup>7</sup> <http://dpreview.com/reviews/>

<sup>8</sup> <http://ebayconnections.com/>

sua reputação (representada em uma escala de valores) definida pela opinião dos demais membros do sistema que tiveram a oportunidade de negociar com um determinado indivíduo. Golbeck [GOL 2005] [GOL 2006] apresenta um trabalho focado em comunidades virtuais baseadas em confiança, sendo esta confiança explicitada pelos membros da comunidade.

Nesta tese não se está abordando a questão da reputação ou nível de confiança de uma comunidade em relação a um indivíduo ou grupo de indivíduos como foco principal, mas sim a relevância da opinião de um dado indivíduo na elaboração de uma recomendação. Procurando sintetizar a idéia apresentada nesta tese, pode-se afirmar que a relevância da opinião é algo que o indivíduo possui e não será alterada diretamente pelas opiniões da comunidade, ou seja, ao contrário da reputação e confiança de um indivíduo (exemplo de trabalhos na área são apresentados em [WAN 2004], [GOL 2005], [GOL 2006], [MAS 2004], [ZIE 2004], [MON 2002], [SAB 2002], [AVE 2005] e [MAS 2004a]) que podem ser alteradas pela opinião da comunidade sobre o indivíduo, a relevância de opinião é inerente ao nível de conhecimento de alguém na área em que atua, não sofrendo influência da opinião dos demais no seu cálculo.

Antes de prosseguir existe a necessidade de uma apresentação mais formal da diferença conceitual entre os termos Autoridade, Confiança, Relevância e Reputação [FER 96] [LAR 2004]:

- **Autoridade:** *s.f. 1...2..3..4.Mando; 5. Crédito; 6. Competência; 7...8. Pessoa de grande competência num assunto.*

- **Confiança:** *s.f. 1. Segurança íntima de procedimento; 2. Crédito fé; 3. Boa fama. Dar confiança a. Depositar confiança em. 1....2.Ter em bom conceito.*

- **Relevância:** *s.f. 1. Qualidade de relevante. 2. ...3. Grande valor, conveniência ou interesse; importância.*

- **Reputação:** *s.f. 1... 2.Fama, celebridade e renome.3. Conceito em que uma pessoa é tida; bom ou mau nome.*

O foco desta tese está em trabalhar a Relevância ou Autoridade da opinião do recomendador através de uma análise automatizada de dados.

### 3.1 O modelo Mo-DROP

Ao considerar-se um *site* de comércio eletrônico que procura recomendar para um usuário um DVD da série de sucesso como “*Sex and the city*”, talvez a relevância da opinião de quem avaliou (pontuou) este DVD não seja muito importante.

Supondo que um sistema de recomendação verifique que este DVD (“*Sex and the city*”) seja de interesse para um usuário, ao recomendar o mesmo, talvez o usuário considere esta recomendação e decida comprar o item. Não se preocupando com quem avaliou ou opinou sobre o produto. Em outros domínios a preocupação é diferente.

Procurando melhor explicar a fundamentação da abordagem adotada nesta tese pode-se imaginar pessoas em posições específicas em diferentes carreiras, por exemplo, advogados, músicos e professores. Pessoas em posições de destaque em carreiras profissionais específicas, usualmente têm suas opiniões muito destacadas no seu meio profissional, ou seja, a comunidade da qual este profissional faz parte, considera a opinião destas pessoas muito relevantes (devido a formação e atuação do profissional na área). A comunidade inclusive pode atribuir um nível de confiança e reputação ao profissional, o que extrapolaria a idéia centrada em somente relevância de opinião.

Pode-se imaginar o cenário no qual existe um maestro experiente e de renome regendo uma orquestra. Esta pessoa está em uma posição de destaque, e não existe dúvida sobre sua competência devido a sua capacidade e conhecimento na área. Sendo assim, todos de sua área de atuação concordam que a opinião do mesmo em assuntos relacionados a esta área é mais relevante do que a opinião de quem não é da mesma área ou não tem tanta experiência. Caso este maestro recomende qualquer conteúdo da área para alguém, a pessoa que recebe a recomendação deverá dar mais importância a esta recomendação recebida.

Outro exemplo pode ser observado na área de pesquisa. Considere um sistema que pode prever o interesse de um usuário em um artigo baseado na opinião de outros usuários sobre o mesmo artigo (Sistema de Recomendação aplicando a técnica de filtragem colaborativa). Supondo que os outros usuários não possuam muita experiência ou conhecimento dos artigos desta área de conhecimento, talvez suas opiniões não sejam tão relevantes para um usuário alvo, mesmo que estes usuários avaliem os itens da mesma maneira que este usuário alvo (*like-minded*). Talvez a recomendação baseada em usuários com maior relevância de opinião possa ser mais acertada e muito mais interessante. Em termos gerais, caso o usuário alvo que receba uma recomendação do Sistema de Recomendação tenha conhecimento da relevância da opinião dos recomendadores este poderia melhor avaliar as recomendações. Desta forma o usuário poderia sentir-se mais estimulado em acessar as recomendações, pois poderia confiar mais na qualidade destas [HER 2004].

Procurando representar a relevância da opinião dos usuários que colaboram na elaboração da recomendação para usuários alvo em um Sistema de Recomendação foi proposto o Modelo para Determinação da Relevância da Opinião do usuário (MO-DROP). Neste modelo é proposta uma métrica denominada Ranqueamento do Recomendador (RR) [CAZ 2005b].

A idéia da disponibilização deste valor de ranqueamento surge da necessidade de indicar ao usuário alvo de um Sistema de Recomendação Colaborativo qual a relevância da opinião de quem participou do processo de seleção e recomendação de um item. Procurando criar uma relação para melhor entendimento pode-se citar o conceito do *pagerank* [PAG 98] apresentado no *site* <http://www.google.com> que procura trabalhar a relevância de um determinado link que é recuperado em uma consulta realizada pelo usuário.



### 3.1.1 A métrica para o Ranqueamento do Recomendador

Os critérios para medir a importância da opinião do usuário certamente são dependentes do domínio que está sendo considerado. Mesmo assim é possível conceber um modelo geral baseado em atributos relevantes, cada um normalizado e com importância (peso) distinta [CAZ 2005b].

Nesta tese propõe-se uma métrica que possibilita medir o grau da relevância da opinião de um usuário, ou seja, permite verificar o quanto esta opinião teria importância em um sistema de recomendação. A métrica proposta constitui-se em uma escala de valores que pode ser parametrizada, visando representar numericamente esta relevância de opinião.

Estes atributos devem ser selecionados de acordo com o domínio específico da aplicação para representar a relevância da opinião do usuário. Por exemplo, em um *site* de comércio eletrônico para venda de livros (*Bookstore*) alguns atributos poderiam ser destacados para calcular o Ranqueamento do Recomendador (RR) para um usuário, por exemplo: formação, número de publicações lidas referentes à área do livro e número de *feedbacks* fornecido pelo usuário por ano, sobre publicações nesta área de conhecimento.

### 3.1.2 Equações

Visando viabilizar o cálculo do RR de um usuário, algumas equações são propostas e serão descritas em detalhes a seguir. Primeiramente, supondo que uma série de atributos do domínio que está sendo trabalhado tenha sido selecionada e que estes atributos apresentam diferentes escalas de valores, a Equação (6) é utilizada no modelo para normalizar os valores dos atributos.

$$an = \text{MinMax}(a) = \frac{a - \text{min}A}{\text{max}A - \text{min}A} \times (\text{newmax}A - \text{newmin}A) + \text{newmin}A \quad (6)$$

A Equação (6) realiza uma transformação linear nos dados originais. Assumindo que *minA* e *maxA* são os valores mínimos e máximos de um atributo. A normalização *Min-Max* procura mapear um valor *a* de *A* para um valor *an* na escala [*newminA*, *newmaxA*]. Conforme o autor Han [HAN 2001], esta normalização é essencial para manter a consistência de escalas entre todos os atributos.

Como exemplificação, pode-se supor que dois atributos tenham sido selecionados para definir a relevância da opinião do usuário, o atributo A1 apresenta valores entre [50-100], e o A2 apresenta valores entre [5-10]. Supondo que a escala de representação desejada esteja entre [0-10]:

$$A1 = \text{MinMax}(73) = ((73-50)/(100-50)) \times (10-0) = 4,6$$

$$A2 = \text{MinMax}(7) = ((7-5)/(10-5)) \times (10-0) = 4,0$$

Observa-se que desta forma dois atributos com escalas distintas foram convertidos para uma escala desejada entre [0-10].

$$RR_{area} = \frac{\sum_{i=1}^n an_i * p_i}{\sum_{i=1}^n p_i} \quad (7)$$

A Equação (7) é utilizada para calcular o Ranqueamento do Recomendador (RR), e representa a média aritmética ponderada dos atributos ( $an_1, an_2, \dots, an_n$ ) selecionados para o cálculo (atributos dependentes do domínio). Esta equação converte toda a informação em um único valor quantitativo dentro de uma escala específica. Representa a relevância geral da opinião do usuário em uma determinada área de conhecimento do domínio (determinada área em que o Sistema de Recomendação pode recomendar itens). Por exemplo, supondo unicamente os dois atributos A1 e A2 utilizados no exemplo anterior com pesos 4,0 e 6,0 respectivamente, o valor de Ranqueamento do Recomendador (em uma escala de [0-10]) para a área específica do Sistema de Recomendação será:

$$RR = (4,6 \times 4 + 4,0 \times 6) / (4 + 6) = 4,24$$

Observa-se que dependendo da regra estabelecida para interpretação dos valores, poder-se-ia considerar que uma relevância de opinião de 4,24 seria referente a uma baixa relevância, dentro de uma escala de [0-10]. Este Ranqueamento do Recomendador (RR) também é designado nesta tese por  $RR_{area}$ , nos casos em que o domínio apresentar segmentação em diferentes áreas de interesse.

$$RR_{total} = \frac{\sum_{i=1}^m RR_{area_i}}{m} \quad (8)$$

A equação (8) calcula o que se denominou de Ranqueamento Total do Recomendador ( $RR_{total}$ ). Este valor constitui-se em uma média aritmética entre todos os  $RR_{area}$  do usuário em questão e que sejam superiores a um determinado limite estabelecido (valor a partir do qual assume-se que o usuário apresenta relevância de opinião, exemplo 5,0 em uma escala de [0-10]).

Esta equação é aplicável em situações de domínios segmentados, ou seja, onde o usuário pode ter um peso de opinião diferente por área de interesse dentro de um Sistema de Recomendação. Esta métrica representa o quanto a opinião do usuário é relevante para todas as áreas que ele apresenta interesse.

Por exemplo, um domínio amplo pode ser subdividido em algumas áreas de interesse, onde um usuário pode apresentar diferente nível de conhecimento e experiência em cada uma. Por exemplo, se forem consideradas duas áreas específicas de interesse de um usuário: “redes neurais” e “sistemas multiagentes”, o  $RR_{\text{area}}$  para cada área específica de interesse será calculado usando a Equação 2, e um  $RR_{\text{total}}$ , aplicando a Equação (8). Supondo que o usuário apresente  $RR_{\text{area}} = 6,0$  para a área de “redes neurais” e  $RR_{\text{area}} = 7,0$  para a área de “sistemas multiagentes”, o seu  $RR_{\text{total}}$  será:

$$RR_{\text{total}} = (6,0 + 7,0) / 2 = 6,5$$

### 3.1.3 Exemplo de Aplicação

Objetivando apresentar uma aplicação prática do modelo proposto (Mo-DROP) este foi aplicado em um caso hipotético de uma livraria (livraria disponível na Internet e na qual os usuários apresentam seus perfis cadastrados). O modelo para representar a relevância da opinião do usuário pode ser aplicado a qualquer escala de valores (sendo parametrizável). Neste exemplo o modelo foi aplicado em uma escala de [0-10], onde 0 significa a menor relevância de opinião possível e 10 representa a mais alta relevância de opinião.

#### 3.1.3.1 Domínio e Atributos

O domínio do exemplo que está sendo descrito refere-se a uma Livraria hipotética disponibilizada na Internet, onde é possível encontrar livros sobre ciência da computação em diversas áreas de interesse dentro deste domínio (exemplo, Área 1= “Aprendizagem de Máquina”, Área 2= “Mineração de Dados” e Área 3= “Modelagem de Perfil de Usuários”). Os atributos selecionados para calcular o Ranqueamento do Recomendador (RR) visando representar a relevância da opinião no exemplo são descritos na Tabela 3.1 e Tabela 3.2.

Tabela 3.1: Atributos qualitativos dos usuários

Atributos / Acrônimo	Valores
Formação / FO	Para cada valor de FO foi atribuído um peso específico em uma escala de [0-10]: - Não graduado= 0; - Graduando = 1; - Graduado = 2; - Especialista = 3; - Mestrando = 4; - Mestre = 5; - Doutorando = 6; - Doutor =9; - Pós-doutor =10.

Atributos / Acrônimo	Valores
Áreas de Interesse /AI	Esta lista constitui-se em uma listagem das áreas de interesse cadastradas na livraria.

Tabela 3.2: Atributos quantitativos dos usuários

Atributos / Acrônimo	Valores
Feedback pontuação /FP	Este valor é medido em unidades FP.
Feedback explícito (texto)/ FE	Este valor é medido em unidades FE.
Livros acessados / LA	Este valor é medido em unidades LA.
Livros lidos/ LL	Este valor é medido em unidades LL.

Todas as métricas quantitativas (FP, FE, LA e LL) apresentam valores disponíveis na base de dados sintética da Livraria em questão, e estes foram normalizados para uma escala específica de valores. A Equação (6) foi aplicada no processo de normalização dos valores, e o valor máximo definido para cada atributo foi obtido através de uma base de dados sintética gerada randomicamente e unicamente para esta exemplificação e que se compõe de 50 registros de clientes hipotéticos.

O valor mínimo definido para cada atributo foi igual a 0 e máximo igual a 10. Todos os atributos analisados apresentavam diferentes escalas de valores (por exemplo, de acordo com a base de dados sintética, o número máximo de livros acessados em um ano foi igual a 100, mas o número máximo de livros lidos foi igual a 40).

Quando foi aplicada a Equação (7) para calcular o Ranqueamento do Recomendador, os pesos empíricos atribuídos aos atributos foram: FO = 2,0; FP= 3,0; FE= 2,0; LL= 2,0; e LA= 1,0 (atributos descritos anteriormente nas tabelas 3.1 e 3.2).

### 3.2 Aplicando o Mo-DROP

A tabela 3.3 apresenta as métricas (atributos) e valores respectivos extraídos do banco de dados referentes a dez usuários hipotéticos, entretanto, os atributos não se encontram normalizados.

Neste exemplo, objetivando ilustrar a aplicação do modelo proposto foram disponibilizados apenas 10 usuários. Na linha 1 da tabela 3.3, apresenta-se um usuário que é doutor; apresentou alguns *feedbacks* sobre livros em uma área específica;

apresentou alguns *feedbacks* textuais; leu alguns livros e somente acessou alguns livros no site de comércio eletrônico da livraria.

A pergunta é — Qual a relevância da opinião deste usuário para a área de interesse do mesmo dentro de um sistema de recomendação que aplique o modelo MO-DROP?

Tabela 3.3: Valores não normalizados dos atributos

Usuário	Atributos				
	FO	FP	FE	LL	LA
U1	9	11	1	19	11
U2	9	21	8	22	32
U3	9	34	5	36	65
U4	10	13	8	25	34
U5	6	24	0	16	4
U6	9	3	0	11	31
U7	4	21	1	30	3
U8	9	34	3	34	34
U9	10	11	12	40	100
U10	9	14	2	24	63

A tabela 3.4 apresenta todos os atributos normalizados após a aplicação da Equação (6). Pode-se observar que todos os atributos foram colocados na escala desejada [0-10]. O atributo FO permaneceu com os mesmos valores, uma vez que já se encontrava na faixa desejada.

Tabela 3.4: Atributos normalizados para o cálculo do RR

Usuário	Atributos				
	FO	FP	FE	LL	LA
U1	9	3	1	3	1
U2	9	6	7	4	3
U3	9	10	4	9	6
U4	10	3	7	5	3
U5	6	7	0	2	0
U6	9	0	0	0	3
U7	4	6	1	7	0
U8	9	10	3	8	3
U9	10	3	10	10	10
U10	9	4	2	4	6

Finalmente, foi calculado o  $RR_{\text{area}}$  aplicando a Equação (7) para todos os atributos selecionados e usando os respectivos pesos. A tabela 3.5 apresenta o valor calculado para cada usuário.

A interpretação final dos valores obtidos mostra que a opinião dos usuários U3 e U9 são as mais relevantes entre os usuários interessados nesta área de interesse do Sistema de Recomendação da Livraria. Neste exemplo, foi apresentado o cálculo do  $RR_{area}$  por usuário em uma área específica de interesse (exemplo, “mineração de dados”), mas o mesmo cálculo pode ser realizado para cada área de interesse do usuário.

Os resultados dos experimentos sugerem que o modelo é flexível e pode ser facilmente aplicado para domínios distintos, uma vez que os atributos e seus pesos são parametrizáveis.

Tabela 3.5: Resultados do Ranqueamento do Recomendador

Usuário	$RR_{area}$
U1	3
U2	6
U3	8
U4	6
U5	4
U6	2
U7	4
U8	7
U9	8
U10	5

O benefício esperado pela aplicação deste modelo é que o mesmo possa auxiliar na representação da relevância da opinião do usuário que participa da geração de uma recomendação para um usuário alvo, e desta forma espera-se estar oferecendo uma informação que venha a facilitar a escolha do usuário sobre o “consumo” do item recomendado. Um exemplo completo da aplicação deste modelo na modelagem da relevância de opiniões de pesquisadores utilizando como fonte os currículos dos mesmos é apresentado na Subseção 6.1.2 desta tese.

### 3.3 Utilizando a métrica em recomendação colaborativa

Uma vez proposta e definida a métrica de Ranqueamento do Recomendador houve a necessidade da inclusão desta métrica no processo de recomendação por filtragem colaborativa, uma vez que se desejava dar mais valor às avaliações de pesquisadores com maior relevância de opinião na área do item a ser recomendado. Desta forma a equação (5) apresentada na Subseção 2.4.2, que se constitui na equação clássica para predição de item a ser recomendado foi adaptada para a utilização da métrica proposta neste capítulo.

A equação (9) apresenta o resultado da adaptação da equação clássica e constitui-se na equação aplicada para a recomendação colaborativa nesta tese. Nesta equação foi incluída a relevância da opinião do recomendador ou recomendadores resultantes da aplicação do Coeficiente de Pearson (equação (4)). Para cada recomendador selecionado como *like-minded* é selecionado o seu RR para a área a qual o item a ser recomendado pertence. Para que o valor do RR não provocasse distorção do valor final de avaliação, este valor que está na exemplificação em uma escala entre [0-10] (lembrando que o valor é parametrizável) é convertido para um valor entre [0-1] visando não causar distorção na escala de avaliação do item que é entre [1-5]. Sendo assim o pesquisador  $X$  (que apresenta gostos semelhantes ao pesquisador  $Z$ , alvo da recomendação) está contribuindo com suas avaliações (*ratings*) do item  $I$  para que o Sistema de Recomendação decida se este item seria de interesse do pesquisador  $Z$ . Este pesquisador  $X$  apresenta um  $RR_{area}$  para a área do item a ser recomendado igual a 7, na equação (9) este valor de  $RR_{u,area(i)}$  (relevância da opinião do recomendador  $u$  na área em que o artigo  $i$  se encontra classificado) será igual a 0,7 na equação.

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u} * RR_{u,area(i)}}{\sum_{u=1}^n |w_{a,u}|} \quad (9)$$

## RESUMO DO CAPÍTULO:

Neste capítulo foi apresentado um modelo que permite modelar a relevância da opinião do usuário. Este modelo denominado de Modelo para Determinação da Relevância da Opinião (Mo-DROP) do usuário propõe-se a ser de aplicação genérica, ou seja, independente de área de conhecimento. A importância deste modelo nesta tese refere-se ao entendimento assumido de que em Sistemas de Recomendação colaborativos faz-se necessário saber a relevância da opinião dos usuários que contribuíram com a elaboração da recomendação. Este modelo foi publicado em [CAZ 2005b] [MIL 2005], e a nova equação proposta foi publicada em [SIL 2005].

## 4 PREDIZENDO INTERESSES DOS USUÁRIOS

Os Sistemas de Recomendação (principalmente os colaborativos) apresentam algumas limitações (Seção 2.6), como por exemplo, o problema do novo usuário [KON 97] e a superespecialização do usuário [BAL 97]. O primeiro problema ou limitação ocorre quando um novo usuário começa a utilizar o Sistema de Recomendação e o sistema não dispõe de muita informação a respeito deste usuário. O desafio está em como recomendar itens sem possuir informação suficiente sobre este usuário. Quais itens deveriam ser recomendados? Em que áreas ou categorias este usuário estaria interessado?

Outra limitação bastante citada na literatura refere-se ao fato dos Sistemas de Recomendação fornecerem recomendações somente baseando-se no perfil do usuário. Desta forma, os usuários não possuem a chance de explorar novos itens, que não sejam similares aos itens incluídos em seu perfil, isto é, o sistema não recomenda itens fora do domínio de atuação do usuário. Esta situação pode criar um descontentamento no usuário, uma vez que os interesses dos usuários tendem a se modificar e se aperfeiçoar com o tempo [BAL 97]. Este problema torna-se mais perceptível em Sistemas de Recomendação baseado em conteúdo, onde a busca de novo item baseia-se na similaridade de conteúdos entre itens já consumidos com satisfação e novos itens.

Por exemplo, poder-se-ia imaginar um usuário pesquisador que estaria interessado em artigos relacionados com o assunto “mineração de dados”, porém este assunto poderia estar relacionado com outras áreas, como “aprendizagem de máquina” e este pesquisador pode não estar muito atento a esta relação. Alguns pesquisadores com grande experiência e com relevância de opinião poderiam indicar este relacionamento entre as áreas de “mineração de dados” e “aprendizagem de máquina”.

A questão maior é a de que um pesquisador pode estar interessado em uma área de conhecimento, e, por uma série de motivos, pode desconhecer a ligação desta área com outras áreas de conhecimento complementares. Baseando-se na experiência de outros pesquisadores, com relevância de opinião, este conhecimento sobre as ligações entre áreas de conhecimento, poderiam emergir e poderiam ser oferecidas para este pesquisador que desconhece estas ligações, de forma a poupar trabalho ao mesmo.

Para extrair e prever estes relacionamentos entre áreas de interesse e prever quais usuários poderiam ser favorecidos por este conhecimento, foi proposta nesta tese



de doutorado um *framework* que aplica informações provenientes do perfil dos usuários de um Sistema de Recomendação, para propiciar a geração de predição.

Uma forma de realizar esta extração constitui-se na aplicação de técnicas e ferramentas de mineração de dados sobre as coleções de dados. Uma das definições de Descoberta de Conhecimento em Banco de Dados - DCBD (do inglês *Knowledge Discovery in Databases* - KDD) encontrada na bibliografia é a que segue:

*“DCBD constitui-se em um processo de extração não trivial de informações potencialmente úteis, as quais não são previamente conhecidas e encontram-se implícitas em grandes coleções de dados”* [FAY 96].

O processo de DCBD foi proposto por volta de 1989, e apresenta o conjunto de etapas necessárias para produzir conhecimentos a partir de dados de um banco de dados, e a mineração de dados constitui-se na etapa da transformação dos dados em informação.

Quando se trabalha com Sistemas de Recomendação na Internet, a mineração de dados torna-se um recurso indispensável para a descoberta de relações entre itens, entre usuários e entre itens e usuários [HER 2004].

A utilização do DCBD está relacionada com diversos domínios de aplicações, como por exemplo: análises corporativas, medicina, biologia, mercado, etc [APT 2002] [SMY 2002]. Diversas tarefas de DCBD podem ser aplicadas dependendo do domínio da aplicação e do interesse do usuário final. De modo geral, cada função de DCBD extrai um tipo diferente de conhecimento do banco de dados [BER 2000] [TWO 99].

Nesta tese aplicou-se a tarefa de mineração de dados denominada de Regras de Associação [AGR 93] para derivar regras a partir da base de dados de perfil de usuários. A aplicação desta tarefa de mineração e não outra como Classificação ou Agrupamento [HAN 2001], deve-se à necessidade de relacionar itens de uma forma de fácil interpretação e que permita identificar a relação de causalidade entre estes itens [TAN 2000].

Normalmente, a mineração de regras de associação apresenta dois problemas principais conforme [TAN 2000]:

1. Um número muito grande de regras é gerado dificultando, desta forma, a seleção das mesmas;
2. Não são todas as regras geradas que se apresentam como interessantes na etapa de pós-processamento.

Se o usuário está procurando por informações ou se o mesmo está explorando uma nova área, seria uma boa estratégia perguntar a alguns colegas ou outros especialistas por suas recomendações (*word of mouth* [SHA 95]). Caso o usuário confie nestes colegas e se os mesmos apresentarem um bom conhecimento (existindo uma forma de avaliar a confiança ou conhecimento) sobre o problema que está sendo

estudado então o usuário teria uma boa chance de obter informação relevante e de grande qualidade. Nesta tese a abordagem diferencia-se dos demais Sistemas de Recomendação, pois se leva em consideração a relevância da opinião do usuário na seleção das regras que poderiam ser mais interessantes. Para especificar se uma regra é aceitável ou não, se aplica a métrica denominada Ranqueamento do Recomendador (descrita em detalhes no Capítulo 3).

No *framework* proposto neste capítulo são aplicadas regras de associação e agentes de software para permitir a descoberta de novas regras para indicar novos relacionamentos entre áreas de interesse para os usuários. Na figura 4.1 é apresentado um esquema macro da proposta para a descoberta de novas relações entre áreas de interesse.

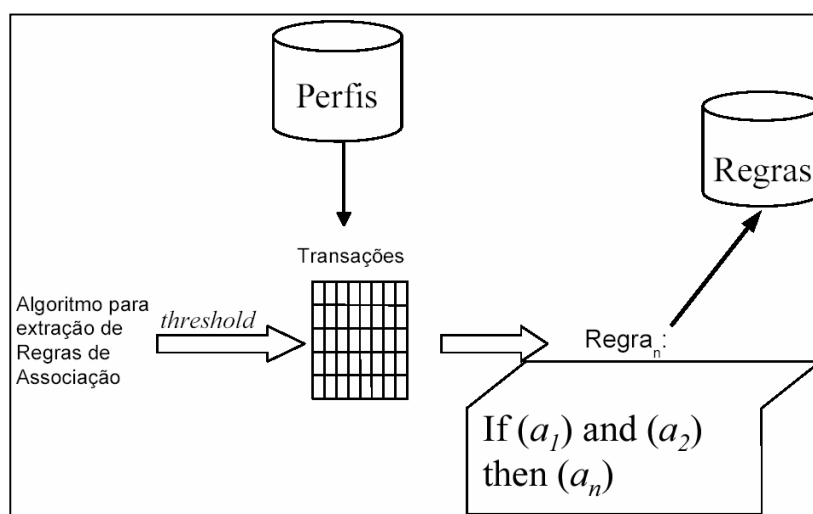


Figura 4.1: Visão macro da descoberta de relações entre áreas

Na figura 4.1 pode-se observar o pré-processamento de dados provenientes de transações da base de dados de perfis dos usuários (usuários x áreas de interesse) e a submissão da matriz resultante a um algoritmo para extração de regras de associação. Na sequência ocorre a seleção de regras para abastecerem a base de dados de Regras a serem aplicadas nos respectivos usuários. Maiores detalhes na subseção 4.2.1.

#### 4.1 Regras de Associação

O problema referente à mineração de regras de associação foi introduzido por Agrawal [AGR 93].

Neste tipo de tarefa cada tupla da base de dados consiste em um conjunto de atributos denominados itens. Cada tupla corresponde a uma transação, e um item pode

assumir um valor verdadeiro ou falso (binário). Uma regra de associação constitui-se em um relacionamento  $X \Rightarrow Y$ , onde  $X$  e  $Y$  são conjuntos de itens e a intersecção entre os mesmos constitui-se em um conjunto vazio.

Cada regra de associação é associada a um fator de suporte (FSup), e a um fator de confiança (FConf). O FSup constitui-se na razão do número de tuplas que satisfazem  $X$  e  $Y$ , sobre o número total de tuplas ( $\text{FSup} = |X \cup Y|/N$ ). O FConf constitui-se na razão do número de tuplas que satisfazem  $X$  e  $Y$  sobre o número de tuplas que satisfazem  $X$  ( $\text{FConf} = |X \cup Y|/|X|$ ) [CAB 97] [AGR 93].

A tarefa de extração de regras a partir de um banco de dados consiste em encontrar todas as regras possíveis com FSup e FConf maiores ou iguais a um valor de FSup e FConf especificado pelo usuário.

### Definição Formal

Formalmente, regras de associação são definidas da seguinte maneira [REZ 2005], seja  $D$  uma base de dados composta por um conjunto de itens  $A = \{a_1, a_2, \dots, a_m\}$  ordenados e por um conjunto de transações  $T = \{t_1, t_2, \dots, t_m\}$ , na qual cada transação  $t_i \in T$  é composta por um conjunto de itens tal que  $t_i \subseteq A$ .

A regra de associação é uma implicação na forma  $LHS \Rightarrow RHS$ , em que  $LHS \subset A$ ,  $RHS \subset A$  e  $LHS \cap RHS = \emptyset$ . A regra  $LHS \Rightarrow RHS$  ocorre no conjunto de transações  $T$  com confiança  $\text{conf}$  se em  $\text{conf}\%$  das transações de  $T$  em que ocorre  $LHS$  ocorre também  $RHS$ . A regra  $LHS \Rightarrow RHS$  ocorre no conjunto de transações  $T$  com suporte  $\text{sup}$  se em  $\text{sup}\%$  das transações em  $D$  ocorre  $LHS \cup RHS$ .

O problema de descobrir todas as regras de associação a partir de um conjunto de transações  $T$  consiste na geração de regras que tenham um suporte e uma confiança superiores a um dado limite. Estas regras são denominadas pelos autores de regras fortes. Neste trabalho aplicam-se métricas de suporte e de confiança, devido à primeira refletir a importância da regra e a segunda a confiabilidade de predição realizada utilizando a mesma.

Por exemplo, uma regra denominada R1 apresenta o seguinte conhecimento descoberto, R1: A1= “Mineração de dados”  $\Rightarrow$  A2= “Aprendizagem de máquina”, sugere que de acordo com as transações em nossa base de dados, a maioria dos pesquisadores que estão interessados em artigos relacionados com a área de interesse A1=“ Mineração de dados”, também apresentam a tendência de estarem interessados em artigos relacionados com a área de interesse A2= “Aprendizagem de máquina”. Se esta regra for considerada útil, conforme critérios previamente especificados, então poderá ser aplicada como conhecimento novo para prever o interesse dos usuários.

## 4.2 Um framework para prever o interesse dos usuários (F-UIP)

Nesta seção é descrito em detalhes o framework (F-UIP) proposto, e implementado para permitir a descoberta de conhecimento novo, e a utilização do mesmo para prever interesses de usuários. O F-UIP inclui a relevância da opinião dos usuários na seleção final das regras de associação, tendo sido publicado na sua íntegra em [CAZ 2005c]. A proposta da criação deste *framework* surgiu da necessidade de criar novas recomendações que viessem a prover os usuários alvos com recomendações inesperadas, que procurem propiciar a identificação da evolução dos interesses dos usuários e aprimoramento do perfil dos usuários do sistema de recomendação. A Figura 4.2 apresenta uma visão macro deste *framework* (F-UIP).

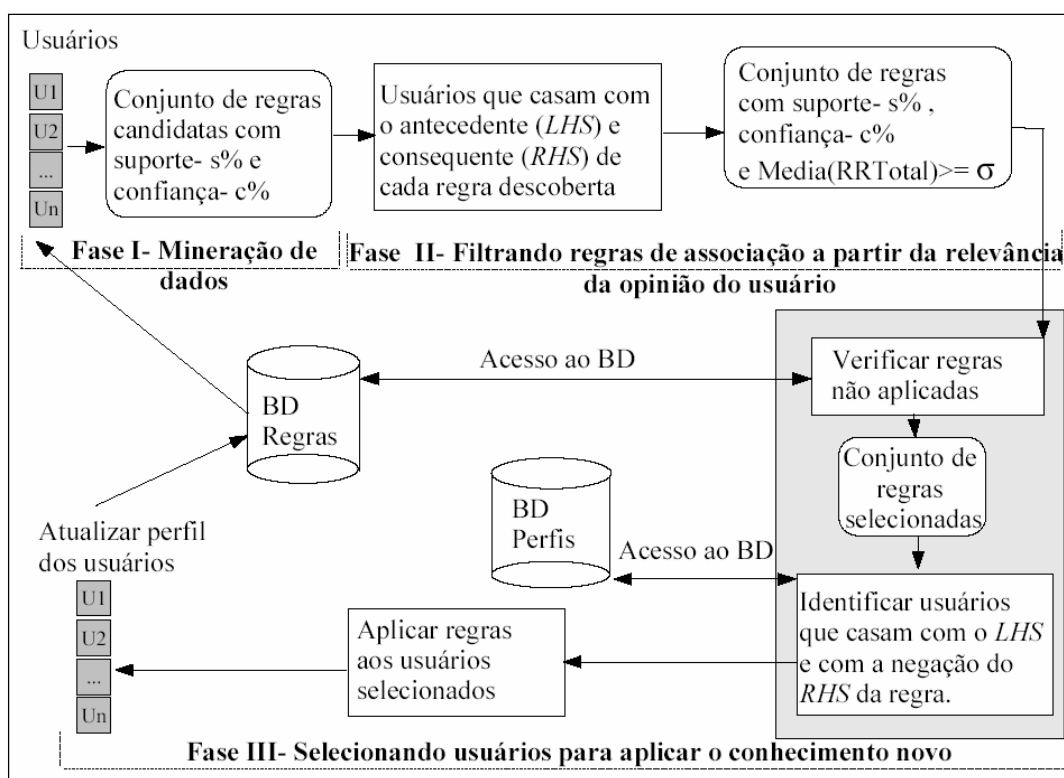


Figura 4.2: *Framework* para prever o interesse dos usuários

O F-UIP encontra-se dividido em três fases descritas a seguir.

### 4.2.1 Fase I -- Mineração de dados

Nesta fase, o *framework* utiliza informações provenientes da base de dados de perfil de usuários do Sistema de Recomendação. Nesta base de dados é possível encontrar

algumas informações ditas de caráter demográfico [PAZ 99] e informações sobre cada área de interesse do usuário.

Pode-se citar, por exemplo, um usuário que está interessado nas áreas de “aprendizagem de máquina” e “mineração de dados”, o valor da métrica de relevância da opinião do usuário para cada uma destas áreas de interesse (Ranqueamento do Recomendador por área conforme explicado no Capítulo 3 —  $RR_{\text{área}}$ ), e a média de todos os  $RR_{\text{área}}$  deste usuário com valor superior a um limite estabelecido (Ranqueamento Total do Recomendador—  $RR_{\text{total}}$ ).

Com o objetivo de descobrir conhecimento novo entre as diversas áreas de interesse (conhecimento este representado por regras), utiliza-se de uma base de dados de perfis de usuários, dados sobre áreas de interesse e o valor do Ranqueamento do Recomendador ( $RR_{\text{área}}$ ). Primeiramente, estes dados são extraídos e transformados em uma matriz bidimensional de áreas de interesse por usuários. Nesta matriz, as colunas representam a totalidade de áreas de interesse disponíveis em um Sistema de Recomendação, conforme a tabela 4.1. Nesta matriz cada transação representa um usuário registrado no sistema (linhas da matriz), e as colunas são as áreas de interesse disponíveis no sistema. Na matriz as transações são representadas usando valores binários (o valor 1 significa que o usuário possui interesse na respectiva área; 0 significa ausência de interesse na respectiva área). Na tabela 4.1, por exemplo, o usuário U6 está interessado em receber recomendações referentes a artigos científicos das áreas de interesse A2, A3 e A5, mas apresenta ausência de interesse nas áreas de interesse A1 e A4.

Tabela 4.1: Usuário e áreas de interesse

Usuários	Áreas de Interesse				
	A1	A2	A3	A4	A5
U1	1	1	1	0	0
U2	0	1	0	0	1
U3	1	0	1	1	1
U4	1	1	1	1	0
U5	1	0	1	1	1
U6	0	1	1	0	1
U7	0	0	0	1	1
U8	0	1	1	1	0
U9	1	1	0	1	1
U10	1	0	1	1	1

Após esta matriz ter sido completada com as informações provenientes da base de dados, o conteúdo da mesma é submetido a um algoritmo de Regras de Associação. Na fase I, para decidir quais são as regras que irão participar da fase seguinte do *framework* (fase II), são aplicadas métricas de suporte e confiança. Tan e Kumar [TAN

2000] apresentam uma ampla discussão sobre as métricas para mensuração de qualidade de regras de associação.

#### 4.2.2 Fase II – Filtrando regras de associação a partir da relevância da opinião do usuário

Nesta fase, procura-se encontrar na base de dados todos os usuários que casam com cada uma das regras encontradas e selecionadas na Fase I. Por exemplo, caso a seguinte regra – R1:  $A1 \Rightarrow A2$  – seja encontrada, significando que A2 (área 2) será uma área de interesse do usuário caso A1 (área 1) tenha sido apontada como área de interesse no perfil deste usuário, será necessário encontrar todos os usuários da base de dados de perfis de usuários que casam com esta regra, ou seja, que se interessam por A1 e A2.

A razão para esta busca é identificar a relevância da opinião dos usuários que casam com a regra. Métricas clássicas como suporte e confiança podem mostrar que esta regra é interessante baseando-se nas transações, mas não se baseiam na relevância da opinião do usuário para selecionar as regras. Nesta tese verificou-se a necessidade de incluir mais esta métrica para garantir que o peso da opinião dos indivíduos fosse levado em consideração no momento da seleção das regras descobertas.

Calcula-se uma média aritmética entre os Ranqueamentos Totais dos Recomendadores que casam com a regra ( $RR_{total}$  de cada usuário – calculado sobre as áreas cobertas pela regra). Caso a média do RR for  $\geq \sigma$  a regra é considerada relevante, caso contrário é descartada.  $\sigma$  é um parâmetro parametrizável.

Para filtrar as regras resultantes da Fase II foi construída uma matriz onde cada linha corresponde a um usuário na base de dados de perfis do Sistema de Recomendação e as colunas representam cada regra descoberta na Fase I (com exceção da primeira coluna que representa o usuário). A relação entre usuários e regras é então apresentada na forma binária, conforme apresentado na tabela 4.2. Cada regra pode ser tratada como uma variável binária, significando que, caso a regra em questão case com o usuário então o valor será 1; caso contrário o valor será 0. Por exemplo, as regras R1, R2, R6 e R7 da tabela 4.2 casam com as áreas de interesse do Usuário 1 (U1).

Tabela 4.2: Regras e Ranqueamento do Recomendador

Usuários	Regras									
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
U1	1	1	0	0	0	1	1	0	0	0
U2	0	1	0	1	1	1	1	0	0	0
U3	0	1	1	1	1	1	0	1	1	1
U4	1	1	1	0	0	1	1	1	0	1
U5	0	1	1	1	1	0	1	1	0	1
U6	0	1	0	1	1	0	1	1	0	0
U7	0	0	1	1	1	0	0	1	0	0
U8	0	1	1	0	0	1	1	0	1	0
U9	1	0	1	1	0	1	1	0	1	1
U10	0	1	1	1	1	1	0	0	0	1

Por exemplo, a regra – R1:  $A1 \Rightarrow A2$  , casou com os usuários U1, U4 e U9 , e se estes usuários apresentam um  $RR_{total}$  para estas áreas (A1 e A2) igual a 7,5; 8,0; 8,5; sendo assim a regra foi selecionada pois a média dos  $RR_{totais}$  dos três é igual a 8,0, ou seja, esta será aplicada na fase III, se for considerado  $\sigma = 7,0$  , por exemplo.

#### 4.2.3 Fase III – Selecionando usuários para aplicar o conhecimento novo

Esta fase faz uso direto do conteúdo produzido pela Fase II. A idéia é que exista uma base de dados para preservar as regras usadas no passado, e esta base é utilizada para verificar se alguma das regras descobertas já foi aplicada anteriormente, visando não perder tempo com a mesma regra já aplicada antes. Caso a regra já tenha sido utilizada anteriormente, esta não será aplicada nos usuários (pois já foi utilizada) e não será incluída na base de novas regras (para não gerar duplicações de regras).

A próxima ação constitui-se em verificar quais usuários casam com a pré-condição da regra (LHS) e não casam com a pós-condição da regra (RHS). A razão desta seleção é identificar usuários que tenham declarado interesse em áreas que constam no LHS da regra, mas não tenham declarado interesse nas áreas constantes no RHS da regra.

Uma vez que se tenha identificado todos os usuários que atendam a esta condição, o sistema irá prever novas áreas para os usuários selecionados. Após estes usuários terem sido identificados, seus perfis serão atualizados de forma temporária. Esta ação é feita visando permitir obter certeza do interesse do usuário pela nova área identificada. Como se está trabalhando com Sistemas de Recomendação, necessita-se ter certeza que este usuário tem real interesse na nova área e esta certeza virá através do feedback do usuário.

Supondo que o sistema tenha predito que de acordo com a regra R9, o usuário U1 poderia ter interesse na área A2, serão recomendados itens relacionados a esta área de interesse para este usuário. Para verificar o real interesse do usuário e se o perfil do mesmo deveria ser atualizado definitivamente, incluindo esta nova área, o sistema aguarda o *feedback* do usuário aos itens recomendados. Caso o *feedback* seja positivo, o perfil será atualizado.

### 4.3 Algoritmo para predição de interesses

Nesta seção é descrito o algoritmo para descrever o processo de descoberta de regras (figura 4.3). A entrada para este algoritmo consiste de uma lista com todas as regras descobertas (RegrasD) durante a Fase I do *framework*, o perfil dos usuários e, da base de dados de regras, um conjunto de regras previamente descoberto (RegrasBR). A saída consiste em uma lista de regras selecionadas (RegrasS) para serem aplicadas, e os perfis dos usuários temporariamente atualizados. Todas as regras descobertas durante a Fase I (RegrasD) são consideradas.

No primeiro passo (1), o algoritmo procura descobrir todos os usuários que casam com o LHS e RHS da regra. Com o objetivo de atender este, cada regra descoberta em RegrasD é comparada com as áreas que cada usuário disse ter interesse. Todos os valores dos  $RR_{total}$  (para as áreas da regra em questão) dos usuários que casaram com as áreas da regra são acumulados na variável RR. Caso a média deste RR seja igual ou maior do que um determinado limite (exemplo, 7,0), então esta regra é selecionada como uma possível regra de interesse.

No passo (2), o sistema verifica se cada regra pré-selecionada é realmente nova, ou seja, não é um conhecimento já obtido anteriormente. Buscando analisar a condição de novidade da regra (*novelty*), o banco de dados de regras é verificado. Caso a regra realmente seja uma novidade, então esta será selecionada como uma regra final para ser usada na predição de áreas de interesse dos usuários.

Finalmente, no passo (3), as regras finalistas são casadas com os usuários. Neste momento são avaliados quais usuários casam com o LHS da regra e com a negação do RHS da regra. Os usuários selecionados poderão receber (e avaliar) as recomendações de itens relacionados com o RHS da regra.

Deve-se observar que o algoritmo aqui apresentado está na forma inicialmente proposta, não havendo preocupação com a otimização do mesmo. O trabalho de otimização e melhoria constitui-se em trabalho futuro.

<p>Entradas:</p> <ul style="list-style-type: none"> <li>- Limiar de aceitação da regra (<math>\sigma</math>);</li> <li>- Conjunto de regras descobertas na Fase I (RegrasD);</li> <li>- Perfil dos Usuários com suas áreas de interesse e ID (<math>Up</math>);</li> <li>- Conjunto de Regras da Base de Regras (RegrasBR).</li> </ul> <p>Saídas:</p> <ul style="list-style-type: none"> <li>- Conjunto de regras selecionadas na Fase III (RegrasS);</li> <li>- Perfil dos Usuários atualizados temporariamente nas áreas de interesse (<math>Up_u</math>).</li> </ul> <p>Início</p> <p>Inteiros: <math>rr = 0, k = 0, i = 0; m = 0, n = 0;</math></p> <p>(1) Para cada <math>regra_m \subseteq RegrasD</math> faça        Para cada <math>usuario_n \subseteq Up</math> faça            Se <math>((Buscar\_Areas(RegrasD[m].antecedente, Up[n].areas)=True) \wedge</math>                <math>(Buscar\_Areas(RegrasD[m].consequente, Up[n].areas)=True))</math>            então                <math>rr \leftarrow rr + Calcular\_RRtotal\_Regra(RegrasD[m], Up[n]);</math>            Final Se            <math>k=k+1;</math>        Final faça        Se <math>(Media(rr, k) \geq \sigma)</math>            então                <math>RegrasTemp[i] \leftarrow RegrasD[m];</math></p>
--



```

        i = i+1;
    Final Se
    k =0;
    rr = 0;
Final faça
m = 0;
(2) Para cada regram ⊆ RegrasTemp faça
    Para cada regran ⊆ RegrasBR faça
        Se (Buscar_Regra(RegrasTemp[m], RegrasBR[n])≠ True)
            então
                RegrasS[k] ←RegrasTemp[m];
                k=k+1;
        Final Se
    Final faça
Final faça
m = 0;n = 0;k = 0;
(3) Para cada regram ⊆ RegrasS faça
    Para cada usuarion ⊆ Up faça
        Se ((Buscar_Areas(RegrasS[m].antecedente, Up[n].areas)=True) ^
            (Buscar_Areas(¬RegrasS[m].consequente, Up[n].areas)=True))
            então
                Upu[k].id ← Up[n].id;
                Upu[k].areas ← RegrasS[m].consequente;
                k =k +1;
        Final Se
    Final faça
Final faça
Fim

```

Figura 4.3: Algoritmo de predição de possíveis interesses dos usuários

Um exemplo completo da aplicação deste *framework* é apresentado na Subseção 6.1.5 desta tese.

## RESUMO DO CAPÍTULO:

Neste capítulo foi apresentado um *framework* constituído por três fases e baseado em mineração de regras de associação e no Ranqueamento do Recomendador. Este *framework* denominado de *Framework* para Predizer o Interesse de Usuários (F-UIP) propõe-se a auxiliar na descoberta de regras que permitam associar áreas de interesse baseado no peso da opinião dos usuários da comunidade de um Sistema de Recomendação. As regras descobertas e que atendam aos requisitos especificados são aplicadas para minimizar o problema da superespecialização. Este *framework* com os respectivos resultados foi publicado na íntegra em [CAZ 2005c].

## 5 O PROTÓTIPO W-RECMAS

Visando consolidar e validar as proposições desta tese, foi construído um protótipo de sistema cujo objetivo é recomendar artigos científicos para pesquisadores e estudantes. O sistema utiliza a base de dados do CV-Lattes do CNPq e do sistema Qualis da Capes para a obtenção das informações necessárias para o cálculo do RR dos usuários. Neste Capítulo o protótipo do sistema W-RECMAS (*Web-RECommender system based on Multi-Agent System for academic paper recommendation*) é descrito em detalhes.

O sistema W-RECMAS constitui-se em um Sistema de Recomendação, baseado em Sistemas Multiagentes e mineração de dados, projetado para recomendar artigos científicos e indivíduos para pesquisadores [CAZ 2003] [CAZ 2004] [CAZ 2005a]. Através deste sistema, os pesquisadores recebem recomendações de acordo com seu perfil e terão conhecimento de qual a relevância da opinião do grupo ou do indivíduo que participou da recomendação. Espera-se que esta combinação da métrica de relevância e predição de itens diminua a sobrecarga de informação, pois esta deverá servir de parâmetro para auxiliar o usuário alvo na escolha de qual artigo (item) acessar primeiro.

Tendo em vista a importância da relevância da opinião do usuário, e sua aplicabilidade em Sistemas de Recomendação, as próximas seções apresentam mais detalhadamente a estrutura e o funcionamento do sistema W-RECMAS.

Primeiramente faz-se necessário enfatizar que no contexto deste trabalho utiliza-se o conceito “usuários” para representar os acadêmicos, “itens” para os artigos científicos e áreas para as grandes áreas nas quais se concentram os artigos. Essa nomenclatura – usuário x item – é o padrão utilizado no Sistema de Recomendação baseado em FC [HER 99] [ADO 2005].

O W-RECMAS pode ser especificado como sendo:

- um conjunto  $U$  de  $k$  distintos usuários  $U = \{u_1, u_2, \dots, u_k\}$ ;
- um conjunto  $I$  de  $n$  distintos itens  $I = \{i_1, i_2, \dots, i_n\}$ ;
- um conjunto  $AI$  de  $m$  distintas áreas de interesse  $AI = \{ai_1, ai_2, \dots, ai_m\}$ ;
- os artigos científicos (itens) bem como os usuários encontram-se relacionados a estas áreas de interesse, ou seja:
  - cada artigo científico (item) corresponde a uma ou mais áreas de interesse de tal forma que  $F: i_n \rightarrow AI$ ;

- cada usuário está relacionado a uma ou mais áreas de interesse de tal forma que  $F: u_k \rightarrow AI$  ;
- um conjunto  $R$  de distintas avaliações (*ratings*) dadas por usuários de  $U$  a itens de  $I$ ,  $R(u_k, i_n)$ , ou seja,  $F: Users \times Items \rightarrow Ratings$ .
- o sistema apresenta uma peculiaridade no que se refere a quantificação da relevância da opinião destes usuários nas áreas em que este possui interesse, ou seja:
  - um conjunto  $O$  de distintas relevâncias de opinião de usuários de  $U$  por área de interesse  $AI$ ,  $O(u_k, ai_m)$ , ou seja,  $F: Users \times \text{Áreas de Interesse} \rightarrow \text{Relevância de opinião}$ .

A figura 5.1 apresenta uma visão macro da arquitetura multiagente do sistema W-RECMAS.

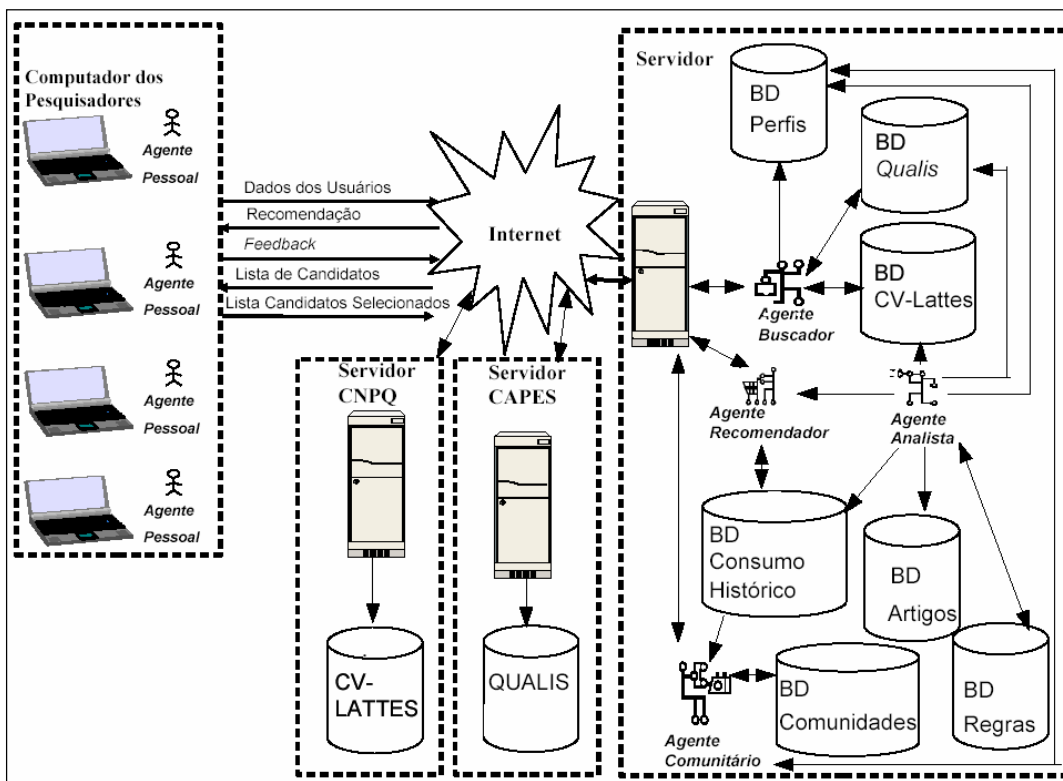


Figura 5.1: Visão macro da arquitetura do protótipo W-RECMAS

Na figura 5.1 é possível visualizar os agentes realizando as suas tarefas específicas. A seguir será apresentada uma explanação resumida do funcionamento do sistema.

O Agente Buscador procura obter informações sobre o usuário que se inscreve no Sistema de Recomendação, acessando uma base de dados que disponibilize estas informações, bem como uma base de dados que apresente uma classificação das produções científicas destes pesquisadores. Utilizando as informações fornecidas pelo usuário e pelo Agente Buscador, o Agente Analista calcula o Ranqueamento do Recomendador (RR) para este usuário, e desta forma um perfil completo é criado. De acordo com o explanado no Capítulo 3, o perfil trata-se de um elemento fundamental para o Sistema de Recomendação funcionar [PAZ 99].

Aplicando uma técnica de recomendação o Agente Analista, baseado no perfil do usuário, procura predizer um item para ser recomendado. Uma vez que exista uma recomendação, o Agente Recomendador entrega esta recomendação ao Agente Pessoal (este agente permanece ativo na máquina do usuário observando-o e esperando um contato do Agente Recomendador ou do Agente Comunitário). O Agente Pessoal lembra o usuário que este deve verificar suas novas recomendações. O Agente Comunitário verifica as informações do usuário buscando identificar membros com potencial para definir uma comunidade virtual comum, sendo que o quesito privacidade deve ser observado [CAZ 2005d]. Estes agentes serão descritos em detalhes na Seção 5.1.

Deve-se observar que para a construção do protótipo e para o cálculo do Ranking do Recomendador foram utilizadas informações disponibilizadas pelo servidor do CNPQ<sup>9</sup> e do servidor da Capes<sup>10</sup>, mais precisamente informações do sistema CV-Lattes e do sistema Qualis. A seguir, estes sistemas serão rapidamente descritos.

## 5.1 CV-Lattes e Qualis

O Sistema CV-Lattes é o componente da Plataforma Lattes desenvolvido para ser um sistema de informação curricular. No âmbito do CNPq as informações do CV-Lattes são aplicadas em avaliações como:

1. na avaliação da competência de candidatos à obtenção de bolsas e auxílios;
2. na seleção de consultores, de membros de comitês e de grupos assessores;
3. no subsídio à avaliação da pesquisa e da pós-graduação brasileiras.

Todos os bolsistas de pesquisa (mestrado, doutorado e de iniciação científica), orientadores credenciados e outros clientes do Conselho devem possuir o currículo Lattes cadastrado no CNPq. Sendo assim a base de currículos Lattes torna-se uma excelente fonte de informações referentes aos pesquisadores brasileiros nas diversas áreas de conhecimento.

A figura 5.2 apresenta os indicadores de produção do CV-Lattes de um pesquisador.

---

<sup>9</sup> <http://www.cnpq.br>

<sup>10</sup> <http://www.capes.gov.br>

Indicadores de produção	
Produção bibliográfica	Produção técnica
Orientações concluídas	Dados complementares
<b>Produção bibliográfica</b>	
Artigos publicados em periódicos	
Completos	2
Trabalhos em eventos	
Completos	20
Livros e capítulos	
Livros publicados ou organizados	0
Capítulos de livros publicados	1
Textos em jornais ou revistas (magazines)	
Jornal de notícias	1
Demais tipos de produção bibliográfica	5
<b>Produção técnica</b>	
Softwares	
Softwares sem registro ou patente	1
Trabalhos técnicos	4
Demais tipos de produção técnica	14
<b>Orientações concluídas</b>	
Aperfeiçoamento/Especialização	3
Graduação	26
<b>Dados complementares</b>	
Participação em bancas examinadoras	27
Participação em bancas de comissões julgadoras	7
Participação em eventos	34
Orientações em andamento	3

Figura 5.2: Indicadores de produção no CV-Lattes do pesquisador

No currículo Lattes de cada pesquisador é possível encontrar todas as publicações, áreas de atuação e indicadores de produção. Nesta tese optou-se em fazer uso desta importante e completa fonte de informações para compor o perfil dos pesquisadores no W-RECMAS. Ao trabalhar com o conteúdo das produções acadêmicas de cada pesquisador, observou-se a necessidade de analisar cada publicação de forma a diferenciá-las quanto ao impacto das mesmas, desta forma optou-se por trabalhar com uma listagem de relevância de publicações, denominada Qualis<sup>11</sup> apresentada pela Capes. Esta lista apresenta uma categorização sobre a relevância (impacto) entre as revistas científicas, jornais e conferências.

Esta lista foi utilizada no protótipo descrito, apesar da mesma ter criado bastante controvérsia no mundo acadêmico brasileiro. A adoção da mesma nesta aplicação foi feita por comodidade e não por esta ter sido superior em qualidade a outras existentes. A Figura 5.3 apresenta uma visualização da lista de periódicos classificados no Qualis. Maiores informações sobre o Qualis da Ciência da Computação podem ser obtidas no *site* da Unicamp<sup>12</sup> (Figura 5.4).

<sup>11</sup> <http://qualis.capes.gov.br/>

<sup>12</sup> <http://qualis.ic.unicamp.br/>

Qualis - Sistema de Classificação de Periódicos, Anais e Revistas - Microsoft Internet Explorer

Address: http://qualis.capes.gov.br/

Ministério da Educação Destaque do Governo

**QUALIS - Classificação de Periódicos, Anais, Jornais e Revistas**  
versão 1.0

Classificação relativa a dados de 2003

Consulta

Critérios de Classificação do Qualis por Área

ISSN	Título	Classificação	Circulação
0002-9890	American mathematical monthly	B	Internacional
1531-4278	ACM Journal Of Education Resources In Computing	B	Internacional
1084-6654	ACM Journal of Experimental Algorithmics	B	Internacional
0163-5808	ACM Sigmod Record	A	Internacional
0362-1340	ACM Sigplan Notices	B	Internacional
0163-5948	ACM Software Engineering Notes	B	Internacional
1084-4309	ACM Transactions on Design Automation of Electronic Systems	A	Internacional
1539-9087	ACM Transactions on Embedded Computing Systems	B	Internacional
1046-8188	ACM Transactions on Information Systems	A	Internacional
0104-6188	ACM Transactions On Information Systems	B	Internacional
0098-3500	ACM Transactions on Mathematical Software	A	Internacional
0798-4545	Acta microscopica - Interamerican Committee of Societies for Electron Microscopy	C	Internacional
0001-706X	Acta Tropica	C	Internacional
0065-2458	Advances in Computers	B	Internacional
0965-9978	Advances in Engineering Software	B	Internacional
0273-1177	Advances in Space Research	C	Internacional
0030-9708	Advances in water resources	Sem Classificação	Internacional
0308-521X	Agricultural Systems	C	Internacional

Figura 5.3: Classificação Qualis

Qualis Ciência da Computação - Portal Qualis/CC - Mozilla Firefox

Address: http://qualis.ic.unicamp.br/

Qualis-CC

página inicial | conferências | periódicos | arquivos

you are here: home

navigation

- Home
- Perguntas Frequentes
- Atualizações e Revisões
- Conferências
- Periódicos
- Arquivos

**Qualis Ciência da Computação**  
by rapido - last modified 30-11-2005 15:42

O Qualis é um instrumento concebido pela Capes para a classificação de veículos de divulgação da produção científica, tecnológica, artística etc, dos programas de pós-graduação. Ele foi implantado em 1998 e desde então vem sendo utilizado para a composição de indicadores fundamentais para o processo de avaliação dos programas.

Os critérios adotados para a construção do Qualis são descritos no Documento de Área, definido pelo Comitê de Área no início de cada triênio e divulgado no site da Capes. Na área computação, essa descrição toma a forma de um *Conjunto de Regras* que define a classificação dos veículos (periódicos e congressos).

O *Conjunto de Regras* da área de computação foi especificado pela primeira vez durante a gestão do prof. Virgílio Almeida como representante de área (2001-2003), e significa um primeiro passo na construção do Qualis de Ciência da Computação.

Além do *Conjunto de Regras*, a Capes divulga também uma *Lista de Veículos*. No caso da computação, assim como de outras áreas, a *Lista de Veículos* é ampliada anualmente, a partir dos relatórios dos programas de pós-graduação (processo referenciado pela Capes como 'Coleta' -- derivado do nome do aplicativo que os coordenadores utilizam para compor o relatório). Assim, é importante entender que a Lista de Veículos publicada no site da Capes não é uma lista completa de classificação de todos os veículos na área de computação, especialmente no caso de congressos. Ela é apenas a lista dos veículos nos quais algum pesquisador de um programa de pós-graduação publicou no período de referência.

A construção de uma Lista de Veículos para a área de computação, que sinalize para os programas os veículos de maior relevância para a comunidade nacional e internacional, mesmo que não tenham sido ainda citados no Coleta de Dados, é um desafio que a comunidade de

Figura 5.4: Classificação Qualis - Ciência da Computação

A figura 5.5 apresenta o esquema geral do cálculo do Ranqueamento do Recomendador dos usuários, instanciando o modelo do capítulo 3 para aplicação no CV-Lattes e utilizando a lista Qualis. Para calcular o RR por área de interesse do usuário, primeiramente o usuário deve informar ao W-RECMAS quais são estas áreas, e sua chave de acesso do CV-Lattes no CNPQ. O conteúdo dos CV-Lattes é decomposto e o Ranqueamento do Recomendador por área de interesse é calculado baseado nos valores dos indicadores de produção e nos artigos publicados, constantes no currículo do usuário.

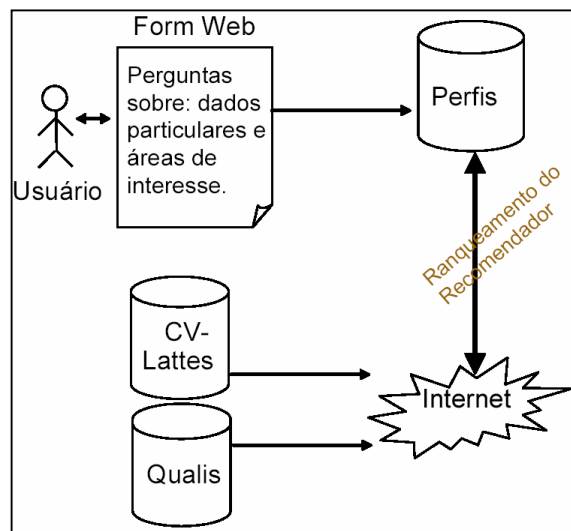


Figura 5.5: Esquema do cálculo do RR

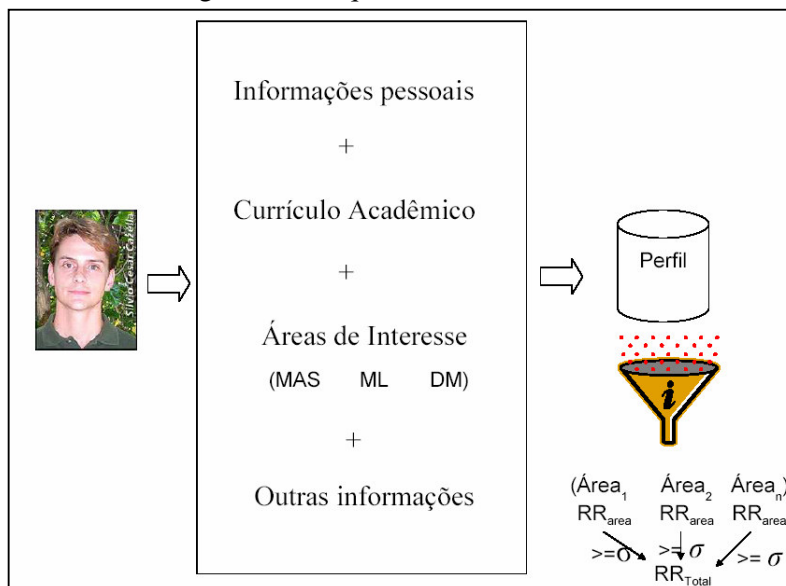


Figura 5.6: Informações para o cálculo do RR dos usuários

O modelo que representa a relevância da opinião do usuário pode ser aplicado utilizando qualquer escala de valores. Por exemplo, a escala para o Ranqueamento do Recomendador poderia ser de [0-10], onde 0 é o menor valor de relevância de opinião e 10 é o maior valor de relevância de opinião. Deve-se destacar que os usuários possuirão um RR por área de interesse, conforme apresentado na figura 5.6.

## **5.2 Agentes de software**

No contexto desta tese um agente é entendido como uma entidade autônoma e independente, a qual possui um objetivo (meta) específico, possui a habilidade de realizar uma tarefa específica em um ambiente dito computacional [CAZ 97][RUS 2002]. Estes agentes apresentam as seguintes características [WEI 99]: autonomia, habilidade social, continuidade, capacidade de comunicar-se e capacidade de raciocínio.

As funcionalidades inerentes ao sistema são suportadas basicamente pelos seguintes agentes de software: Agente Buscador, Agente Analista, Agente Pessoal, Agente Recomendador e Agente Comunitário.

### **5.2.1 Agente Buscador**

A principal tarefa deste agente é suprir o Sistema de Recomendação com informações dos usuários que se cadastram no mesmo, buscando permitir a criação de um perfil detalhado do usuário.

Uma vez que o Agente Buscador receba o nome completo e a chave de registro do usuário no CV-Lattes, este agente verifica se o usuário é realmente novo no sistema e caso o nome não conste na base de dados do sistema, será incluído e o agente irá buscar o CV-Lattes do usuário na base de dados do CNPQ. Este agente também é responsável por acessar a lista do Qualis. Uma vez que estas informações estiverem disponíveis, o Agente Buscador envia uma mensagem para o Agente Analista.

O Agente Buscador monitora periodicamente (a periodicidade é um valor parametrizável), o Sistema CV-Lattes e Qualis, procurando verificar alguma atualização que tenha sido feita no CV-Lattes do Usuário ou na lista do Qualis. Caso tenha ocorrido alguma atualização, o agente copia o currículo ou a lista do Qualis novamente e envia uma mensagem para o Agente Analista, que se encarregará de realizar o parse destes conteúdos e atualizar o perfil do usuário do sistema. A figura 5.7 apresenta uma visão da arquitetura interna do Agente Buscador e suas interfaces.



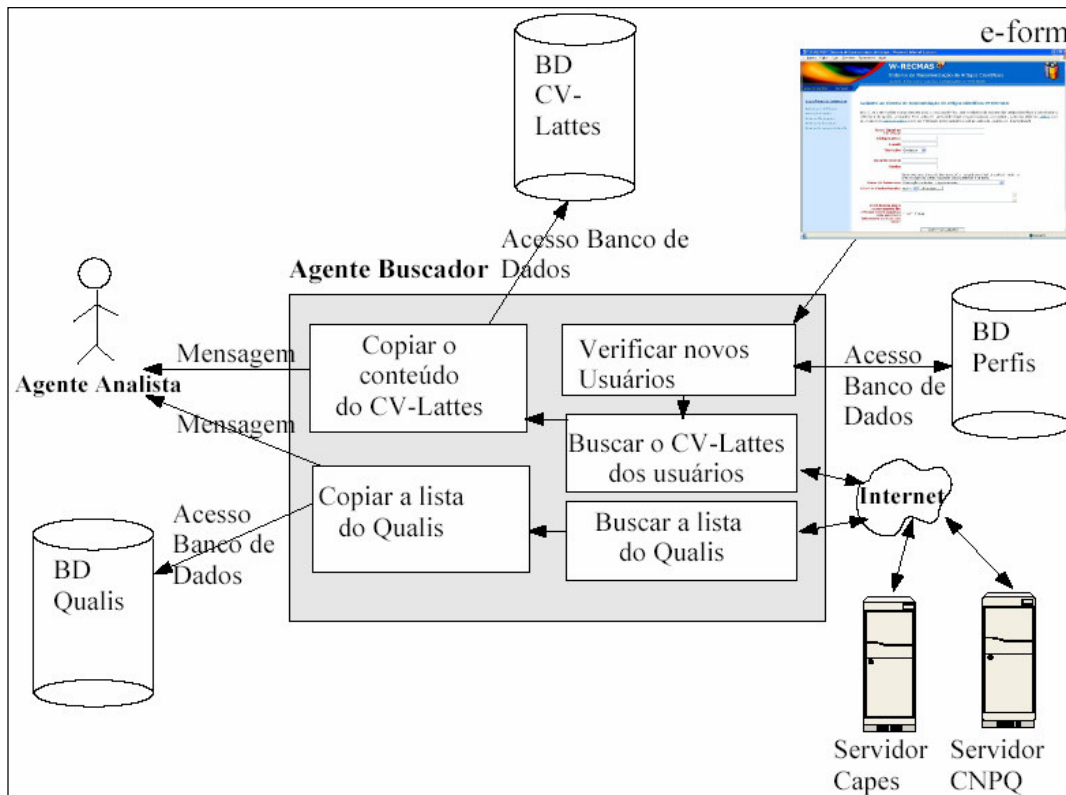


Figura 5.7: Arquitetura interna e interfaces do Agente Buscador

### 5.2.2 Agente Analista

O Agente Analista constitui-se no mais complexo agente do sistema, uma vez que o mesmo realiza muitas tarefas que estão descritas a seguir:

a) Decompor o CV-Lattes do usuário e a lista do Qualis: Quando o Agente Analista recebe uma mensagem proveniente do Agente Buscador requisitando a análise de um CV-Lattes ou da lista do Qualis, o Agente Analista inicia a análise do conteúdo do currículo ou lista, colocando o conteúdo em um outro objeto, gravando em uma base de dados auxiliar para posteriormente redefinir o perfil do usuário.

b) Calcular a métrica de Ranqueamento do Recomendador do usuário: O Agente Analista calcula a relevância da opinião do usuário (Capítulo 3) para as respectivas áreas de conhecimento em que o usuário apresenta interesse. Por exemplo, suponha que o usuário seja um Doutor em Ciência da Computação e uma de suas áreas de interesse seja “mineração de dados” e que aplicando o cálculo do  $RR_{area}$ , este valor ficou em 7,9 em uma escala de [0-10], significando que caso este usuário colabore em alguma processo de recomendação, a sua opinião apresentará um peso igual a 7,9 na área de interesse, ou seja, a opinião apresenta alta relevância (assumindo-se alta relevância valores entre 7,0 e 10,0).

c) Casar o perfil do usuário e os itens disponíveis: Buscando realizar este casamento entre perfil e itens, o Agente Analista precisa decidir qual técnica de filtragem é melhor para cada usuário em um momento específico – filtragem colaborativa ou baseada em conteúdo. Na primeira técnica, é assumida a existência de um domínio de itens para ser recomendado e que cada usuário do sistema pode fornecer opiniões sobre os itens (geralmente pontuações). Este tipo de técnica propõe alguma personalização por explorar similaridades e dissimilaridades entre as preferências dos usuários. O sistema compara as pontuações fornecidas pelos usuários e encontra usuários com similaridade de opinião (*like-minded*), baseado em algum critério e similaridades, o sistema irá recomendar itens para estes usuários.

A filtragem baseada em conteúdo, foca na análise do conteúdo do item e não na pontuação que este recebeu de outros indivíduos. No caso da solução desenvolvida a recomendação é feita pela relação área de interesse e itens da referida área. Na filtragem baseada em conteúdo, a idéia principal é avaliar o conteúdo de um item e casar este com os itens consumidos no passado pelo usuário em questão. Caso o conteúdo seja “similar”, este é possível de ser recomendado como novo item. O agente neste caso precisa decidir qual técnica deve ser aplicada em um momento específico. Por exemplo, caso existam pontuações suficientes, bem como um número suficiente de usuários, o sistema poderá aplicar um coeficiente para determinar quem é similar [KON 97], ou seja, o agente irá atuar usando a técnica de filtragem colaborativa.

Outra forma de atuação do agente implica na utilização de filtragem baseada em conteúdo, isto significa que o agente analisa os interesses do usuário e os itens acessados pelo mesmo no passado, para predizer se um item específico será de interesse para o usuário alvo (maiores detalhes no Capítulo 2).

d) Predizer novas áreas de interesse para o usuário baseado em abordagem colaborativa: Nesta tarefa o Agente Analista aplica a tarefa de mineração denominada regras de associação visando descobrir conhecimento novo para melhorar a lista de interesses do usuário, baseado na opinião dos outros usuários que participam do sistema. Para descobrir estas novas áreas o agente faz uso do *framework* F-UIP descrito no capítulo 4.

A figura 5.8 apresenta a arquitetura interna do Agente Analista e suas respectivas interfaces.

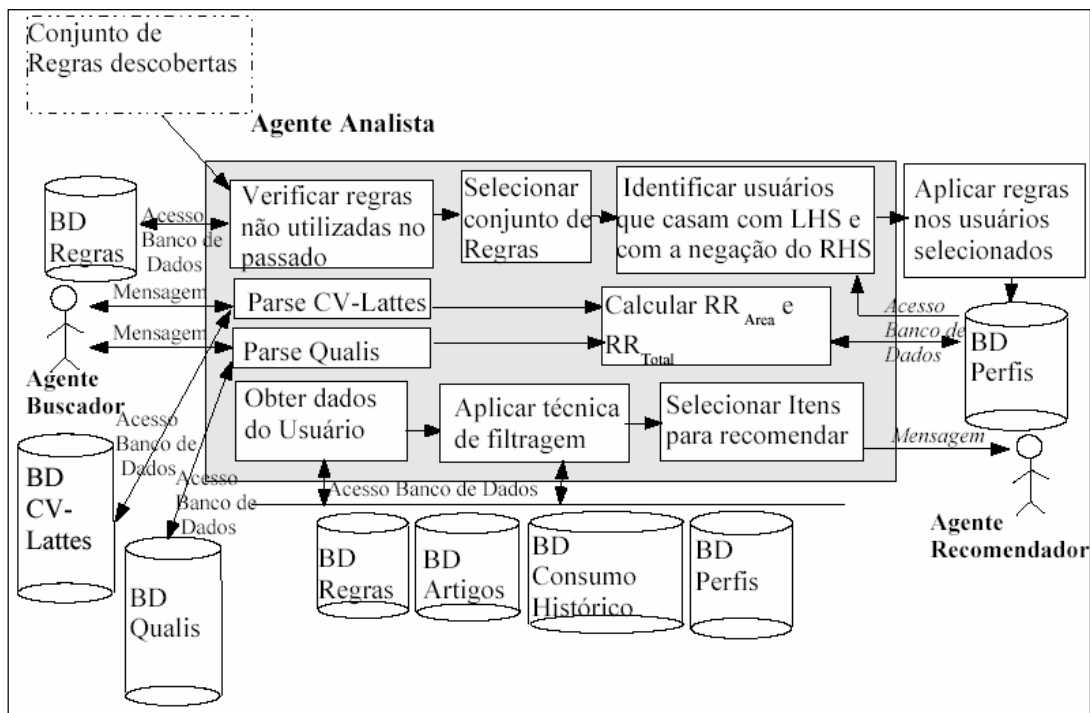


Figura 5.8: Arquitetura interna e interfaces do Agente Analista

### 5.2.3 Agente Recomendador

Este agente foi projetado para comunicar a lista de recomendações para o Agente Pessoal do usuário. O Agente Recomendador recebe a lista de recomendações do Agente Analista para um usuário específico, a partir desta lista, este seleciona um total de itens (valor parametrizável) baseado no valor do Ranqueamento do Recomendador (seleciona os itens recomendados com maiores valores de RR). O restante dos itens que não forem recomendados são salvos em uma base de dados de consumos históricos, com um atributo definido como RF (recomendações futuras). Outros valores possíveis são RCF (recomendações com *feedback*) e RSF (recomendações sem *feedback*).

A figura 5.9 apresenta uma visão da arquitetura interna do Agente Recomendador e suas respectivas interfaces.

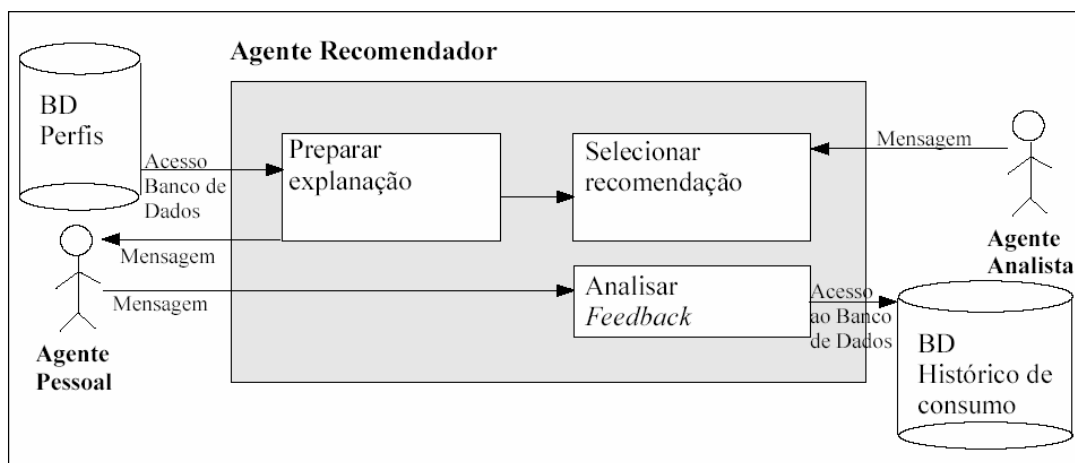


Figura 5.9: Arquitetura interna e interfaces do Agente Recomendador

#### 5.2.4 Agente Pessoal

Este agente foi projetado para permanecer fisicamente localizado na máquina do usuário e sua função fundamental é apresentar as recomendações para o usuário alvo (aplicando tecnologia *push*<sup>13</sup> [SCH 99] [SCH 2001]). A figura 5.10 apresenta uma visão da arquitetura interna do Agente Recomendador e suas respectivas interfaces, e a figura 5.11 inclui a tecnologia *push* aplicada.

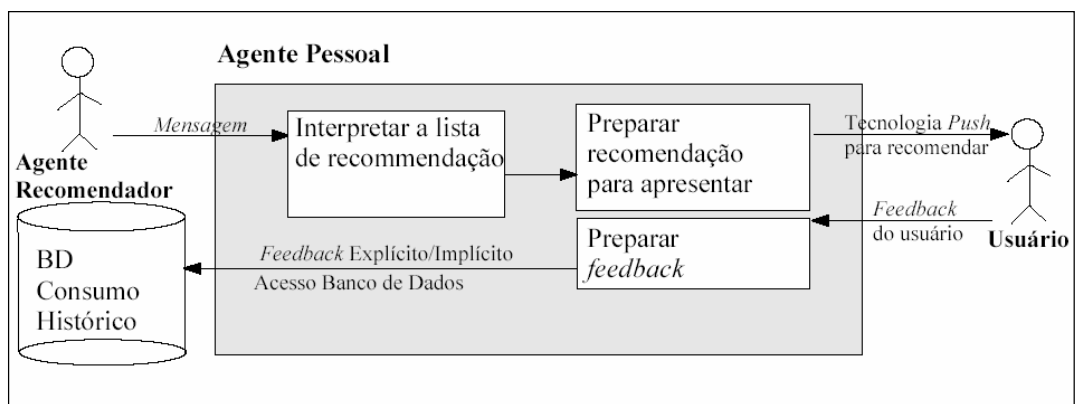


Figura 5.10: Arquitetura interna e interfaces do Agente Pessoal

<sup>13</sup> Tecnologia *push* ou *webcasting* representa a transmissão, de informações específicas para pessoas, de forma individual.

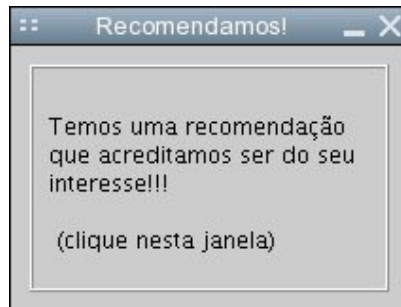


Figura 5.11: Aplicando a tecnologia *push* para recomendar

### 5.2.5 Agente Comunitário

Outra preocupação nesta tese refere-se à existência de usuários nos quais outros usuários poderiam confiar para terem auxílio na busca de informações. Neste contexto, não se trabalha com relações explícitas de confiança, ou seja, aquelas relações onde o usuário descreve a sua confiança em um dos demais usuários do sistema (como exemplo, tem-se o sistema e-bay<sup>14</sup>, onde os usuários explicitamente declaram o quanto satisfatória foi a transação realizada com outros usuários do sistema).

Neste trabalho relações implícitas de confiança são alvos de análise. Partindo-se do princípio de que a troca direta de informação entre usuários pode ser mais proveitosa do que uma busca constante por informações em *sites* de busca, optou-se por trabalhar com a identificação de usuários similares e confiança envolvida. As relações de confiança representam uma maneira simples de construir comunidades, ainda mais se for possível identificar indivíduos com interesses em comum e que sejam *like-minded* [CAS 2001]. Autores com Abdul-Rahman e Hailes [ABD 2000] demonstraram que em um contexto predefinido (exemplo, filmes) os usuários tendem a desenvolver conexões sociais com pessoas com as quais apresentam interesses e preferências similares. Estes resultados foram estendidos no trabalho de Ziegler e Lausen [ZIE 2004], onde foi apresentado através de um estudo empírico de uma comunidade real (*online*) a correlação entre confiança e similaridade de usuários.

Isto significa que o sistema é capaz de identificar este tipo de relação de confiança implícita e uma vez feita esta identificação este procura colocar estes usuários em contato (respeitando a questão de privacidade apresentado no Capítulo 2 desta tese). Caso o sistema identifique esta relação entre os usuários, este pode recomendar usuários como um contato interessante, visando que os mesmos troquem informações.

Com o objetivo de colocar estas pessoas em contato, o agente não poderia simplesmente enviar os dados de cada usuário para outro usuário, uma vez que os dados dos usuários de um Sistema de Recomendação devem ser protegidos conforme explicado no Capítulo 2. Por esta razão, o Agente Comunitário inicia a comunicação com o Agente Recomendador e este contata os respectivos Agentes Pessoais dos usuários. Esta comunicação se faz necessária para verificar se um dos usuários deseja entrar em contato com os demais que estão sendo “recomendados” pelo sistema. Por

<sup>14</sup> <http://www.e-bay.com>

exemplo, o Agente Pessoal do usuário 1 recebe uma lista de usuários que poderiam participar de uma possível comunidade, sendo que cada usuário é designado por seu  $RR_{\text{area}}$  para cada área de interesse comum.

Esta comunicação visa apresentar a relevância da opinião dos usuários em suas respectivas áreas. O Agente Pessoal apresenta uma lista de possíveis usuários (recomendados) que poderiam ser incluídos na comunidade, e requisita ao proprietário do agente que este selecione quais usuários da lista gostaria de entrar em contato para trocar idéias. Neste mesmo momento, o Agente Pessoal dos demais usuários (membros da lista), estarão realizando a mesma tarefa.

Uma vez que o Agente Comunitário tiver recebido a resposta de todos os Agentes Pessoais dos usuários envolvidos na montagem desta possível comunidade, este analisará as respostas. Por exemplo, Usuário1 será colocado em contato com Usuário2, se e somente se o Usuário1 apresentar interesse explícito em contatar com Usuário2, e o Usuário2 apresentar o mesmo interesse. Caso este interesse mútuo não exista, eles não serão colocados em contato.

Após a verificação de todas as respostas enviadas pelos envolvidos na possível comunidade, o Agente Comunitário fornece o contato entre os usuários para que estes estabeleçam contato entre si. A figura 5.12 apresenta uma visão da arquitetura interna do Agente Comunitário e suas interfaces.

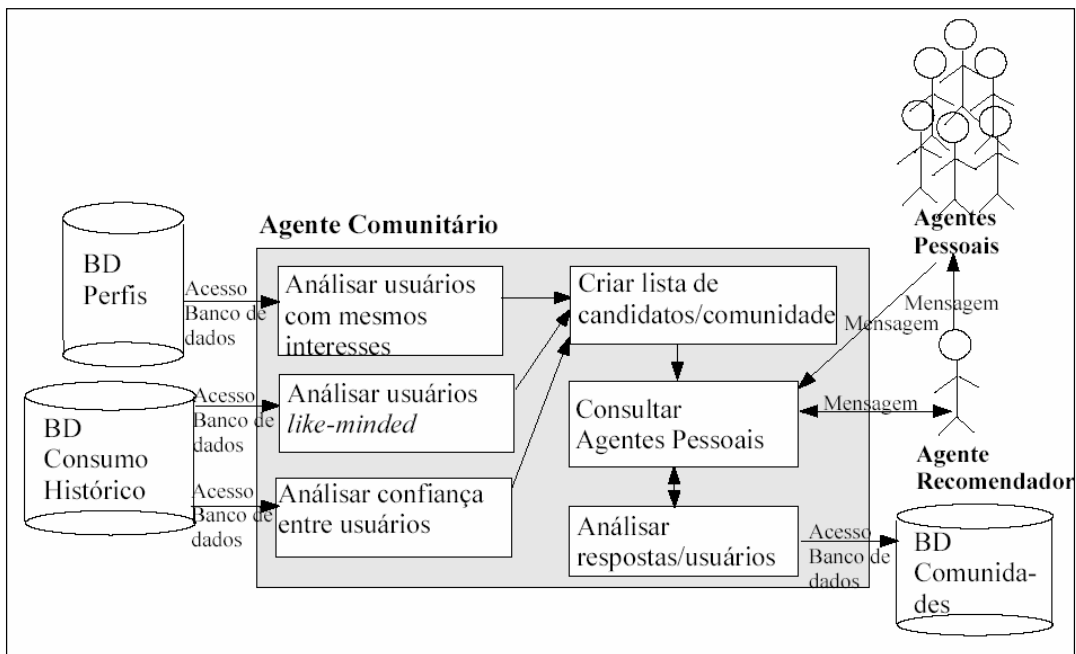


Figura 5.12: Arquitetura interna e interfaces do Agente Comunitário

A seguir define-se a regra completa que é aplicada pelo Agente Comunitário para definir os candidatos a uma dada comunidade. Sendo  $U = \{u_0, \dots, u_{n-1}\}$  um conjunto de  $n$  usuários e sendo  $AI = \{a_0, \dots, a_{m-1}\}$  um conjunto de  $m$  áreas de possível interesse.

Seja  $C$  uma comunidade onde  $C \subseteq U$ . O Agente Comunitário considera usuários como sendo candidatos a uma nova comunidade caso a seguinte regra seja satisfeita:

*Iff* ( $related\_area(u_n, AI) \cap related\_area(u_{n+1}, AI) \neq \emptyset$  and  
 $like\_minded(u_n, u_{n+1}) = true$  and  
 $trustworthy(u_n, u_{n+1}) = true$   
 then candidates

Aplicando esta regra entende-se que os usuários que satisfizerem a mesma são realmente candidatos a uma comunidade se e somente se:

- a) os usuários apresentam pelo menos uma área de interesse em comum;
- b) são *like-minded* (conforme definido no capítulo 2);
- c) os usuários apresentam algum tipo de relação de confiança (*trust*).

Nesta tese assume-se que caso um usuário receba recomendações de um grupo de pessoas e o usuário alvo fornecer um *feedback* positivo para um mínimo de três recomendações (este número constitui-se em uma variável parametrizável) produzidas com a participação destes usuários, existe uma relação de confiança.

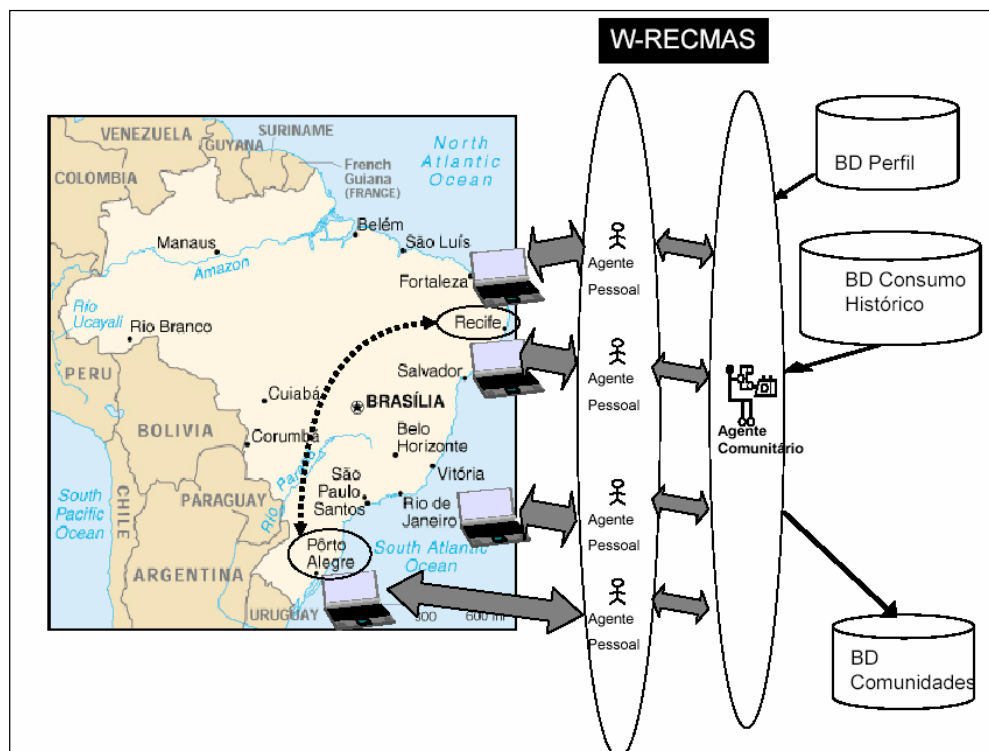


Figura 5.13: Estabelecendo uma comunidade

A figura 5.13 apresenta a criação de uma pequena comunidade virtual composta por dois pesquisadores que se encontram em localidades distintas do território nacional, possuem interesses em comum e talvez pesquisem as mesmas áreas, porém não conhecem o trabalho um do outro. Através da aplicação de Sistemas de Recomendação e da identificação da existência de áreas de interesse em comum e gostos semelhantes, é possível a construção de uma “ponte” para comunicação entre estes pesquisadores (com a devida apresentação de interesse por ambos).

Esta verificação de que um outro pesquisador está desenvolvendo trabalho semelhante e que possui interesse em área comum, pode facilitar a junção de esforços futuros, ou seja, a criação de pesquisas únicas que unirão esforços e conhecimentos de ambos (maiores detalhes no Anexo G).

### 5.3 Arquitetura do protótipo W-RECMAS

A figura 5.14 apresenta como todos os componentes do protótipo comunicam-se a fim de calcular a relevância da opinião dos pesquisadores, realizar as recomendações de artigos científicos e indivíduos usando esta relevância, identificar novos interesses e definir possíveis comunidades.

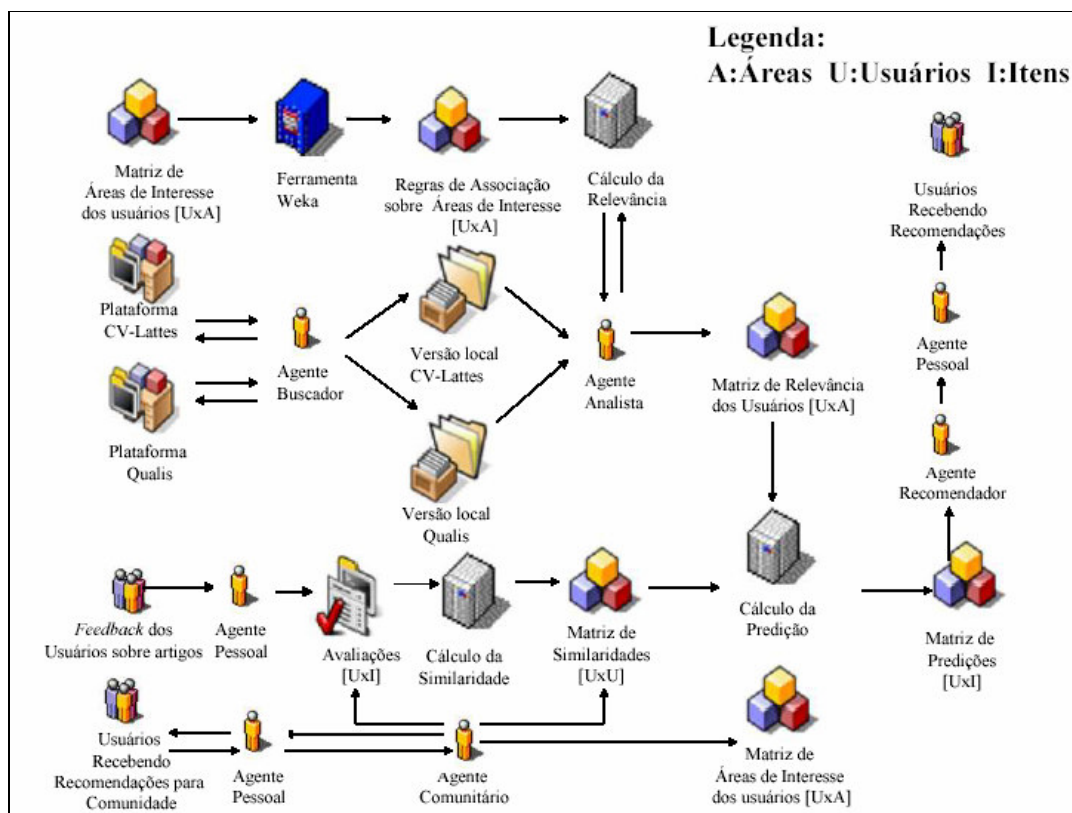


Figura 5.14: Visão geral da arquitetura do protótipo W-RECMAS



O fluxo de informação, partindo de uma base de itens e usuários já cadastrados, funciona da seguinte forma:

1. O Agente Buscador acessa o CV-Lattes dos usuários na plataforma Lattes e acessa a lista do Qualis, e armazena uma cópia local de cada um. O acesso e a verificação da atualização, tanto dos currículos como da lista do Qualis, é definida em um valor parametrizável (exemplo, pode ocorrer semanalmente).
2. Em seguida, o Agente Analista decompõe o CV-Lattes local e realiza o cálculo da relevância da opinião do usuário. Neste passo, a relevância da opinião dos usuários está associada ao seu perfil (o qual apresenta áreas de interesse do usuário). A mesma decomposição ocorre com a lista do Qualis.
3. Paralelamente, os usuários fazem avaliações (feedback) sobre artigos que lhes foram recomendados previamente, sendo que estas avaliações são fornecidas ao Agente Pessoal. Através destas avaliações é realizado o cálculo da similaridade já demonstrado no Capítulo 2, para a identificação dos usuários.
4. Com a relevância e a similaridade é realizado o cálculo da predição dos itens para os usuários (recomendações), segundo equação 9 do Capítulo 3, Seção 3.3.
5. As recomendações selecionadas são encaminhadas ao Agente Recomendador que se obriga a encaminhar ao correto Agente Pessoal do usuário.
6. Periodicamente, a matriz de áreas de interesses dos indivíduos é submetida à mineração por regras de associação através da aplicação do F-UIP e auxílio da ferramenta Weka<sup>15</sup> [WIT 2000] (Capítulo 4). Esta aplicação busca descobrir regras de associação que apresentem relações novas entre as áreas de possível interesse dos usuários (descoberta baseada na relevância da opinião de cada usuário).
7. Enquanto o processo de recomendação permanece ocorrendo o Agente Comunitário encontra-se em *background* em busca de usuários para serem aproximados (recomendados). Somente serão alvo de aproximação os usuários que aceitaram esta função a cadastrarem-se no sistema.
8. Uma vez que o Agente Comunitário identifica os usuários que poderiam ser aproximados, este encaminha um convite para o Agente Pessoal dos candidatos, e aguarda a resposta de cada usuário candidato. Caso ambos candidatos informem que possuem interesse em manter contato, ocorrerá a inclusão dos mesmos em uma determinada comunidade. Conforme figura 5.15.

---

<sup>15</sup> <http://www.cs.waikato.ac.nz/~ml/weka>

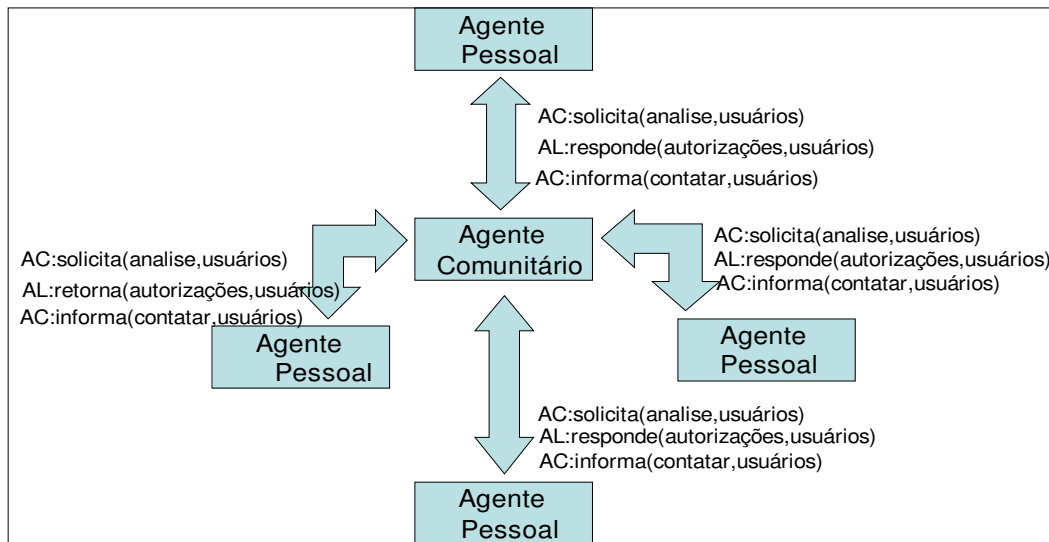


Figura 5.15: Comunicação para estabelecer uma comunidade no W-RECMAS

### 5.3.1 Implementação do W-RECMAS

Abaixo seguem alguns detalhes referentes à implementação feita.

#### 5.3.1.1 Linguagem de Programação

Para realizar a implementação do protótipo do sistema foi utilizada a linguagem Java<sup>16</sup>, mais especificamente a versão J2SE 1.4.2. Java é uma tecnologia que proporciona a independência de plataforma, logo o protótipo desenvolvido também possui esta característica, sendo desenvolvido e testado em ambiente Linux (distribuição Fedora Core<sup>17</sup>).

Para a implementação do W-RECMAS optou-se por um projeto *open-source* desenvolvido em Java, chamado JADE<sup>18</sup> (*Java Agent DEvelopment Framework*) [BEF 99], que segue as especificações da FIPA<sup>19</sup> (*Foundation for Intelligent Physical Agents*) para a interoperabilidade de Sistemas Multiagentes. Com ele foi possível implementar as tarefas dos agentes e permitir a troca de mensagens seguindo as especificações ACL da FIPA.

#### 5.3.1.2 Camada de Persistência

A camada de persistência constitui-se na ligação entre o ambiente de execução do protótipo e o banco de dados, ou seja, é a tarefa de obter, manipular e armazenar as informações num banco de dados.

<sup>16</sup> <http://java.sun.com>

<sup>17</sup> <http://fedora.redhat.com>

<sup>18</sup> <http://jade.tilab.com>

<sup>19</sup> <http://www.fipa.org>

Para viabilizar a persistência dos dados, foi utilizada a tecnologia denominada Hibernate<sup>20</sup>, na versão 2.1.6. Sua principal vantagem é a facilidade de realizar o mapeamento entre as classes do protótipo (paradigma orientado a objetos) e as tabelas do banco de dados (paradigma relacional), onde é apenas necessária a criação de um arquivo XML<sup>21</sup> com a definição do mapeamento. Com o seu uso, é possível abstrair totalmente a tecnologia de banco de dados que se está utilizando, pois o acesso a eles é realizado sempre a instâncias das classes mapeadas, ou seja, não há acesso direto às tabelas do banco de dados – isto é realizado automaticamente pelo Hibernate.

#### 5.3.1.3 Banco de Dados

O sistema gerenciador de banco de dados (SGBD) escolhido foi o MySQL<sup>22</sup> considerado um dos mais populares na comunidade *open-source*<sup>23</sup>.

#### 5.3.1.4 CoFE – Collaborative Filtering Engine

O CoFe constitui-se em tecnologia desenvolvida pelo grupo de pesquisa “*Intelligent Information Systems*” da Universidade do Estado de Oregon, EUA, sendo uma implementação de uma engenharia para recomendação, feita em Java, usando a técnica de filtragem colaborativa explicada no capítulo 2. O projeto é *open-source*, e pode ser acessado em <http://eecs.oregonstate.edu/iis/CoFE>. Neste protótipo, utilizou-se a versão 0.3.

O objetivo do CoFE é facilitar a evolução da tecnologia de Sistemas de Recomendação, tanto no meio acadêmico como no comercial, disponibilizando um *framework* de fácil uso e adaptação. O CoFE é utilizado para realizar as recomendações, sendo que neste protótipo o mesmo sofreu uma adaptação para considerar a relevância da opinião dos usuários.

### 5.3.2 Interface do protótipo

Nesta seção é apresentada a interface do protótipo W-RECMAS. A Figura 5.16 apresenta a interface de *login* do protótipo W-RECMAS. Nesta interface observa-se que caso o usuário ainda não esteja cadastrado no sistema, ele pode iniciar o cadastramento clicando em novo usuário e a interface apresentada na Figura 5.17 será disponibilizada.

---

<sup>20</sup> <http://www.hibernate.org>

<sup>21</sup> <http://www.w3.org/XML/>

<sup>22</sup> <http://www.mysql.com>

<sup>23</sup> <http://www.opensource.org>

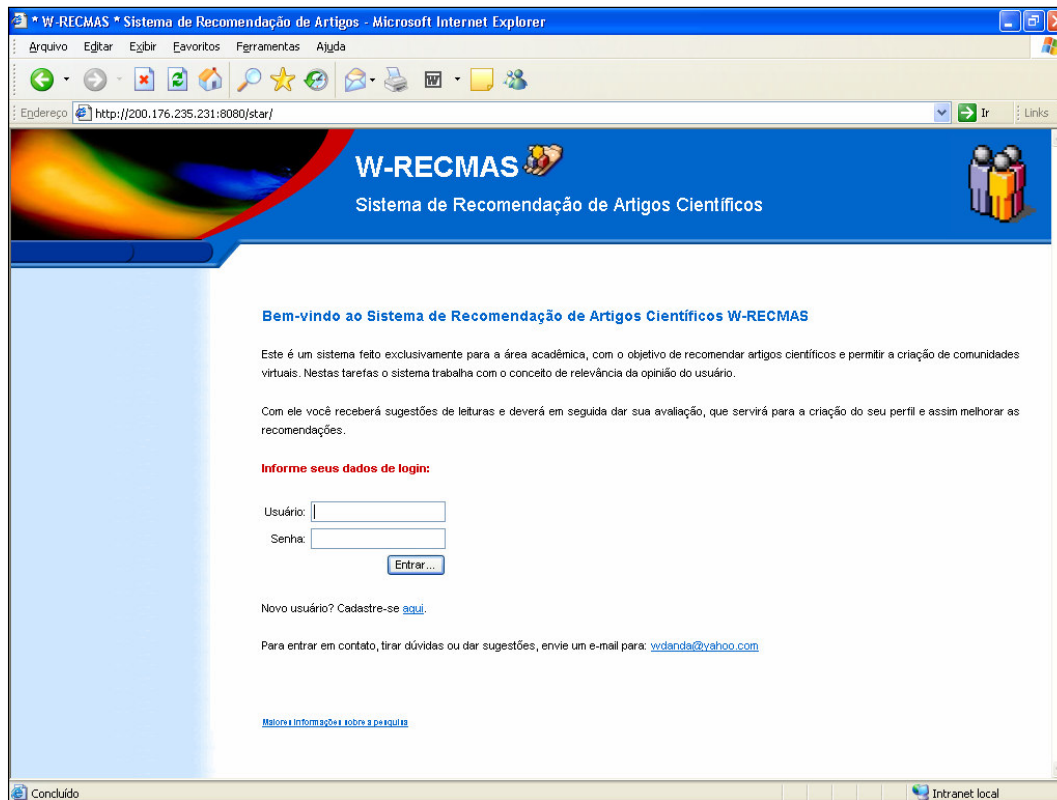


Figura 5.16: Interface de *login* do protótipo W-RECMAS

A Figura 5.17 apresenta a interface para cadastramento no sistema. Neste cadastramento são solicitadas informações como o nome completo do novo usuário bem como o seu código no CV-Lattes (por exemplo, K4793452Z8), visando viabilizar o *download* do currículo do usuário.

Pode se observar nesta interface a solicitação de e-mail de contato (buscando permitir a comunicação futura do sistema com o usuário), titulação (graduando, graduado, mestrando, especialista, mestre, doutorando, doutor ou pós-doutor), nome de usuário e senha, seleção de áreas de interesse entre as disponíveis atualmente no sistema - Inteligência Artificial (Mineração de dados: Agrupamento, Mineração de dados: Classificação, Mineração de dados: Regras de Associação, Sistemas Multiagentes, Representação de Conhecimento), Computação Gráfica, Redes de Computadores (Redes *peer-to-peer*, Sistemas Distribuídos e Segurança de Sistemas Computacionais) e Engenharia de Software (Métodos Ágeis, Modelos de Qualidade), nível de conhecimento na área de interesse, e autorização do usuário para que o Agente Comunitário procure incluí-lo em alguma comunidade através da identificação de pessoas com interesses em comum aos seus.

Através dos dados fornecidos neste cadastramento o sistema construirá um perfil para o usuário através da atuação dos agentes de software.

**W-RECMAS**  
Sistema de Recomendação de Artigos Científicos

Encerrar Sessão Principal

### Cadastro ao Sistema de Recomendação de Artigos Científicos W-RECMAS

Este é um sistema feito exclusivamente para a área acadêmica, com o objetivo de recomendar artigos científicos e considerar a relevância do opinião do usuário. Para tanto, é imprescindível que o usuário possua um cadastro junto à plataforma [Lattes](#) com seu respectivo [Currículo Lattes](#). Deve ser informado neste cadastro o Código Lattes do usuário, ex: **K479345228**.

**Nome (igual ao CVLattes):**

**Código Lattes:**

**E-mail:**

**Titulação:** Graduado

**Usuário (login):**

**Senha:**

Selecione suas áreas de interesse, informe qual seu nível de conhecimento na área e clique no botão *Adicionar* para confirmar a seleção.

**Áreas de Interesse:** Mineração de dados :: Agrupamento

**Nível de Conhecimento:** Baixo

**Você deseja que o nosso agente lhe informe sobre usuários com possíveis interesses comuns aos seus?**  Sim  Não

Figura 5.17: Interface de cadastramento no protótipo W-RECMAS

A Figura 5.18 apresenta a interface para busca de artigo e posterior inclusão do mesmo caso este não exista na base de artigos do sistema. No cadastramento de um artigo novo o usuário deverá fornecer informações como: título do artigo, autor(res), informações sobre local de publicação, ano da publicação, url, informar se quem está cadastrando o artigo é um dos autores, e local para *upload* do arquivo em formato pdf ou ps.

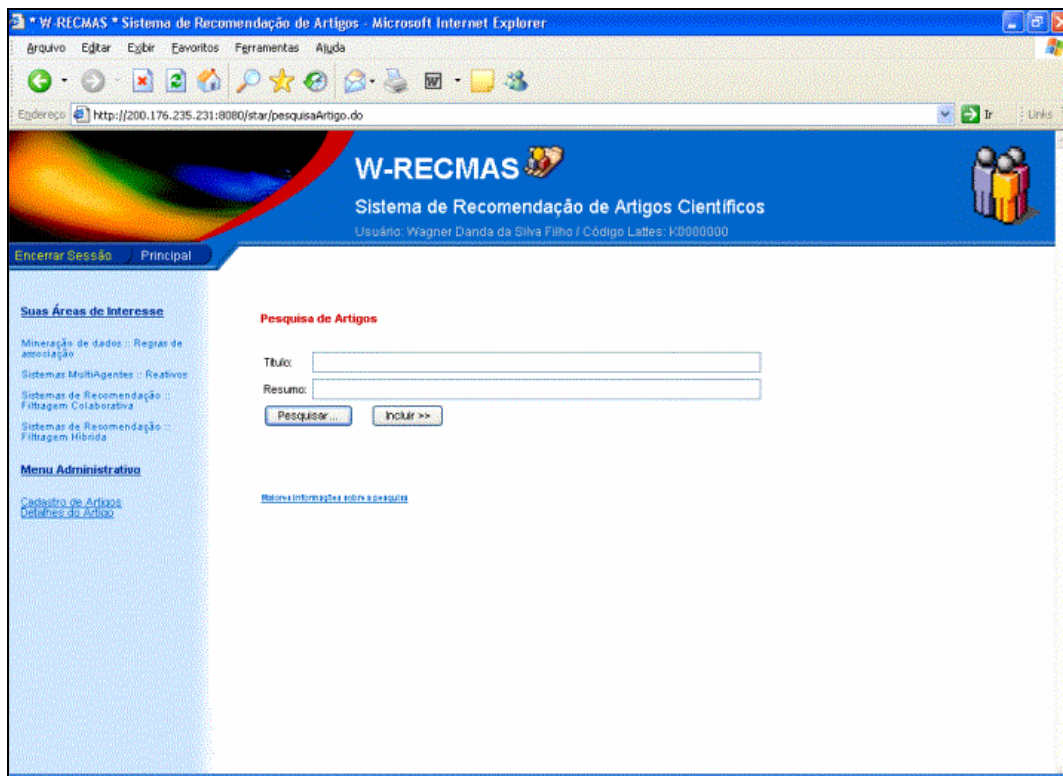


Figura 5.18: Interface de busca e posterior inclusão de artigos no protótipo W-RECMAS

A Figura 5.19 apresenta a interface de recomendação de artigos do protótipo. Nesta interface pode-se observar que o usuário está recebendo uma recomendação baseada em um grupo de recomendadores com relevância de opinião média (entre 4-6) na área de interesse da qual o usuário possui interesse. Pode-se observar que o sistema apresenta a predição calculada de opinião do usuário pelo artigo, sendo que o sistema “acredita” que o usuário irá considerar este artigo Bom (4 estrelas).

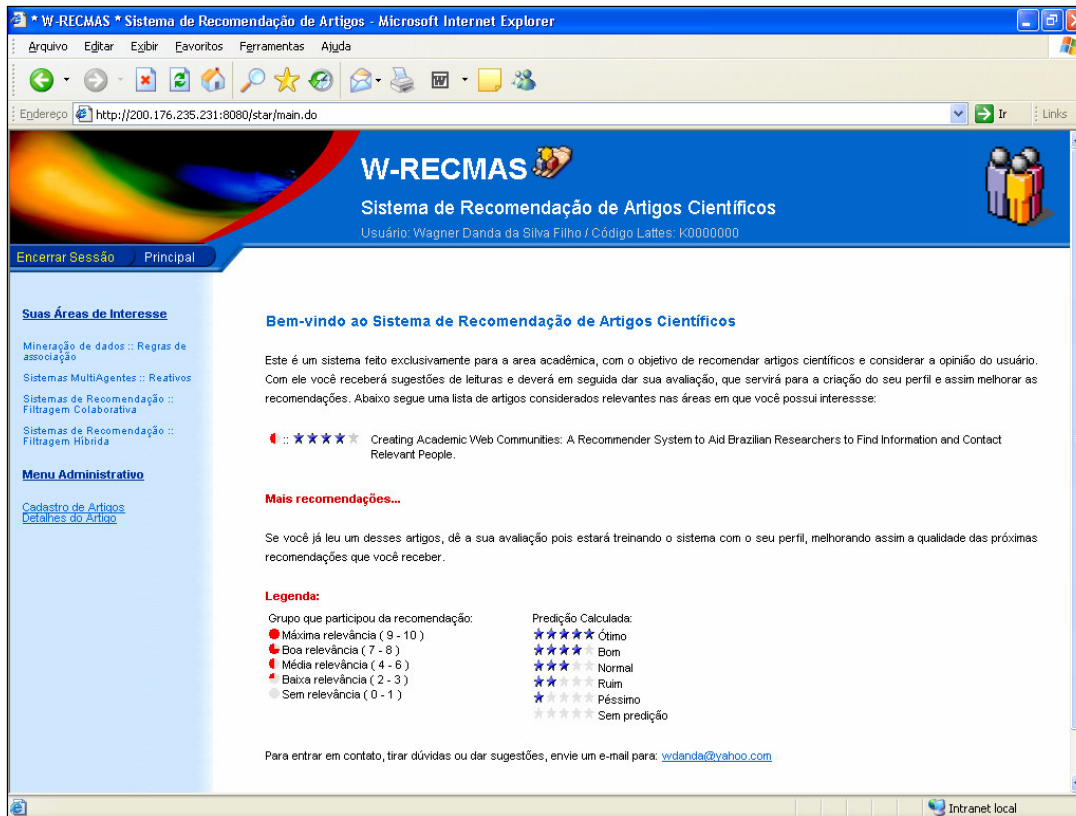


Figura 5.19: Interface de recomendação de artigos no protótipo W-RECMAS

A Figura 5.20 apresenta as legendas utilizadas na interface de recomendação. Em (a) pode-se observar a classificação da relevância da recomendação (Ranqueamento do Recomendador) do grupo que ajudou na formação da recomendação. Em (b) observa-se a classificação das possíveis predições de opinião do usuário alvo da recomendação quanto ao artigo recomendado.

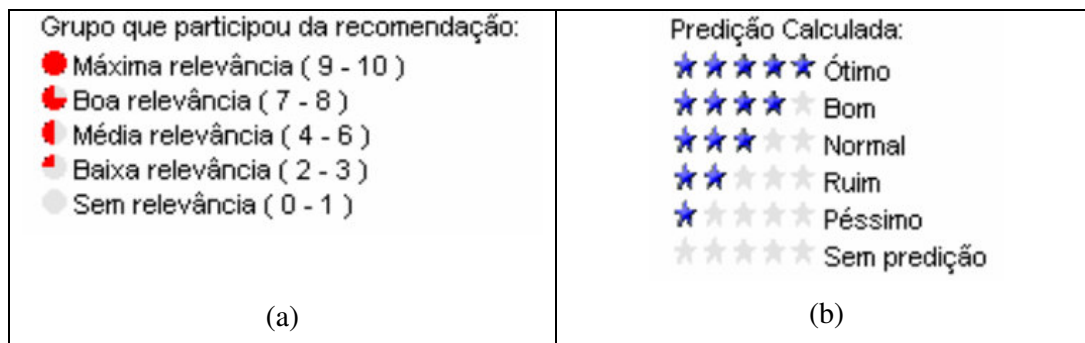


Figura 5.20: (a) Legenda da relevância dos recomendadores; (b) Legenda da predição calculada para recomendação





Figura 5.21: Interface de feedback da recomendação no protótipo W-RECMAS

A Figura 5.21 apresenta a interface para fornecimento do *feedback* do usuário quanto ao artigo recomendado. Observa-se que a interface é igual a interface apresentada em 5.19, isto deve-se a funcionalidade de clicar sobre as estrelas e assim fornecer ao sistema o *feedback* desejado. Observa-se que o sistema havia predito que este usuário acharia o artigo Bom, porém o *feedback* demonstrou uma opinião de indiferença pelo artigo, este usuário o considerou normal, ou seja, nem bom e nem ruim.

As próximas interfaces apresentadas referem-se ao módulo denominado comunidade virtual do usuário. Serão descritas algumas interfaces que permitem que os usuários - que aceitaram que o Agente Comunitário procurasse por pessoas de interesse do usuário em questão – interajam com estes usuários.

Na figura 5.22 observa-se que o usuário usu0119 foi reconhecido pelo sistema como tendo um perfil semelhante ao do usu0706 na área de interesse de redes neurais. Para manter a privacidade dos usuários, nesta tela são exibidos somente o nome e a relevância de opinião do contato dentro da área. Os demais dados somente poderão ser visualizados se ambos os usuários aceitarem se conhecer (ambos devem responder “Sim” na caixa de seleção situada abaixo do ponto de interrogação). Por exemplo, caso o usuário usu0119 tenha interesse em conhecer o usuário usu0706, este responderá “Sim”, o mesmo se aplicará para o segundo usuário. Havendo concordância de ambas as partes, o contato do novo usuário irá compor a comunidade de contatos de interesse



destes usuários na área em que seus perfis são semelhantes (neste caso a área seria redes neurais).



Figura 5.22: Interface de Recomendação de Usuários

Na figura 5.23 é apresentada a interface de Mensagens, a qual permite que o usuário possa ler as mensagens destinadas a ele enviadas por outros contatos.



Figura 5.23: Interface de Mensagens

O usuário registrado no sistema poderá somente enviar mensagens para os contatos que estão cadastrados na sua comunidade. Clicando no ícone “✉” ao lado da fotografia de um contato, a mensagem é enviada em particular para este contato. Este mesmo ícone é também visualizado no lado direito do nome da área de interesse, desta forma ao clicá-lo o usuário estará enviando a mensagem para todos os contatos cadastrados dentro desta área em sua comunidade. Já o ícone localizado no cabeçalho do quadro permite enviar mensagens para todos os contatos, independente da área cadastrada.

Conforme pode-se observar na figura 5.23 o usuário usu0054 recebeu duas mensagens. A primeira foi enviada pela usuária usu0797 com o título “Encontro de doutores”, e a segunda enviada pelo usuário usu0706 solicitando a referência de um artigo. Pode-se também verificar que ambos usuários remetentes estão cadastrados na comunidade dos contatos de interesse do usuário usu0054, ou seja, o Agente Comunitário contactou os Agentes Pessoais destes usuários, e estes consultaram os seus proprietários sobre o interesse comum dos usuários em entrarem em contato. Deve-se destacar que esta descoberta de pessoas que teriam interesse em entrar em contato foi feita sem a ajuda direta dos usuários, ou seja, sem esforço de busca dos mesmos. A figura a seguir apresenta a interface do perfil do usuário em comunidades (figura 5.24).



Figura 5.24: Interface do Perfil e Comunidade

## 5.4 Limitações e Dificuldades

Durante o período de elaboração deste protótipo algumas dificuldades e limitações foram percebidas:

a) Após o desenvolvimento do Agente Buscador (parte integrante da solução apresentada por esta tese), ocorreu uma modificação na forma de acesso ao Servidor de CV-Lattes. A mudança implementada pelo CNPQ barrou o acesso e consulta do Agente Buscador (um *crawler*) na forma como foi projetada, pois foi incluída uma imagem que deveria ser digitada para ser feito o acesso aos dados. Desta forma foi incluído no trabalho mais uma informação necessária para a correta busca – a chave de acesso do currículo no CV-Lattes, ou seja, o código de acesso passou a ser informado pelo usuário;

c) O uso do CV-Lattes como fonte de dados mostrou-se limitado por alguns motivos como: a base de dados é alimentada pelo próprio acadêmico e não sofre uma verificação prévia para averiguar tanto a veracidade das informações como a qualidade das mesmas. Na tentativa de avaliar os artigos publicados por cada acadêmico esbarrou-se na limitação da qualidade do dado, pois muitos estavam incorretamente digitados, por exemplo, SBIA ocorreu como SBA e SAI em alguns casos. Estes erros de alimentação causaram perdas de pontuação por alguns acadêmicos. Seria necessário investir esforços na construção de um dicionário de sinônimos para auxiliar na correta classificação;

### RESUMO DO CAPÍTULO:

Neste capítulo apresentou-se uma descrição detalhada do protótipo W-RECMAS. O maior detalhamento foi dado a descrição dos agentes de software que participam da abordagem multiagente adotada nesta tese. Foram apresentados e descritos quanto a suas funcionalidades os agentes Buscador, Analista, Pessoal, Recomendador e Comunitário. O conteúdo deste Capítulo foi publicado em [CAZ 2003], [CAZ 2004], [CAZ 2005a], [SIL 2005], [BRE 2005] e [CAZ 2006], apresentando progressão de detalhes conforme o andamento da tese.

## 6 EXPERIMENTOS E RESULTADOS

Neste Capítulo são apresentados os experimentos realizados nesta tese e que tiveram os seguintes objetivos:

- 1) verificar quais indicadores de produção (bem como a importância de cada um) do CV-Lattes a comunidade acadêmica considera como sendo relevantes para definir a relevância da opinião do usuário pesquisador;
- 2) verificar o uso do modelo Mo-DROP em uma aplicação real;
- 3) verificar qual a diferença da aplicação ou não da relevância da opinião do usuário em Sistema de Recomendação de acordo com a percepção da comunidade acadêmica;
- 4) verificar se o agente de software fazendo uso do modelo Mo-DROP (Capítulo 3), conseguiria igualar-se ou aproximar-se da avaliação humana de relevância de opinião;
- 5) verificar a aplicabilidade do *framework* F-UIP em uma situação real;
- 6) verificar junto a pesquisadores se as recomendações feitas por usuários detentores de RR maiores, recairia em maior satisfação dos recomendados.

Desta forma, foram conduzidos seis experimentos para verificar os objetivos listados acima, sendo alguns realizados de maneira *offline* (sem a participação de acadêmicos), e outros tiveram participação de acadêmicos. Os experimentos que envolveram acadêmicos (amostra de acadêmicos da área da Ciência da Computação e Sistemas de Informação) foram realizados no período de julho de 2005 até junho de 2006.

### 6.1 Descrição dos experimentos realizados

Alguns dos experimentos realizados foram conduzidos com a participação direta (consulta) a um grupo de acadêmicos ou base de dados reais.

#### 6.1.1 Primeiro Experimento: Ponderação de atributos do CV-Lattes

O primeiro experimento teve como objetivo verificar junto a uma amostra de acadêmicos quais indicadores de produção do CV-Lattes deveriam ser levados em consideração na aplicação do Mo-DROP para o cálculo da relevância da opinião dos usuários e qual peso cada indicador deveria assumir.

Para a realização deste experimento havia a necessidade da participação de profissionais que tivessem um conhecimento maior da proposta e constituição do CV-

Lattes, desta forma procurou-se trabalhar com uma amostra de doutores em ciência da computação.

Aos acadêmicos (amostra) foi solicitado selecionar e pontuar os indicadores de produção do CV-Lattes que consideravam importantes para definir a relevância da opinião de um usuário que possui CV-Lattes (o formulário preenchido pelos acadêmicos respondentes encontra-se no Anexo B).

#### 6.1.1.1 Método de pesquisa

A forma de abordagem adotada para a realização do experimento foi a de enviar um e-mail pessoal solicitando ao acadêmico a análise de currículos e retorno das considerações pelo respondente. Os acadêmicos receberam um e-mail particular e neste foi explicado em detalhes a proposta desta tese e do experimento, e os acadêmicos foram solicitados a informar alguns dados pessoais como: área principal de conhecimento na qual atuavam e a existência de atuação prévia como avaliador de CV-Lattes. No texto de explicação do experimento foi afirmado que os respondentes não seriam identificados no relato final desta tese, preservando desta forma a privacidade dos mesmos.

Ao acadêmico respondente, também, foi solicitada permissão do uso de seu CV-Lattes na realização do segundo experimento descrito em detalhes na Seção 6.1.2.

#### 6.1.1.2 Perfil da amostra

A amostra utilizada no primeiro e segundo experimento constituiu-se em uma amostra não probabilística, mais precisamente uma amostra por conveniência, ou seja, amostra onde o pesquisador seleciona membros mais acessíveis da população.

Por tratar-se de uma amostragem dita não probabilística deve-se ressaltar que os resultados obtidos não podem ser generalizados e constituindo-se, desta forma, um indicativo do entendimento da amostra selecionada, sendo passível de nova pesquisa para aprofundamento e generalização dos resultados apresentados.

A amostra selecionada constituiu-se de um total de 30 doutores da área de ciência da computação, os quais foram solicitados a selecionar e ponderar os indicadores de produção do CV-Lattes que consideravam importantes para a definição da relevância da opinião de um pesquisador. Neste experimento os acadêmicos não foram solicitados a justificar suas respostas e nem foi alvo de pesquisa as regras particulares de avaliação dos mesmos, pois o foco do experimento foi buscar as ponderações de importância de cada indicador de produção selecionado.

Como respondentes obteve-se um total de 25 acadêmicos dos 30 convidados a participar, ou seja, 83% dos convidados retornaram suas avaliações. Destes 25 acadêmicos respondentes, 15 já haviam participado de processos em que tiveram de analisar o conteúdo dos CV-Lattes de outros acadêmicos, o que demonstra que os mesmos já apresentavam conhecimento em análise de CV-Lattes (figura 6.1).

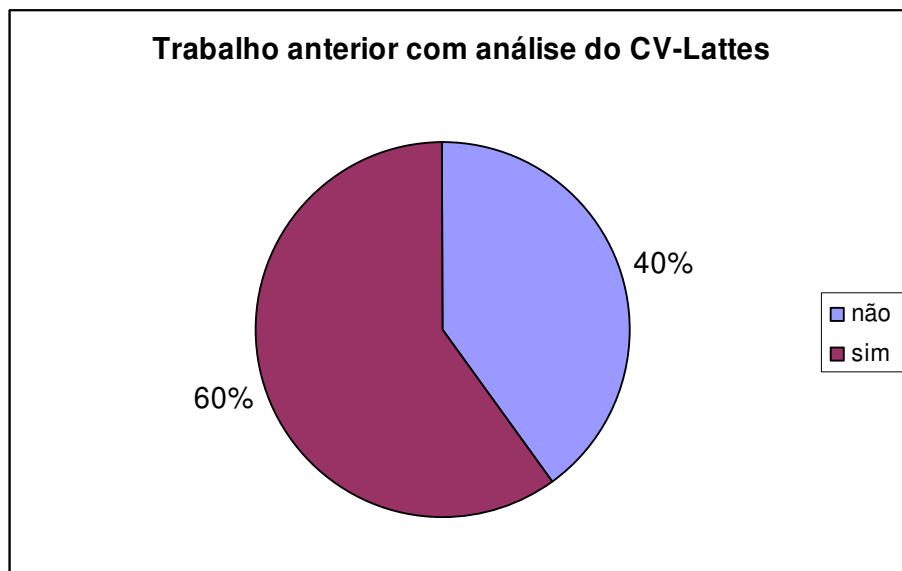


Figura 6.1: Acadêmicos e atuação com análise prévia de CV-Lattes

A figura 6.2 apresenta a distribuição dos respondentes quanto ao tempo de doutoramento em anos. Destes 25 acadêmicos a maioria apresenta doutoramento de no máximo cinco anos (48% dos respondentes).

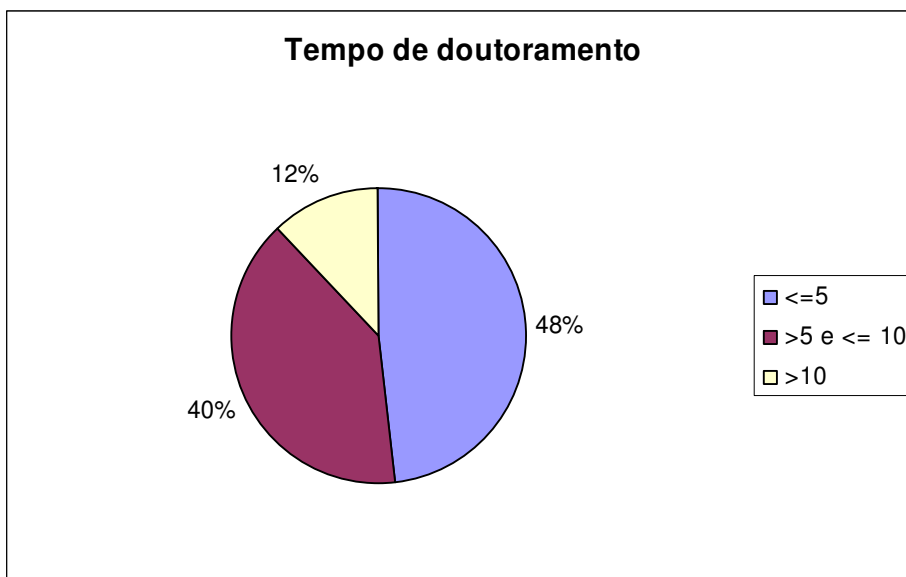


Figura 6.2: Acadêmicos e tempo de doutoramento

A figura 6.3 apresenta a distribuição dos respondentes quanto as suas principais áreas de atuação dentro da ciência da computação. Verifica-se pela figura que a maior

parcela dos respondentes concentra seus esforços de pesquisa na área de Inteligência Artificial.

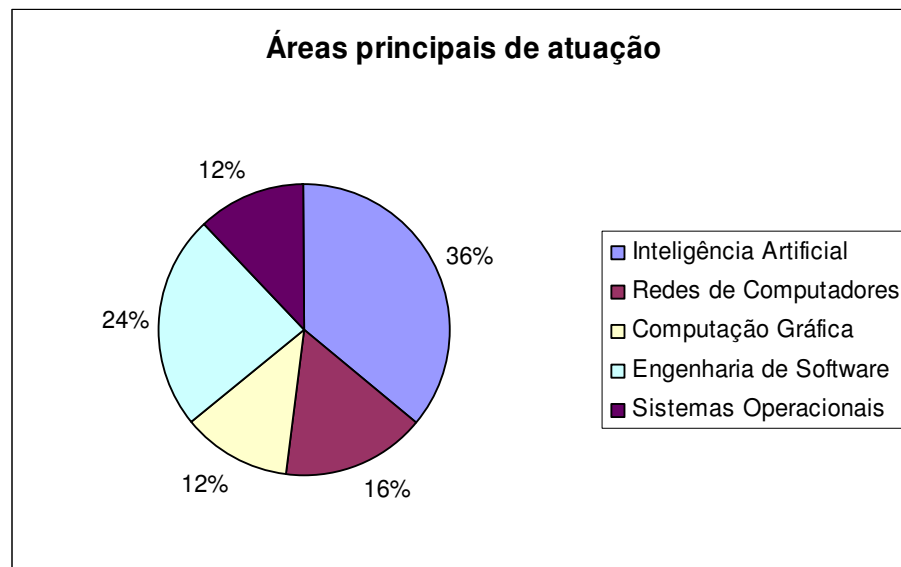


Figura 6.3: Acadêmicos e áreas de atuação

#### 6.1.1.3 Resultados

Conforme as respostas recebidas dos acadêmicos pode-se observar que o número mínimo de indicadores do CV-Lattes ressaltados como importantes para a definição da relevância da opinião de um usuário foram oito (08) e o máximo foram vinte e um (21), sendo que a média foi de dezesseis (16) indicadores.

A tabela 6.1 apresenta por categorias os indicadores de produção disponíveis no CV-Lattes. Nesta tabela é apresentado o percentual de respondentes da amostra que consideraram o indicador em questão relevante (% Consid. Ind. Rel.), bem como a média das respostas de importância do indicador (% Imp. do Ind.) e o percentual total de relevância da categoria (% Imp. da Cat.). Por exemplo, na categoria Produção Bibliográfica, o indicador de produção “Artigos publicados em periódicos” foi considerado como relevante por 100% dos avaliadores, e sua importância dentro da categoria ficou sendo de 36% (neste caso foi considerado o indicador mais importante se comparado com os demais indicadores apresentados nesta categoria). A importância desta categoria, em relação ao total, ficou em 46%.

De acordo com as repostas apresentadas pode-se aprimorar as ponderações dos indicadores utilizados na aplicação do modelo Mo-DROP (Capítulo 3) utilizando como fonte de dados o CV-Lattes.

Tabela 6.1: Relevância dos Indicadores de Produção

Categoria: Produção Bibliográfica			% Imp.da Cat.
Indicadores de Produção	% Consid. Ind. Rel.	% Imp.do Ind.	46
1) Artigos publicados em periódicos	100	36	
2) Trabalhos em anais de eventos	100	28	
3) Livros ou Capítulos de livros	100	29	
4) Textos em jornais ou revistas	48	4	
5) Demais tipos de produção bibliográfica	40	3	
TOTAL:		100	
Categoria: Produção Técnica			% Imp.da Cat.
Indicadores de Produção	% Consid. Ind. Rel.	% Imp.do Ind.	15
1) Software	95	17	
2) Produtos tecnológicos	85	18	
3) Trabalhos técnicos	93	20	
4) Demais tipos de produção técnica (organização de eventos)	93	22	
5) Demais tipos de produção técnica (relatórios de pesquisa)	93	14	
6) Demais tipos de produção técnica (apresentações de trabalhos)	85	9	
TOTAL:		100	
Categoria: Orientação Concluída			% Imp.da Cat.
Indicadores de Produção	% Consid. Ind. Rel.	% Imp.do Ind.	29
1) Tese doutorado	100	45	
2) Dissertação de mestrado	100	26	
3) Trabalho de conclusão	96	12	
4) Especialização /Aperfeiçoamento	93	8	
5) Iniciação científica	93	9	
TOTAL:		100	
Categoria: Informações Complementares			% Imp.da Cat.
Indicadores de Produção	% Consid. Ind. Rel.	% Imp.do Ind.	08
1) Participações em banca de trabalhos de conclusão	85	28	
2) Participações em eventos	78	19	
3) Participações em banca de comissões julgadoras	93	37	
4) Orientações em andamento	93	16	
TOTAL:		100	
Categoria: Demais Ttrabalhos Relevantes			% Imp.da Cat.
Indicadores de Produção	% Consid. Ind. Rel.	% Imp.do Ind.	02
Demais trabalhos relevantes	19	100	
TOTAL:		100	
TOTAL Geral:			100



#### 6.1.1.4 Comentários dos Respondentes

A seguir são apresentados alguns comentários emitidos pelos respondentes e considerados relevantes no experimento realizado:

1) “.....os indicadores do CV-Lattes são extremamente vagos, uma vez que são quantitativos e não levam em conta o perfil do pesquisador em questão. Existem pesquisadores que atuam em áreas mais teóricas, então publicações em periódicos Qualis A são visadas. Já outros pesquisadores atuam em áreas aplicadas que geram produtos tecnológicos e então nestes casos outros parâmetros devem ser levados em conta para avaliar o pesquisador. O erro está em querer avaliar todos utilizando os mesmos critérios e os mesmos pesos.....”

2) “.....além do tipo de produção é importante observar o Qualis.....”

3) “.....artigos publicados em periódicos de relevância nacional e artigos publicados em periódicos de relevância internacional devem ser considerados de forma distinta.....”

A primeira observação descreve um cuidado que deve ser feito na interpretação das competências e autoridade dos pesquisadores, uma vez que o CV-Lattes reflete somente o perfil de pesquisador (acadêmico), deixando descoberto o perfil dos pesquisadores que trabalham na pesquisa mais aplicada junto à indústria. As observações apresentadas foram relevantes, e as observações 2 e 3 reforçaram o correto uso do Qualis e distinção entre os tipos de publicações de cada acadêmico no Segundo Experimento (Seção 6.1.2).

### 6.1.2 Segundo Experimento: Aplicando o modelo Mo-DROP

Objetivando apresentar uma aplicação prática do modelo proposto (Mo-DROP) este foi aplicado no protótipo W-RECMAS. Nesta instanciação do modelo foram utilizados dados provenientes de CV-Lattes de pesquisadores (pesquisadores que autorizaram o uso dos mesmos no experimento 1), bem como os resultados obtidos no experimento anteriormente descrito.

#### 6.1.2.1 Domínio e Atributos

O domínio do experimento que está sendo descrito é referente a acadêmicos da área de ciência da computação, e neste experimento existem diversas áreas de interesse para os usuários do sistema. Objetivando representar a relevância da opinião do usuário são descritos na Tabela 6.2 os atributos quantitativos.

Os atributos da tabela 6.2 foram extraídos do CV-Lattes como texto. Cada atributo selecionado para o experimento foi decomposto e analisado. Todas as áreas de interesse e publicações foram extraídas e cada área foi separadamente analisada. Os atributos constantes na tabela 6.2 foram selecionados com base no experimento 1, e o peso de cada atributo foi definido com base na avaliação dos pesquisadores que participaram do experimento (exemplo, o atributo APP foi avaliado no experimento 1 com um peso individual de 36% na sua categoria, que recebeu um peso de categoria igual a 46%, desta forma foi calculada a relação deste atributo e todos os demais avaliados, ou seja,  $\text{Peso APP} = 36 * 46 / 100 = 16,56$ ).

Tabela 6.2: Atributos quantitativos do CV-Lattes e seus pesos

Atributos	Acrônimo	%Peso	Valores
Artigos publicados em periódicos	APP	16,56	Este valor é medido em APP unidades.
Trabalhos em anais de eventos	TAE	12,98	Este valor é medido em TAE unidades.
Livros ou Capítulos de livro	LCL	13,34	Este valor é medido em LCL unidades.
Textos em jornais ou revistas	TJR	1,84	Este valor é medido em TJR unidades.
Demais tipos de produção bibliográfica	DTP	1,38	Este valor é medido em DTP unidades.
Software	SW	2,55	Este valor é medido em SW unidades.
Produtos tecnológicos	PT	2,70	Este valor é medido em PT unidades.
Trabalhos técnicos	TT	4,00	Este valor é medido em TT unidades.
Demais tipos de produção técnica (organização de eventos)	OE	3,30	Este valor é medido em OE unidades.
Demais tipos de produção técnica (relatórios de pesquisa)	RP	2,10	Este valor é medido em RP unidades.
Demais tipos de produção técnica (apresentação de trabalhos)	AT	1,35	Este valor é medido em AT unidades.
Tese de doutorado	TD	13,05	Este valor é medido em TD unidades.
Dissertação de mestrado	DM	7,54	Este valor é medido em DM unidades.
Trabalho de conclusão	TC	3,48	Este valor é medido em TC unidades.
Especialização/Aperfeiçoamento	EA	2,32	Este valor é medido em EA unidades.
Iniciação científica	IC	2,61	Este valor é medido em IC unidades.
Participação em banca de trabalho de conclusão	PBTC	2,24	Este valor é medido em PBTC unidades.
Participação em eventos	PE	1,52	Este valor é medido em PE unidades.

Atributos	Acrônimo	%Peso	Valores
Participação em banca de comissões julgadoras	PBCJ	2,96	Este valor é medido em PBCJ unidades.
Orientações em andamento	OA	1,28	Este valor é medido em OA unidades.
Demais trabalhos relevantes	DTR	0,90	Este valor é medido em DTR unidades.

Todas as métricas quantitativas (APP, TAE, LCL, TJR, DTP, SW, PT, TT, OE, RP, AT, TD, DM, TC, EA, IC, PBTC, PE, PBCJ, OA e DTR) foram normalizadas para uma escala específica de valores. A Equação (6) do Capítulo 3, foi aplicada no processo de normalização dos valores, e o valor máximo definido para cada atributo foi obtido através de uma amostragem de 50 Currículos Vitae Lattes de pesquisadores da ciência da computação provenientes do sistema CV-Lattes do CNPQ. Estes currículos foram selecionados por conveniência, utilizando nomes de pesquisadores da UFRGS (Universidade Federal do Rio Grande do Sul), UNISINOS (Universidade do vale do Rio dos Sinos), PUC-RS (Pontifícia Universidade Católica do Rio Grande do Sul), UNB (Universidade de Brasília), UCS (Universidade de Caxias do Sul), UFPA (Universidade Federal do Pará) e UFBA (Universidade Federal da Bahia).

O valor mínimo definido para cada atributo foi igual a zero. Todos os atributos analisados apresentavam diferentes escala de valores (por exemplo, de acordo com a amostragem de currículos o número máximo de trabalhos publicados em eventos pelos pesquisadores selecionados foi igual a 100, mas o número de livros e Capítulos publicados foi igual a 25, ou seja, valores com escalas muito distintas).

Quando foi aplicada a Equação (7) do Capítulo 3, para calcular o Ranqueamento do Recomendador, os pesos atribuídos aos atributos foram exatamente os listados na tabela 6.2. Uma importante consideração deve ser feita: no experimento que está sendo apresentado existe distinção em importância entre publicações Internacionais e Nacionais. Esta distinção baseia-se na listagem de importância de publicações, denominada Qualis (descrita no Capítulo 5). Esta lista foi utilizada no experimento aqui descrito, apesar da mesma ter criado bastante controvérsia no mundo acadêmico. A adoção da mesma nesta aplicação foi feita por comodidade e não por esta ter sido superior em qualidade a outras existentes.

Devido a disponibilidade desta lista, trabalhou-se com valorações distintas para as publicações de cada pesquisador no nível internacional e nacional/local conforme descrito na tabela 6.3.

Tabela 6.3: Peso das categorias de publicações

Categorias Qualis/Atributos	A	B	C/D
Internacionais			
APP	3,0	2,0	1,0

Categorias Qualis/Atributos	A	B	C/D
Internacionais			
TAE	3,0	2,0	1,0
Nacionais/Locais			
APP	2,0	1,5	1,0
TAE	2,0	1,5	1,0

Para tornar a explicação anterior um pouco mais clara, apresenta-se um exemplo prático: um pesquisador X possui as seguintes publicações em seu CV-Lattes nos últimos 3 anos: 2 publicações internacionais na categoria A, 2 internacionais na categoria B e 1 internacional na categoria C, sendo assim o atributo  $APP = 2*3.0 + 2*2.0 + 1*1.0 = 11$ .

Aplicou-se uma distinção entre orientações concluídas (OC = TD, DM e TC) e orientações em andamento (OA). Nestes atributos foi realizada uma distinção entre alunos de graduação, e pós-graduação. Foi definida uma classificação com pesos específicos:

- Estudante de Doutorado = 5,0;
- Estudante de Mestrado = 3,0;
- Estudante de Graduação = 1,0.

Por exemplo, para um pesquisador com 15 orientações: 5 de doutorado, 5 de mestrado e 5 de graduação,  $OC = 5*5.0 + 5*3.0 + 5*1.0 = 45$ .

#### 6.1.2.2 Aplicando o Mo-DROP

A tabela 6.4 apresenta as métricas de produção (atributos) e valores respectivos extraídos do CV-Lattes de dez pesquisadores, entretanto, os atributos não se encontram normalizados.

Na linha 1 da tabela 6.4 é apresentado um usuário que é doutor em ciência da computação; apresenta duas publicações em periódicos (publicações nacionais e internacionais), vinte trabalhos publicados em eventos (publicações nacionais e internacionais) e um livro publicado; vinte e dois estudantes orientados, e sete estudantes orientados atualmente. A pergunta é — Qual a relevância da opinião deste usuário para a área de interesse do mesmo dentro de um sistema de recomendação que aplique o modelo Mo-DROP?

A tabela 6.5 apresenta todos os atributos normalizados após a aplicação da Equação (6). Pode-se observar que todos os atributos foram colocados na escala desejada [0-10].

Tabela 6.4: Atributos não normalizados para o cálculo do RR

Usuário	Atributos																				
	APP	TAE	LCL	TJR	DTP	SW	PT	TT	OE	RP	AT	TD	DM	TC	EA	IC	PBTC	PE	PBCJ	OA	DTR
U1	2	20	1	1	5	1	0	4	0	0	14	0	0	22	3	0	27	34	7	7	0
U2	3	38	6	2	1	2	0	0	0	0	7	0	5	6	4	11	19	8	37	2	0
U3	2	18	3	0	0	1	0	0	0	0	2	0	0	10	0	0	0	3	0	1	6
U4	0	2	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	2	0	0	1
U5	13	86	8	1	2	7	1	38	0	0	36	0	7	20	0	13	46	13	4	7	0
U6	3	37	9	0	0	0	0	0	0	0	0	1	21	8	0	0	44	0	0	0	0
U7	0	2	0	0	0	1	0	0	0	0	0	0	0	3	0	0	1	3	0	0	0
U8	0	36	0	0	0	0	0	0	0	1	0	0	0	10	1	3	0	1	3	0	4
U9	9	48	14	0	0	0	0	42	0	0	33	0	3	11	0	16	37	0	5	7	6
U10	12	175	19	0	3	5	0	10	0	0	7	11	27	0	0	0	0	6	1	12	7

Tabela 6.5: Atributos normalizados para o cálculo do RR

Usuário	Atributos																				
	APP	TAE	LCL	TJR	DTP	SW	PT	TT	OE	RP	AT	TD	DM	TC	EA	IC	PBTC	PE	PBCJ	OA	DTR
U1	2,1	2,2	0,8	9,5	10,0	3,2	3,3	3,3	3,3	3,3	6,6	3,2	1,6	10,0	10,0	1,5	7,6	10,0	5,6	9,2	0,8
U2	3,3	3,8	4,9	10,0	4,7	5,5	3,3	2,0	3,3	3,3	3,9	3,2	4,3	2,9	10,0	10,0	5,4	5,5	10,0	3,0	0,8
U3	2,1	1,9	2,5	2,0	1,6	3,2	3,3	2,0	3,3	3,3	2,0	3,2	1,6	5,7	2,1	1,5	0,3	3,0	2,3	1,8	10,0
U4	1,5	0,9	0,8	2,0	1,6	3,2	3,3	2,0	10,0	3,3	1,2	3,2	1,6	0,0	2,1	1,5	0,3	2,4	2,3	0,6	2,5
U5	10,0	9,9	6,5	9,5	7,7	10,0	10,0	10,0	3,3	3,3	10,0	3,2	5,4	10,0	2,1	10,0	10,0	8,1	4,2	9,2	0,8
U6	2,7	3,2	7,4	2,0	1,6	0,9	3,3	2,0	3,3	3,3	1,2	4,7	10,0	4,3	2,1	1,5	10,0	1,4	2,3	0,6	0,8
U7	1,5	0,9	0,0	2,0	1,6	3,2	3,3	2,0	3,3	3,3	1,2	3,2	1,6	0,8	2,1	1,5	0,6	3,0	2,3	0,6	0,8
U8	1,5	4,0	0,0	2,0	1,6	0,9	3,3	2,0	3,3	10,0	1,2	3,2	1,6	5,7	5,7	3,9	0,3	1,9	3,7	0,6	7,8
U9	7,3	5,1	10,0	2,0	1,6	0,9	3,3	10,0	3,3	3,3	10,0	3,2	3,2	6,4	2,1	10,0	10,0	1,4	4,7	9,2	10,0
U10	10,0	10,0	10,0	2,0	10,0	10,0	3,3	5,2	3,3	3,3	3,9	10,0	10,0	0,0	2,1	1,5	0,3	4,5	2,8	10,0	10,0

Finalmente, foi calculado o  $RR_{area}$ , aplicando a Equação (7) do Capítulo 3 para todos os atributos selecionados e usando os respectivos pesos. A tabela 6.6 apresenta o valor calculado para cada usuário.

Tabela 6.6: Resultados do Ranqueamento do Recomendador

Usuário	$RR_{area}$
U1	3,8
U2	5,0
U3	2,5
U4	2,3
U5	8,4
U6	4,5
U7	1,7
U8	2,8
U9	6,5
U10	10,0

A interpretação final dos valores obtidos demonstra que a opinião do usuário U10 é mais relevante entre os usuários interessados nesta área de pesquisa. Neste exemplo, foi apresentado o cálculo do  $RR_{area}$  por usuário em uma área específica de interesse (ex, inteligência artificial), mas o mesmo cálculo pode ser realizado para cada área de interesse do usuário.

Os resultados do experimento sugerem que o modelo é flexível e pode ser facilmente aplicado para domínios distintos, uma vez que os atributos e seus pesos podem ser trocados e parametrizados.

O benefício esperado pela aplicação deste modelo é que o mesmo possa auxiliar na representação da relevância da opinião do usuário que participa de uma recomendação, e desta forma espera-se estar oferecendo uma informação que venha a facilitar a escolha do usuário alvo sobre o “consumo” do item recomendado. Buscando verificar se a apresentação do Ranqueamento do Recomendador junto ao item recomendado faria diferença para o usuário alvo, o terceiro experimento (Subseção 6.1.3) foi realizado.

### 6.1.3 Terceiro Experimento: Interfaces de Recomendação

O terceiro experimento teve como objetivo verificar junto a uma amostra de acadêmicos de Ciência da Computação e profissionais de Sistemas de Informação qual a diferença percebida ao apresentar-se uma interface de um sistema de recomendação com a inclusão do valor do Ranqueamento do Recomendador e outra interface sem o valor do Ranqueamento do Recomendador.

Os acadêmicos selecionados (amostra) foram solicitados a avaliarem duas interfaces de sistemas de recomendação com suas devidas recomendações e responderem a um questionário (conforme apresentado no Anexo C).

### 6.1.3.1 Método de pesquisa

Os participantes da amostra receberam um e-mail pessoal, no qual foram descritos alguns termos importantes no contexto do experimento que se pretendia realizar, tais como: Sistemas de Recomendação, Sistemas de Recomendação Colaborativos e Relevância da Opinião do Usuário.

O participante foi solicitado a inserir-se em um cenário hipotético e responder três questões objetivas após a visualização das alternativas de interface do sistema de recomendação (apresentadas na figura 6.4 e na figura 6.5).

Na Alternativa 1 (figura 6.4) o usuário visualizava somente a predição do sistema de recomendação ao artigo recomendado, ou seja, o quanto o sistema predizia que o usuário alvo gostaria da recomendação.



The screenshot shows the W-RECMAS web interface in Microsoft Internet Explorer. The page title is "W-RECMAS - Sistema de Recomendação de Artigos Científicos". Below the header, there is a welcome message and a brief description of the system. A table lists recommended articles with their predicted ratings and download links.

Artigos Recomendados	Predição [1;5]	Download	Avaliação (feedback) [1;5]
An Economic Model of User Rating in an Online Recommender System	4.5		1
Improving Recommendation Lists Through Topic Diversification	4.0		1
Influence in Ratings-Based Recommender Systems: An Algorithm-Independent Approach	4.0		1
Combining Data Mining Technique and Users' Relevance Opinion to Build an Efficient Recommender System	3.5		1
Analysis of Recommendation Algorithms for E-Commerce	3.2		1

Figura 6.4: Interface 1 do terceiro experimento

Na Alternativa 2 (figura 6.5) o usuário visualizava o Ranqueamento do Recomendador que auxiliou na recomendação, ou seja, o quão importante era a opinião de quem auxiliou na recomendação dentro de um sistema colaborativo.





Figura 6.5: Interface 2 do terceiro experimento

### 6.1.3.2 Hipótese de pesquisa

A hipótese que foi levantada para a realização deste experimento é a de que a inclusão do RR dos recomendadores na interface do sistema de recomendação propiciaria uma melhor aceitação das recomendações realizadas que envolvessem especialistas nas áreas de interesse.

### 6.1.3.3 Perfil da amostra

A amostra utilizada no quarto experimento constituiu-se em uma amostra não probabilística de um total de 70 acadêmicos, entre acadêmicos da área de ciência da computação e da área de sistemas de informação. Neste experimento os participantes foram solicitados a apresentarem sugestões sobre possíveis áreas de aplicação da solução proposta.

Como respondentes obteve-se um total de 62 acadêmicos. Destes 62 respondentes 27 são doutores, 10 doutorandos, 6 mestres, 10 graduandos e 9 graduados conforme apresentado na figura 6.6.

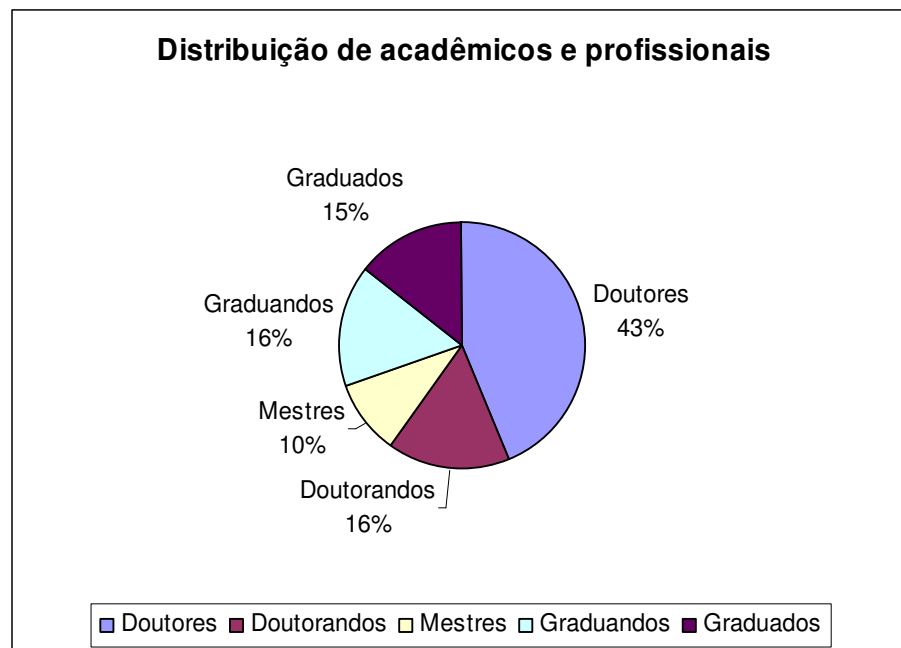


Figura 6.6: Distribuição dos acadêmicos e profissionais do terceiro experimento

#### 6.1.3.4 Resultados

Os respondentes tendo em mente o cenário hipotético apresentado no experimento e abaixo descrito:

*“...supondo que você seja um pesquisador iniciando estudos na Área de Conhecimento referente a Sistemas de Recomendação, ou seja, você iniciou a busca por material relevante e confiável para que possa obter conhecimento desta área. Você poderia reportar-se a um sistema de recuperação de informação (exemplo, google) e começar a busca de forma autônoma, o que seria exaustivo pois você não possui conhecimento de que material seria relevante para ler. Visando auxiliar seu início de busca lhe foi ofertado um Sistema de Recomendação, o qual apresenta duas alternativas de apresentação de seu conteúdo (Alternativa 1 e Alternativa 2)”* .

Responderam as seguintes questões:

Q1) A inclusão do Ranqueamento do Recomendador neste contexto seria um fator que facilitaria a seleção de itens recomendados ao usuário?

Observou-se que 61 acadêmicos de um total de 62 responderam que sim a esta questão, ou seja, 98,4% dos respondentes.

Q2) Você ao verificar a recomendação com a inclusão da relevância da opinião do usuário sentiu-se mais seguro ao decidir que item ler / acessar primeiro devido ao sistema mostrar que este indivíduo tem maior ou menor autoridade na área?

Observou-se que 60 acadêmicos de um total de 62 responderam que sim a esta questão, ou seja, 96,7% dos respondentes.

Q3) Em um contexto de recomendação de itens este fator (Ranqueamento do Recomendador) poderia ser um direcionamento para seleção de itens, diminuindo a sensação de sobrecarga de informação comum em sistemas de recuperação de informações (exemplo, google)?

Observou-se que 61 acadêmicos de um total de 62 responderam que sim a esta questão, ou seja, 98,4% dos respondentes.

A tabela 6.7 sumariza estes valores obtidos como resultados.

Tabela 6.7: Respostas do terceiro experimento

Questões	Respostas		
	% Sim	% Não	% Total
Q1	98,4	1,6	100
Q2	96,7	3,3	100
Q3	98,4	1,6	100

Com os resultados obtidos foi possível confirmar a hipótese apresentada no item 6.1.3.2. O retorno que os respondentes enviaram foram levados em consideração na elaboração do experimento final descrito na Seção 6.1.6.

#### 6.1.3.5 Comentários dos Respondentes

Segundo os respondentes as possíveis aplicações do Ranqueamento do Recomendador (solução proposta e apresentada na figura 6.5) seriam:

- 1) Alunos de pós-graduação iniciando em uma área de pesquisa;
- 2) Orientadores em busca de novas áreas;
- 3) Qualquer situação que demande escolha de uma área de conhecimento onde o pesquisador não possua familiaridade;
- 4) Sistemas onde a avaliação técnica é fundamental (exemplo, avaliação de câmeras digitais);
- 5) Sistemas de comunidades virtuais e fóruns (exemplo, verificar a relevância da opinião de quem emite avaliação ou idéia sobre algo);

Problemas relevantes quanto à aplicação do Ranqueamento do Recomendador segundo os respondentes:

- 1) “...como garantir que os rankeados são realmente especialistas...”
- 2) “...uma lista de 1.000 artigos listados pelo sistema, contra uma lista de 10.000 artigos no google talvez não fossem vistas como coisas diferentes pelo usuário, nos dois casos existirá sobrecarga de informação...”
- 3) “...como garantir que o especialista não terá ganhos financeiros se avaliar melhor um item ou invés de outro?”

4) “...na minha opinião a opinião de um recomendador com notória experiência seria importante em áreas técnicas, nas quais não possuo conhecimento. Em situações de recomendação cultural ou artística, preferiria recomendações feitas por pessoas c/ gostos semelhantes aos meus...”

5) “...quem possui muita experiência pode avaliar bem artigos mais aprofundados e mal os básicos. Como ficaria a recomendação para os iniciantes neste caso?”

#### **6.1.4 Quarto Experimento: Avaliação humana x agente de software**

O quarto experimento teve como objetivo verificar junto a uma amostra de acadêmicos qual a relevância da opinião de um grupo de acadêmicos (que concordaram em disponibilizar seus CV-Lattes para análise) em algumas áreas de interesse.

Os acadêmicos selecionados (amostra) foram solicitados a avaliarem dois CV-Lattes de outros acadêmicos. A avaliação constituiu-se em ponderar a relevância da opinião do acadêmico que estava sendo avaliado para uma área de interesse onde este seria especialista e em uma área em que este supostamente estaria interessado a ingressar (não especialista). O enunciado completo do experimento encontra-se disponível no Anexo D.

Neste experimento havia a necessidade da participação de profissionais que tivessem um conhecimento maior da proposta e constituição do CV-Lattes, desta forma, procurou-se trabalhar com uma amostra de doutores em ciência da computação.

##### **6.1.4.1 Método de pesquisa**

A forma de abordagem adotada para a realização do experimento foi a mesma descrita anteriormente no primeiro experimento (Subseção 6.1.1.1). Neste experimento cada acadêmico recebeu dois CV-Lattes (autorizados previamente pelos seus responsáveis para uso neste experimento) para avaliar. Esta avaliação foi realizada considerando duas áreas de interesse para cada um dos currículos: uma das áreas consistia da área de especialização do currículo do pesquisador e outra em uma área hipotética de interesse. Foram avaliados 13 currículos.

No texto de explicação do experimento foi afirmado que os respondentes não seriam identificados no relato final da tese, preservando desta forma a privacidade dos mesmos.

##### **6.1.4.2 Hipótese de pesquisa**

A hipótese que foi levantada para realização deste experimento é a de que o agente computacional (Agente Analista) que faz uso do modelo Mo-Drop (Capítulo 5) poderia chegar a valores próximos dos fornecidos pela amostra que avaliou os CV-Lattes de alguns acadêmicos, e utilizando a escala de avaliação do RR seriam gerados valores entre [0-4,9] para áreas em que os acadêmicos não são especialistas e valores entre [5,0-10] para as áreas em que os acadêmicos são considerados especialistas.

##### **6.1.4.3 Perfil da amostra**

A amostra utilizada neste experimento foi a mesma do primeiro experimento (Seção 6.1.1.2), sendo que como respondentes obteve-se um total de 26 acadêmicos dos 30 convidados a participar.

#### 6.1.4.4 Resultados

Na tabela 6.8 são apresentadas as avaliações realizadas pelos acadêmicos (Avaliação Humana) a alguns CV-Lattes (Currículo dos acadêmicos). As avaliações foram feitas para uma área em que o acadêmico avaliado era especialista e uma área na qual o acadêmico avaliado supostamente estaria disposto a iniciar pesquisa, ou seja, não especialista. Na tabela 6.8 são apresentadas as avaliações realizadas pelo agente computacional (Avaliação do agente) e as avaliações realizadas pelos humanos (Currículo dos acadêmicos) aos mesmos currículos, fazendo aplicação do Mo-DROP instanciando o CV-Lattes.

Tabela 6.8: Resultados das avaliações

Currículo dos acadêmicos avaliados	Avaliação humana		Avaliação do agente		Diferença entre avaliação humana e avaliação do agente	
	Especialista (AHE)	Não Especialista (AHNE)	Especialista (AAE)	Não Especialista (AANE)	Especialista (DAE)	Não Especialista (DANE)
AC1	6,0	3,0	5,5	2,0	0,5	1,0
AC2	10,0	3,0	9,5	2,8	0,5	0,2
AC3	8,0	5,0	8,0	3,9	0,0	1,1
AC4	7,0	3,0	4,0	2,0	3,0	1,0
AC5	9,0	3,0	8,8	2,1	0,2	0,9
AC6	10,0	5,0	9,3	2,5	0,7	2,5
AC7	10,0	6,0	9,5	5,3	0,5	0,7
AC8	7,0	3,0	2,4	2,0	4,6	1,0
AC9	9,0	3,0	6,1	2,5	2,9	0,5
AC10	10,0	4,0	9,3	2,5	0,7	1,5
AC11	8,0	3,0	5,0	2,6	3,0	0,4
AC12	8,0	3,0	2,7	2,5	5,3	0,5
AC13	8,0	3,0	5,2	2,0	2,8	1,0
<b>Média:</b>	<b>8,5</b>	<b>3,6</b>	<b>6,6</b>	<b>2,7</b>	<b>1,9</b>	<b>0,9</b>
<b>Desvio Padrão:</b>	<b>1,3</b>	<b>1,0</b>	<b>2,6</b>	<b>0,9</b>	<b>1,8</b>	<b>0,6</b>

Observando-se a diferença entre a avaliação humana e a avaliação realizada pelo agente de software verifica-se que a média das diferenças das avaliações foi, respectivamente, 1,9 para currículos na área em que o pesquisador avaliado era

especialista e 0,9 para currículos na área em que o pesquisador avaliado não era especialista. Interpretando estes valores observa-se que quando se trata de uma área de não especialização do acadêmico, a avaliação humana e a avaliação do agente de software apresentaram-se muito próximas conforme disposto na figura 6.7.

Na figura 6.7 pode-se observar que os valores referentes a avaliação humana de um currículo acadêmico na área em que o avaliado não é especialista (AHNE) e a avaliação de um agente de software de um currículo de um acadêmico na área em que o avaliado não é especialista (AANE) foram muito similares.

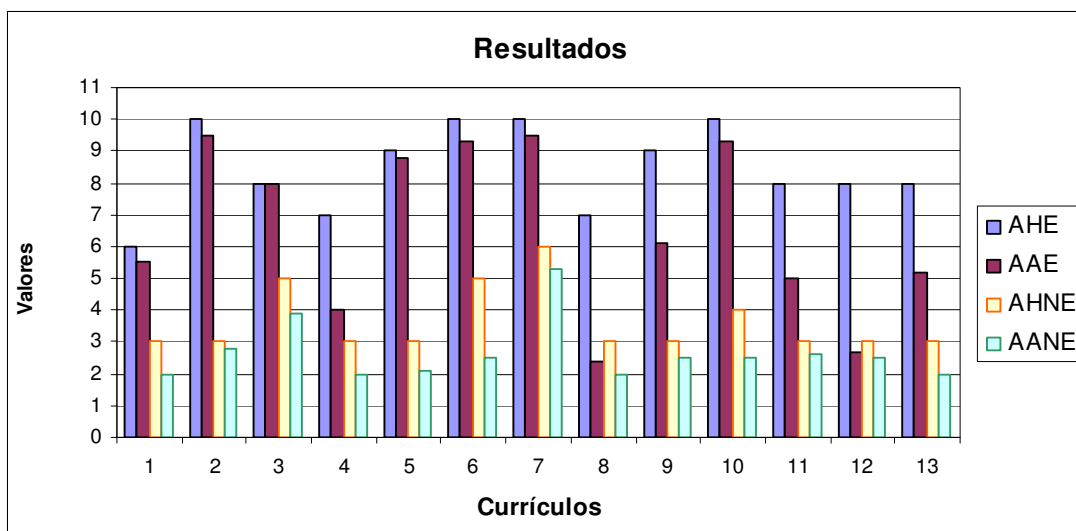


Figura 6.7: Avaliação humana x avaliação agente software

Da mesma forma é possível observar que os valores referentes a avaliação humana de um currículo na área em que o avaliado é especialista (AHE) e a avaliação de um agente de software de um currículo na área em que o avaliado é especialista (AAE) apresentaram alguns pontos de maior diferenciação.

Buscando avaliar a hipótese descrita para este experimento e assumindo-se que valores de  $RR \geq 5,0$  e  $\leq 10$  deveriam ser gerados para as áreas em que os currículos dos acadêmicos avaliados são especialistas e valores de  $RR \geq 0$  e  $< 5$  deveriam ser gerados para as áreas em que os currículos de acadêmicos avaliados não são especialistas, foi realizada a seguinte verificação:

a) analisou-se o percentual de acadêmicos, na área em que o acadêmico avaliado é especialista, onde o valor do RR ficou na faixa almejada: dos 13 currículos analisados, um total de 10 ficaram na faixa esperada, sendo assim 77% dos currículos ficaram dentro do esperado, e 23% ficaram fora do esperado.

b) analisou-se o percentual de acadêmicos, na área em que o acadêmico avaliado não é especialista, onde o valor do RR ficou na faixa almejada: dos 13 currículos analisados, um total de 12 ficaram na faixa esperada, sendo assim 92% ficaram na faixa esperada, e 8% ficaram fora do esperado.

Com os resultados obtidos, evidencia-se a comprovação parcial da hipótese formulada (Subseção 6.1.4.2) devido a limitações (Subseção 6.1.4.5) nos dados fornecidos aos agentes computacionais.

#### 6.1.4.5 Limitações

Analisando os resultados procurou-se explicar o porquê das distorções de valores que ocorreram em alguns dos resultados. Por exemplo, o AC12 (acadêmico 12) obteve somente 2,7 para uma área em que este é notadamente especialista. Como explicar este resultado? Analisando-se os dados do CV-Lattes fornecido ao agente, observou-se que este valor foi gerado devido a vários veículos em que este acadêmico publicou não terem sido encontrados no Qualis pelo agente, o que ocasionou uma baixa na sua pontuação.

Estes veículos não foram encontrados por apresentarem erros na alimentação dos nomes no próprio CV-Lattes do acadêmico. Observou-se que os valores de RR obtidos fora da faixa esperada, foram consequência, na maioria dos casos, de falha na alimentação do currículo pelo acadêmico, o que ocasionou perda de dados que ajudariam na avaliação dos currículos pelo agente.

### 6.1.5 Quinto Experimento: Aplicando o *framework* F-UIP

Neste experimento apresenta-se uma aplicação prática do *framework* F-UIP proposto no Capítulo 4.

#### 6.1.5.1 Base de dados

O *framework* proposto foi testado utilizando-se uma base de dados sintética. Para a realização deste experimento foram utilizados 80 CV-Lattes de pesquisadores brasileiros da área de ciência da computação. Estes currículos foram utilizados para gerar a base de dados sintética.

A amostra de currículos foi coletada por conveniência, direto do site do CNPQ no período de agosto a setembro de 2004, sendo formada por 4 currículos de graduandos, 6 currículos de graduados, 2 currículos de especialistas, 16 currículos de mestrandos, 19 currículos de mestres, 10 currículos de doutorandos, 15 currículos de doutores e 8 currículos de pós-doutores.

A base sintética gerada inclui 10.000 registros (com representação proporcional de cada nível de formação acadêmica selecionada e descrita no parágrafo anterior). Cada registro da base de dados corresponde a um pesquisador e suas respectivas áreas de interesse relacionadas à ciência da computação, incluindo seus respectivos  $RR_{\text{area}}$ . A figura 6.8 apresenta a distribuição final por formação na base de dados gerada. Esta base de dados sintética era composta por cinco campos caracterizando as áreas de interesse dos pesquisadores.

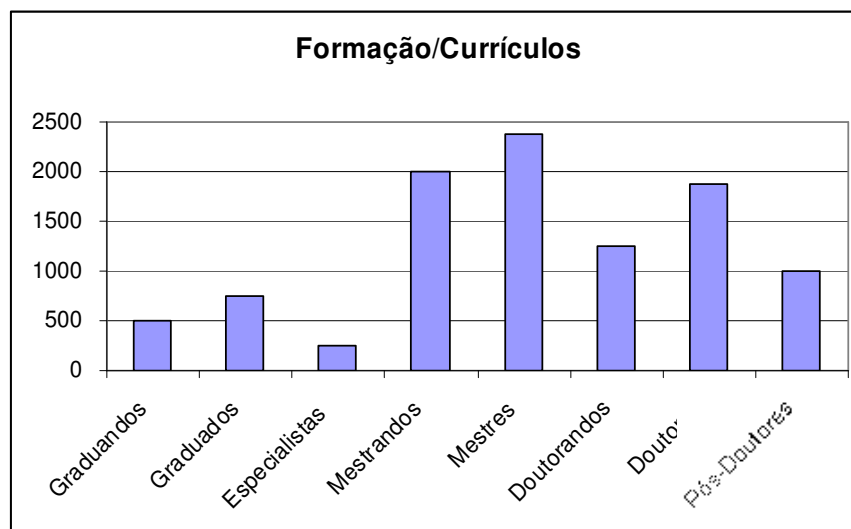


Figura 6.8: Distribuição das formações na base de dados sintética

A figura 6.9 apresenta estas áreas de interesse e os respectivos números de usuários interessados e o número de usuários que não explicitaram interesse em cada uma das áreas: A1 com 9.100 interessados e 900 que não explicitaram interesse; A2 com 6.532 interessados e 3.468 que não explicitaram interesse; A3 com 1.701 interessados e 8.299 que não explicitaram interesse; A4 com 6.938 interessados e 3.062 que não explicitaram interesse; A5 com 3.001 interessados e 6.999 que não explicitaram interesse.

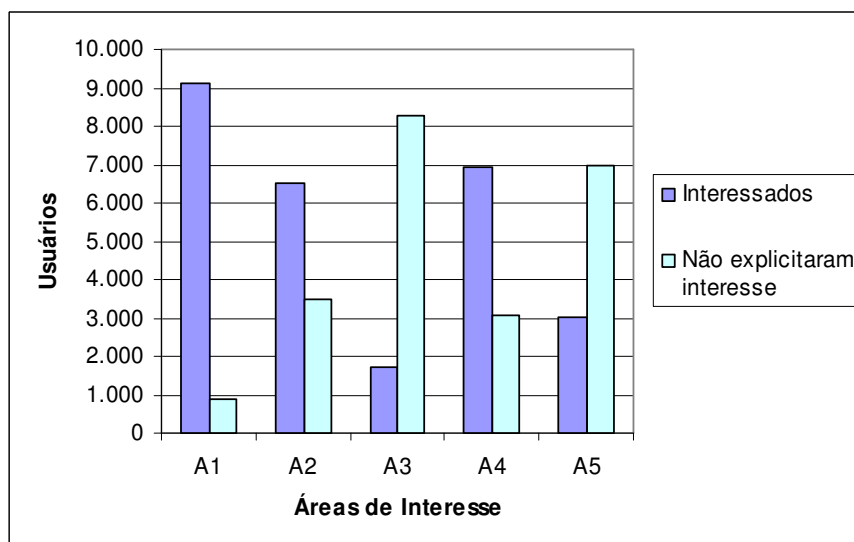


Figura 6.9: Áreas de interesse e usuários

A base de dados sintética apresentava a seguinte distribuição de  $RR_{\text{totais}}$  (valores entre 0 até 10, representando a relevância da opinião dos usuários): 0 registros entre [0,0-2,9]; 1.151 registros entre [3,0-4,9]; 654 registros entre [5,0-6,9]; 5.611 registros



entre [7,0-8,9] e 2.584 registros entre [9,0-10,0], conforme figura 6.10. Deve ser observado que estes valores representam a média aritmética entre todos os  $RR_{\text{áreas}}$  de um usuário, porém ao aplicar-se o *framework* na fase II o  $RR_{\text{total}}$  dos usuários foi calculado para as áreas envolvidas nas regras a serem analisadas.

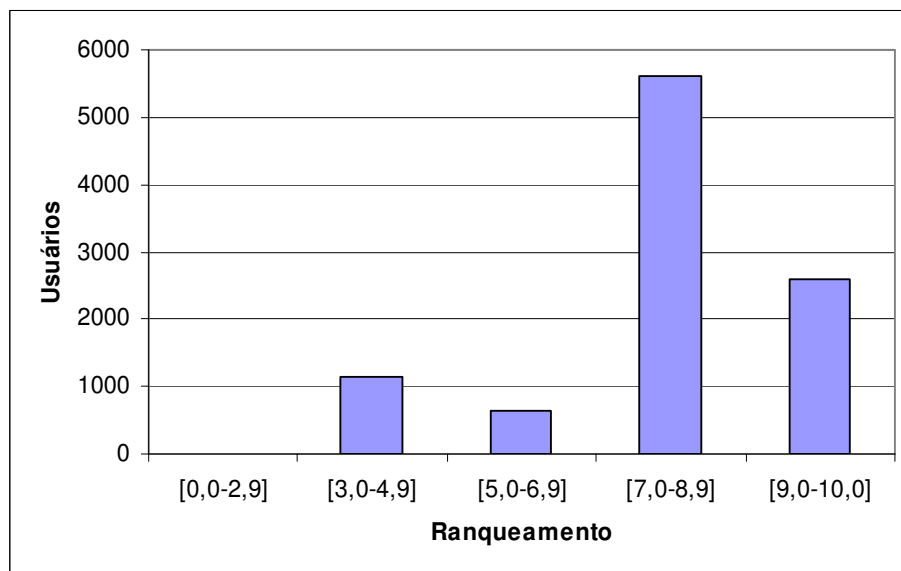


Figura 6.10: Distribuição dos valores do Ranqueamento do Recomendador

Neste experimento utilizou-se o algoritmo Apriori [AGR 94] (versão implementada na ferramenta de mineração Weka [WIT 2000]), parametrizada com fator de suporte mínimo (*lowerBoundMinSupport*) igual a 0,1 e máximo igual a 1 (*upperBoundMinSupport*) e mínima confiança igual a 0,8 ; sendo solicitada a geração de um total de 10 regras. A Tabela 6.10 apresenta um total de cinco regras positivas obtidas através da aplicação da mineração e selecionadas para exemplificar o *framework*.

Por exemplo, pode-se analisar a regra R4 da tabela 6.9: esta regra apresenta um valor de confiança igual a 1,0, sendo selecionada na Fase I; esta por sua vez foi selecionada na Fase II, devido a média entre o  $RR_{\text{totais}}$  dos usuários (1.701 usuários) que casam com o LHS e RHS da regra ser igual a 7,3 (este valor representa que a opinião destes usuários são relevantes para as áreas participantes na regra em questão); esta regra foi selecionada para a Fase III, sendo encontrado um total de 103 usuários que casaram com o LHS e casaram com a negação do RHS desta regra (usuários que seriam alvo de sua aplicação). Desta forma esta regra foi selecionada, porém caso não fossem encontrados usuários que casassem com a regra, esta por ter sido considerada relevante seria preservada em base de dados para uma futura aplicação (caso ocorrido com a regra R5).

As regras R1, R2 e R3 apresentam um valor de confiança adequado, porém as médias entre os  $RR_{\text{totais}}$  dos usuários que casam com o LHS e RHS foi menor do que 7,0 (indicando que a relevância de opinião da comunidade em questão não é muito significativa dentro da parametrização aplicada); por esta razão as regras não foram selecionadas para Fase III.

Tabela 6.9: Regras Geradas

Regras	Confiança	LHS/RHS	Média dos RRs	LHS/¬RHS	Selecionada?	
R1	If A2=1 then A1=1	0,98	6431	6,7	101	N
R2	If A2=1 and A4=1 then A1=1	0,97	3369	6,0	101	N
R3	If A4=1 then A1=1	0,87	6038	4,6	900	N
R4	If A1=1 and A3=1 then A2=1	1,00	1701	7,3	103	S
R5	If A5=1 then A1=1	1,00	3001	8,5	0	S

Como finalização do experimento realizado verifica-se a viabilidade da aplicação do *framework* na minimização do problema da superespecialização uma vez que com base na colaboração de usuários com alto valor de opinião nas áreas foi possível sugerir novas áreas aos usuários alvo. Esta minimização surge com a aplicação das novas regras selecionadas com o auxílio das informações constantes no banco de dados e as informações referentes à opinião da comunidade.

Certamente ainda existe a possibilidade de inadequação na recomendação feita ao usuário alvo, pois o mesmo pode não aceitar ou não se contentar com a recomendação em uma área que ele não tenha pedido ou se inscrito previamente. Como uma solução a esta consideração existe o feedback do usuário em Sistemas de Recomendação, que se apresenta como uma informação de alta relevância neste caso, pois somente através deste feedback é possível verificar se o usuário ficou satisfeito ou não com as recomendações. Caso o usuário tenha demonstrado satisfação com a recomendação então se entende que a regra deve ser aplicada definitivamente no perfil deste usuário (diminuindo a superespecialização), caso contrário, esta regra não será aplicada no perfil do usuário.

Uma vez que a superespecialização se configura como um “engessamento” das recomendações feitas a um usuário, ao aplicar este Framework observa-se que novas possibilidades de recomendação irão surgir, e uma vez que o usuário deverá apresentar seu feedback sobre novas recomendações geradas a partir da regra, será possível verificar seu interesse ou não, e desta forma o seu perfil irá evoluir no sistema e não permanecerá como algo estático.

#### 6.1.6 Sexto Experimento: Recomendando itens

O objetivo deste sexto experimento foi o de verificar se as recomendações geradas por indivíduos com maior valor de Ranqueamento do Recomendador resultariam em valores

de avaliações mais altos (*feedback*) por parte dos pesquisadores que receberam as recomendações.

#### *6.1.6.1 Hipótese da Pesquisa*

A hipótese que foi levantada para realização deste experimento é a de que a média do valor das avaliações (*feedback*) dos recomendados as recomendações geradas com base em recomendadores com maior nível de Ranqueamento do Recomendador apresentaria maior valor se comparada com a média do valor das avaliações das recomendações geradas com base em recomendadores com menor nível de Ranqueamento do Recomendador.

#### *6.1.6.2 Método de Pesquisa*

Para a realização deste experimento foi selecionada uma amostra de pesquisadores doutores da área de ciência da computação. Estes pesquisadores concentram suas pesquisas em áreas de interesse como Engenharia de Software, Computação Gráfica, Redes de Computadores e Inteligência Artificial. Para cada pesquisador foi calculado o Ranqueamento do Recomendador (RR) com base no CV-Lattes do pesquisador na sua área de interesse.

O experimento foi dividido em três fases descritas a seguir, e documentadas no Anexo E desta tese:

#### **Fase I: Coleta de Artigos**

Visando popular uma base de artigos científicos no sistema W-RECMAS foi solicitado aos pesquisadores por e-mail que fornecessem os artigos que consideravam melhor representar a sua área de pesquisa e que os mesmos consideravam interessantes e recomendáveis a seus estudantes / colegas.

Cada pesquisador recomendou um mínimo de 05 artigos científicos relacionados a sua área de interesse, não ocorrendo duplicidade de conteúdo, e poucos pesquisadores recomendaram artigos de sua própria autoria (um total de 2% dos artigos constituiu-se em conteúdo produzido pelo próprio pesquisador colaborador). Com o auxílio destes pesquisadores populou-se a base de artigos do W-RECMAS com um total de 141 artigos científicos, obtendo-se uma média de 35 artigos científicos por área de interesse.

Procurando evitar a possível geração de uma base de dados sem itens avaliados (problema da esparsialidade descrito na subseção 2.6.6) considerou-se cada um dos artigos inseridos pelos pesquisadores como já avaliados com um valor de *rating* igual a 5 (em uma escala de 1-5), ou seja, para evitar uma base muito esparsa, assumiu-se que o artigo recomendado deveria ser guardado com uma pontuação igual a 5 para o pesquisador que o recomendou nesta fase. Para o sistema estes artigos são entendidos como itens já consumidos pelos pesquisadores que os incluíram (histórico de consumo).

### **Fase II: Coleta de Avaliações Preliminares**

Nesta fase os pesquisadores receberam algumas recomendações iniciais de itens (artigos científicos), para permitir a coleta de avaliações dos pesquisadores viabilizando, desta forma, a realização do processo de filtragem de informação colaborativa.

Cada pesquisador recebeu um total de 05 artigos científicos recomendados, a seleção destes artigos baseou-se em filtragem por conteúdo, ou seja, os artigos inseridos e considerados de relevância por cada pesquisador foram utilizados como histórico de consumo para viabilizar a filtragem. Cada artigo foi analisado pelas palavras-chave, e os primeiros cinco que fecharam com o perfil e histórico de consumo dos pesquisadores foram os recomendados. No total foram feitas 135 recomendações, alguns itens foram recomendados para mais de um pesquisador visando viabilizar a terceira fase do experimento.

Uma vez recebida via e-mail a recomendação destes 5 artigos científicos, os pesquisadores foram convidados e orientados na avaliação a ser realizada sobre a recomendação feita. Os pesquisadores retornaram o e-mail com a avaliação dos artigos utilizando uma escala de 1-5. As avaliações retornadas foram inseridas na base de dados e permitiram a realização da terceira fase do experimento.

### **Fase III: Recomendações**

Nesta fase os pesquisadores receberam algumas recomendações geradas pela aplicação da filtragem colaborativa, utilizando a equação (9) apresentada na Seção 3.3.

Cada pesquisador recebeu um total de 02 artigos científicos recomendados, a seleção destes artigos baseou-se em filtragem colaborativa aplicando-se o Ranqueamento do Recomendador neste cálculo. Deve ser destacado que um pesquisador recebeu somente uma recomendação e isto ocorreu pelo fato de seu modo de avaliar os itens ser muito diferente dos demais pesquisadores (fugiu ao padrão do grupo recaindo no problema da similaridade descrito na Subseção 2.6.2). No total foram feitas 53 recomendações, alguns itens foram recomendados para mais de um pesquisador.

Uma vez recebida via e-mail a recomendação destes artigos científicos, os pesquisadores foram convidados e orientados na avaliação a ser realizada sobre a recomendação feita. No e-mail enviado propositalmente não foi indicado o nível de relevância de opinião (Ranqueamento do Recomendador) dos colaboradores na recomendação. Os pesquisadores retornaram um e-mail com a avaliação aos artigos utilizando uma escala de 1-5.

#### *6.1.6.3 Perfil da amostra*

A amostra utilizada neste experimento constituiu-se em uma amostra não probabilística de um total de 30 doutores da área de ciência da computação.

Os pesquisadores selecionados concentram suas pesquisas em áreas de interesse como Engenharia de Software, Computação Gráfica, Redes de Computadores e Inteligência Artificial. A amostra selecionada encontra-se distribuída nestas áreas de interesse da seguinte maneira: 07 pesquisadores trabalham na área de Inteligência Artificial, 07 na área de Engenharia de Software, 07 na área de Redes de Computadores e 06 na área de Computação Gráfica.

Como respondentes obteve-se um total de 27 acadêmicos dos 30 convidados a participar, ou seja, 90% dos convidados aceitaram participar das três fases do experimento.

#### 6.1.6.4 Base de dados utilizada

Para a realização do experimento foi construída uma base de dados real com artigos fornecidos por pesquisadores em suas devidas áreas de atuação.

A base de dados resultante apresentou a seguinte constituição:

1. um total de 141 artigos científicos recomendados por especialistas em suas áreas de interesse;
1. artigos científicos classificados em quatro áreas de interesse (inteligência artificial, computação gráfica, redes de computadores e engenharia de software);
2. uma média de 35 artigos científicos por área de interesse;
3. um total de 27 usuários participando do experimento;
4. 329 avaliações de 27 usuários para 141 artigos científicos;
  - a. cada usuário forneceu um mínimo de 5 avaliações na primeira fase;
  - b. cada usuário realizou exatas 5 avaliações na segunda fase;
  - c. cada usuário (com exceção de um pesquisador) realizou 2 avaliações na terceira fase.

#### 6.1.6.5 Resultados

Avaliando-se o *feedback* obtido na terceira fase do experimento (na qual foi utilizada a filtragem colaborativa) observou-se que a média das pontuações retornadas pelos pesquisadores às 53 recomendações geradas pela colaboração de pesquisadores com Ranqueamento do Recomendador elevado (28 recomendações com valor elevado, considerando-se RR elevado valores entre 7-10) ficou em 4,17 (em uma escala de 1-5) e a média das pontuações retornadas pelos pesquisadores às recomendações geradas pela colaboração de pesquisadores com Ranqueamento do Recomendador mais baixo (25 recomendações com valor baixo, considerando-se RR mais baixo os valores entre 0-6) ficou em 3,08 (em uma escala de 1-5). Desta forma evidencia-se a comprovação da hipótese de pesquisa apresentada na Subseção 6.1.6.1.

Algumas considerações devem ser feitas, uma vez que se realizou um experimento que envolvia usuários, constituindo-se um experimento em um ambiente real de recomendação:

- 1) os pesquisadores que participaram da amostra tiveram que realizar a avaliação de alguns artigos recomendados por filtragem baseada em conteúdo e outro recomendado por filtragem colaborativa, observou-se que em muitos casos os amostrados avaliaram superficialmente o artigo, seja por falta de tempo ou por desinteresse devido ao artigo não apresentar as palavras-chave exatamente dentro da área de pesquisa do pesquisador;
- 2) alguns pesquisadores mostraram uma abordagem diferente para verificar a relevância, centraram o foco em quem eram os autores, ou seja, se os autores não eram conhecidos na área o pesquisador já inferia que o artigo não seria interessante. Este comportamento foi observado em pelo menos três pesquisadores que estavam avaliando artigos recomendados em áreas que eles são especializados;
- 3) procurar avaliar o interesse de alguém por algo, apresentou-se uma tarefa difícil pois o processo de decisão e o aspecto cognitivo de cada pesquisador são extremamente particulares e em muitos casos não apresentam um padrão, mesmo tratando-se de pesquisadores especialistas na mesma área, com o mesmo nível de formação, formação realizada na mesma instituição (o que poderia auferir uma “cultura“ de análise de conteúdo muito semelhante);
- 4) em alguns casos, ao analisar-se as avaliações fornecidas pelos pesquisadores, verificou-se que em uma mesma amostra onde alguns pesquisadores apresentavam o mesmo perfil, e foram expostos aos mesmos itens, pelo menos um apresentava um comportamento de análise totalmente diverso dos demais, ou seja, todos os demais concordavam com o quão interessante e recomendável o item seria, com exceção de um pesquisador.

#### 6.1.6.6 Comentários dos Respondentes

A seguir são apresentados alguns comentários emitidos pelos respondentes e considerados relevantes no experimento realizado:

- 1) “.....os autores do artigo não são conhecidos na área, sendo assim pontuei com valores menores.....”
- 2) “.....o artigo não esta exatamente na minha área de pesquisa, sendo assim considero recomendação inadequada.....”
- 3) “.....pontuei bem alguns artigos, apesar de não estarem na minha área, por ter curiosidade.....”
- 4) “.....a recomendação me surpreendeu, sendo altamente relevante e adequada, estava falando com meus orientandos sobre este tópico e autor no dia que li a recomendação.....”

## 6.2 Conclusões

Os seis experimentos realizados foram de extrema importância para a finalização desta tese uma vez que permitiram:

- 1) verificar segundo uma amostra de acadêmicos quais os indicadores de produção em suas categorias deveriam ser aplicados com quais pesos no cálculo da relevância da opinião do usuário (Ranqueamento do Recomendador) ao utilizar-se o CV-Lattes. Este experimento auxiliou na correta instanciação do modelo Mo-DROP;
- 2) verificar a utilização do Mo-DROP em uma aplicação real;
- 3) verificar a aceitação da comunidade acadêmica (envolvida na amostra selecionada) e coletar idéias quanto a importância da inclusão do RR como informação extra para seleção de item a ser analisado;
- 4) verificar se com a aplicação real do modelo através do agente analista seria possível reproduzir a avaliação (pontuação) dos especialistas humanos quanto a relevância da opinião de um usuário. O resultado demonstrou que os valores produzidos com a aplicação do modelo chegou próximo aos valores da avaliação humana (principalmente nos casos em que os currículos não apresentavam erros de preenchimento de conteúdo);
- 5) verificar a aplicabilidade do *framework* F-UIP com dados provenientes de CV-Lattes;
- 6) verificar se recomendações produzidas por pesquisadores com maior relevância de opinião ocasionariam pontuação mais elevada na avaliação retornada pelos recomendados.

### 6.2.1 Limitações dos experimentos

Nestes experimentos realizados foram percebidas algumas limitações. Uma das limitações percebidas refere-se aos experimentos terem sido realizados com uma amostra não-probabilística o que torna os resultados interessantes, porém não generalizáveis.

O uso de uma base de dados textual como o CV-Lattes, com alimentação livre pelos acadêmicos (o que resulta em falta de padronização) constituiu-se em uma outra limitação, pois muitos dados estavam incorretos nos currículos avaliados, ou com falta de precisão (erros de preenchimento).

### RESUMO DO CAPÍTULO:

Neste capítulo apresentou-se uma descrição detalhada dos experimentos realizados nesta tese bem como os resultados obtidos e as limitações dos experimentos.

## 7 CONCLUSÕES

A busca pela informação que venha realmente a auxiliar o usuário na realização de suas tarefas constitui-se em uma atividade tipicamente solitária. Durante esta busca o usuário deve saber o que procura de forma a deixar claro aos mecanismos de busca (recuperação de informação) o que devem procurar.

Quando esta realidade refere-se a Internet verifica-se um aumento considerável na dificuldade de encontrar o que venha a satisfazer o usuário, pois existe uma grande quantidade de recursos e caminhos possíveis de serem percorridos. Qual destes caminhos é mais confiável? Qual destes caminhos resulta na informação realmente importante?

Buscando facilitar o processo de busca pela informação surgiram os Sistemas de Recomendação. O fato dos Sistemas de Recomendação ofertarem itens considerados potencialmente relevantes para um usuário faz com que o usuário não tenha de despender esforço procurando. Existem várias pesquisas focadas em Sistemas de Recomendação conforme apresentado na Seção 2.5 e no Anexo A desta tese. Nas pesquisas e nos *sites* comerciais descritos nesta tese, observou-se a não aplicação da relevância da opinião de um usuário no processo de recomendação e nem na apresentação do conteúdo recomendado ao usuário alvo, considerando-se esta preocupação um diferencial no trabalho desenvolvido.

Alguns domínios específicos apresentam uma peculiaridade para recomendação, intuitivamente parece ser relevante ter conhecimento do grau de maturidade ou autoridade do usuário que participou da recomendação do item. Um caso típico é a recomendação de artigos científicos para pesquisadores, foco principal deste trabalho de tese.

A presente tese foi iniciada partindo da observação desta peculiaridade de alguns domínios, onde a inclusão da relevância de opinião do usuário no processo de recomendação constituiria em um facilitador na recomendação. O que se apresentou nesta tese foi um estudo para modelar a relevância da opinião de um usuário em Sistemas de Recomendação e a utilização desta métrica para aprimorar a recomendação, pela oferta de itens que tenham sido avaliados por pessoas com maior relevância de opinião. Visando viabilizar a aplicação dos Sistemas de Recomendação colaborativos e a relevância da opinião dos usuários, foram utilizadas tecnologias como: mineração de dados, sistemas multiagentes, modelagem de perfil de usuários e comunidades virtuais. Através deste estudo foi oferecida uma arquitetura que permite que agentes de *software* auxiliem e facilitem na busca da informação desejada pelos usuários cadastrados, recomendando o que for mais indicado ao perfil dos mesmos.



Isto posto, cabe revisitar as contribuições da tese citadas na Seção 1.4, apresentando comentários que elucidam como os problemas foram tratados nos experimentos realizados e no texto.

- *Uma grande questão em Sistemas de Recomendação refere-se à definição da relevância da opinião de quem participa de um Sistema de Recomendação, uma vez que em filtragem colaborativa, a opinião de um usuário sobre dado item é fundamental.*

Nesta tese foi apresentado um modelo completo que viabilizou modelar e definir a relevância da opinião dos usuários dentro de um Sistema de Recomendação para pesquisadores. Este modelo mostrou-se flexível e foi instanciado para utilização na realidade brasileira, fazendo uso de informações provenientes de base de dados governamentais e públicas. Com o auxílio de pesquisadores foi possível definir os atributos que deveriam ser levados em consideração, bem como a importância de cada, para calcular a relevância da opinião de um pesquisador. Através dos resultados dos experimentos realizados evidenciou-se a importância da inclusão desta métrica de relevância da opinião no processo de recomendação, sendo que como resultado final verifica-se que a avaliação de itens recomendados por pesquisadores com maior relevância de opinião realmente causou uma maior satisfação nos recomendados.

- *Uma outra questão de relevância para a comunidade que pesquisa os Sistemas de Recomendação recai sobre a dificuldade do sistema criar o perfil atualizado dos usuários, uma vez que na grande maioria dos sistemas o usuário tende a possuir um perfil estático, ou seja, não progride com o passar do tempo, causando desinteresse dos usuários.*

Visando acompanhar a evolução do perfil do usuário houve a preocupação de propor um *framework* que permitisse identificar novos possíveis interesses dos usuários. A solução proposta faz uso de mineração de dados e da métrica de relevância da opinião dos demais membros da comunidade para identificar estes possíveis interesses. Com os resultados dos experimentos realizados verificou-se a aplicabilidade desta solução de forma a surpreender o usuário com recomendações inesperadas, procurando identificar possíveis mudanças de interesse do usuário. Desta forma, evidencia-se esta solução como uma possibilidade de evitar ou reduzir a superespecialização em Sistemas de Recomendação.

- *A apresentação e especificação de uma arquitetura completa de Sistema de Recomendação baseada em agentes de software, mineração de dados e na inclusão da relevância da opinião do recomendador no processo de recomendação constituiu-se em uma outra contribuição deste trabalho (Capítulo 5).*

Nesta tese foi apresentada uma arquitetura baseada em sistemas multiagentes e mineração de dados. A utilização de agentes visou permitir uma melhor decomposição de tarefas dentro do processo de recomendação com a inclusão da relevância da opinião do usuário. Estes agentes desempenham tarefas como calcular o RR dos usuários, identificar e realizar a recomendação, atualizar os dados externos ao sistema e permitir a integração da comunidade de pesquisadores. Através dos resultados dos experimentos realizados verificou-se a capacidade dos agentes desempenharem seus papéis e atuarem

como avaliadores de conteúdo para definir o RR dos usuários do sistema utilizando ponderações fornecidas por pesquisadores sobre o CV-Lattes.

- *Uma nova abordagem no cálculo da recomendação de um item ao usuário do sistema de recomendação. A inclusão da relevância da opinião do usuário que avalia um item no cálculo de predição de um item para um usuário alvo constitui-se em contribuição para a área de Sistemas de Recomendação (Capítulo 3).*

Nos experimentos realizados e nas recomendações feitas pelo sistema, foi aplicada a equação (9) apresentada na Seção 3.3. Nesta equação procurou-se incluir a relevância da opinião dos usuários como um valor que viesse a reforçar as avaliações de quem possuísse maior valor de RR. Através dos resultados dos experimentos realizados evidenciou-se a importância da inclusão deste valor na geração da predição de recomendações.

## 7.1 Trabalhos Futuros

Esta tese permitiu que fossem vislumbrados possíveis trabalhos de pesquisa a serem realizados no futuro, visando complementá-la e auxiliar de forma direta na área de sistemas de recomendação. Abaixo serão descritas algumas das possibilidades identificadas:

a) Através da aplicação do *framework* F-UIP proposto no Capítulo 4 desta tese, verificou-se a necessidade de desenvolvimento de um agente de software que auxilie no gerenciamento do conhecimento gerado através da aplicação de mineração de dados. O processo de descoberta de conhecimento em banco de dados apresenta a geração de conhecimento como saída deste processo e no caso da solução proposta, procura-se identificar novas áreas de interesse para um usuário baseando-se nas regras geradas. Com o tempo o número de regras geradas e aceitas tende a ser bem significativo, desta forma seria necessário um agente que procurasse gerenciar todo este conhecimento novo e atentasse para as regras que já foram testadas com êxito ou sem êxito nos usuários. Este agente seria denominado “Agente Gerenciador de Conhecimento” (este trabalho de pesquisa já se encontra em andamento em um trabalho de conclusão de curso).

b) A relevância da opinião de um usuário pode não ser suficiente na solução de ajudar um usuário a “confiar” em uma recomendação que esta sendo predita. Desta forma seria interessante a elaboração de uma outra métrica que pudesse representar a “reputação” deste usuário frente à comunidade de usuários. Por exemplo, poder-se-ia pensar em soluções como a adotada pelo *site* de comércio eletrônico e-bay (<http://www.e-bay.com>), onde um usuário pode comentar como foi a sua relação de compra e venda em uma transação realizada com outro usuário do *site*. Esta descrição pode ajudar outros usuários a confiarem ou não neste usuário que vendeu um produto. Outro exemplo apresentado na pesquisa da Profa. Dra Julita Vassileva da Universidade de Saskatchewan em uma rede P2P cada usuário poderia ter sua reputação calculada de acordo com o *feedback* dos demais usuários quanto a qualidade dos itens disponibilizados por este usuário na rede. Desta forma a pesquisadora propõe uma visualização de todos os *peers* em forma de constelação, onde os mais confiáveis frente

a opinião da comunidade são apresentados com maior destaque, ou seja, um usuário pode facilmente identificar qual o usuário com melhor reputação na comunidade. Este tipo de informação, também poderia ser incluída na solução proposta por esta tese. A idéia poderia ser uma solução baseada em recompensa e punição, onde todos os usuários iniciariam no sistema com um mesmo valor de reputação, e este valor de reputação sofreria variação devido ao *feedback* dos demais usuários alvos das recomendações realizadas com a participação de um dado usuário recomendador.

A cada *feedback* positivo as suas recomendações, este usuário recomendador poderia receber um prêmio que viesse aumentar a sua reputação frente à comunidade, por sua vez cada *feedback* negativo resultaria em uma punição. Como os usuários não têm como descobrir quem são os indivíduos que participam da elaboração de uma recomendação, não existiria preocupação com ações que procurassem prejudicar de forma deliberada um determinado indivíduo. Este trabalho de pesquisa já se encontra em andamento em um trabalho de conclusão de curso.

c) Aplicação de Sistemas de Recomendação e Ranqueamento do Recomendador para criação de Comunidades Virtuais (trabalho de pesquisa já iniciado e com resultados preliminares descritos no Anexo G desta tese).

d) Pesquisa sobre a influência dos estilos decisórios, sobre a decisão do que acessar, bem como a influência de quesitos referentes a estilo cognitivo na avaliação de conteúdo de um artigo científico. Durante o experimento final desta tese evidenciou-se que cada pesquisador possui um conjunto de critérios pessoais para avaliar a importância ou não de um conteúdo científico. Estes critérios vão muito além da simples questão do conteúdo ser informativo, pois em muitos casos por conversas informais com os participantes da amostra, verificou-se que estes buscavam avaliar o conteúdo do artigo com a visão de um avaliador de artigo para congresso científico ou revista científica.

e) Aplicação do CV-Lattes para determinar se o usuário é um iniciante na área de pesquisa, se possui experiência avançada ou se é do nível intermediário. Esta classificação permitiria casar os artigos científicos (classificados em básicos, intermediários ou avançados) com o nível de conhecimento do usuário.

f) Aplicação do Qualis para determinar se o artigo científico é para um público iniciante na área de pesquisa, intermediário ou avançado. Esta classificação do artigo poderia ser realizada utilizando as categorias do Qualis e a antiguidade das publicações. Por exemplo, artigos científicos de conferências atuais Qualis A seriam conteúdo para pesquisadores avançados na área, e os artigos mais antigos poderiam ser para usuários intermediários ou iniciantes na área.

g) Atualização automática do perfil do usuário. Assim como é proposta a inclusão de novas áreas de interesse nos perfis para atualizá-los, pode ser trabalhada a retirada de áreas de interesse em que o usuário está apresentando baixo interesse.

h) Otimização do algoritmo referente ao F-UIP apresentado na seção 4.3.

## 7.2 Limitações

O modelo Mo-Drop apresentado nesta tese apresenta algumas limitações quanto a sua aplicabilidade. Ele foi concebido para domínios onde o conhecimento sobre os itens a serem recomendados é fundamental. Observou-se que em domínios onde o gosto é mais importante como recomendação de filmes, músicas a relevância do recomendador não se faz necessária, porém em domínios onde o conhecimento técnico faz diferença, como a recomendação de artigos científicos, o modelo é aplicável com sucesso. A mesma limitação é observada no *framework* F-UIP.

## 7.3 Considerações Finais

Como conclusão final do trabalho destaca-se que com base nos experimentos realizados com a participação de uma amostra de pesquisadores foi possível evidenciar que a inclusão da métrica de relevância da opinião do usuário em um Sistema de Recomendação de artigos científicos apresenta-se como um diferencial, seja na interface que irá permitir a seleção de qual recomendação acessar, como nos resultados de satisfação da recomendação fornecida. Contudo deve-se observar que os resultados provem de uma amostra de pesquisadores, portando não são generalizáveis, constituindo-se em uma evidência.

Destaca-se como a principal contribuição desta tese para a área de Sistemas de Recomendação a definição de como determinar e incluir no processo de recomendação a autoridade (relevância da opinião) de um usuário, uma vez que os sistemas atuais não trabalham com este conceito.

Por fim deve-se observar que no processo de recomendação atual do sistema W-RECMAS está-se verificando o Ranqueamento do Recomendador para a geração da recomendação, porém não é observado o Ranqueamento do Recomendador no momento da entrega de uma recomendação, ou seja, não se está observando o quanto o recomendado sabe ou não da área de interesse (somente sabe-se que este possui interesse na área). Seria interessante classificar os itens (artigos científicos) a serem recomendados em categorias como, por exemplo, em artigos básicos para área de interesse, artigos intermediários e artigos avançados, e no momento de entregar a recomendação verificar o nível de conhecimento do pesquisador alvo da recomendação. Esta verificação seria interessante para evitar a recomendação de itens que sejam considerados interessantes para recomendação, porém sejam muito básicos ou avançados para o perfil do recomendado.

## REFERÊNCIAS

- [ABD 2000] ABDUL-RAHMAN, A.; HAILES, S. Supporting trust in virtual communities. In: INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 2000. **Proceedings...** [S.l. : s.n.], 2000.
- [ADO 2005] ADOMAVICIUS, G. et al. Incorporating contextual information in recommender systems using a multidimensional approach. **ACM Transactions on Information Systems**, New York, v.23, n. 1, p. 103-145, Jan. 2005.
- [ADO 2005a] ADOMAVICIUS, G.; TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. **IEEE Transactions on Knowledge and Data Engineering**, New York, v. 17, n. 6, p. 734-749, June. 2005.
- [AGG 99] AGGARWAL, C.C.; GATES, S. C. **Recommender systems: knowledge from mining user experiences**. EUA: Divisão de Pesquisa, IBM, 1999. 23 f. (Relatório Técnico RT 99A000195).
- [AGR 93] AGRAWAL, R. et al. Mining Association Rules between sets of items in large databases. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, SIGMOD, 1993, Washington. **Proceedings...** [S.l.: s.n.], 1993. p. 207-216.
- [AGR 94] AGRAWAL, R. et al. Fast algorithms for mining association rules in large databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, VLDB, 20., 1994, Santiago. **Proceedings...** Hove: Morgan Kaufmann, 1994. p. 478-499.
- [AGU 2002] AGUZZOLI, S.; AVESANI, P.; MASSA, P. Collaborative Case-Based Recommendation System. New York: Springer, 2002. (Lecture Notes in Computer Science, Lecture

Notes in Artificial Intelligence, v. 2416).

- [ANS 2000] ANSARI, A. et al. Internet Recommendation Systems. **Journal of Marketing Research**, [S.l.], v.37, n.3, p. 363-375, Aug. 2000.
- [APT 2002] APTE, C. et al. Business Application of Data Mining. **Communications of the ACM**, New York, v.45, n.8, p. 49-53, Aug. 2002.
- [AVE 2005] AVESANI, P.; MASSA, P.; TIELLA, R. A trust-enhanced recommender system application: Moleskiing. In: ACM SYMPOSIUM ON APPLIED COMPUTING. **Proceedings...**, New York: ACM, 2005. p. 1589 - 1593.
- [BAL 97] BALABANOVIC, M.; SHOHAM, Y. Fab: Content-Based, Collaborative Recommendation. **Communications of the ACM**, New York, v.40, n.3, p. 66-72, Mar. 1997.
- [BEA 2002] BERGHOLZ, A. Coping with Sparsity in a Recommender System. In: WEB KNOWLEDGE DISCOVERY ON DATABASE, **Proceedings...**[S.l.: s.n.], 2002. p. 86-99.
- [BEC 2005] BELLINI, C. G. P.; MANSSOUR, A. B. B. Understanding the success of an internet-mediated community in Brazil. In: ANNUAL GLOBAL INFORMATION TECHNOLOGY MANAGEMENT WORLD CONFERENCE, GITMA, 2005. **Proceedings...** [S.l.: s.n.], 2005. p. 105-108.
- [BEF 99] BELLIFEMINE, F., POGGI, A.; RIMASSA, G. JADE – A FIPA-compliant agent framework. CSELT internal technical report. In: THE PRACTICAL APPLICATION OF INTELLIGENT AGENTS AND MULTI-AGENTS, PAAM, 1999. **Proceedings...**[S.l.: s.n.], 1999. p.97-108.
- [BEJ 92] BELVIN, N. J.; CROFT, W. B. Information Filtering and Information Retrieval: two sides of the same coin? **Communications of the ACM**, New York, v.35, n.12, p. 29, Dec. 1992.
- [BEL 2000] BELKIN, N. J. Helping people find what they don't know. **Communications of the ACM**, New York, v.43, n.8, p. 58-61, Aug. 2000.

- [BER 2000] BERRY, M. J. A. **Mastering Data Mining**: the art and science of customer relationship management. New York: John Wiley & Sons, 2000. 494p.
- [BOL 98] BOLLACKER, K. D. et al. Citeseer: An autonomous Web agent for automatic retrieval and Identification of Interesting publications. In: INTERNACIONAL CONFERENCE OF AUTONOMOUS AGENTS, 1998, New York. **Proceedings...** Los Alamitos, CA: ACM Press, 1998. p. 116-123.
- [BRA 97] BRADSHAW, J. M. **Software agents**. London: Mit Press, 1997.
- [BRE 2005] BRESSLER, F. ; CAZELLA, S. C.; MACHADO, C. P. Deciding what software to offer: PopRecom – a recommender system of software products. In: INTERNATIONAL CONFERENCE ON DECISION SUPPORT SYSTEMS, ISDSS, 8., 2005. **Proceedings...** [S.l.: s.n.], 2005.
- [BRW 98] BRENNER, W. et al. Intelligent software agents. Heidelberg: Springer-Verlag, 1998.
- [CAB 97] CABENA, P. et al. Discovering data mining: from concept to implementation. USA: Prentice-Hall, 1997.
- [CAS 2001] CASE, S. et al. Enhancing E- Communities with agent- based systems. **Computer**, Los Alamitos, v.34, n.7, p. 64- 69, 2001.
- [CAZ 2003] CAZELLA, S. C.; ALVARES, L. O. C. W-RECMAS: Um sistema de recomendação híbrido baseado em sistemas multiagentes para recomendação de artigos científicos. Porto Alegre: Instituto de Informática, Universidade Federal do Rio Grande do Sul, 2003. 86 f. (Relatório Técnico RT 339).
- [CAZ 2004] CAZELLA, S. C.; ALVARES, L. O. C. Creating virtual web communities through a hybrid recommender system. In: WORKSHOP DE TESES E DISSERTAÇÕES EM INTELIGÊNCIA ARTIFICIAL, WTDIA, 2., 2004. **Anais...** [S.l.: s.n.], 2004. p.51-60.
- [CAZ 2005a] CAZELLA, S. C.; ALVARES, L. O. C. Creating Academic Web Communities: A Recommender System to Aid Brazilian Researchers to

Find Information and Contact Relevant People. In: ANNUAL GLOBAL INFORMATION TECHNOLOGY MANAGEMENT WORLD CONFERENCE, GITMA, 2005. **Proceedings...** [S.l.: s.n.], 2005. p. 45-48.

- [CAZ 2005b] CAZELLA, S. C.; ALVARES, L. O. C. Modeling user's opinion relevance to recommending research papers. In: INTERNATIONAL CONFERENCE ON USER MODELING, UM, 10., 2005. **User Modeling**: proceedings. Berlin: Springer-Verlag, 2005. p. 337-341. (Lecture Notes in Artificial Intelligence, 3338).
- [CAZ 2005c] CAZELLA, S. C.; ALVARES, L. O. C. Combining Data Mining Technique and Users' Relevance Opinion to Build an Efficient Recommender System. **Revista Tecnologia da Informação**, Brasília, v. 5, n. 1, p.9-20, 2005.
- [CAZ 2005d] CAZELLA, S.C.; REATEGUI, E.B. Sistemas de Recomendação, Mini-Curso, In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 2005. **Anais...** [S.l.: s.n.], 2005. p. 306 - 348.
- [CAZ 2006] CAZELLA, S. C.; ALVARES, L. O. C. An Architecture based on multi-agent system and data mining for recommending research papers and researchers. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, 2006. **Proceedings...** [S.l.: s.n.], 2006. p. 67-72.
- [CAZ 97] CAZELLA, S. C. **Uma arquitetura para coordenar a interação de agentes na Internet**. 1997. 114 f. Dissertação (Mestrado em Ciências da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- [DEN 82] DENNING, P. J. Electronic Junk. **Communications of the ACM**, New York, v.25, n.3, p. 163-165, Mar. 1982.
- [ETZ 95] ETZIONI, O.; WELD, D. S. Intelligent agents on the Internet: fact, fiction, and forecast. **IEEE Expert**, New York, v. 10, n.4, p. 44-49, Aug. 1995.
- [FAY 96] FAYYAD, U. M. et al. (Ed.). **Advances in knowledge discovery and Data Mining**. Menlo Park: AAAI, 1996.



- [FER 96] FERREIRA, A. B. de H. **Novo Dicionário Aurélio da Língua Portuguesa**. 2. ed. Rio de Janeiro: Nova Fronteira, 1986.
- [FIG 98] FIGALLO, C. **Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and maintaining a Competitive Edge**. New York: Wiley Computer, 1998.
- [FOL 92] FOLTZ, P. W. ; DUMAIS, S. T. Personalized Information Delivery: An Analysis of Information Filtering Methods. **Communications of the ACM**, New York, v.35, n.12, p. 51-60, Dec. 1992.
- [FRE 2002] FREITAS, F. L. **Sistemas Multiagente cognitivos para Recuperação, Classificação e Extração Integradas de Informação da Web**. 2002. 201 f. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Santa Catarina, Florianópolis.
- [GAR 2005] GARDEN, M. et al. Semantic feedback for hybrid recommendations in Recommendz. In: IEEE INTERNATIONAL CONFERENCE ON E-TECHNOLOGY, E-COMMERCE AND E-SERVICE, 2005. **Proceedings...** [S.l.: s.n.], 2005.
- [GLO 2001] GLOVER, E. J. et al. Improving Web Searching with User Preferences. **Communications of the ACM**, New York, v.44, n.12, p. 97-102, Dec. 2001.
- [GOD 92] GOLDBERG, D. et al. Using collaborative filtering to weave an information Tapestry. **Communications of the ACM**, New York, v.35, n.12, p. 61-70, Dec. 1992.
- [GOL 2005] GOLBECK, J. **Computing and applying trust in web-based social networks**. 2005. 184 f. Tese (Doutorado em Ciência da Computação) – Computer Science Department, University of Maryland, EUA.
- [GOL 2006] GOLBECK, J. Generating Predictive Movie Recommendations from Trust in Social Networks. In: INTERNATIONAL CONFERENCE ON TRUST MANAGEMENT, 4., 2006, Pisa, Italy. **Proceedings...** [S.l.: s.n.], 2006.
- [HAG 97] HAGEL III, J.; ARMSTRONG, A. **NET Gain: expanding markets through virtual communities**. USA: Harvard Business School, 1997.

- [HAH 2002] HAHSLER, M.; SCHULTZ, A. G. Evaluation of Recommender Algorithms for an Internet Information Broker Based on Simple Association Rules and on the Repeat-Buying Theory. In: WEB KNOWLEDGE DISCOVERY ON DATABASE, **Proceedings...**[S.l.: s.n.], 2002. p. 100-114.
- [HAN 2001] HAN, J.; KAMBER, M. **Data Mining**: concepts and techniques. USA: Morgan Kaufmann, 2001.
- [HAU 97] HAUBEN, M.; HAUBEN, R. **Netizens**: On the History and Impact of Usenet and the Internet. Los Alamitos, CA: IEEE Computer Society, 1997.
- [HER 2000] HERLOCKER, J. L. **Understanding and Improving Automated Collaborative Filtering Systems**. 2000. 220 f. Tese (Doutorado em Ciência da Computação) - University of Minnesota, Minnesota.
- [HER 2004] HERLOCKER, J. L. et al. Evaluating Collaborative Filtering Recommender Systems. **ACM Transactions on Information Systems**, New York, v.22, n.1, p. 5-53. Jan. 2004
- [HER 99] HERLOCKER, J. et al. An algorithmic framework for performing collaborative filtering. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, **Proceedings...**[S.l.: s.n.], 1999. p. 230-237.
- [KAU 97] KAUTZ, H. et al. Referral Web: combining social networks and Collaborative filtering. **Communications of the ACM**, New York, v.40, n.3, p. 63-65, Mar. 1997.
- [KON 97] KONSTAN, J. A. et al. Grouplens: Applying Collaborative Filtering to Usenet News. **Communications of the ACM**, New York, v.40, n.3, p. 77-87, Mar. 1997.
- [LAR 2004] LAROUSSE . Dicionário Português Larousse. 3. ed. Rio de Janeiro: Ática, 2004.
- [LOE 92] LOEB, S.; TERRY, D. Information Filtering. **Comunications of ACM**, New York, v.35, n.12, p.26, Dec. 1992.

- [LUE 98] LUEG, C. Considering Collaborative Filtering as Groupware: Experiences and Lessons Learned. In: INTERNATIONAL CONFERENCE ON PRACTICAL ASPECTS OF KNOWLEDGE MANAGEMENT, 1998. **Proceedings...** [S.l.: s.n.], 1998.
- [MAE 94] MAES, P. Agents that reduce work and information overload. **Communications of the ACM**, New York, v.37, n.7, p. 30-40, July 1994.
- [MAS 2004] MASSA, P. et al. Trust-aware Collaborative Filtering for Recommender System. In: INTERNATIONAL CONFERENCE ON COOPERATIVE INFORMATION SYSTEM, 2004, Agia Napa, Cyprus. **Proceedings...** Berlin: Springer-Verlag, 2004. p. 492-? (Lecture Notes in Artificial Intelligence, v. 3290 ).
- [MAS 2004a] MASSA, P. et al. Using Trust in Recommender Systems: an experimental analysis. In: INTERNATIONAL CONFERENCE ON TRUST, iTrust, 2004. **Proceedings...**[S.l.: s.n.], 2004.
- [MCC 2005] MCCAREY, F. et al. Rascal: A Recommender Agent for Agile Reuse. **Artificial Intelligence Review**, [S.l.], n. 24, p. 253–276, 2005.
- [MIL 2005] MILANI, F.; CAZELLA, S.C. Um modelo para determinar a autoridade de usuários em Sistemas de Recomendação. In: FÓRUM DE INTELIGÊNCIA ARTIFICIAL, 3., 2005. **Anais...** Canoas: ULBRA, 2005. p. 31-39.
- [MON 2002] MONTANER, M.; LÓPEZ,B.; LA ROSA, J. L. Developing Trust in Recommender Agents. In: INTERNATIONAL JOINT CONFERENCE ON AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS, AAMAS, 1., 2002, Bologna. **Bringing people and agents together: proceedings**. New York: ACM, 2002. p 304 – 305.
- [MON 2003] MONTANER, M.; LÓPEZ,B.; LA ROSA, J. L. A Taxonomy of Recommender Agents on the Internet. **Artificial Intelligence Review**, [S.l.], v.19, n. 4, p. 285–330, 2003.
- [MOU 98] MOUKAS, A.; MAES, P. Amalthea: An Evolving Multi-Agent Information Filtering and Discovery System for the WWW. In: AUTONOMOUS AGENTS AND MULTI-AGENT SYSTEMS, 1998. **Proceedings...** [S.l.],1998. p. 59 – 88.

- [MUR 2004] O'MURCHU, I.; BRESLIN, J.G.; DECKER, S. Online Social and Business Networking Communities. In: WORKSHOP ON APPLICATION OF SEMANTIC WEB TECHNOLOGIES TO WEB COMMUNITIES, ECAI, 2004. **Proceedings...** [S.l.: s.n.], 2004.
- [PAG 98] PAGE, L. et al. The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library Technologies Project. Stanford, CA: Departamento de Ciência da Computação, Stanford University, 1998. 17 f. (RT SIDL-WP-1999 0120).
- [PAZ 99] PAZZANI, M. J. A Framework for collaborative, content-based and demographic filtering. **Artificial Intelligence Review**, Hingham, v.13, n. 5, p. 393-408, 1999.
- [PRE 99] PRETSCHNER, A. ; GAUCH, S. Ontology-Based Personalized Search. In: IEEE INTL. CONF. ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 11., 1999. **Proceedings...** [S.l.]: IEEE, 1999. p. 391 – 398.
- [RES 97] RESNICK, P.; VARIAN, H. R. Recommender Systems. **Communications of the ACM**, New York, v.40, n.3, p. 55-58, Mar. 1997.
- [REZ 2005] REZENDE, S. O. Mineração de dados, Mini-Curso, In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 2005. **Anais...** [S.l.: s.n.], 2005, p. 397 - 433.
- [RHE 2000] RHEINGOLD, H. **The virtual community**: homesteading on the electronic frontier. Massachusetts, USA: MIT Press, 2000.
- [RIE 98] RIEDL, J. et al. Using filtering agents to improve prediction quality in the GroupLens Research Collaborative Filtering System. In: ACM CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 1998. **Proceedings...** New York: ACM Press, 1998. p. 345 – 354.
- [RIE 99] RIEDL, J. et al. Combining Collaborative Filtering with Personal Agent for Better Recommendations. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16.; INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE CONFERENCE, 11., 1999. **Proceedings...** Menl Park, CA: American Association for Artificial

Intelligence [S.l.], 1999. p. 439-446.

- [RUS 2002] RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence**: a modern approach. Upper Saddle River, N.J.: Prentice Hall, USA, 2002.
- [SAB 2002] SABATER, J.; SIERRA, C. Reputation and social network analysis in multi-agent systems. In: INTERNATIONAL JOINT CONFERENCE ON AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS, AAMAS, 1., 2002, Bologna. **Bringing people and agents together**: proceedings. New York: ACM, 2002. p. 475-482.
- [SAR 2000] SARWAR, B. M. et al. Analysis of Recommendation Algorithms for e-commerce. In: ACM CONFERENCE ON E-COMMERCE, 2., 2000, Minneapolis. **Proceedings...** New York: ACM Press, 2000. p. 158-167.
- [SCH 2001] SCHAFER, J.B.; KONSTAN, J.; RIEDL, J. E-Commerce recommendation applications. **Data Mining and Knowledge Discovery**, [S.l.], v.5, n. 1-2, p. 115-153, Jan. 2001.
- [SCH 99] SCHAFER, J.B.; KONSTAN, J.; RIEDL, J. Recommender Systems in E-Commerce. In: ACM CONFERENCE ON E-COMMERCE, 2., 1999, Minneapolis. **Proceedings...** New York: ACM Press, 1999. p. 158-166.
- [SHA 95] SHARDANAND, U.; MAES, P. Social information filtering: Algorithms for automating "word of mouth". In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, CHI, 1995, Denver, US. **Proceedings...** New York: ACM, 1995. p. 210-217.
- [SIL 2005] SILVA FILHO, W. D. da; CAZELLA, S. C. STAR: Um Framework para recomendação de artigos científicos baseado na relevância da opinião dos usuários e em filtragem colaborativa. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, ENIA, 2005. **Anais...** [S.l.: s.n.], 2005. p. 1042-1051.
- [SIN 99] SINGH, M. P. ; JOSHI, A. Multiagent Systems on the Net. **Communications of the ACM**, New York, v.42, n.3, p. 39-40, Mar. 1999.
- [SLY 2002] SLYKE, C. V. et al. Gender Differences in perceptions of Web-based Shopping. **Communications of the ACM**, New York, v.45, n.45, p. 82-

86, Aug. 2002.

- [SMY 2002] SMYTH, P. Business Applications of Data Mining. **Communications of the ACM**, New York, v.45, n.8, p. 49-53, Aug. 2002.
- [TAN 2000] TAN,P.; KUMAR,V. Interestingness Measures for Association Patterns: A perspective. In: WORKSHOP ON POSTPROCESSING IN MACHINE LEARNING AND DATA MINING, KDD, 2000. **Proceedings...** [S.l.: s.n.], 2000. p. 417-423.
- [TER 2001] TERVEEN, L. G. ; HILL, W. Beyond Recommender Systems: Helping People Help Each Other. In: CARROL, J. (Ed.). **HCI in the New Millennium**. [S.l.]: Addison Wesley, 2001. p. 1-21.
- [TOR 2004] TORRES, R. **Personalização na Internet**. [S.l.]: Novatec, 2004.
- [TWO 99] TWO CROWS CORPORATION. Introduction to Data Mining and Knowledge Discovery. 3rd ed. 1999, Disponível em: <<http://www.twocrows.com/intro-dm.pdf>>. Acesso em: 27 abr. 2002.
- [WAN 2004] WANG Y.; VASSILEVA J. Trust-Based Community Formation in Peer-to-Peer File Sharing Networks, In: IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, WI, 2004, Beijing, China. **Proceedings...** [S.l.: s.n.], 2004.
- [WEI 99] WEISS, G. **Multiagent Systems: a modern approach to distributed artificial intelligence**. EUA: MIT Press, 1999.
- [WEL 2002] WELLMAN, B. Designing the Internet for a Networked Society. **Communications of the ACM**, New York, v.45, n.8, p. 91-96, Aug. 2002.
- [WIT 2000] WITTEN, I. H. **Data Mining: practical machine learning tools and techniques with java implementations**. San Francisco: Morgan Kaufmann, 2000. 369 p.
- [YOS 99] YOSHIDA, S. et al. SocialWare: Multiagent Systems for Supporting Network Communities. **Communications of the ACM**, New York, v.42, n.3, p. 55-61, Mar. 1999

- [ZAI 2000] ZAIANE, O. R. Web Mining: Concepts, Practices and Research. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, SBBD, 15., 2000, João Pessoa. **Tutorial...** João Pessoa: CEFET-PB; Porto Alegre: PUCRS, 2000. p. 410-474.
- [ZIE 2004] ZIEGLER, C. ;LAUSEN, G. Analyzing correlation between trust and user similarity in online communities. In: INTERNATIONAL CONFERENCE ON TRUST MANAGEMENT, 2004, Maui, Hawaii. **Proceedings...** [S.l.: s.n.], 2004.

## **ANEXO A ANÁLISE DOS SISTEMAS DE RECOMENDAÇÃO E W-RECMAS**

Neste Anexo é apresentada uma comparação entre uma lista de sistemas de recomendação encontrados na literatura e descritos em [MON 2003] [ADO 2005] e o protótipo proposto e implementado.



Tabela A.1. Análise dos Sistemas de Recomendação

Nome	Representação	Inicial	Aprendizado	Feedback	Adaptação	Filtragem	Casamento	Método	Relevância da opinião
ACRNews	Agrupamento	Conjunto de Treinamento	Regras, Agrupamento	Implícito (histórico de navegação)	Adição de novo item	Conteúdo	Agrupamento	Filtragem baseada em conteúdo	Não Aplicado!
Amalthea	Vetor de Características	Manual	Indexação de Informação	Explícito (pontuação)	Seleção natural, GFF	Conteúdo	Similaridade (coseno)	Filtragem baseada em conteúdo	Não Aplicado!
Amazon	Histórico	Vazio	Não aplicado	Explícito (pontuação) Implícito (histórico de compras)	Adição de novo item	Híbrido	Desconhecido	Híbrido	Não Aplicado!
Anatagonomy	Vetor de Características	Vazio	Indexação de Informação	Explícito (pontuação) Implícito (scrolling)	Adição de novo item	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!

Nome	Representação	Inicial	Aprendizado	Feedback	Adaptação	Filtragem	Casamento	Método	Relevância da opinião
Beehive	Agrupamento	Vazio	Agrupamento	Implícito (histórico de mail)	Adição de novo item	Colaborativo	Agrupamento	Filtragem Colaborativa	Não Aplicado!
Bellcore Vídeo	Matriz de Avaliação	Conjunto de Treinamento	Não aplicado	Explícito (pontuação)	Adição de novo item	Colaborativa	Vizinho mais próximo	Filtragem Colaborativa	Não Aplicado!
CDNow	Histórico	Vazio	Não aplicado	Explícito (pontuação) Implícito (histórico de compras)	Adição de novo item	Híbrido	Desconhecido	Híbrido	Não Aplicado!
Fab	Vetor de Características	Vazio	Indexação de Informação	Explícito (pontuação)	Seleção natural	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!
ifWeb	Rede Semântica	Conjunto de Treinamento	Seleção de Características	Explícito (pontuação)	GFF	Conteúdo	Casamento de palavra-chave	Filtragem baseada em conteúdo	Não Aplicado!
INFOrmer	Rede Associativa	Conjunto de Treinamento	Seleção de Características	Explícito (pontuação)	Adição de novo item	Conteúdo	Comparação Gráfica	Filtragem baseada em conteúdo	Não Aplicado!
Krakatoa Chronicle	Vetor de Características	Vazio	Indexação de Informação	Explícito (pontuação) Implícito (scrolling, tempo de despendido)	Adição de novo item	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!
Let's Browse	Vetor de Características	Conjunto de Treinamento	Indexação de Informação	Implícito (links, tempo de despendido)	Adição de novo item	Conteúdo	Similaridade (coseno)	Filtragem baseada em conteúdo	Não Aplicado!

Nome	Representação	Inicial	Aprendizado	Feedback	Adaptação	Filtragem	Casamento	Método	Relevância da opinião
Letizia	Vetor de Características	Vazio	Indexação de Informação	Implícito (links, tempo de despendido)	Adição de novo item	Conteúdo	Similaridade (coseno)	Filtragem baseada em conteúdo	Não Aplicado!
LifeStyle Finder	Características demográficas	Estereótipo	Não aplicado	Explícito (pontuação) Implícito (histórico de compras)	Adição de novo item	Demográfico	Demográfico	Filtragem Demográfica	Não Aplicado!
MovieLens	Vetor de Características, Regras	Conjunto de Treinamento	Indexação de Informação, regras	Explícito (pontuação)	Adição de novo item	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!
News Dude	Vetor de Características	Conjunto de Treinamento	Indexação de Informação	Explícito (gosto/não gosto, Eu já conheço, Fale me mais )	Modelos de curta/longa duração	Conteúdo	Vizinho mais próximo, classificadores Bayesianos	Filtragem baseada em conteúdo	Não Aplicado!
NewsWeeder	Vetor de Características	Conjunto de Treinamento	Indexação de Informação	Explícito (pontuação)	Adição de novo item	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!

}

Nome	Representação	Inicial	Aprendizado	Feedback	Adaptação	Filtragem	Casamento	Método	Relevância da opinião
NewT	Vetor de Características	Conjunto de Treinamento	Indexação de Informação	Explícito (gosto/não gosto)	Seleção natural	Conteúdo	Similaridade (coseno)	Filtragem baseada em conteúdo	Não Aplicado!
Personal WebWatcher	Vetor de Características	Manual	Indexação de Informação	Implícito (links)	Adição de novo item	Híbrido	Classificador Bayesiano	Híbrido	Não Aplicado!
Recommender	Regras	Conjunto de Treinamento	Regras	Explícito (pontuação)	Adição de novo item	Híbrido	Regras	Híbrido	Não Aplicado!
Ringo/FireFly	Matriz de Avaliação	Conjunto de Treinamento	Não aplicado	Explícito (pontuação)	Adição de novo item	Híbrido	Vizinho mais próximo	Filtragem colaborativa	Não Aplicado!
SIFT Netnews	Vetor de Características, Árvore de Decisão	Conjunto de Treinamento	Indexação de Informação	Explícito (gosto/não gosto)	Manual	Colaborativo	Comparação Gráfica	Filtragem baseada em conteúdo	Não Aplicado!

Nome	Representação	Inicial	Aprendizado	Feedback	Adaptação	Filtragem	Casamento	Método	Relevância da opinião
Syskill & Weibert	Vetor de Características, Árvore de Decisão	Manual	Seleção de Características, Indexação de Informação, ID3	Explícito (pontuação)	Adição de novo item	Conteúdo	Classificador Bayesiano, Similaridade (coseno), árvore de decisão	Filtragem baseada em conteúdo	Não Aplicado!
Webmate	Vetor de Características	Vazio	Indexação de Informação	Explícito (gosto/não gosto)	Adição de novo item	Conteúdo	Similaridade (coseno)	Filtragem baseada em conteúdo	Não Aplicado!
WebSail	Vetor de Características	Vazio	Indexação de Informação	Explícito (gosto/não gosto)	Adição de novo item	Conteúdo	TW2( <i>The Tailored Winnow2</i> )	Filtragem baseada em conteúdo	Não Aplicado!
WebSell	Histórico	Vazio	Não aplicado	Desconhecido	Adição de novo item	Híbrido	CBR	Híbrido	Não Aplicado!
Websift	Regras, Padrões	Conjunto de Treinamento	Regras	Implícito (histórico de navegação)	Adição de novo item	Híbrido	Regras, padrões	Híbrido	Não Aplicado!
WebWatcher	Vetor de Características	Manual	Indexação de Informação	Implícito (links)	Adição de novo item	Híbrido	Similaridade (coseno)	Híbrido	Não Aplicado!
<b>W-RECMAS/ Artigos Científicos</b>	<b>Histórico/ Matriz de Avaliação</b>	<b>Formulário/ de Informações sobre usuário</b>	<b>Mineração/ Multiagentes/ Aprendizado de Máquina</b>	<b>Explícito (pontuação) Implícito (histórico de compras)</b>	<b>Baseado em feedback/ Novo assunto</b>	<b>Híbrido</b>	<b>Similaridade (Person)/ Vizinho mais próximo</b>	<b>Híbrido</b>	<b>Aplicado!</b>

## ANEXO B TEXTO DO PRIMEIRO EXPERIMENTO

O texto a seguir refere-se ao texto enviado a amostra selecionada para participação do Experimento 1.

---

Caros colegas,

primeiramente, agradeço a disponibilidade de vocês em auxiliar nesta fase final da minha tese.

A tese procura identificar o peso da opinião de um usuário em um Sistema de Recomendação, e tenta quantificar este valor para as áreas de interesse do mesmo. Por exemplo, um pesquisador com muitos anos de pesquisa na área de Redes de Computadores teria um grande valor na sua opinião (autoridade) ao recomendar um artigo científico para outros usuários do sistema, mas um estudante de iniciação científica possuiria uma relevância (autoridade) menor de opinião.

As perguntas a seguir referem-se ao Currículo Lattes, e as respostas prestadas serão de grande utilidade para aprimorar e analisar o modelo que proponho na minha tese de Doutorado.

Conforme informado em e-mail anterior este é o primeiro e-mail de um número total de 4 E-mails que serão enviados para coletar as opiniões dos acadêmicos.

Por favor em caso de dúvidas entrem em contato:

- MSN: silvio\_cazella@hotmail.com
- Skype: silvio\_cazella
- E-mail: silvio.cazella@gmail.com

- COMPROMETIMENTO:

Toda a informação aqui prestada é de caráter totalmente sigiloso, em nenhum momento as pessoas que fizeram parte da amostra serão identificadas, bem como suas opiniões aqui esboçadas.

- Perfil:

Por favor, preencha as questões abaixo referentes ao seu perfil:

- 1) Nome completo:
- 2) Já trabalhou com avaliação de currículos Lattes? ( ) Sim ( ) Não

3) Caso a resposta anterior seja sim, com qual finalidade?

4) Qual sua área de Pesquisa Principal (grande Área):

- a.  IA
- b.  Redes de Computadores
- c.  Engenharia de SW
- d.  Computação Gráfica
- e.  Sistemas de Informação
- f.  Sistema Operacionais
- g.  Outra: \_\_\_\_\_

- Pesquisa:

Objetivo da pesquisa:

A pesquisa que será realizada é de natureza exploratória e tem por objetivo aprofundar o entendimento da importância dos indicadores de produção do CV-LATTES na avaliação de currículos.

Este entendimento visa auxiliar na correta construção de um modelo para o cálculo de autoridade da opinião de acadêmicos em suas respectivas áreas de interesse.

Nesta etapa o acadêmico é solicitado a analisar os indicadores de produção e informar quais destes são relevantes (na sua opinião) para analisar o peso de opinião de um acadêmico (autoridade) e atribuir o respectivo peso para cada indicador e grupo de indicadores. A questão aqui é verificar se os indicadores (quantitativos inicialmente), poderiam servir como apoio preliminar na decisão da relevância (autoridade da opinião).

Para tanto, solicito o preenchimento da tabela disponibilizada a seguir (pág 5). Por favor, observe que nesta tabela a soma do percentual de importância dos macros indicadores (Produção Bibliográfica, Produção Técnica, Orientação concluída, Informações complementares e Demais trabalhos relevantes) deve ser igual a Total Geral= 100%, e a soma dos itens dentro dos macros Indicadores deverá ser 100% também.

Exemplo:

Produção Bibliográfica			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Artigos publicados em periódicos	(X) SIM      () NÃO	30%	40%
2) Trabalhos em anais de eventos	(X) SIM      () NÃO	20%	
3) Livros ou Capítulos de livros	(X) SIM      () NÃO	20%	
4) Textos em jornais ou revistas	(X) SIM      () NÃO	30%	
5) Demais tipos de produção bibliográfica	() SIM      (X) NÃO	0%	
TOTAL:		100%	
Produção Técnica			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Software	() SIM      () NÃO	%	30%
2) Produtos tecnológicos	() SIM      () NÃO	%	
3) Trabalhos técnicos	() SIM      () NÃO	%	
4) Demais tipos de produção técnica (organização de eventos)	() SIM      () NÃO	%	
5) Demais tipos de produção técnica (relatórios de pesquisas)	() SIM      () NÃO	%	
6) Demais tipos de produção técnica (apresentações de trabalhos)	() SIM      () NÃO	%	
TOTAL:		100%	
Orientação concluída			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) tese doutorado	() SIM      () NÃO	%	30%
2) Dissertação de mestrado	() SIM      () NÃO	%	
3) Trabalho de conclusão	() SIM      () NÃO	%	
4)Especialização/Aperfeiçoamento	() SIM      () NÃO	%	
5) Iniciação científica	() SIM      () NÃO	%	
TOTAL:		100%	
Informações complementares			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Participações em banca de trabalhos de conclusão	() SIM      () NÃO	%	0%
2) Participações em eventos	() SIM      () NÃO	%	
3) Participações em banca de comissões julgadoras	() SIM      () NÃO	%	
4) Orientações em andamento	() SIM      () NÃO	%	
TOTAL:		100%	
Demais trabalhos relevantes			
	() SIM      () NÃO	%	0%
TOTAL Geral:			100%

Favor completar a Tabela Indicadores com a sua avaliação e caso deseje apresentar alguma observação, por favor, utilize o espaço reservado na página 6:



Produção Bibliográfica			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Artigos publicados em periódicos	( ) SIM ( ) NÃO	%	%
2) Trabalhos em anais de eventos	( ) SIM ( ) NÃO	%	
3) Livros ou Capítulos de livros	( ) SIM ( ) NÃO	%	
4) Textos em jornais ou revistas	( ) SIM ( ) NÃO	%	
5) Demais tipos de produção bibliográfico	( ) SIM ( ) NÃO	%	
<b>TOTAL:</b>		<b>100%</b>	
Produção Técnica			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Software	( ) SIM ( ) NÃO	%	%
2) Produtos tecnológicos	( ) SIM ( ) NÃO	%	
3) Trabalhos técnicos	( ) SIM ( ) NÃO	%	
4) Demais tipos de produção técnica (organização de eventos)	( ) SIM ( ) NÃO	%	
5) Demais tipos de produção técnica (relatórios de pesquisas)	( ) SIM ( ) NÃO	%	
6) Demais tipos de produção técnica (apresentações de trabalhos)	( ) SIM ( ) NÃO	%	
<b>TOTAL:</b>		<b>100%</b>	
Orientação concluída			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) tese doutorado	( ) SIM ( ) NÃO	%	%
2) Dissertação de mestrado	( ) SIM ( ) NÃO	%	
3) Trabalho de conclusão	( ) SIM ( ) NÃO	%	
4) Especialização/Aperfeiçoamento	( ) SIM ( ) NÃO	%	
5) Iniciação científica	( ) SIM ( ) NÃO	%	
<b>TOTAL:</b>		<b>100%</b>	
Informações complementares			
Indicadores de Produção	O Indicador é relevante para definir a autoridade do pesquisador?	Grau de Importância (%)	
1) Participações em banca de trabalhos de conclusão	( ) SIM ( ) NÃO	%	%
2) Participações em eventos	( ) SIM ( ) NÃO	%	
3) Participações em banca de comissões julgadoras	( ) SIM ( ) NÃO	%	
4) Orientações em andamento	( ) SIM ( ) NÃO	%	
<b>TOTAL:</b>		<b>100%</b>	
Demais trabalhos relevantes			
	( ) SIM ( ) NÃO	%	%
<b>TOTAL Geral:</b>			<b>100%</b>

## ANEXO C TEXTO DO TERCEIRO EXPERIMENTO

O texto a seguir refere-se ao texto enviado a amostra selecionada para participação do Experimento 3.

---

Caros (as) colegas,

estou enviando para voces um experimento referente a minha tese. Gostaria muito de contar com sua contribuicao.

Este experimento possui uma duracao aproximada de 15 minutos (máximo), para poder responde-lo algumas consideracoes sao necessarias, por favor leia as mesmas com atencao:

Definicoes:

### 1) Sistemas de Recomendacao:

Tratam-se de sistemas que procuram recomendar itens que seriam de interesse de um usuario conforme o perfil do mesmo. Procuram filtrar informacoes para o usuario. Exemplo: sites de comércio eletrônico que procuram recomendar itens para o cliente, baseando-se no perfil de consumo do mesmo.

### 2) Sistemas de Recomendacao colaborativos:

Tratam-se de sistemas que procuram recomendar itens a um usuário baseando-se na opinião de outros usuários consumidores que ja avaliaram um determinado item (levam em conta a opiniao da comunidade). Este sistema basicamente procura verificar usuarios com mesmos gostos por itens semelhantes (usando para tanto a pontuacao (rating) fornecida pelos usuarios avaliadores).

Neste tipo de aplicacao, com base na opiniao dos usuarios que possuem gostos semelhantes ao usuario alvo (aquele que espera por uma recomendacao), o sistema procura predizer a nota que este usuário alvo daria para um item ainda nao avaliado pelo mesmo.

Exemplo, supondo que o item X não tenha sido avaliado/recomendado para o usuário Silvio, e que os Usuários Maria, João e Paulo possuem gosto muito parecido com o Silvio (pontuam os itens de forma semelhante) e já avaliaram este item X. Desta forma o sistema procura prever a nota que o Silvio daria ao item, supondo que a nota calculada em uma escala Likert de 1-5 pontos tenha sido 4.5, o sistema entendera que o Silvio estaria interessado em receber a recomendação deste item X e o recomenda.

Neste tipo de aplicação não existe nenhuma preocupação com a relevância da opinião de Maria e João ao analisarem o item X, ou seja, o sistema não analisa o quanto estes são competentes para avaliarem este tipo de item.

### 3) Relevância da opinião do usuário (Ranqueamento do Recomendador):

Este valor constitui-se em uma métrica que procura analisar a autoridade do recomendador (utilizando uma série de atributos selecionados de seu perfil) ao pontuar um item X, ou seja, o quanto a opinião do avaliador é relevante naquela área de conhecimento.

Com esta métrica procura-se permitir que o usuário alvo de recomendação consiga identificar as recomendações feitas por pessoas com maior conhecimento da área do item (autoridade).

Experimento:

- Cenário:

Supondo que você seja um pesquisador iniciando estudos na Área de Conhecimento referente a Sistemas de Recomendação, ou seja, você iniciou a busca por material relevante e confiável para que possa obter conhecimento desta área. Você poderia reportar-se a um sistema de recuperação de informação (exemplo, google) e começar a busca, o que seria exaustivo pois você não tem conhecimento de que material seria relevante para ler.

Visando auxiliar seu início de busca lhe foi ofertado um Sistema de Recomendação, o qual apresenta duas alternativas de apresentação de seu conteúdo (Alternativa 1 e Alternativa 2).

#### 1) Alternativa 1 para Recomendações:

Nesta primeira alternativa as recomendações são apresentadas para você da seguinte forma: Nome do artigo recomendado, predição (escala de 1 até 5 significando o quanto o sistema "acredita" que você gostará do item; não gostará do artigo=1 e gostará muito do artigo=5) feita pelo sistema baseando-se em avaliações (hipotéticas) feitas por você a outros itens já recomendados previamente. Você dispõe da opção de fazer download do artigo e depois é solicitado a contribuir com o sistema fornecendo um feedback pessoal sobre o mesmo (escala de 1 (não gostei) até 5 (adorei) o artigo).

Por favor, acesse a página de recomendação 1 abaixo, antes da segunda alternativa, e analise bem a mesma:

[http://www.inf.unisinos.br/~cazella/exp/recomendacoes\\_prev.html](http://www.inf.unisinos.br/~cazella/exp/recomendacoes_prev.html)

#### 2) Alternativa 2 para Recomendações:

Nesta segunda alternativa as recomendações são apresentadas para você da seguinte forma: Nome do artigo recomendado, Ranqueamento do Recomendador (uma média da relevância da opinião (autoridade) dos usuários que auxiliaram o sistema a calcular esta recomendação; a relevância constitui-se em um valor de 0 (nenhuma relevância de opinião para esta área de conhecimento) até 10 (relevância de opinião muito alta para esta área de conhecimento)). Em outras palavras, o sistema mostra que o grupo que avaliou este artigo possui grande experiência e conhecimento desta área de Sistemas de Recomendação.

Você dispõe da opção de fazer download do artigo e depois é solicitado a contribuir com o sistema fornecendo um feedback pessoal sobre o mesmo (escala de 1 (não gostei) até 5 (adorei) o artigo).

Por favor, acesse a página de recomendação 2 abaixo e analise bem a mesma:

<http://www.inf.unisinos.br/~cazella/exp/recomendacoes.html>

---

#### QUESTÕES A SEREM RESPONDIDAS:

Tendo visualizado as duas alternativas de recomendação por favor, responda as questões a seguir usando a sua percepção pessoal: ( lembre-se que você é um iniciante nesta área de conhecimento buscando por material relevante)

#### QUESTÕES:

1) A inclusão do Ranqueamento do Recomendador neste contexto seria um fator que facilitaria a seleção de itens recomendados ao usuário?

Sim

Não

2) Você ao verificar a recomendação com a inclusão da relevância da opinião do usuário sentiu-se mais seguro ao decidir que item ler/acessar primeiro devido ao sistema mostrar que este indivíduo tem maior ou menor autoridade na área?

Sim

Não

3) Em um contexto de recomendação de itens este fator (Ranqueamento do Recomendador) poderia ser um direcionamento para seleção de itens, diminuindo a sensação de sobrecarga de informação comum em sistemas de recuperação de informação (exemplo, google):

Sim

Não

4) Onde você verifica aplicação deste tipo de solução com ganho para o usuário final?

Outras considerações que você considera relevante:

Grato pela contribuição, Silvio Cesar Cazella

## ANEXO D TEXTO DO QUARTO EXPERIMENTO

O texto a seguir refere-se ao texto enviado a amostra selecionada para participação do Experimento 4.

---

Caro(a) colega,

grato por ter auxiliado na realização (com sucesso) do primeiro experimento da tese. Estou enviando o segundo e-mail (serão um total de 4) com a segunda parte do experimento.

Nesta etapa, solicito que você analise dois CV-LATTES (em anexo), e forneça uma avaliação para a relevância da opinião do acadêmico para as ÁREAS DE INTERESSE abaixo. A avaliação deverá ser um valor entre 0 (nenhuma relevância) e 10 (total relevância) para a opinião dos acadêmicos.

Por favor, verifique o CV-LATTES e não a pessoa, ou seja, se conhecer a mesma tente focar somente no CV-LATTES desta e desconsidere aspectos pessoais.

O tempo necessário para responder a esta consulta fica em torno de 15 minutos.

Agradeceria muito se a resposta me fosse enviada em uma semana, se possível e claro;-).

ÁREAS DE INTERESSE:

Caso o acadêmico em questão fosse avaliar um Artigo Científico pertencente às áreas abaixo, ao meu ver, o peso da opinião deste acadêmico seria:

Acadêmico 1:

- ENGENHARIA DE SOFTWARE => Peso=??
- INTELIGÊNCIA ARTIFICIAL=> Peso=??

Acadêmico 2:

158

- ENGENHARIA DE SOFWTARE => Peso=??
- INTELIGENCIA ARTIFICIAL=> Peso=??

---

As informacoes aqui prestadas serao de aspecto sigiloso, ou seja, de forma alguma serao divulgadas as relacoes de avaliadores e avaliados.

Atenciosamente,

--

Silvio Cesar Cazella

2 attachments — Download all attachments

Currículo do Sistema de Currículos Lattes (ACADEMICO).htm  
71K View Download

Currículo do Sistema de Currículos Lattes (ACADEMICO).htm  
300K View Download

## ANEXO E TEXTO DO SEXTO EXPERIMENTO

### Fase I: Coleta de Artigos

Caro(a) colega,

primeiramente gostaria de agradecer a sua colaboração nesta finalização da minha tese.

Este experimento que estou iniciando com sua participação visa coletar algumas informações sobre Artigos Científicos que você indicaria como de leitura fundamental aos seus alunos iniciantes (suponha este aluno como sendo aluno de mestrado) na sua área de pesquisa, bem como sua preferência com relação a alguns Artigos Científicos.

Este experimento ocorrerá em duas etapas:

- 1) nesta etapa você será solicitado a indicar alguns artigos que você consideraria ser de importância fundamental para seus alunos iniciantes em sua área de pesquisa;
- 2) em uma segunda etapa você será solicitado a avaliar uma pequena amostra de artigos quanto a relevância/qualidade destes segundo sua percepção.

Desta forma solicito seu auxílio na coleta destas informações:

Etapa 1:

- 1) Informar um número mínimo de CINCO artigos que você considera de leitura obrigatória aos seus estudantes iniciantes de mestrado (você pode indicar mais de cinco, se assim o desejar):

Área de Pesquisa:

Título do Artigo:

Autor(es):

Fonte: Caso você possua o mesmo em meio digital, por favor, envie para que possamos abastecer a base de dados do sistema.

Se for possível informar estes artigos em uma semana da data deste e-mail, eu agradeceria muito.

PS: Maiores informações sobre a tese em: <http://www.inf.unisinos.br/~cazella/phd.html>

Grato pela colaboracao.

Silvio Cesar Cazella

**Fase II: Coleta de Avaliações Preliminares**

Caro(a) colega,

gostaria de agradecer a sua colaboracao nesta finalizacao da Tese. Em e-mail anterior voce foi solicitado(a) a indicar Artigos Cientificos que voce recomendaria na sua area. Com o seu auxilio foi possivel a construcao de uma base significativa de artigos científicos (obrigado).

Atraves deste e-mail estou dando continuidade ao ultimo experimento da tese.

a) Objetivo do experimento:

Coletar avaliacoes aos artigos científicos, buscando permitir a criacao de uma base de artigos com a avaliacao da comunidade de pesquisadores.

b) Este experimento sera conduzido da seguinte maneira:

- 1) Neste e-mail voce esta recebendo a recomendacao de alguns artigos que poderiam ser de seu interesse (alguns são indiretamente relacionados a sua área de pesquisa e alguns procuram "testar" o seu interesse para melhorar seu perfil no sistema) para avaliar utilizando uma escala de [1-5] pontos (sendo 1 indicativo de que voce nao gostou do artigo e nao recomendaria o mesmo, e 5 indicando que voce gostou muito do artigo e certamente recomendaria o mesmo).

Estes artigos deverao ser avaliados quanto ao interesse gerado em voce, adequacao da recomendação para voce e se voce recomendaria o mesmo para que outros pesquisadores. Por favor não avalie o artigo como em uma analise de artigo para congresso ou periodico, limite-se a avaliar o quanto este artigo foi interessante para voce, e voce acredita ser interessante para outros . Por exemplo, considere que voce buscou conteudo no google e este lhe apresenta links, ao acessar um dos links voce avalia o quanto este link foi interessante quanto a disponibilizacao das informações que você precisava/buscava (nem sempre o conteudo eh perfeito e completo mas nos auxilia em termos de entendimento de um dado assunto). Voce nao estaria avaliando apresentacao de conteudo, disposicao da informacao, recursos graficos, e sim o conteudo em sim quanto a adequacao ao que voce precisava ou se interessava.

- 2) Em um segundo e-mail (apos o retorno de suas avaliacoes dos artigos constantes neste e-mail) voce estara recebendo a recomendacao de novos artigos para avaliar utilizando a mesma escala e os mesmos criterios.

A razao destes dois e-mails (gerados nas etapas acima) deve-se a necessidade da utilizacao de tecnicas distintas de recomendacao para avaliacao do experimento.



Por favor acesse os arquivos attached e retorne suas avaliacoes em uma escala de [1-5] ate a semana que vem (19/05/2006).

Gostaria de ressaltar que as avaliacoes, bem como comentarios que voce venha a disponibilizar nao serao publicados vinculando voce aos mesmos, ou seja, sua privacidade sera preservada.

RESPOSTAS:

Arquivo	Avaliacao [1-5]

PS:Maiores informacoes sobre a tese em: <http://www.inf.unisinos.br/~cazella/phd.html>

Grato pela colaboracao,

--

Silvio Cesar Cazella

### **Fase III: Recomendações**

Caro(a) colega,

gostaria de agradecer a sua colaboracao nesta finalizacao da Tese. Em e-mail anterior voce foi solicitado(a) a avaliar um total de cinco Artigos Cientificos, e com base em suas avaliacoes e as avaliacoes de outros colegas foi possivel ao sistema W-RECMAS predizer artigos que seriam considerados de seu interesse.

Atraves deste e-mail estou dando finalização ao ultimo experimento desta tese e solicito, mais uma vez, a sua colaboracao. Junto deste e-mail encontram-se attached alguns artigos considerados de possivel interesse para voce, ou seja, o sistema esta lhe recomendando estes artigos.

a) Objetivo do experimento:

Avaliar a qualidade da recomendacao gerada pelo sistema W-RECMAS (predicao) atraves do feedback dos recomendados.

b) Este experimento sera conduzido da seguinte maneira:

Neste e-mail voce esta recebendo a recomendacao de alguns artigos que foram considerados de seu interesse pelo sistema. Por favor, avalie os mesmos fornecendo

um retorno sobre a recomendação feita pelo sistema. Avalie utilizando uma escala de [1-5] pontos (sendo 1 indicativo de que voce nao gostou do artigo e nao recomendaria o mesmo, e 5 indicando que voce gostou muito do artigo e certamente recomendaria o mesmo).

Estes artigos deverao ser avaliados quanto ao interesse gerado em voce, adequacao da recomendação para voce e se voce recomendaria o mesmo para que outros pesquisadores. Por favor nao avalie o artigo como em uma analise de artigo para congresso ou periodico, limite-se a avaliar o quanto este artigo foi interessante para voce.

Por favor acesse os arquivos atachados e retorne suas avaliacoes em uma escala de [1-5] ate o dia 18/06/2006.

Gostaria de ressaltar que as avaliacoes, bem como comentarios que voce venha a disponibilizar nao serao publicados vinculando voce aos mesmos, ou seja, sua privacidade sera preservada.

RESPOSTAS:

---

Arquivo	Avaliacao [1-5]
---------	-----------------

---

PS: Maiores informacoes sobre a tese em: <http://www.inf.unisinos.br/~cazella/phd.html>

Grato pela colaboracao,

--

Silvio Cesar Cazella

## ANEXO F BANCO DE DADOS

Neste Anexo é apresentada a descrição das tabelas que compõem o banco de dados do protótipo implementado.

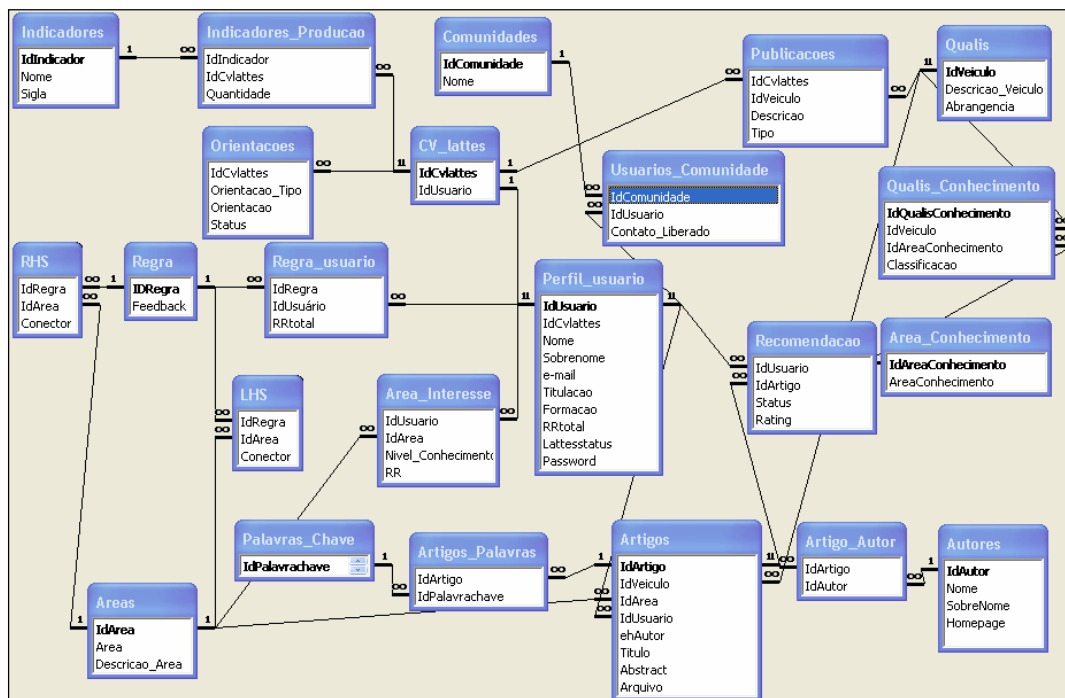


Figura A.1: Banco de dados do protótipo

## Descrição das Tabelas:

Tabela		Indicadores		
Definição		Representa os indicadores de produção apresentados no CV-Lattes.		
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdIndicador	Código único do identificador, chave primária.	Sim	Sim	Não
Nome	Nome do Indicador no CV-Lattes.	Não	Não	Não
Sigla	Sigla de representação do Indicador no CV-Lattes.	Não	Não	Não

Tabela		Indicadores_Producao		
Definição		Representa os valores dos indicadores de produção apresentados no CV-Lattes de cada usuários do sistema.		
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdIndicador	Código do identificador, chave estrangeira.	Sim	Não	Sim
IdCvlattes	Código do CV-Lattes, chave estrangeira.	Sim	Não	Sim
Quantidade	Número de itens em cada indicador, conforme o CV-Lattes de cada usuário.	Não	Não	Não

Tabela		CV_lattes		
Definição		Representa a referência ao CV-Lattes do usuário (pesquisador).		
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdCvlattes	Código único do identificador, chave primária.	Sim	Sim	Não

Tabela	CV_lattes				
Definição	Representa a referência ao CV-Lattes do usuário (pesquisador).				
Atributos:					
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira	
IdUsuario	Código único do identificador, estrangeira. chave	Sim	Não	Sim	

Tabela	Orientações				
Definição	Representa as orientações realizadas e que se encontram descritas no CV-Lattes do usuário (pesquisador).				
Atributos:					
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira	
IdCvlattes	Código do identificador do CV-Lattes na Base de dados, chave estrangeira.	Sim	Não	Sim	
Orientacao_Tipo	Especifica se a orientação e de graduação ou pós.	Não	Não	Não	
Orientacao	Descrição textual da orientação.	Não	Não	Não	
Status	Especifica se a orientação foi finalizada ou está em andamento.	Não	Não	Não	

Tabela	Qualis				
Definição	Representa o conteúdo da classificação Qualis da Capes.				
Atributos:					
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira	
IdVeiculo	Código único do veículo de publicação na Base de dados, chave primária.	Sim	Sim	Não	
Descricao_Veiculo	Descrição textual do veículo de publicação.	Não	Não	Não	

Tabela	Qualis			
Definição	Representa o conteúdo da classificação Qualis da Capes.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
Abrangencia	Especifica se o veículo é de alcance local, nacional ou internacional.	Não	Não	Não

Tabela	Qualis_Conhecimento			
Definição	Representa o conteúdo da classificação Qualis da Capes.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdQualisConhecimento	Código único da relação do veículo do qualis e suas áreas de conhecimento na Base de dados, chave primária.	Sim	Sim	Não
IdVeiculo	Código único do veículo de publicação na Base de dados, chave estrangeira.	Sim	Não	Sim
IdAreaConhecimento	Código único da área de conhecimento na Base de dados, chave estrangeira.	Sim	Não	Sim
Classificacao	Apresenta a classificação atual do veículo que pode ser A, B, C ou D.	Sim	Não	Não

Tabela	Area_Conhecimento			
Definição	Representa as áreas de conhecimento em que os veículos de publicação são cadastrados.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdAreaConhecimento	Código único da área do conhecimento na Base de dados, chave primária.	Sim	Sim	Não
AreaConhecimento	Descrição textual da área de conhecimento.	Sim	Não	Não

<b>Tabela</b>	Publicacoes			
<b>Definição</b>	Representa as publicações realizadas e que se encontram descritas no CV-Lattes do usuário (pesquisador).			
<b>Atributos:</b>				
<b>Nome</b>	<b>Definição</b>	<b>Requerido</b>	<b>Chave Primária</b>	<b>Chave Estrangeira</b>
IdCvlattes	Código do identificador do CV-Lattes na Base de dados, chave estrangeira.	Sim	Não	Sim
IdVeiculo	Código do identificador do Veículo de publicação na Base de dados, chave estrangeira.	Sim	Não	Sim
Descricao	Descrição textual da publicação.	Não	Não	Não
Tipo	Informa se a publicação foi feita em livro, congresso, etc.	Não	Não	Não

<b>Tabela</b>	Regra			
<b>Definição</b>	Representa a regra que relaciona áreas de interesse (são as regras geradas pela aplicação do F-IUP).			
<b>Atributos:</b>				
<b>Nome</b>	<b>Definição</b>	<b>Requerido</b>	<b>Chave Primária</b>	<b>Chave Estrangeira</b>
IdRegra	Código único da regra na Base de dados, chave primária.	Sim	Sim	Não
Feedback	Representa o retorno da comunidade sobre a regra, ou seja, ela foi testada e o feedback pode representar sucesso ou não.	Não	Não	Não

Tabela	RHS			
Definição	Representa as áreas de interesse que compõem o Right Hand Side da regra.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdRegra	Código único da regra na Base de dados, chave estrangeira.	Sim	Não	Sim
IdArea	Código único da área na Base de dados, chave estrangeira.	Sim	Não	Sim
Conector	Operador lógico da regra (or, not, and).	Não	Não	Não

Tabela	LHS			
Definição	Representa as áreas de interesse que compõem o Left Hand Side da regra.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdRegra	Código único da regra na Base de dados, chave estrangeira.	Sim	Não	Sim
IdArea	Código único da área na Base de dados, chave estrangeira.	Sim	Não	Sim
Conector	Operador lógico da regra (or, not, and).	Não	Não	Não

Tabela	Regra_usuario			
Definição	Representa o RRtotal de um usuário de acordo com as áreas de uma determinada regra.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdRegra	Código único da regra na Base de dados, chave estrangeira.	Sim	Não	Sim



Tabela		Regra_Usuario		
Definição		Representa o RRtotal de um usuário de acordo com as áreas de uma determinada regra.		
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdUsuario	Código único do usuário na Base de dados, chave estrangeira.	Sim	Não	Sim
RRtotal	Ranqueamento do recomendador para a regra em questão.	Não	Não	Não

Tabela		Area_Interesse		
Definição		Representa as áreas de interesse de um determinado usuário e o seu respectivo RR para a área.		
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdUsuario	Código único do usuário na Base de dados, chave estrangeira.	Sim	Não	Sim
IdArea	Código único da área na Base de dados, chave estrangeira.	Sim	Não	Sim
Nível_Conhecimento	Representa o quanto o pesquisador conhece uma determinada área (Iniciante, Intermediário ou Avançado).	Não	Não	Não
RR	Ranqueamento do recomendador em uma determinada área de interesse.	Não	Não	Não

<b>Tabela</b>	Perfil_usuario			
<b>Definição</b>	Representa o perfil completo do pesquisador.			
<b>Atributos:</b>				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdUsuario	Código único do pesquisador na Base de dados, chave primária.	Sim	Sim	Não
IdCvlattes	Código único do CV-Lattes do pesquisador na Base de dados, chave estrangeira.	Sim	Não	Sim
Nome	Nome do pesquisador.	Não	Não	Não
Sobrenome	Sobrenome do pesquisador.	Não	Não	Não
e-mail	E-mail de contato do pesquisador.	Não	Não	Não
Titulação	Titulação acadêmica do pesquisador. Pré-definida por Graduando, Graduado, Especialista, Mestrando, Mestre, Doutor, Doutorando e Pós-doutorado.	Não	Não	Não
FO	Curso de Formação acadêmica.	Não	Não	Não
RRtotal	Ranqueamento do Recomendador para as áreas de interesse do pesquisador.	Não	Não	Não
Lattesstatus	Atualizado ou não atualizado.	Não	Não	Não
Password	Senha de acesso.	Não	Não	Não

<b>Tabela</b>	Usuarios_Comunidade			
<b>Definição</b>	Representa as comunidades virtuais nas quais o usuário encontra-se inserido.			
<b>Atributos:</b>				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdComunidade	Código único da comunidade na Base de dados, chave estrangeira.	Sim	Não	Sim

Tabela	Usuarios_Comunidade			
Definição	Representa as comunidades virtuais nas quais o usuário encontra-se inserido.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdUsuario	Código único do usuário na Base de dados, chave estrangeira.	Sim	Não	Sim
Contato_Liberado	Significa que o usuário permitiu que um determinado usuário entre em contato com ele (constitui-se em um código de usuário).	Não	Não	Não

Tabela	Comunidades			
Definição	Representa as várias comunidades criadas e disponíveis no sistema.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdComunidade	Código único da comunidade na Base de dados, chave primária.	Sim	Sim	Não
Nome	Nome da comunidade.	Não	Não	Não

Tabela	Recomendacao			
Definição	Representa as recomendações realizadas pelo sistema ao pesquisador.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdUsuario	Código único do pesquisador na Base de dados, chave estrangeira.	Sim	Não	Sim

<b>Tabela</b>	Recomendacao			
<b>Definição</b>	Representa as recomendações realizadas pelo sistema ao pesquisador.			
<b>Atributos:</b>				
<b>Nome</b>	<b>Definição</b>	<b>Requerido</b>	<b>Chave Primária</b>	<b>Chave Estrangeira</b>
IdArtigo	Código único do artigo na Base de dados, chave estrangeira.	Sim	Não	Sim
Status	Informa se o item está para ser recomendado - RF (recomendações futuras), ou se o item foi recomendado e já foi avaliado pelo usuário - RCF (recomendações com feedback), ou está aguardando uma avaliação - RSF (recomendações sem feedback).	Não	Não	Não
Rating	Valor da avaliação do item.	Não	Não	Não

<b>Tabela</b>	Areas			
<b>Definição</b>	Representa as áreas de interesse oferecidas pelo sistemas.			
<b>Atributos:</b>				
<b>Nome</b>	<b>Definição</b>	<b>Requerido</b>	<b>Chave Primária</b>	<b>Chave Estrangeira</b>
IdArea	Código único da área de interesse na Base de dados, chave primária.	Sim	Sim	Não
Area	Nome da área de interesse.	Não	Não	Não
Descricao_Area	Descrição detalhada da área de interesse.	Não	Não	Não

Tabela	Palavraschave			
Definição	Representa as palavras-chave que podem identificar os artigos.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
Id Palavrachave	Código único da palavra na Base de dados, chave primária.	Sim	Sim	Não
Palavrachave	Palavra.	Não	Não	Não

Tabela	Artigos_Palavras			
Definição	Representa as palavras-chave que identificam um ou mais artigos.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdArtigo	Código único do artigo na Base de dados, chave estrangeira.	Sim	Não	Sim
IdPalavra	Código único da palavra na Base de dados, chave estrangeira.	Sim	Não	Sim

Tabela	Artigos			
Definição	Representa os artigos disponíveis no sistema e que serão os itens a serem recomendados para cada pesquisador.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdArtigo	Código único do artigo na Base de dados, chave primária.	Sim	Sim	Não

<b>Tabela</b>	Artigos			
<b>Definição</b>	Representa os artigos disponíveis no sistema e que serão os itens a serem recomendados para cada pesquisador.			
<b>Atributos:</b>				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdVeiculo	Código único do veículo na Base de dados, chave estrangeira. Relaciona o artigo ao veículo de publicação e a área de conhecimento (ciência da computação, engenharia ou multidisciplinar, por exemplo).	Sim	Não	Sim
IdUsuario	Código único do usuário na Base de dados, chave estrangeira.	Sim	Não	Sim
IdArea	Código único da área na Base de dados, chave estrangeira.	Sim	Não	Sim
ehAutor	Informa se a pessoa que cadastrou o artigo no sistema é um dos autores.	Sim	Não	Não
Titulo	Título completo do artigo.	Sim	Não	Não
Resumo	Resumo completo do artigo.	Sim	Não	Não
Arquivo	Arquivo em formato pdf ou ps.	Sim	Não	Não

<b>Tabela</b>	Artigo_Autor			
<b>Definição</b>	Representa todos os autores de um determinado artigo.			
<b>Atributos:</b>				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdArtigo	Código único do artigo na Base de dados, chave estrangeira.	Sim	Não	Sim
IdAutor	Código único do autor na Base de dados, chave estrangeira.	Sim	Não	Sim

Tabela	Autores			
Definição	Representa a lista de todos os autores dos artigos.			
Atributos:				
Nome	Definição	Requerido	Chave Primária	Chave Estrangeira
IdAutor	Código único do autor na Base de dados, chave primária.	Sim	Sim	Não
Nome	Nome do autor.	Sim	Não	Não
SobreNome	Sobrenome do autor.	Sim	Não	Não
HomePage	Homepage do Autor.	Não	Não	Não

## ANEXO G COMUNIDADES VIRTUAIS

Neste Anexo apresenta-se o início de um dos trabalhos futuros propostos por esta tese de doutorado que se constitui em utilizar Sistemas de Recomendação para identificação de comunidades virtuais. Este conteúdo foi aqui incluído por tratar-se de um trabalho inicial que depende de estudo mais aprofundado, tendo sido desenvolvido em forma de pesquisa pelo doutorando e seus orientandos. O objetivo da inclusão deste conteúdo aqui é apresentar a continuidade do trabalho em questão. Primeiramente busca-se definir o que se entende pelo termo Comunidade. Segundo o autor Case [CAS 2001], pode-se definir o termo Comunidade como sendo:

*“...um local onde pessoas com interesses em comum reúnem-se para trocar experiências, responder perguntas, promover colaboração, obter conhecimentos e compartilhar casos de sucessos e insucessos”.*

As primeiras formas de Comunidades Virtuais foram criadas baseadas em usuários que poderiam encontrar-se e discutir temas não interessando a distância que estes encontravam-se [HAU 97] [FIG 98]. Conforme Rheingold [RHE 2000], uma Comunidade Virtual representa o surgimento de comunidades socialmente motivadas pelo interesse mútuo na Internet. O mesmo autor descreve comunidades virtuais como sendo:

*"...comunidade virtuais são agregações sociais que emergem da rede, quando um número suficiente de pessoas envolvem-se nesta discussão pública por um determinado período, visando formar redes de relacionamento utilizando o cyberspace".*

Uma comunidade torna-se uma excelente fonte de conhecimento e informação, e talvez o mais importante, esta auxilia na diminuição da duplicação de esforço, uma vez que se pode basear na experiência que os membros tiveram no passado. Por exemplo, pode-se imaginar uma situação hipotética onde um estudante da graduação necessita iniciar uma pesquisa para realizar seu trabalho de conclusão de curso, este estudante poderá perder muito tempo nesta busca por material, passando por caminhos que levarão ao sucesso ou não, caminhos talvez trilhados anteriormente por outros estudantes, ou seja, esforço já despendido por outros. Entretanto, caso este estudante estivesse escrito em uma comunidade com indivíduos com interesses em comum, estes poderiam fornecer-lhe alguns atalhos nesta busca por informação, talvez diminuindo o tempo perdido pelo aluno na busca por informação [CAZ 2004].

Segundo Case [CAS 2001], para se criar uma comunidade é necessário enfrentar alguns desafios específicos:



- os membros da comunidade deveriam ser direcionados diretamente para pessoas que pensam de forma semelhante a sua (*like-minded*) e para informações relevantes;
- os membros da comunidade deveriam ter acesso a informação requerida sem sentirem-se sobrecarregados;
- os membros da comunidade *deveriam ser informados sobre outros membros que pensam de forma semelhante a sua (like-minded)*;
- os membros deveriam disseminar informação para as pessoas apropriadas.

Algumas comunidades virtuais são amplamente divulgadas na Internet, possuindo um número significativo de membros segundo apresentado na Tabela G.1, proposta por Golbeck [GOL 2005]:

Tabela G.1: Número de membros em comunidades virtuais

Website	URL	Número de membros
Adult Friend Finder	<a href="http://adultfriendfinder.com">http://adultfriendfinder.com</a>	15.700.000
Alt.com	<a href="http://alt.com">http://alt.com</a>	2.600.000
Amigos	<a href="http://amigo.com">http://amigo.com</a>	3.500.000
Asia Friend Finder	<a href="http://asiafriendfinder.com">http://asiafriendfinder.com</a>	6.000.000
Black Planet	<a href="http://blackplanet.com">http://blackplanet.com</a>	14.000.000
Fotolog	<a href="http://fotolog.net">http://fotolog.net</a>	1.000.000
Friend Finder	<a href="http://friendfinder.com">http://friendfinder.com</a>	3.600.000
Friendster	<a href="http://friendster.com">http://friendster.com</a>	17.000.000
gradFinder	<a href="http://www.grafinder.com">http://www.grafinder.com</a>	3.000.000
Hi5	<a href="http://hi5.com">http://hi5.com</a>	6.100.000
LinkedIn	<a href="http://linkedin.com">http://linkedin.com</a>	1.500.000
LiveJournal	<a href="http://livejournal.com">http://livejournal.com</a>	5.700.000
My Space	<a href="http://myspace.com">http://myspace.com</a>	6.000.000
Orkut	<a href="http://orkut.com">http://orkut.com</a>	3.000.000
Out Personals	<a href="http://outperonals.com">http://outperonals.com</a>	1.050.000
The Face Book	<a href="http://thefacebook.com">http://thefacebook.com</a>	1.000.000
Tickle	<a href="http://tickle.com">http://tickle.com</a>	18.000.000
ZeroDegrees	<a href="http://zerodegrees.com">http://zerodegrees.com</a>	1.300.000

Fonte: [GOL 2005], p. 16.

## **G1. Criação de comunidades virtuais**

Conforme descrito na Subseção 5.2.5 a identificação e criação das comunidades virtuais (ou de contato) nesta tese, ficou a cargo do Agente Comunitário. Para identificar os possíveis candidatos as comunidades o Agente Comunitário procura verificar na base de perfis os usuários que apresentam interesses em comum, na base de consumo histórico os que são *like-minded* e que apresentam um certo grau de confiança entre si (indicação não explícita e sim implícita) conforme apresentado pela regra da Subseção 5.2.5. Na subseção seguinte será descrito um experimento inicial, realizado sobre uma base de dados sintética, e que demonstra a aplicabilidade da solução descrita na subseção sobre o Agente Comunitário na construção destas comunidades virtuais.

### **G1.1 Experimento com base de dados sintética**

Este experimento teve como objetivo demonstrar a viabilidade da criação de comunidades virtuais a partir da aplicação de uma base de dados sintética gerada a partir de uma base de dados real proveniente de um Sistema de Recomendação. Neste trabalho foi utilizada a base de dados do Movielens adaptada e acrescida de um campo referente a relevância da opinião de cada usuário [SIL 2005].

#### *G1.1.1 Hipótese da Pesquisa*

A hipótese que foi levantada para realização deste experimento é a de que com a utilização de uma base de dados proveniente de um Sistema de Recomendação real e adaptada, seria possível a criação das comunidades virtuais pela identificação de usuários com interesses em comum, usuários com gostos em comum (*like-minded*) e usuários que apresentassem implicitamente uma relação de confiança.

#### *G1.1.2 Método de Pesquisa*

O método aplicado neste experimento constituiu-se na análise *offline* utilizando a base sintética gerada a partir da base de dados do Movielens.

#### *G1.1.3 Experimento realizado*

Neste experimento havia a necessidade de medir a similaridade entre usuários para permitir a construção de comunidades virtuais, para tanto foi utilizado o coeficiente de Pearson utilizado na filtragem colaborativa e descritos da Subseção 2.4.2.

##### *G1.1.3.1 Gerando comunidades virtuais*

Para que dois ou mais usuários sejam identificados como usuários com potencial (candidatos) para a criação de uma comunidade virtual, três critérios devem ser satisfeitos conforme definido na subseção 5.2.5 e rerepresentado a seguir:

- possuírem interesse comum em pelo menos uma área de interesse;
- serem *like-minded*;
- apresentar relação de confiança (implícita).

Para a criação das comunidades virtuais de usuários neste experimento, primeiramente foi preciso encontrar na base de perfis da base sintética quais usuários possuíam áreas de interesse em comum com outros usuários. Esta análise retornou um conjunto de relações usuário x usuário com áreas de interesse em comum. Por exemplo,

supondo uma relação entre os usuários U1 e U3 na área de interesse A1, e uma relação entre os usuários U1 e U4 na área de interesse A2, surge a seguinte relação “usuário X usuário X área de interesse em comum” como apresentado na tabela G.2.

Tabela G.2: Relação de usuários e áreas de interesse comum

Usuário1	Usuário2	Área Comum
U1	U3	A1
U1	U4	A2

Cabe ressaltar que se caso dois usuários possuíssem mais de uma área de interesse em comum, esta relação também seria incluída no resultado (como exemplo, supondo que U1 e U3, além da área A1, também tivessem a área A2).

Na base de dados sintética foram encontradas 670.760 relações para 643 usuários em 18 áreas de interesse.

Após descobrir estas relações, foi preciso calcular a correlação entre estes usuários e suas respectivas áreas de interesse em comum através da aplicação do coeficiente de Pearson. A idéia ao se aplicar o coeficiente de Pearson foi manter na base somente os usuários, que além de possuírem áreas de interesse em comum, também possuíssem forte correlação (coeficiente  $> 0$ ), descartando os de fraca correlação e os sem correlação (coeficiente  $\leq 0$ ) – lembrando que este coeficiente é calculado de acordo com os itens em comum pontuados pelos usuários. Assumiu-se que para a segunda fase, os usuários deveriam possuir pelo menos 3 itens em comum pontuados dentro da mesma área de ineteresse, caso contrário esta relação seria descartada.

Na base de dados sintética a área de interesse “A1” é a que concentrou o maior número de correlações fortes entre os usuários (aproximadamente 11.000 relações), seguida da área de “A2”. Das 670.760 relações encontradas anteriormente, apenas 49.762 (7,42% do total) foram consideradas fortes pela análise feita. Esta diferença de valores entre a análise anterior e a atual, ocorreu devido ao fato de que as relações iniciais entre os usuários eram fracas de acordo com o coeficiente de Pearson ( $\geq 0$ ), ou os itens pontuados em comum em uma mesma área de interesse na base de dados sintética não atingiam o parâmetro mínimo configurado de “3” itens.

A área de interesse “A3” foi a que apresentou o melhor coeficiente de correlação dos usuários, podendo-se concluir que as pontuações dadas pelos usuários aos itens relacionados a esta área foram bem parecidas, se comparadas as pontuações realizadas para os itens da área de “A2”.

Após descobrir as correlações fortes presentes na base, foi possível gerar as comunidades de contato para cada usuário. Nesta etapa foi simulada a recomendação de 3 artigos (itens) para os usuários alvos da recomendação e de acordo com o feedback dado foi possível verificar se a mesma estava dentro do conjunto predito para aquele item (assumindo-se que no domínio de valores possíveis para uma avaliação os valores de avaliações entre [1-2] indicam que o usuário não gostou da recomendação, [3] indica que o usuário ficou indiferente, e [4-5] indica que o usuário gostou da recomendação),

isto é, foi verificada a existência ou não de relação de confiança entre o usuário alvo e os usuários participantes do processo de recomendação.

Como resultado observou-se que o número de comunidades de contatos geradas para a área de interesse de “A1” foi muito superior às demais áreas, sendo criadas aproximadamente 530 comunidades de um total de 1421. Vale ressaltar que nesta etapa a quantidade de comunidades de contatos geradas para cada área pode variar de acordo com os itens selecionados para gerar a predição para um usuário alvo. Neste experimento, foram selecionados 3 itens aleatoriamente. Para a confirmação da criação da nova comunidade, por questões de privacidade, seria solicitado que os usuários envolvidos respondessem se realmente desejam conhecer o contato recomendado. Caso uma das partes discorde, o contato não seria estabelecido, permitindo assim, que o usuário tenha o controle sobre sua privacidade.

#### *GI.1.3.2 Avaliação dos Resultados*

Através dos resultados apresentados verificou-se a viabilidade de identificar comunidades virtuais utilizando dados provenientes de um Sistema de Recomendação. Pode-se observar uma limitação clara nesta abordagem no momento em que a base de dados sintética apresenta uma esparcialidade muito elevada ou um número de itens muito maior do que o número de usuários.

Em um trabalho futuro envolvendo usuários reais pretende-se utilizar o Ranqueamento do Recomendador como uma informação que permita facilitar a decisão do usuário sobre o estabelecimento de contato ou não com outros usuários e conseqüentemente a formação de comunidades com estes membros. Também seria interessante através de um uso prolongado do sistema verificar a satisfação dos usuários com as comunidades estabelecidas, ou seja, verificar o quanto à comunidade em que o usuário decidiu inserir-se trouxe benefícios e facilitou a troca de informações.