



RECDOC: UM SISTEMA DE RECOMENDAÇÃO PARA UMA BIBLIOTECA  
DIGITAL NA WEB

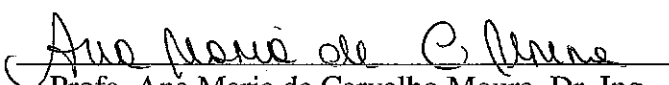
Catarina Carneiro Rocha

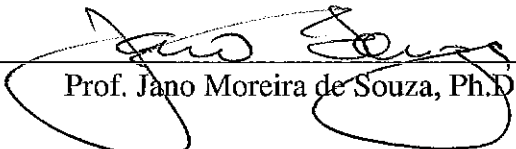
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

  
Prof. Geraldo Bonorino Xexéo, D.Sc

  
Profa. Ana Regina Cavalcanti da Rocha, D.Sc.

  
Profa. Ana Maria de Carvalho Moura, Dr. Ing.

  
Prof. Jano Moreira de Souza, Ph.D.

ROCHA, CATARINA CARNEIRO

RecDoc: um sistema de recomendação  
para uma biblioteca digital na Web [Rio de  
Janeiro] 2003

IX, 114 p. 29,7 cm (COPPE/UFRJ,  
M.Sc., Engenharia de Sistemas e  
Computação, 2003)

Tese - Universidade Federal do Rio de  
Janeiro, COPPE

1. Sistemas de Recomendação
2. Filtragem Colaborativa
3. Bibliotecas Digitais

I. COPPE/UFRJ II. Título ( série )

# **AGRADECIMENTOS**

---

À minha família, por seu irrestrito apoio, compreensão e dedicação. Por estar presente nos momentos felizes e também nos mais difíceis.

Aos meus orientadores Geraldo Xexéo e Ana Regina pela orientação, trocas de conhecimento e apóio.

Aos professores Jano e Ana Maria por aceitarem participar da banca e por suas sugestões para melhoria do trabalho.

Ao CNPQ pelo apoio financeiro.

À todos os amigos que compartilharam os momentos que passei ao longo do trabalho.

À Mariella e Mariano pelos momentos de companheirismo regados a cafés e filmes franceses.

À Carlinha, Ale e Ana Paula, por sua amizade, por sua presença e por todos os momentos compartilhados.

À Gabriela e Nicolaas pela amizade e apóio em diversos momentos.

À Carmen pelas revisões e sugestões que permitiram aprimorar o trabalho.

Ao pessoal administrativo da COPPE/Sistemas, Ana Paula, Patrícia e Solange.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

RECDOC: UM SISTEMA DE RECOMENDAÇÃO PARA UMA BIBLIOTECA  
DIGITAL NA WEB

Catarina Carneiro Rocha

Março/2003

Orientadores: Geraldo Bonorino Xexéo  
Ana Regina Cavalcanti da Rocha

Programa: Engenharia de Sistemas e Computação

Com a crescente disponibilidade de informação em ferramentas de busca na Web, bibliotecas digitais e bancos de dados on-line, é cada vez mais importante o desenvolvimento de sistemas personalizados que ajudem o usuário a encontrar informações relevantes. Sistemas de recomendação são serviços personalizados que vêm sendo desenvolvidos para ajudar as pessoas com a abundância e sobrecarga de informação, pois possibilitam que elas compartilhem opiniões e experiências.

Essa dissertação apresenta um sistema de recomendação RecDoc, desenvolvido para uma biblioteca digital virtual na Web. Recdoc indica documentos baseados na própria experiência do usuário, na similaridade entre documentos e na colaboração entre usuários.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## RECDOC: A RECOMMENDER SYSTEM TO A WEB DIGITAL LIBRARY

Catarina Carneiro Rocha

March/2003

Advisors: Geraldo Bonorino Xexéo

Ana Regina Cavalcanti da Rocha

Department: System and Computing Engineering

*With the increasing availability of information from search engines, digital libraries and online databases, it becomes more important to have personalized systems that help users find relevant information. Recommender systems have been developed to help people deal with information abundance and overload. They enable people to share their opinions and benefit from each other's experience.*

This thesis presents a recommender system, RecDoc, developed to a Web digital virtual library. Recdoc makes recommendations based on the user's experience, similarity between documents and collaboration with other users.

# Índice

<b>CAPÍTULO 1</b> .....	<b>1</b>
<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1 MOTIVAÇÃO .....	1
1.2 OBJETIVO DA TESE .....	3
1.3 ORGANIZAÇÃO DA TESE.....	3
<b>CAPÍTULO 2</b> .....	<b>5</b>
<b>SISTEMAS DE RECOMENDAÇÃO</b> .....	<b>5</b>
2.1 RECUPERAÇÃO DE INFORMAÇÃO (RI) .....	5
2.2 RECUPERAÇÃO DE INFORMAÇÃO NA WEB .....	7
2.3 FILTRAGEM DE INFORMAÇÃO (FI).....	10
2.3.1 <i>Comparativo entre Recuperação e Filtragem de Informação</i> .....	10
2.3.2 <i>Filtragem Baseada em Perfil</i> .....	11
2.3.3 <i>Filtragem com Agentes</i> .....	13
2.3.4 <i>Trabalhos Relacionados a Filtragem de Informação</i> .....	13
2.4 SISTEMAS DE RECOMENDAÇÃO .....	14
2.4.1 <i>Abordagens de Sistemas de Recomendação</i> .....	16
2.4.2 <i>Processo de Filtragem Colaborativa</i> .....	18
2.4.3 <i>Algoritmos de Filtragem Colaborativa</i> .....	19
2.4.4 <i>Trabalhos Relacionados a Sistemas de Recomendação</i> .....	21
2.4.5 <i>Aplicação de Sistemas de Recomendação em Sistemas de Comércio Eletrônico</i> .....	24
2.4.6 <i>Problemas e Desafios dos Sistemas de Recomendação</i> .....	26
2.5 CONSIDERAÇÕES FINAIS .....	29
<b>CAPÍTULO 3</b> .....	<b>30</b>
<b>BIBLIOTECAS DIGITAIS</b> .....	<b>30</b>
3.1 INTRODUÇÃO.....	30
3.2 SISTEMAS DE BUSCA EM BIBLIOTECAS DIGITAIS .....	32
3.3 GERENCIAMENTO DA COLEÇÃO .....	32
3.4 SERVIÇOS PERSONALIZADOS .....	33
3.5 TRABALHOS RELACIONADOS .....	35
3.6 CONSIDERAÇÕES FINAIS .....	48
<b>CAPÍTULO 4</b> .....	<b>49</b>
<b>RECDOC: UM SISTEMA DE RECOMENDAÇÃO PARA UMA BIBLIOTECA DIGITAL VIRTUAL NA WEB</b> .....	<b>49</b>
4.1 CONSIDERAÇÕES INICIAIS .....	49
4.2 CONSTRUÇÃO DE UMA BIBLIOTECA DIGITAL VIRTUAL NA WEB.....	50
4.3 BIBLIOTECA DIGITAL VIRTUAL DE ENGENHARIA DE SOFTWARE .....	51
4.3.1 <i>Sistema de busca da biblioteca</i> .....	52
4.3.2 <i>Gerenciamento da coleção</i> .....	54
4.3.3 <i>Serviços personalizados</i> .....	55
4.3.4 <i>Perfil do usuário definido na biblioteca</i> .....	56
4.4 RECDOC: SISTEMA DE RECOMENDAÇÃO.....	60
4.4.1 <i>Recomendação por conteúdo</i> .....	60
4.4.2 <i>Recomendação por colaboração</i> .....	61
4.4.3 <i>Arquitetura</i> .....	63
4.4.4 <i>Algoritmo de filtragem colaborativa</i> .....	64
4.4.5 <i>Justificativa e avaliação da recomendação</i> .....	68
4.5 BASE DE DADOS DA BIBLIOTECA.....	70
4.6 COMPARATIVO COM TRABALHOS RELACIONADOS .....	73
4.6.1 <i>Comparativo entre a Biblioteca de Engenharia de Software e demais bibliotecas</i> .....	73

4.6.2	Comparativo entre RecDoc e demais sistemas de recomendação .....	75
4.7	CONSIDERAÇÕES FINAIS .....	77
<b>CAPÍTULO 5</b>	.....	<b>78</b>
<b>ESTUDOS DE CASO</b>	.....	<b>78</b>
5.1	CONSIDERAÇÕES INICIAIS .....	78
5.2	ESTUDO DE CASO: PRECISÃO DO ALGORITMO DE PREDIÇÃO .....	78
5.3	ESTUDO DE CASO: DESEMPENHO DA ABORDAGEM COLABORATIVA .....	84
5.4	ESTUDO DE CASO: DESEMPENHO DO SISTEMA DE RECOMENDAÇÃO .....	88
<b>CAPÍTULO 6</b>	.....	<b>91</b>
<b>CONCLUSÕES E PERSPECTIVAS FUTURAS</b>	.....	<b>91</b>
6.1	CONCLUSÕES .....	91
6.2	PERSPECTIVAS FUTURAS .....	92
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	.....	<b>94</b>
<b>CASOS DE USO DA BIBLIOTECA DIGITAL VIRTUAL DE ENGENHARIA DE SOFTWARE</b>	.....	<b>102</b>

# Índice de Figuras

FIGURA 1 - PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO.....	6
FIGURA 2 – FERRAMENTA DE META-BUSCA.....	9
FIGURA 3 – PROCESSO DE FILTRAGEM COLABORATIVA .....	18
FIGURA 4 – FORMULÁRIO DE AVALIAÇÃO DE FILMES DO MOVIELENS.....	23
FIGURA 5 – AVALIAÇÃO DE ITENS NA AMAZON.COM. ....	25
FIGURA 6 - JUSTIFICATIVA DA RECOMENDAÇÃO APRESENTADA NA AMAZON.COM. ....	27
FIGURA 7 - REFERÊNCIA BIBLIOGRÁFICA NA BIBLIOTECA DIGITAL ACM. ....	37
FIGURA 8 – DEFINIÇÃO DE UM BINDER NA BIBLIOTECA DIGITAL ACM. ....	38
FIGURA 9 – RESULTADO DA BUSCA NA BIBLIOTECA DIGITAL DA IEBE. ....	40
FIGURA 10 - BUSCA AVANÇADA DA NCSTRL. ....	41
FIGURA 11 - RESULTADO DA BUSCA NA NCSTRL .....	42
FIGURA 12 - BUSCA POR CAMPO BIBLIOGRÁFICO NA SPRINGER. ....	43
FIGURA 13 - REFERÊNCIA BIBLIOGRÁFICA NA BIBLIOTECA DIGITAL CITESBER. ....	45
FIGURA 14 – DEFINIÇÃO DA BUSCA NA BIBLIOTECA VIRTUAL DE SAÚDE (BVS).....	46
FIGURA 15 - ARQUITETURA DA BIBLIOTECA.....	51
FIGURA 16 - BUSCA POR PALAVRA-CHAVE. ....	52
FIGURA 17 - RESULTADO DA BUSCA. ....	53
FIGURA 18 - CADASTRO DE DOCUMENTO.....	54
FIGURA 19 - REFERÊNCIA BIBLIOGRÁFICA E SERVIÇOS PERSONALIZADOS. ....	56
FIGURA 20 - PERFIL DO USUÁRIO.....	58
FIGURA 21 - FORMULÁRIO DE AVALIAÇÃO DO DOCUMENTO. ....	58
FIGURA 22 – DEFINIÇÃO DO FILTRO PERSONALIZADO.....	59
FIGURA 23 - ESQUEMA DE RECOMENDAÇÃO POR CONTEÚDO. ....	60
FIGURA 24 - PSEUDO-CÓDIGO DA RECOMENDAÇÃO POR CONTEÚDO.....	61
FIGURA 25 - ESQUEMA DE RECOMENDAÇÃO POR COLABORAÇÃO.....	62
FIGURA 26 – ARQUITETURA DO SISTEMA DE RECOMENDAÇÃO RECDoc. ....	63
FIGURA 27 - PSEUDO-CÓDIGO DO ALGORITMO DE RECOMENDAÇÃO. ....	66
FIGURA 28 – COEFICIENTES DE CORRELAÇÃO ENTRE USUÁRIOS U1 E U2, U1 E U3.....	68
FIGURA 29 – PREDIÇÃO DA AVALIAÇÃO DO USUÁRIO U1 PARA O ITEM I6. ....	68
FIGURA 30 – RECOMENDAÇÃO DE DOCUMENTOS NO PERFIL DO USUÁRIO.....	69
FIGURA 31 - JUSTIFICATIVA DA RECOMENDAÇÃO COLABORATIVA. ....	70



# Índice de Tabelas

TABELA 1 - COMPARATIVO ENTRE SRI E SFI.....	11
TABELA 2 – COMPARATIVO ENTRE SISTEMAS DE FILTRAGEM DE INFORMAÇÃO. ....	14
TABELA 3 - COMPARATIVO ENTRE SISTEMAS DE RECOMENDAÇÃO.....	24
TABELA 4 – COMPARATIVO ENTRE SITES DE COMÉRCIO ELETRÔNICO. ....	26
TABELA 5 – ESQUEMA DE PERFIL DE USUÁRIO PARA .....	34
TABELA 6 - DEFINIÇÃO DO PERFIL DO USUÁRIO NO SISTEMA HERMES.....	35
TABELA 7 - DEFINIÇÃO DE UM BINDER NA BIBLIOTECA DA ACM.....	39
TABELA 8 - INFORMAÇÕES UTILIZADAS NO SERVIÇO DE ALERTA DA BIBLIOTECA SPRINGER.....	43
TABELA 9 - COMPARATIVO ENTRE AS BIBLIOTECAS DIGITAIS. ....	47
TABELA 10 – DEFINIÇÃO DO PERFIL DO USUÁRIO.....	57
TABELA 11 - MATRIZ DE AVALIAÇÕES – “USUÁRIOS X ÍTENS” .....	67
TABELA 12 - MÉDIA ARITMÉTICA DAS NOTAS PARA OS USUÁRIOS.....	67
TABELA 13 - BIBLIOTECA DE ENGENHARIA DE SOFTWARE X DEMAIS BIBLIOTECAS. ....	74
TABELA 14 - RECDOC X SISTEMAS DE RECOMENDAÇÃO.....	76
TABELA 15 - LISTA DE DOCUMENTOS DO CURSO APRENDIZADO ORGANIZACIONAL. ....	79
TABELA 16 – AVALIAÇÕES DOS DOCUMENTOS DO CURSO APRENDIZADO ORGANIZACIONAL.....	80
TABELA 17 - MÉDIA ARITMÉTICA DAS AVALIAÇÕES POR USUÁRIO.....	80
TABELA 18 – MATRIZ DE CORRELAÇÃO ENTRE OS USUÁRIOS.....	81
TABELA 19 - MATRIZ DE PREDIÇÃO PARA AS NOTAS APAGADAS NO ESTUDO .....	81
TABELA 20 – NOTAS REAIS E NOTAS OBTIDAS NA PREDIÇÃO.....	82
TABELA 21 – COMPARATIVO ENTRE RESULTADOS DO ALGORITMOS DE FILTRAGEM COLABORATIVA.....	83
TABELA 22 - RESULTADOS OBTIDOS NA RECOMENDAÇÃO COLABORATIVA. ....	86
TABELA 23 – DESEMPENHO DO SISTEMA DE RECOMENDAÇÃO.....	89

# Capítulo 1

## Introdução

---

*Neste capítulo são apresentadas as questões que motivaram a realização deste trabalho, seu objetivo principal e a forma como esta dissertação está organizada.*

### 3.1 Motivação

Com o crescimento contínuo do volume de informações disponíveis na Web, há uma certa dificuldade em encontrar informações relevantes para o usuário. A abundância e sobrecarga de informação na Web são problemas que impulsionam pesquisas em diversas áreas do conhecimento. A área de Recuperação de Informação (RI) tem um papel fundamental nesse contexto, provendo técnicas que possibilitam buscas eficientes, ou seja, que retornam o maior número de documentos relevantes a uma consulta realizada pelo usuário (BAEZA-YATES & RIBEIRO-NETO, 1999).

Contudo, muitos itens retornados pelos sistemas de busca na Web podem não ser úteis ao usuário. Este fato ocorre devido ao volume de dados a ser processado, a heterogeneidade e distribuição das fontes de informação, a não utilização de um padrão de metadados para descrever a semântica dos dados contidos nas páginas Web e a dificuldade do usuário em expressar sua necessidade de informação através de uma consulta por palavras-chaves (MIZZARO, 1997, LAWRENCE, 2000, MOURA, 2002). Outro aspecto a ser considerado, é o fato destes sistemas não armazenarem informações sobre o usuário, nem sobre o contexto no qual a busca é realizada (LAWRENCE, 2000). Uma alternativa é representar necessidades de informação do usuário e oferecer sistemas de busca personalizados, como a iniciativa do My Yahoo (MANBER & ROBISON, 2000).

A área de Filtragem de Informação (FI) surgiu para tratar o problema da sobrecarga da informação, propondo a utilização de filtros para eliminar informações não relevantes. A área teve um grande crescimento com o advento dos grupos de discussão, onde milhares de mensagens são enviadas diariamente aos usuários. A

filtragem geralmente utiliza perfis de usuários para representar a necessidade de informação do usuário e agentes inteligentes que realizam tarefas para eliminar itens do fluxo de informação.

Uma outra proposta mais recente para o problema, é a utilização de sistemas de recomendação. O enfoque principal dessa solução é encontrar itens de interesse ao usuário, ao invés de eliminar itens irrelevantes.

Sistemas de recomendação procuram automatizar o processo social de indicação boca a boca (SHARDANAD & MAES, 1995). As pessoas geralmente recorrem a indicações de amigos, especialistas ou publicações especializadas quando precisam tomar uma decisão, escolher um produto que vão comprar ou até para decidir qual filme vão assistir no cinema. Assim como recorremos a esses recursos no dia a dia, saber a opinião de outras pessoas ajudaria a selecionar informações relevantes, em situações onde se tem um grande volume de informações. Os sistemas de recomendação automatizam esse processo, pois permitem que as pessoas compartilhem opiniões e possam se beneficiar com a experiência uma das outras (TERVEEN & HILL, 2001).

Sistemas de recomendação têm sido utilizados em sites de entretenimento, de jogos, fóruns de discussão e em lojas de comércio eletrônico (SCHAFER et al., 2001). A Amazon.com, por exemplo, usa uma técnica de recomendação como parte da sua política de CRM (*Customer Relationship Management*). O objetivo é transformar simples navegadores em compradores, através de recomendações de produtos que o usuário esteja interessado em comprar.

Sistemas de recomendação têm sido utilizados também em bibliotecas digitais. As bibliotecas são provedoras de informação semelhantes aos sistemas de busca na Web, compartilhando, portanto, alguns dos seus problemas, como a sobrecarga de informação. Trabalhos na área de bibliotecas digitais têm explorado a questão de que as bibliotecas são provedoras de informações e de serviços. A visão de que uma biblioteca é um repositório de informação onde os usuários apenas acessam e fazem busca no acervo, é uma visão ultrapassada. Bibliotecas digitais precisam oferecer, além de um sistema de busca, serviços úteis a seus usuários (PAEPCKE, 1996). Elas têm o papel de disponibilizar um espaço onde as pessoas possam se comunicar, compartilhar e produzir conhecimento (BORGMAN, 1999).

Essa tese apresenta uma experiência nessa linha, procurando integrar sistemas de recomendação com bibliotecas digitais.

## **3.2 Objetivo da Tese**

O objetivo fundamental deste trabalho é definir um sistema de recomendação para uma biblioteca digital virtual na Web.

Para atingir este objetivo, foi definida e implementada uma biblioteca genérica na Web. A biblioteca possui um ambiente personalizado onde os usuários compartilham informações e experiências. Neste ambiente, os usuários armazenam referências dos documentos lidos, avaliam e comentam documentos, além de receberem recomendações personalizadas.

Para validar a biblioteca desenvolvida, uma customização foi realizada para a linha de pesquisa de Engenharia de Software da COPPE/UFRJ. A biblioteca foi implantada e utilizada em um ambiente de pesquisa, onde os estudantes acessaram a biblioteca para consultar o material dos cursos, buscar documentos e cadastrar referências bibliográficas utilizadas nas suas dissertações.

O sistema de recomendação proposto foi desenvolvido para apoiar os usuários a encontrar documentos relevantes na biblioteca. O sistema age de forma pró-ativa, procurando por itens de interesse, recomendando-os sem que seja necessário o usuário requisitar o serviço. RecDoc indica documentos ao usuário baseado nas referências bibliográficas armazenadas no seu perfil de usuário e na colaboração com outros usuários.

Com o intuito de validar o sistema de recomendação proposto, foram realizados estudos de casos utilizando a biblioteca de Engenharia de Software.

## **3.3 Organização da Tese**

Esta dissertação contém mais 5 capítulos, além desta Introdução.

No segundo capítulo, é apresentado o estudo realizado sobre sistemas de recomendação. São discutidos, inicialmente, conceitos e problemas da área de recuperação e filtragem de informação na Web. Estes problemas serviram de motivação para o desenvolvimento e pesquisa dos sistemas de recomendação.

No terceiro capítulo, os conceitos relacionados a bibliotecas digitais são discutidos e trabalhos da área são apresentados.

No quarto capítulo, a biblioteca Web genérica, a customização para a Biblioteca Digital de Engenharia de Software e o sistema de recomendação RecDoc são apresentados.

No quinto capítulo, estudos de caso realizados para validar o sistema de recomendação proposto e os resultados obtidos são apresentados.

No sexto capítulo, as principais conclusões e perspectivas futuras para a continuidade deste trabalho são apresentadas.

# Capítulo 2

## Sistemas de Recomendação

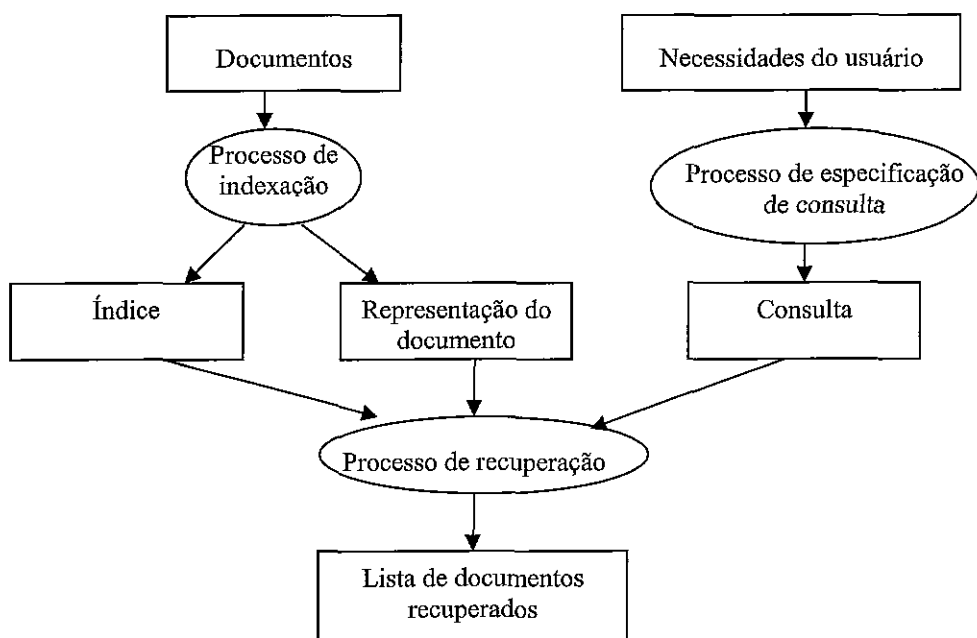
---

*Este capítulo apresenta seções introdutórias a conceitos e técnicas da área de recuperação e filtragem de informação, com enfoque principal na Web. A seguir, é mostrado o estudo realizado sobre Sistemas de Recomendação, que teve sua origem nas áreas citadas acima. Por fim, as possíveis abordagens e algoritmos de recomendação são analisados, juntamente com os principais trabalhos da área.*

### 2.1 Recuperação de Informação (RI)

Recuperação de Informação (RI) é a área de pesquisa que trata da representação, armazenamento e acesso a itens de informação. O problema principal em um sistema de recuperação é encontrar o subconjunto de objetos relevantes à consulta especificada pelo usuário. A representação e o armazenamento devem prover acesso fácil a essas informações (BAEZA-YATES & RIBEIRO-NETO, 1999).

Um sistema de recuperação de informação (SRI) é composto por um processo de indexação, um processo de especificação de consultas e um processo de recuperação (GEY, 1996). A **Figura 1** ilustra a interação entre os três processos. O processo de indexação gera estruturas de dados (índices) para os documentos. No processo de especificação, o usuário define uma consulta que represente as suas necessidades de informação. O processo de recuperação consiste em retornar uma lista de documentos que satisfaça a consulta do usuário.



**Figura 1 - Processo de recuperação de informação.**

O processo de indexação envolve a criação de estruturas de dados associadas a documentos constituídos de textos, ou estruturas associadas à parte textual dos documentos contendo textos, figuras, etc. Estas estruturas são denominadas de índice e contêm uma coleção de termos que indicam o local onde a informação relacionada pode ser localizada. Estes termos devem ser organizados de forma a facilitar a busca (SILVA, 1998). Os índices podem conter ainda dados sobre características dos termos na coleção de documentos, tais como a frequência de termos em um documento e o inverso da frequência de termos na coleção.

O processo de especificação da consulta é geralmente uma tarefa difícil. Há frequentemente uma distância semântica entre a real necessidade do usuário e o que ele expressa na consulta formulada (BAEZA-YATES & RIBEIRO-NETO, 1999). Esta distância é provocada principalmente pelo limitado conhecimento do usuário no universo de pesquisa.

O processo de recuperação consiste em gerar uma lista de documentos que satisfaçam a consulta formulada pelo usuário. Para acelerar essa tarefa, utilizam-se os índices construídos para a coleção.

## 2.2 Recuperação de Informação na Web

Na última década, a área de RI tem atingido um alcance muito maior do que seus objetivos iniciais, que eram *indexar texto e buscar documentos em uma coleção*. O fato que impulsionou as pesquisas na área foi o crescimento da Web, a partir dos anos 90. A crescente complexidade dos objetos armazenados e o grande volume de dados exigem processos de recuperação cada vez mais sofisticados (BAEZA-YATES & RIBEIRO-NETO, 1999).

O aumento contínuo de documentos disponíveis na Web tem tornado à tarefa de encontrar a informação desejada cada vez mais difícil. Muitas ferramentas de busca e técnicas de indexação têm sido desenvolvidas para resolver esse problema (LAWRENCE, 2000).

As ferramentas de busca (*search engines*) encontradas na Web são essencialmente uma retrospectiva dos sistemas de recuperação da informação. Estes sistemas criam um índice de palavras do documento e retornam uma lista ordenada de itens que correspondem à consulta do usuário (LAWRENCE, 2000). No caso da Web, os documentos são indicados por URLs, inicialmente correspondentes a páginas HTML, porém atualmente também podem corresponder a páginas em outros formatos como PDF, PS (*postscript*), etc.

A ferramenta de busca Yahoo é um catálogo (ou diretório) de páginas Web similar a páginas amarelas com taxonomias hierárquicas que classificam distintos temas ou áreas de conhecimento (YAHOO, 2003). A inserção de páginas no catálogo é feita através de requisição ao sistema, onde o próprio usuário indica uma classificação, revisada manualmente. Segundo BAEZA-YATES (2000), a vantagem principal deste método é que se o usuário encontra algo, certamente será útil. Outra vantagem levantada pelo autor é a intervenção manual na classificação, o que muitas vezes garante uma melhor qualidade se comparada com abordagens de classificação automática. Como desvantagem, tem-se o fato de que nem tudo que existe na Web está classificado.

A ferramenta Google (GOOGLE, 2003), desenvolvida na Stanford University, disponibiliza uma busca baseada no paradigma de recuperação por texto completo (*full text*) e uma busca por categorias hierárquicas semelhante ao Yahoo (YAHOO, 2003). A indexação na Web é feita através de um programa chamado *crawler* ou *robot* que percorre a Web compilando páginas novas ou atualizando páginas. A principal



vantagem em relação a abordagem manual, é a capacidade de coletar e classificar um grande volume de informações. Entretanto, essa tarefa de percorrer a Web atualizando e acrescentando páginas, é uma tarefa demorada e difícil devido ao crescimento contínuo da mesma.

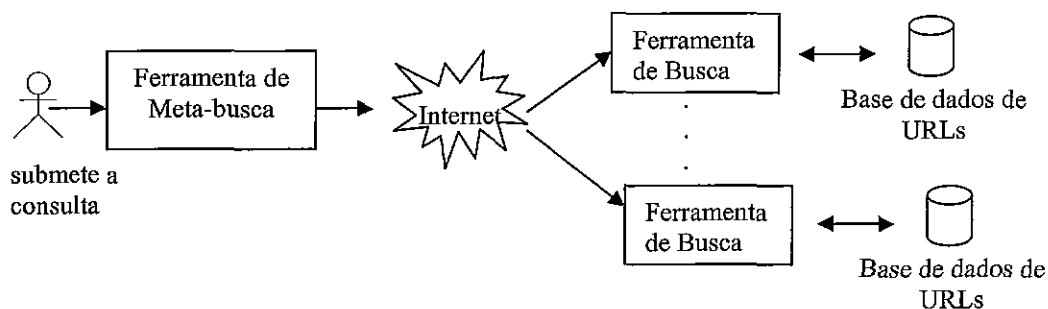
As ferramentas de busca WebCrawler (WEBCRAWLER, 2003), Cade (CADE, 2003) e Achei (ACHEI, 2003) oferecem busca por categoria, como o catálogo do Yahoo (YAHOO, 2003). As duas últimas ferramentas realizam a busca somente na Web brasileira. A ferramenta AltaVista (ALTAVISTA, 2003) oferece busca por texto completo, como o Google (GOOGLE, 2003), mas a inserção de páginas é feita de forma manual.

Um dos problemas envolvendo a qualidade do resultado dos sistemas de busca na Web é a definição incompleta da consulta por parte do usuário. Os usuários raramente pensam em alternativas para melhorar o resultado da busca e geralmente utilizam somente uma ou duas palavras para definir a consulta (LAWRENCE, 2000).

Outro problema é que as ferramentas de busca na Web geralmente tratam requisições de consulta em isolado. Poucos resultados retornados podem ser realmente úteis ao usuário, pois os resultados de uma consulta são idênticos e independentes do usuário ou do contexto da busca (MIZZARO, 1997).

Segundo (LAWRENCE, 2000), o valor que o documento tem para o usuário depende do contexto e essa informação poderia ser obtida do usuário na forma de palavras-chave adicionadas à consulta. É necessário que as ferramentas incrementem seus sistemas de busca com informação sobre o contexto, usando informação implícita ou explícita dos usuários. Outra abordagem para solucionar o problema, citada pelo autor, é o desenvolvimento de ferramentas de busca restritas a um domínio específico. Essa abordagem introduz um outro problema: o usuário precisa saber qual ferramenta de busca é mais apropriada para o contexto da busca que deseja fazer.

Uma das soluções encontradas na literatura para melhorar os sistemas de busca na Web, são os sistemas de meta-busca (SMB). Estes sistemas fazem a requisição da consulta, em paralelo, em várias ferramentas de buscas diferentes. O SMB retorna ao usuário os resultados mais relevantes, combinados e classificados. A **Figura 2** ilustra o funcionamento de um SMB.



**Figura 2 – Ferramenta de Meta-busca.**

SynSeek (KAPUR & ZHANG, 2000) é uma ferramenta de meta-busca que oferece um sistema de busca booleana direcionada pelo uso de uma lista de conceitos e sinônimo. O usuário seleciona os termos da consulta nessa lista. A consulta é submetida às ferramentas de busca AltaVista, InfoSeek e WebCrawler (ALTAVISTA, 2003, INFOSEEK, 2003, WEBCRAWLER, 2003). A lista de conceitos e sinônimos foi criada manualmente pelo departamento de ciências cognitivas da Princeton University e é gerenciada pela ferramenta *tesauro* WordNet (WORDNET, 2002).

O MetaMiner é uma ferramenta de meta-busca que possibilita ao usuário escolher os sistemas de busca que deseja submeter a consulta (METAMINER, 2003). Esta pode ser aplicada a ferramentas de busca na Web brasileira, como Achei e Radar UOL, ou na Web mundial como LookSmart e AOL. No resultado da consulta, o MetaMiner informa em qual sistema o resultado foi encontrado (ACHEI, 2003, RADAR UOL, 2003, LOOKSMART, 2003, AOL, 2003).

A ferramenta de meta-busca Inquirer2 (GLOVER et al., 2001), desenvolvida no NEC Research Center, requisita ao usuário o contexto da busca através da seleção de uma categoria. Por exemplo, o usuário submete uma consulta com a palavra “jaguar” na ferramenta. Antes de executar a consulta, o usuário tem que especificar a categoria de interesse, por exemplo, automóvel ou animal. Esta abordagem reduz o escopo da busca e consequentemente a torna mais eficiente.

Um problema comum tanto nas ferramentas de busca na Web, quanto nos SMBs, é o fato de que estes não armazenam as buscas realizadas. Eles não retêm informações sobre o usuário, como suas preferências, para uma utilização posterior e auxílio nas buscas, melhorando-as e/ou refinando-as. Outro serviço não disponibilizado por estas ferramentas, é um serviço de aviso de que novos documentos foram encontrados para

uma consulta específica. Esse fato deve-se ao alto custo da própria ferramenta (LAWRENCE, 2000).

Quando um usuário, ou grupo de usuários, têm sua área de interesse bem definida e relativamente estável em períodos prolongados (semanas, meses, anos), essas ferramentas começam a não apresentar resultados suficientes pois não apresentam novos resultados relevantes para as mesmas consultas. Foi preciso desenvolver ferramentas que permitissem registrar, tanto as informações de seus usuários e suas consultas individuais, como identificar grupos de usuários com áreas de interesse compartilhadas.

Em (MANBER & ROBISON, 2000), é apresentada a experiência do Yahoo em inserir características e informações sobre o usuário no sistema de busca, desenvolvendo assim, uma busca personalizada.

## **2.1 Filtragem de Informação (FI)**

Com a explosão de documentos eletrônicos na Web, a filtragem de informação (FI) tem sido uma das mais importantes áreas de aplicação da comunidade de recuperação de informação (KIM et al., 2000). FI é a área de pesquisa que oferece ferramentas para separar informações relevantes e irrelevantes (BELKIN & CROFT, 1992). Ainda segundo o autor, FI envolve a entrega de informação para pessoas que precisam dela.

Filtragem de informação não é um conceito novo e não é limitado a documentos eletrônicos (FOLTZ & DUMAIS, 1992). Quando selecionamos artigos para ler ou compramos apenas revistas de assuntos que nos interessam, estamos fazendo filtragem.

FI é necessária em situações de sobrecarga de informação, em bibliotecas digitais ou na Internet (BELKIN & CROFT, 1992). Filtragem automática parece ser uma boa alternativa para selecionar informações relevantes, mas existem dificuldades em determinar qual informação o usuário deseja ver.

### **2.4.1 Comparativo entre Recuperação e Filtragem de Informação**

Muitas técnicas desenvolvidas em RI são aplicadas a Sistemas de Filtragem de Informação (SFI), mas as duas áreas se diferenciam em três pontos básicos (BELKIN & CROFT, 1992): o interesse do usuário, o conjunto de documentos e o tipo de seleção. A **Tabela 1** exibe um quadro comparativo com as características de cada área. A principal

diferença, dentre os três pontos, é o interesse do usuário. Em SFI, os usuários têm interesse que duram por muito tempo e são representados geralmente em perfis de usuário. Em SRI, o interesse do usuário muda a cada consulta (*ad-hoc*). Apesar dessas diferenças, os autores consideram que filtragem e recuperação de informação têm o mesmo objetivo: encontrar informação para quem precisa dela.

**Tabela 1 - Comparativo entre SRI e SFI.**

<b>Características</b>	<b>SRI</b>	<b>SFI</b>
<b>Interesse do usuário</b>	Consultas tipicamente representam uma meta a curto prazo que pode ser satisfeita recuperando um conjunto particular de documentos.	O perfil do usuário representa um interesse a longo prazo.
<b>Conjunto de documentos</b>	Aplicações assumem que o conjunto de documentos não muda constantemente com o tempo.	Fluxo constante de documentos dependente do tempo.
<b>Tipo de seleção</b>	Atuam encontrando itens relevantes nos Bancos de Dados.	Atuam removendo do fluxo, itens irrelevantes para um determinado usuário.

Os SFI possuem abordagens baseadas em perfil do usuário e agentes inteligentes. As duas abordagens são geralmente combinadas.

#### **2.4.1 Filtragem Baseada em Perfil**

Em SFI, as necessidades do usuário são, geralmente, expressas como perfis do usuário. O perfil pode ser definido como uma representação estruturada de informação que o usuário precisa (AMATO & STRACCIA, 1999).

Os sistemas de filtragem fundamentados no perfil do usuário procuram extrair padrões através de observações do comportamento do usuário a fim de predizer quais itens serão selecionados ou descartados (MALTZ & EHRLICH, 1995).

Existem dois tipos principais de perfil de usuário (KUFLIK & SHOVAL, 2000):

- a) perfil baseado em conteúdo: a necessidade do usuário é expressa como uma consulta, ou seja, como uma lista de termos ou palavras-chave;
- b) perfil colaborativo: essa abordagem utiliza avaliação de padrões de usuários similares.

Segundo o autor, outra maneira de representar a necessidade de informação do usuário é através de regras de filtragem (Se <condição> então <valor>), que podem ser utilizadas em conjunto com o perfil baseado em conteúdo.

Possíveis métodos para criação e atualização de perfis de usuário em sistemas de filtragem de informação são discutidos em (KUFLIK & SHOVAL, 2000). Os autores sugerem especificar o perfil das seguintes maneiras:

- a) perfil criado pelo usuário: essa é a abordagem mais simples, onde o usuário define a área de interesse através de termos (palavras-chave) e atribui pesos a cada termo. Os termos são utilizados no processo de filtragem;
- b) perfil criado pelo sistema através de indexação automática: um conjunto de dados, julgados relevantes pelo usuário, são analisados por um software que extrai termos mais freqüentes. Os termos extraídos recebem pesos, de acordo com a freqüência nas quais aparecem, e são utilizados para definir o perfil;
- c) perfil criado pelo usuário e pelo sistema: consiste na combinação das duas abordagens acima. Inicialmente, o perfil é criado automaticamente e o usuário revisa os termos selecionados e os pesos atribuídos a estes;
- d) estereótipos: esse método assume que o sistema de filtragem já tenha estereótipos pré-definidos. O estereótipo é representado como um perfil baseado em conteúdo, através de uma lista de usuários que tenham informações em comum e o mesmo comportamento de filtragem.

Os problemas relacionados com perfis do usuário são (MALTZ & EHRLICH, 1995): como gerar um perfil inicial para um usuário novo, denominado partida a frio e como atualizar um perfil ao longo do tempo. Um perfil impróprio causa um desempenho ruim: ou o usuário recebe muita informação não relevante; ou deixa de receber informação relevante. O autor acredita que a qualidade da definição do perfil é um fator importante no desempenho de um SFI.

A atualização do perfil (*relevance feedback*) pode ser feito de duas formas: explícita (o próprio usuário avalia o item) ou implícita (usuário não diz diretamente se gostou ou não de um item, mas o sistema observa o usuário e determina sua avaliação).

## 2.4.2 Filtragem com Agentes

Os sistemas de filtragem baseados em agentes apostam em processos cooperativos entre humanos e agentes, onde estes trabalham como um assistente pessoal (MAE, 1994). Segundo o autor, dois problemas devem ser considerados para construir agentes eficientes: competência e confiança. A competência se refere a quando um agente adquire conhecimento suficiente para ajudar ao usuário. O agente se torna mais competente à medida que ele aprende sobre o comportamento do usuário. O segundo problema se refere a confiança do usuário em delegar tarefas para um agente, que as execute automaticamente.

Um agente aprende sobre o usuário das seguintes maneiras (MAE, 1994):

- a) observando o usuário, registrando suas ações e descobrindo um padrão para seu comportamento;
- b) perguntando ao usuário avaliações sobre determinadas ações e obtendo resposta direta, ou indireta;
- c) recebendo treinamento através de exemplos dados pelo usuário;
- d) pedindo conselhos a outros agentes sobre uma determinada tarefa.

## 2.4.3 Trabalhos Relacionados a Filtragem de Informação

A ferramenta SIFT (*Stanford Information Filtering Tool*) (PAEPCKE & GARCIA-MOLINA, 1999) tem o objetivo de filtrar artigos USENET. O usuário descreve um perfil através de um conjunto de palavras-chave, usadas para determinar o filtro de seleção dos artigos a serem indicados ao usuário. Sempre que um novo artigo é inserido no grupo de notícias, ele é checado com o perfil do usuário, verificando-se a similaridade entre eles. Se a similaridade for superior ao limite, então o documento é enviado por correio eletrônico ao usuário.

O sistema LIRA (*Learning Information Retrieval Agent*) (BALABANOVIC et al., 1995), desenvolvido na Stanford University, observa como o usuário atribui pesos a várias páginas visitadas na Web. O sistema recupera, automaticamente, novas páginas que sejam similares ao perfil do usuário. A heurística de busca na Web é alterada à medida que o usuário vai respondendo a indicação de páginas.

O sistema WebLearner (WEBLEARNER, 2002), desenvolvido na University of California at Irvine, observa a avaliação do usuário ao ler uma nova página Web. Se o usuário considera a página boa, esta é colocada em uma *HotList*, caso contrário a página vai ser enviada para uma *ColdList*. O sistema tenta analisar os exemplos para avaliar páginas futuras.

O sistema Amatheia (MOUKAS, 1997) utiliza uma abordagem de múltiplos agentes para busca e filtragem de informação na Internet. O Amatheia baseia-se na busca realizada por mecanismos de busca convencionais como AltaVista (ALTAVISTA, 2003) e Yahoo (YAHOO, 2003), filtrando a informação retornada pelos mesmos. A filtragem utiliza técnicas de aprendizado de máquina.

A **Tabela 2** mostra um comparativo das características dos sistemas de filtragem discutidos.

**Tabela 2 – Comparativo entre sistemas de filtragem de informação.**

Sistema de filtragem	Alternativa	Características
SIFT	Filtragem de artigos USENET.	Perfil baseado em palavra-chave.
WebLearner	Filtragem de páginas Web. Analisa os exemplos para avaliar páginas futuras.	Páginas Web são colocadas em uma <i>HotList</i> ou uma <i>ColdList</i> .
LIRA	Sugere novas páginas Web com base em pesos atribuídos a páginas visitadas.	Agente pessoal de navegação ou interface.
Almathea	Filtragem com agentes e técnicas de aprendizado de máquina.	Agentes para busca e filtragem de informação.

## 2.4 Sistemas de Recomendação

Os sistemas de recomendação vêm sendo desenvolvidos para mediar, apoiar e automatizar o processo cotidiano de compartilhar recomendações (TERVEEN & HILL, 2001).

Os sistemas de recomendação surgiram da evolução dos sistemas de filtragem de informação (SFI). Mesmo com a filtragem de itens não relevantes, o volume de informação apresentado aos usuários ainda continuava muito grande. A partir dessa necessidade, iniciou-se o desenvolvimento de sistemas de recomendação cujo processo-chave é encontrar itens relevantes para o usuário, ao invés de eliminar os não relevantes, como na filtragem.

Sistemas de recomendação provêm o acesso a produtos e informações pertinentes fazendo sugestões personalizadas baseadas em exemplos prévios do que um usuário gosta e desgosta, ou usa métodos de filtragem colaborativa que fazem as recomendações baseadas em preferências de outros usuários (MOONEY & ROY, 2000).

Inicialmente, os sistemas de recomendação foram chamados de “Sistemas de Filtragem Colaborativa”, devido à utilização desse termo no primeiro sistema de recomendação Tapestry (GOLDBERG et al., 1992), desenvolvido pela Xerox, onde os usuários do sistema colaboravam entre si através de indicações de documentos. Já o grupo que desenvolveu o sistema Ringo (SHARDANAD & MAES, 1995) do MIT Media-Lab, por sua vez, utilizou o termo “Filtragem Social” por considerar a recomendação como uma automação do processo social de indicação boca a boca (*automatic word of mouth*). A determinação do termo mais genérico “Sistema de Recomendação” foi feita por RESNICK & VARIANT (1997), editores da seção especial de setembro de 1997 (v. 40, n. 3) da Communications of ACM. Segundo os editores, o termo recomendação era mais apropriado do que filtragem colaborativa, por dois motivos: primeiro, os usuários não precisam colaborar explicitamente uns com os outros ou ao menos se conhecerem; segundo, recomendações sugerem indicação de itens relevantes mais do que filtrar itens não relevantes. Mesmo com a tentativa de RESNICK & VARIANT (1997), o termo filtragem colaborativa tem sido bastante usado até os dias atuais, sendo comum encontrar os dois termos nos trabalhos da área.

Os sistemas de recomendação foram desenvolvidos inicialmente no âmbito acadêmico. Um dos principais grupos de pesquisa da área é o GroupLens, da University of Minnesota, que centraliza diversos projetos de filtragem colaborativa desde 1992 (GROUPLENS, 2002). Os trabalhos do grupo começaram a serem reconhecidos na comunidade de RI na Web a partir de 1996, com o desenvolvimento do sistema de recomendação de artigos para fóruns eletrônicos GroupLens (RESNICK et al., 1994, KONSTAN et al., 1997). Utilizando as mesmas técnicas e algoritmos, o grupo desenvolveu em 1998 um sistema online de recomendação de filmes denominado MovieLens (MOVIELENS, 2002).

Sistema de recomendação tem se mostrado uma ferramenta poderosa a ser utilizada em comércio eletrônico, bibliotecas digitais e gerência de conhecimento. Nos últimos anos, estes sistemas passaram a serem utilizados comercialmente em lojas na Web. Em (SCHAFER et al., 2001), é mostrado como o uso crescente desses sistemas tem possibilitado o aumento das vendas, por ser uma abordagem que ajuda o



consumidor a encontrar o produto desejado. Além disso, possibilita que a loja seja personalizada para cada consumidor. Uma das empresas que comercializam sistemas de recomendação é a Net Perceptions, fundada pelos pesquisadores do GroupLens (NETPERCEPTIONS, 2002). A empresa utiliza a tecnologia desenvolvida pelo grupo e tem como um dos principais clientes, a loja de comércio eletrônico Amazon.com. Outros sites na Web que dispõem de sistemas de recomendação são CDNow.com e MovieFinder.com (CDNOW, 2002, MOVIEFINDER, 2002).

Foram realizados dois *workshops* sobre sistemas de recomendação durante a conferência anual ACM SIGIR (*Conference on Research and Development in Information Retrieval*) nos anos de 1999 e 2001. O primeiro *workshop* teve o enfoque voltado para algoritmos e métodos de avaliação (WRS, 1999, SOBOROFF, 1999). Já o segundo *workshop*, teve o objetivo de reunir diversas iniciativas e projetos da área (WRS, 2001).

#### **2.4.1 Abordagens de Sistemas de Recomendação**

BALABANOVIC (1997) definiu duas abordagens principais para os sistemas de recomendação: abordagem baseada em conteúdo e abordagem colaborativa. Na abordagem baseada em conteúdo, a recomendação de itens ou documentos é feita através da análise de documentos similares a aqueles que o usuário já tenha visto e avaliado. Na abordagem baseada em colaboração, a recomendação tem como base a análise de usuários similares para indicar documentos que estes tenham gostado.

A abordagem de recomendação baseada em conteúdo utiliza a premissa de que o usuário gostaria de ver itens semelhantes aos que já viu no passado (BALABANOVIC & SHOHAM, 1997). Para isso, os itens a serem recomendados são selecionados através da comparação entre o conteúdo dos mesmos e o perfil do usuário, ou seja, comparando-os com os itens de interesse do usuário armazenados no perfil. Essa abordagem emprega técnicas de filtragem de informação por conteúdo como por exemplo, filtragem por palavra-chave e análise semântica latente (DEERWESTER et al, 1990). Sistemas que utilizam essa abordagem são InfoFinder (KRULWICH & BURKEY, 1997) e NewsWeeder (LANG, 1995).

Um dos principais problemas da abordagem por conteúdo é especializar muito a recomendação, ou seja, o sistema só vai recomendar documentos similares aos que o usuário já viu (BALABANOVIC & SHOHAM, 1997). Além disso, não são

consideradas as demais características do item a recomendar, como por exemplo, a qualidade do mesmo. Outro fator relevante é a dificuldade em fazer análise de conteúdo para itens não textuais, como mídias de música e vídeo (SHARDANAD & MAES, 1995).

A abordagem colaborativa é baseada na premissa de que pessoas procurando por informação gostariam de saber o que outras já encontraram e avaliaram (RESNICK et al., 1994). Segundo os autores, sistemas de filtragem colaborativos trabalham no sentido de incluir as pessoas nos sistemas de filtragem, pois acreditam que estas possam avaliar melhor os documentos do que uma função computacional.

Os sistemas de recomendação colaborativos agrupam os interesses compartilhados dos usuários e apresentam facilidades que permitem uma maior interatividade. O processo chave desta abordagem é comparar usuários e encontrar interesses similares, fazendo uso das avaliações ou anotações fornecidas por eles (MOTTA, 1999).

Sistemas desenvolvidos em projetos de pesquisa que utilizam essa abordagem são GroupLens (RESNICK et al., 1994, KONSTAN et al., 1997), Ringo (SHARDANAD & MAES, 1995) e MovieLens (DAHLEN et al., 1998). Comercialmente, os sites na Web são Amazon.com, CDNow.com e MovieFinder.com.

Na abordagem colaborativa, é possível resolver o problema encontrado na abordagem de recomendação por conteúdo, onde o usuário só recebe documentos de conteúdos similares. Entretanto, essa abordagem não resolve outros problemas como a inserção de novos documentos na base, que só vão ser recomendados após um certo número de usuários ter lido e avaliado. Outro problema é o tratamento de usuários que não tem interesses similares ao restante da população. Dessa forma, esse usuário diferente não terá usuários similares no qual o sistema de recomendação colaborativa possa se basear (BALABANOVIC, 1997).

Podemos ver que cada abordagem apresenta algumas vantagens e restrições. BALABANOVIC & SHOHAM (1997) propõem uma abordagem híbrida na qual possam ser combinadas características de ambas abordagens, que são complementares. As vantagens dessa abordagem híbrida, segundo os autores, são:

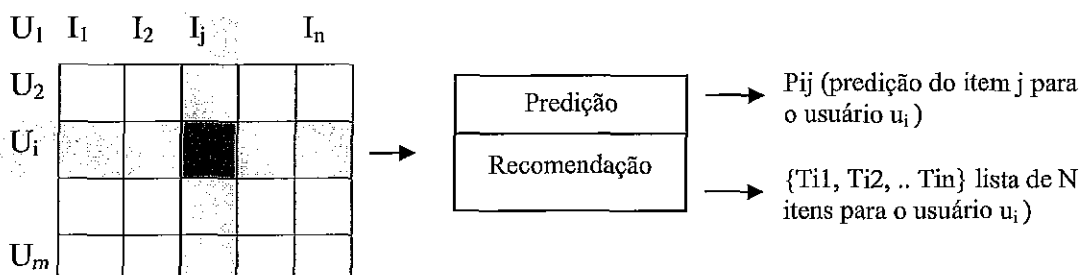
- a) com a recomendação colaborativa é possível utilizar experiências de outros usuários, mostrando ser uma boa alternativa para complementar a abordagem por conteúdo, que é incompleta e imprecisa.

- b) com a recomendação por conteúdo é possível recomendar itens que ainda não tenham sido vistos por outros usuários.
- c) utilizar o perfil com o conteúdo dos documentos para fazer boas recomendações mesmo que estes não tenham sido avaliados por usuários similares.

Estudos experimentais sobre o desempenho das abordagens por conteúdo, por colaboração e híbrida mostraram que o desempenho da abordagem híbrida foi superior ao das duas abordagens isoladamente (HUANG et al., 2002).

#### 2.4.4 Processo de Filtragem Colaborativa

SARWAR et al. (2001) definiram o processo de filtragem colaborativa em duas etapas: uma etapa de predição e outra de recomendação. Na primeira etapa, o algoritmo de predição procura por itens que o usuário não avaliou e calcula uma nota predição para cada um dos itens encontrados. Na etapa de recomendação, são selecionados itens resultantes da etapa anterior que obtiveram notas superiores a um limite. Outros critérios de seleção também podem ser utilizados. A **Figura 3** ilustra este processo. A figura mostra a matriz de usuários e itens, as duas etapas do processo e as saídas geradas em cada uma delas.



**Figura 3 – Processo de filtragem colaborativa**

Na etapa predição, é calculada a similaridade entre os usuários com base na matriz de “usuários x itens”. Considere a matriz ilustrada na **Figura 3**. Cada linha da matriz representa os itens avaliados e não avaliados pelo usuário. Para um determinado usuário  $U_i$ , são verificados itens que ele não avaliou. Para cada um desses itens, são selecionados os usuários que o avaliaram. São calculadas as similaridades entre estes usuários e  $U_i$ . Em seguida, a nota predição é calculada a partir das avaliações e similaridades.

Na etapa recomendação, os itens são selecionados de acordo com algum critério. Geralmente, um primeiro critério aplicado é a nota mínima obtida na predição. O conteúdo do item pode ser utilizado nesta etapa também.

#### **2.4.5 Algoritmos de Filtragem Colaborativa**

BREESE et al. (1998) classificaram os algoritmos de filtragem colaborativa em dois grupos: baseados em memória e baseados em modelo. Os algoritmos baseados em memória calculam a predição do voto para um item, para um determinado usuário, levando em consideração o voto de todos os usuários que avaliaram este item. Os algoritmos que utilizam o modelo do vizinho mais próximo são os mais utilizados neste grupo. Este tipo de algoritmo usa uma variedade de medidas de distância entre os usuários, como distância euclidiana ou coeficiente de correlação Pearson. O grupo de algoritmos classificados como baseados em modelo desenvolve, primeiramente, um modelo das avaliações dos usuários em um subconjunto de dados, chamado conjunto de treinamento, e após aprender sobre o usuário, faz a recomendação com o conjunto de dados restante. O processo de construção do modelo utiliza algoritmos de redes bayesianas, agrupamento (*clustering*) ou baseado em regras, aplicadas ao conjunto de treinamento.

Os algoritmos baseados em modelo foram aplicados a sistemas de recomendação com o objetivo de diminuir o tempo e o alto custo de processamento da etapa de predição dos algoritmos baseados em memória (HERLOCKER et al., 1999). Mas os algoritmos baseados em memória são os mais utilizados tanto em sistemas de pesquisa, quanto comerciais. Em (SARWAR et al., 2001), é proposto um outro tipo de algoritmo baseado na relação “item-item”, ao invés de buscar a relação de similaridade “usuário-usuário”, o grande gargalo dos algoritmos colaborativos, por terem que calcular a correlação entre usuários com base em matrizes esparsas, contendo poucas avaliações.

BREESE et al. (1998) apresentam um estudo comparativo sobre diversos algoritmos de filtragem colaborativa de ambos os grupos, utilizando bases de teste EachMovie e MS Web. O método de correlação (baseado em memória) e o algoritmo de redes bayesianas com árvores de decisão em cada nó (baseado em modelo) foram os que obtiveram melhor desempenho. Segundo o autor, a escolha do melhor algoritmo depende de fatores como a natureza dos dados de teste e da aplicação, disponibilidade dos votos, tamanho da base de testes, tempo de predição e treinamento. Por exemplo, o

algoritmo de redes bayesianas exige menos recursos de memória e menos tempo para fazer a predição do que algoritmos de correlação, mas requer uma etapa de aprendizado que pode levar horas e só após finalizada, pode ser refletida na recomendação.

Em (SHARDANAD & MAES, 1995), foram testados diversos algoritmos de recomendação baseados em memória utilizando diferentes medidas de distância entre os usuários, como coeficiente de Pearson e cosseno entre vetores. Nos experimentos, eles determinaram um subconjunto de dados, chamado *target set* correspondente a 20% dos votos, onde os mesmos foram apagados aleatoriamente. Os algoritmos de predição foram executados para preencher os valores apagados. Foi calculado o coeficiente de erro MAE (*Mean Absolute Error*) para medir a diferença entre as notas obtidas na predição e as avaliações dos usuários. Como resultado, variações do algoritmo de Pearson mostraram ser a alternativa mais eficaz.

BILLSUS & PAZZANI (1998) também realizaram estudos comparativos sobre o desempenho de algoritmos de filtragem colaborativa baseado em coeficientes de correlação (baseado em memória). Segundo os autores, os problemas relacionados a esse grupo de algoritmos podem ser tratados com aplicação de algoritmos de classificação. Neste trabalho, foi mostrado como explorar a estrutura latente da matriz de avaliação pode melhorar o desempenho da etapa de predição. Os experimentos foram realizados com a base de teste *EachMovie* (EACHMOVIE, 2002).

Em (MILLER et al., 1997), foram realizados testes para medir a precisão do algoritmo de predição do GroupLens. O algoritmo utiliza o coeficiente de Pearson para medir a correlação entre os usuários (baseado em memória). Os testes foram realizados em três fóruns de discussão diferentes. Para medir a precisão do algoritmo foram utilizadas as métricas MAE, desvio padrão do erro e coeficiente Pearson entre as notas calculadas pelo algoritmo e as avaliações dos usuários. Os resultados obtidos pelo algoritmo foram comparados com a média aritmética das avaliações. O algoritmo obteve resultados diferentes de um fórum de discussão para outro. Quando a correlação de Pearson obteve valor baixo, a média aritmética obteve resultado mais próximo da avaliação do usuário. Já para os outros dois fóruns, a predição apresentou resultados mais precisos do que a média aritmética.

Em (O'CONNOR & HERLOCKER, 1999), foram realizados testes para medir a precisão de algoritmos de agrupamento (baseados em modelo) como Average Link, ROCK e hMetis. Foi utilizado o coeficiente de Pearson como medida de distância entre os itens durante a formação dos agrupamentos. Os testes foram realizados com a base de

dados do Movielens. As métricas MAE e cobertura (*coverage*) foram aplicadas para medir a precisão dos algoritmos. O algoritmo hMetis foi o que apresentou melhor resultado.

SWAMI é um framework implementado em Java para o desenvolvimento e avaliação de algoritmos de filtragem colaborativa (SWAMI, 2002). O framework é disponível para estudos, sendo possível acessar e utilizar sua biblioteca de funções (API). Em (FISHER et al., 2000), são descritos estudos experimentais com o framework e a base de filmes EachMovie, onde foram avaliados 2 algoritmos: o método de correlação Pearson e o método de correlação associado a agrupamento. O método de correlação Pearson é o mais utilizado devido a sua simplicidade de desenvolvimento, mas requer um tempo de processamento alto para calcular a correlação de todos os usuários da matriz “usuários x itens”. Foi testada, também, a utilização de agrupamento (*clustering*) com o coeficiente Pearson. A associação das duas técnicas tornou o algoritmo mais eficaz e escalável, por ter reduzido o número de usuários a serem processados na etapa de predição, uma vez que esta passou a ser realizada entre usuários do mesmo agrupamento.

A base de filmes EachMovie foi criada pela Compaq Systems Research Center durante um período de 18 meses de uso, com o objetivo de ser utilizada em experimentos de algoritmos de filtragem colaborativa. As descrições dos filmes foram cedidas pela IMDB. Durante esse tempo, cerca de 72.000 usuários avaliaram 1.628 filmes, gerando um total de 2.811.983 avaliações. As notas foram dadas em uma escala de 0 a 5. A base foi utilizada em vários estudos experimentais descritos nessa seção. A base está disponível para estudos e experimentos. Para obtê-la, basta escrever um e-mail para a Compaq concordando com os termos de uso (EACHMOVIE, 2002, IMDB, 2002).

#### **2.4.6 Trabalhos Relacionados a Sistemas de Recomendação**

O sistema de recomendação GroupLens (RESNICK et al., 1994) é um sistema que indica se artigos postados em fóruns de discussão são de interesse ou não do usuário. O objetivo principal é diminuir o tempo gasto pelas pessoas para ler notícias em fóruns. Os usuários avaliam os artigos em notas de 1 a 5, de acordo com interesse do mesmo em lê-lo. Quando os artigos são inseridos no fórum, o sistema utiliza as notas dadas por pessoas que costumam concordar nas avaliações, e prediz se o artigo é de interesse.

Apenas um artigo é recomendado por vez. O sistema utiliza um algoritmo de filtragem colaborativa baseado em memória. A distância entre os usuários é medida através do coeficiente de correlação Pearson (KONSTAN et al., 1997).

O sistema de recomendação de músicas Ringo (SHARDANAD & MAES, 1995), desenvolvido no MIT Media-lab, permite que o usuário requisite 3 tipos de recomendação: itens de interesse, itens rejeitados e nota atribuída a um determinado item. Inicialmente, o usuário recebe, ao se cadastrar, uma lista com 126 artistas, os quais avaliam com notas de 1 a 7. Esta abordagem é utilizada para resolver o problema da partida a frio, onde o sistema ainda não armazena informações suficientes sobre o usuário. O sistema de recomendação determina a similaridade dos usuários através de algoritmos baseados em memória, que utilizam coeficientes de correlação estatísticos, inclusive variações do coeficiente Pearson. A recomendação inicial era insatisfatória, mas à medida que os usuários avaliavam mais artistas, o desempenho do sistema melhorou.

Fab (BALABANOVIC & SHOHAM, 1997), desenvolvido na Stanford University, é um sistema de recomendação de páginas Web, que utiliza abordagem de recomendação híbrida baseada em conteúdo e em colaboração. A recomendação de páginas é realizada em duas etapas. Na primeira etapa, é executado o algoritmo de filtragem colaborativa. Na segunda etapa, os itens a serem recomendados passam por uma manipulação de conteúdo, com o objetivo de eliminar itens não relevantes. É feita uma comparação entre a classificação das páginas e as palavras-chaves definidas no perfil do usuário. O usuário avalia as páginas com notas de 1 a 7. Fab utiliza agentes de busca na Web para seleção de páginas e agentes de filtragem que eliminam os documentos já lidos e descartados pelo usuário.

O sistema de recomendação de filmes MovieLens (DAHLEN et al., 1998, HERLOCKER, 2001, MOVIELENS, 2002) está disponível na Internet, sendo necessário apenas cadastro de alguns dados pessoais para ter acesso ao sistema. Ao se cadastrar, sugere-se que o usuário avalie uma lista de filmes com notas de 1 a 5. O número mínimo de avaliações para se obter uma primeira recomendação é de 5 filmes, sendo indicado pelo menos 15 avaliações. Este requisito é necessário devido ao problema da partida a frio. O formulário de avaliação é exibido na **Figura 4**. Neste formulário, o título do filme é um acesso para o site IMDB, direcionado para a página de descrição do filme (IMDB, 2002). Neste caso, o usuário é reportado a um site externo na mesma janela do sistema, o que pode fazer o usuário perder o foco na

avaliação. A recomendação de filmes é realizada em uma lista de cinco melhores itens (“Top 5”), onde estes são apresentados ao lado da nota média. A base do MovieLens tem 3000 filmes e é atualizada semanalmente com os novos lançamentos nos Estados Unidos com a colaboração do IMDB.

MovieLens  
helping you find the right movies

Rating more movies improves your predictions; you've rated 6 so far.

**Please rate more movies**

MovieLens predicts what movies you are likely to enjoy based on your opinion of previously viewed movies. In order to make recommendations, MovieLens typically needs you to rate 5 to 15 movies. Please go through the following movies and rate any you have seen. As soon as you have rated enough movies, you will automatically be taken to your personalized MovieLens home page. Thank you for using MovieLens.

YOUR RATING	GENRE	TITLE	REVIEWS
? unseen	Crime	Any Number Can Win (Mélodie en sous-sol) (1963) (IMDb)	
? unseen	Drama	Valmô (1950) (IMDb)	
? unseen	Horror, Mystery Sci-fi	Die, Monster, Die! (1965) (IMDb)	
? unseen	Comedy, Drama War	Good Morning Vietnam (1987) (IMDb)	
? unseen	Drama	Great Santini, The (1979) (IMDb)	
? unseen	Comedy, Musical	Hole in the Head/A (1959) (IMDb)	
? unseen	Comedy, Drama	My Mother's Castle (Le Château de ma Mère) (1990) (IMDb)	
? unseen	Comedy	Phat Beach! (1996) (IMDb)	
? unseen	Drama, War	Ran (1985) (IMDb)	
? unseen	Drama, Thriller	Taxi Driver (1976) (IMDb)	

Submit ratings & see more

**Figura 4 – Formulário de avaliação de filmes do MovieLens.**

TeamWorks (MOTTA, 1999), desenvolvido na COPPE/UFRJ, utiliza uma abordagem de recomendação baseada em fluxo de informação entre grupos de trabalho cooperativos. Um grupo recomenda documentos para outro grupo, chamado fluxo de informação. Filtros são aplicados ao fluxo para eliminar documentos não relevantes. Essa abordagem é semelhante ao sistema Tapestry onde as pessoas recomendam diretamente itens umas para as outras. Nessa abordagem é necessário que as pessoas se conheçam e cooperem entre si. O sistema foi implementado em Lotus Notes.

QuickStep (MIDDLETON et al., 2002) é um sistema de recomendação híbrido, baseado em conteúdo e colaboração, que recomenda publicações científicas on-line. Utiliza um algoritmo de classificação dos documentos integrado a uma ontologia que fornece as preferências dos usuários.

A **Tabela 3** apresenta um comparativo dos sistemas de recomendação discutidos considerando alternativas, características e abordagens utilizadas.



**Tabela 3 - Comparativo entre sistemas de recomendação.**

Sistema de recomendação	Alternativa	Abordagem	Características
<b>GroupLens</b>	Recomendação de artigos para fórum eletrônico.	Colaborativa.	<ul style="list-style-type: none"> <li>• perfil do usuário;</li> <li>• avaliação de artigos com notas de 1 a 5.</li> </ul>
<b>Movielens</b>	Recomendação de filmes.	Colaborativa.	<ul style="list-style-type: none"> <li>• perfil do usuário;</li> <li>• avaliação de artigos com notas de 1 a 5.</li> </ul>
<b>Ringo</b>	Recomendação de músicas.	Colaborativa.	<ul style="list-style-type: none"> <li>• perfil do usuário;</li> <li>• avaliação de músicas com notas de 1 a 7.</li> </ul>
<b>Fab</b>	Recomendação de páginas Web.	Híbrida.	<ul style="list-style-type: none"> <li>• agentes de busca e filtragem;</li> <li>• perfil do usuário;</li> <li>• avaliação de páginas com notas de 1 a 7.</li> </ul>
<b>TeamWorks</b>	Filtragem para grupos de trabalho cooperativo.	Colaborativa.	<ul style="list-style-type: none"> <li>• perfil do usuário;</li> <li>• fluxo da informação e Lotus Notes.</li> </ul>
<b>QuickStep</b>	Recomendação de artigos científicos on-line.	Híbrida.	<ul style="list-style-type: none"> <li>• ontologia para capturar preferência dos usuários.</li> </ul>

#### 2.4.7 Aplicação de Sistemas de Recomendação em Sistemas de Comércio Eletrônico

Em (SCHAFER et al., 2001), é mostrado como o uso de sistemas de recomendação tem crescido em lojas de comércio eletrônico. A abordagem ajuda o consumidor a encontrar o produto desejado, transformando simples navegadores em consumidores. Além disso, possibilita que a loja seja personalizada para cada consumidor. A recomendação dos produtos pode ser baseada nos itens mais vendidos do site, na localização do consumidor ou na análise do comportamento do usuário em compras anteriores para prever produtos que podem ser comprados no futuro. Neste trabalho também é mostrado um comparativo entre sites de comércio eletrônico, como Amazon.com, CDNow.com, eBay.com, MovieFinder.com e Drugstore.com (AMAZON, 2002, CDNOW, 2002, MOVIEFINDER, 2002, DRUGSTORE, 2002).

Na Amazon.com, cada usuário tem uma loja personalizada onde armazena referências para os itens comprados, para itens que o usuário já possui ou que deseja comprar. Os itens são avaliados em notas de 1 a 5. O usuário informa se o item deve ser utilizado para gerar recomendações. O usuário também pode enviar um comentário relativo ao item. A **Figura 5** ilustra o formulário de avaliação de itens da Amazon.com. A figura exibe a lista de itens relacionados ao usuário, os títulos e imagens associadas aos itens. O título é um acesso a página com a descrição detalhada do item. A figura também mostra 6 botões de avaliação mutuamente exclusivos com notas de 1 a 5 e uma opção para o usuário informar que não sabe a nota a ser dada.

**Figura 5 – Avaliação de itens na Amazon.com.**

O site oferece os seguintes serviços de recomendação de livros:

- Customers who bought:* ao selecionar um livro, o sistema mostra uma lista de livros comprados por clientes que compraram o item selecionado;
- Your recommendations:* o usuário pode requisitar recomendações online após avaliar alguns itens que tenha comprado. O sistema verifica itens que sejam de interesse baseado nos itens avaliados.
- Eyes:* o usuário requisita o envio de novos itens adicionados ao catálogo da loja por correio eletrônico. Para isso, o usuário especifica uma consulta simples ou booleana mais complexa, com base nos campos autor, título, assunto ou ISBN.

- d) *Customer comments*: um usuário pode ver comentários enviados por outros usuários para um determinado item selecionado. O usuário pode ainda avaliar o comentário como útil ou não. Essa avaliação é exibida ao lado do comentário, por exemplo “2/3 dos usuários acharam este comentário útil”.

A loja eletrônica CDNOW.com oferece os serviços de recomendação:

- a) *Customers who bought*: serviço semelhante ao da Amazon.com;
- b) *Related artists*: ao selecionar um determinado CD, o usuário pode requisitar uma lista de artistas que sejam similares ao selecionado;
- c) *My CDNOW*: os usuários selecionam os CDs que estes possuem e agrupam em dois grupos: CDs aprovados e CDs rejeitados. Quando o usuário requisita a recomendação online, o sistema faz a predição de 6 itens que o usuário poderia gostar.

A **Tabela 4** mostra um comparativo dos serviços oferecidos nos sites de comércio eletrônico discutidos. Para cada serviço, é identificada a técnica de recomendação utilizada.

**Tabela 4 – Comparativo entre sites de comércio eletrônico.**

Sites de Comércio eletrônico	Características	Serviço oferecido	Técnica utilizada
Amazon.com		Customers who bought	Correlação item-item.
		Your recommendations	Correlação usuário – usuário.
		Eyes	Baseado em atributo e conteúdo.
		Customer comments	Baseado na própria experiência do usuário.
CDNow.com		Customers who bought	Correlação item-item.
		MyCDNow	Correlação usuário – usuário.

#### 2.4.8 Problemas e Desafios dos Sistemas de Recomendação

Um problema dos Sistemas de Recomendação se refere à confiabilidade das recomendações (RESNICK, 1997). Em (HERLOCKER, 2001), o autor afirma que a justificativa da recomendação é um fator importante para que o usuário confie no sistema e aceite a recomendação. Neste trabalho, o autor discute como e porque explicar

uma recomendação dada ao usuário, bem como alternativas de interface. Foram realizados estudos experimentais no MovieLens onde 86% dos participantes gostariam de receber as explicações. As melhores justificativas e interfaces votadas foram:

- a) histograma com as distribuições das notas dadas pelos usuários similares;
- b) desempenho do sistema, por exemplo, o sistema acertou 75% das indicações;
- c) a similaridade do item recomendado com o perfil do usuário.

No sistema de recomendação da Amazon.com, por exemplo, o usuário pode requisitar uma explicação sobre a recomendação feita. A **Figura 6** apresenta um exemplo de justificativa deste sistema. A figura apresenta detalhes do livro recomendado como título, autor, preço e avaliação média (em forma de estrelas). O usuário indica se já possui o livro ou rejeita a recomendação na própria página da justificativa. É apresentado também o item utilizado para gerar a recomendação e a avaliação do usuário. Na parte inferior da figura, é exibido uma lista de livros comprados por usuários que compraram o item recomendado (*Customers Who Bought*). Este tipo de justificativa corresponde a alternativa c) apresentada anteriormente.

The screenshot shows a page titled "Books Recommendations > Why was I recommended this?". It features a recommended item, a table for rating the item, and a list of similar items.

**We recommended...**

**Rodin : Sculpture & Drawings**  
 by Catherine Lampert  
 Average Customer Rating: ★★★★★  
 Our Price: \$21.00  
 (Own it) (Not interested)

**Usuário rejeita o item recomendado.**

**Because you purchased or rated...**

Purchased or Rated Items	Not Rated	Your Rating:	Exclude Item
		Don't like it < > I love it!	
<b>Rodin</b> by Dominique Jarrassé, et al	?	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input checked="" type="radio"/> 5	<input type="checkbox"/>

**Item utilizado como base da recomendação.**

**Customers who bought the items above also bought:**

- Rodin**  
by Auguste Rodin, et al
- Rodin's Monument**  
to Victor Hugo
- Rodin**  
by Kirk Varnedoe, et al
- LOOK INSIDE Rodin**

**Customers who bought.**

**Figura 6 - Justificativa da recomendação apresentada na Amazon.com.**

Outro problema dos sistemas de recomendação, discutido em (MIDDLETON et al., 2002, RASHID et al., 2002), é o problema da partida a frio onde o sistema não tem dados suficientes para recomendar itens a um novo usuário. Estes autores realizaram estudos sobre como capturar informação inicial dos usuários. Outro fator importante é como manter o perfil do usuário atualizado, uma vez que a eficiência do sistema depende da participação do usuário (BUCKLEY, 1995).

MIDDLETON et al. (2002) propõem o uso de estruturas semânticas de conhecimento, como ontologias, para prover preferências dos usuários e conhecimento do domínio. Foi desenvolvida uma ferramenta de análise de rede baseado em ontologia para identificar comunidades de práticas. Estas comunidades são grupos informais de pessoas que compartilham algum interesse em comum. A ontologia foi integrada ao sistema de recomendação QuickStep.

RASHID et al. (2002) analisaram diferentes abordagens para obter preferências de novos usuários no sistema. Os autores sugerem utilizar uma lista inicial de itens a serem avaliados. Os problemas desta abordagem são: como selecionar os itens da lista inicial, quantos itens são necessários e como apresentá-los. As diversas abordagens foram testadas no sistema de recomendação de filmes MovieLens, com 300 novos usuários.

O tratamento de matrizes “usuários x itens” esparsas pode ser considerado o principal problema dos algoritmos de filtragem colaborativa. Matrizes esparsas fazem com que os usuários sejam comparados uns aos outros, com base em poucas avaliações. Isto pode associar usuários que não sejam realmente similares. Além disso, existe um alto custo de processamento das matrizes quando estas são esparsas, pois aumentam o número de predições a serem realizadas, aumentando conseqüentemente o tempo de processamento.

Segundo SARWAR et al. (2001), os dois grandes desafios dos sistemas de recomendação são a escalabilidade dos algoritmos de filtragem colaborativa e melhoria da qualidade das recomendações. O primeiro desafio trata do tempo gasto na árdua tarefa de comparar usuários similares, que é o gargalo dos algoritmos de filtragem colaborativa. O segundo desafio é tornar as recomendações realmente úteis aos usuários. Segundo os autores, os dois desafios são de uma certa forma conflitantes, pois quanto menor o tempo gasto no cálculo da similaridade entre os usuários, pior vai ser a qualidade da recomendação e vice-versa. Por esse motivo, as duas questões devem ser levadas em consideração simultaneamente.

## 2.5 Considerações Finais

Neste capítulo foram apresentados os conceitos e problemas relacionados a área de recuperação de informação na Web. Em seguida, foi apresentada a proposta de utilização de filtragem de informação para resolver o problema de sobrecarga de informação na Web. O objetivo da filtragem é eliminar itens não relevantes ao usuário. Os sistemas de recomendação surgiram como uma *alternativa para solucionar o problema*. Estes sistemas têm o objetivo de encontrar itens relevantes, ao invés de filtrar os não relevantes. Os sistemas de recomendação são baseados em conteúdo ou em colaboração. A recomendação baseada em conteúdo utiliza as experiências anteriores do usuário para encontrar novos itens de interesse. Já a abordagem por colaboração é focada na idéia de encontrar pessoas similares e recomendar itens bem avaliados.

Este capítulo apresentou um histórico da pesquisa em sistemas de recomendação, e os principais projetos, tanto acadêmicos, quanto comerciais.

No próximo capítulo, conceitos relacionados a bibliotecas digitais serão apresentados. Alguns problemas discutidos neste capítulo são encontrados em bibliotecas digitais, pois estas são provedoras de informação similares a ferramentas de busca na Web. Trabalhos e projetos da área também serão apresentados.

## Capítulo 3

### Bibliotecas Digitais

---

*Este capítulo apresenta o estudo de conceitos e características de bibliotecas digitais e/ou virtuais na Web. Foram selecionadas as bibliotecas de documentos científicos mais significativas ao enfoque do trabalho. Bibliotecas de mídias como música e vídeo não foram abordadas, por apresentarem questões e problemas específicos que não fazem parte do objetivo deste estudo.*

#### 3.1 Introdução

Bibliotecas digitais geralmente são associadas a dois significados distintos. Em geral, pesquisadores têm a visão de que uma biblioteca digital é um repositório de conteúdo que apoia comunidades de usuários. Já para bibliotecários, o termo significa instituições ou serviços. Pesquisadores se preocupam em como desenvolver um sistema de acesso eficiente, enquanto que bibliotecários se preocupam em prover diferentes serviços aos usuários (FOX et al., 1995, BORGMAN, 1999).

Bibliotecas digitais envolvem diversas questões complexas e avançadas de sistemas de informação, como armazenamento de documentos digitais, gerenciamento de banco de dados distribuídos e heterogêneos, recuperação e filtragem de informação, suporte à colaboração, direitos de acesso, dentre outras (FOX & MARCHIONINI, 1998).

A pesquisa e prática na área de bibliotecas digitais tiveram início em 1991. Inicialmente, foi utilizado o termo “bibliotecas eletrônicas”, sendo substituído por “bibliotecas digitais” após um período de dois anos. Em 1993 surgiram as primeiras seções especiais em jornais e periódicos. No ano seguinte, aconteceram diversas conferências, tutoriais e *workshops* na área. Em 1995, a revista *Communications of ACM* dedicou uma seção especial (v. 38, n.4) à discussão sobre história, definições, desafios, projetos e perspectivas futuras de bibliotecas digitais. Nesta edição, a ACM

*Publishing* descreve sua estratégia em montar uma biblioteca digital na Web para oferecer acesso online a seu catálogo, além de prover diversos serviços à comunidade científica. Nesta edição também são apresentadas iniciativas da Indústria, como da Xerox, da NSF (*National Science Foundation*) e da NASA. NSF é a agência americana de fomento a pesquisa científica (NSF, 2003).

A pesquisa na área se desenvolveu rapidamente na última década. Milhares de bibliotecas digitais vêm sendo desenvolvidas, indo desde pequenas coleções restritas a organizações ou comunidades, até grandes bibliotecas nacionais que oferecem material de pesquisa e cultura (WOLFRAM & XIE, 2002). Cada uma dessas bibliotecas tem características específicas, desenvolvidas em tecnologias diferentes e sistemas de busca seguindo modelos variados. O usuário tem que se adaptar às diversas formas de consulta e interfaces ao acessar cada uma delas. Estas questões vêm sendo consideradas desde 1998, quando FOX & MARCHIONINI (1998), editores da seção especial “Toward a WorldWide Digital Library” da revista *Communications of ACM* (v. 41, n.4), discutiram questões relativas à criação de uma biblioteca digital mundial que unificasse essas diversas iniciativas em isolado. Os principais pontos de discussão foram a interoperabilidade das bibliotecas e o suporte a multi-linguagens. Interoperabilidade pode ser compreendida como o problema de integrar diferentes sistemas, considerando problemas de hardware, software, metadados, etc. Suporte a multi-linguagens corresponde tanto ao tratamento do conteúdo dos documentos escritos em línguas diferentes, quanto ao sistema de busca.

Mais recentemente, começaram a ser desenvolvidas bibliotecas virtuais. Essas bibliotecas armazenam referências bibliográficas ou endereços de sites na Web. Uma iniciativa neste sentido é o programa Prossiga do CNPq que tem financiado a criação de diversas bibliotecas virtuais temáticas (PROSSIGA, 2002). Uma outra iniciativa no Brasil é a Biblioteca Virtual de Saúde (BVS) desenvolvida pelo ministério da saúde brasileiro (BVS, 2002).

Uma das principais conferências da área é a JCDL (*The Joint Conference on Digital Libraries*), que em 2001 unificou as conferências anuais da ACM e da IEEE. A conferência anual ECDL (*European Conference in Digital Libraries*) tem o enfoque em projetos e pesquisas realizadas na Europa, apesar de não ser restrito a esta. Em 2001, foi realizado um *workshop* com a cooperação DELOS-NSF sobre personalização e uso de sistemas de recomendação em bibliotecas digitais. DELOS é uma rede de excelência em



bibliotecas digitais que financia projetos na Europa, organiza diversos eventos como fóruns de discussão, *workshops* e cursos temporários em universidades (DELOS, 2003).

### **3.4 Sistemas de Busca em Bibliotecas Digitais**

Um dos principais requisitos para o desenvolvimento de uma biblioteca digital é oferecer um serviço de busca eficiente aos documentos mantidos em seu repositório. Os sistemas de busca em bibliotecas são similares a sistemas de busca na Web, só que o universo de pesquisa é restrito à coleção da biblioteca (WOLFMAN & XIE, 2002).

Geralmente bibliotecas digitais utilizam três tipos de busca: busca por texto completo, por campos da referência bibliográfica e por navegação. Na busca por texto completo, esta é realizada em todo o conteúdo do documento e a interface de consulta geralmente apresenta uma caixa de texto simples, como ferramentas de busca na Web.

Na busca por campos da catalogação, esta é especificada através de campos como título, autor, ano de publicação, dentre outros. Às vezes, a pesquisa pode ser realizada combinando os campos através de conectores lógicos, como AND e OR. Este tipo de busca é na maioria das vezes chamada de busca avançada.

Na busca por navegação, o usuário navega através dos títulos de periódicos, jornais e revistas até encontrar o documento procurado. Esta abordagem se assemelha a busca por categorias disponibilizadas por algumas ferramentas de busca como Yahoo (YAHOO, 2003).

### **3.5 Gerenciamento da Coleção**

A coleção pode ser atualizada de forma manual ou automática. A alimentação manual do repositório requer a seleção e classificação dos documentos de forma manual, bem como um software de cadastro para o gerenciamento do acervo. A alimentação automática utiliza *crawlers* ou *robots* para varrer a Web a procura de publicações científicas.

Um dos problemas da abordagem manual, é o custo e as dificuldades em obter informação formal de catalogação (como título, autor, resumo, etc.). Essas informações são necessárias para prover as mesmas facilidades de busca das bibliotecas físicas, mas requerem esforço humano de catalogação (WITTEN et al., 1996).

A abordagem automática de alimentação do catálogo reduz o custo e esforço na criação e atualização contínua do catálogo. Por outro lado, perde na qualidade do acervo, pois documentos indesejáveis, ou até pornografias, podem ser adicionados automaticamente. De acordo com (ENTLICH et al., 1997), a característica mais importante de uma biblioteca digital é sua coleção, cuja exigência principal é a seleção de materiais qualificados. Estas duas questões, custo e qualidade, devem ser ponderadas ao definir o método de gerenciamento da coleção.

### **3.6 Serviços Personalizados**

O objetivo principal de um provedor de informação, como uma biblioteca digital, é prover ao usuário a informação certa, no momento desejado através dos meios corretos. Para cumprir essa tarefa, bibliotecas oferecem serviços personalizados baseados em perfil de usuário (AMATO & STRACCIA, 1999). Os autores propuseram um esquema de definição de perfil de usuário específicos para esse problema. O esquema é composto por 5 categorias: informações pessoais; interesse do usuário; entrega de informações; ações do usuário e política de segurança.

A categoria informações pessoais representa a identificação do usuário, sua atividade profissional e dados para contato. A categoria interesse do usuário armazena o que o usuário deseja buscar na biblioteca, ou seja, sua necessidade de informação. A categoria entrega de informações representa quando e como as informações de interesse são enviadas ao usuário. A categoria ações do usuário indica quais ações do usuário serão gravadas para uso posterior do sistema. A política de segurança informa se os dados do usuário podem ser compartilhados com outros usuários. A **Tabela 5** apresenta com mais detalhe a definição de um perfil do usuário para uma biblioteca seguindo o esquema proposto. Algumas categorias apresentam subcategorias associadas.

**Tabela 5 – Esquema de perfil de usuário para  
uma biblioteca digital (AMATO & STRACCIA, 1999).**

<b>Categoria</b>	<b>Sub-categoria</b>	<b>Informações associadas</b>
<b>Informações pessoais</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• dados pessoais do usuário (nome, endereço eletrônico, telefone, etc.);</li> <li>• identificação do usuário no sistema (login e senha).</li> </ul>
<b>Interesse do usuário</b>	Conteúdo do documento.	<ul style="list-style-type: none"> <li>• título do documento;</li> <li>• palavras-chaves;</li> <li>• categorias.</li> </ul>
	Estrutura do documento.	<ul style="list-style-type: none"> <li>• formato do documento (PDF, PS, etc.);</li> <li>• tipo do documento (livro, artigo, etc.);</li> <li>• data de publicação.</li> </ul>
	Fontes de busca do documento.	<ul style="list-style-type: none"> <li>• coleções utilizadas na busca dos documentos;</li> <li>• editores;</li> <li>• páginas Web.</li> </ul>
<b>Entrega de informação</b>	Modo de entrega.	<ul style="list-style-type: none"> <li>• forma de entrega das informações (por correio eletrônico, no próprio sistema, etc);</li> <li>• endereço de entrega.</li> </ul>
	Tempo de entrega.	<ul style="list-style-type: none"> <li>• periodicidade da entrega de informação.</li> </ul>
<b>Ações do usuário</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• documentos que o usuário leu;</li> <li>• avaliação do usuário;</li> <li>• páginas que o usuário acessou.</li> </ul>
<b>Política de segurança</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• lista de usuários que pode acessar o perfil;</li> <li>• dados que podem ser compartilhados.</li> </ul>

Um dos serviços personalizados utilizados em bibliotecas digitais é o serviço de alerta, onde o sistema avisa ao usuário novos documentos inseridos no catálogo. Geralmente o usuário recebe uma notificação por correio eletrônico contendo a lista de documentos encontrados.

O sistema HERMES (FAENSEN et al., 2001) é um exemplo de serviço de alerta. Ele detecta o cadastro de novos documentos em diversas bibliotecas digitais de fontes heterogêneas, faz a integração dessas fontes e envia documentos aos usuários que satisfaçam as restrições definidas no perfil do usuário. As fontes de informação varridas pelo sistema são: Springer (*Lecture Notes*), NCSTRL (*Networked Computer Science Technical Reference Library*) e Nature (SPRINGER, 2002, NCSTRL, 2003). O perfil

do usuário é composto por uma consulta booleana e uma política de notificação. A consulta é expressa em sintaxe SQL. Na política de notificação, o usuário escolhe a periodicidade na qual quer receber as mensagens.

Utilizando o esquema proposto em (AMATO & STRACCIA, 1999), a representação do perfil do usuário no sistema HERMES utiliza as categorias interesse do usuário e entrega de informação. A Tabela 6 exibe a modelagem do perfil para esse sistema.

**Tabela 6 - Definição do perfil do usuário no sistema Hermes.**

<b>Categoria</b>	<b>Sub-categoria</b>	<b>Informações representadas na categoria</b>
<b>Interesse do usuário</b>	Conteúdo do documento.	<ul style="list-style-type: none"> <li>• consulta booleana.</li> </ul>
<b>Entrega de informação</b>	Modo de entrega.	<ul style="list-style-type: none"> <li>• por correio eletrônico.</li> </ul>
	Tempo de entrega.	<ul style="list-style-type: none"> <li>• periodicidade definida pelo usuário.</li> </ul>

A arquitetura do sistema Hermes é composta de 3 componentes: observador, filtro e notificador. O componente observador recebe novas entradas dos provedores de informação. O filtro submete a consulta ao conjunto de novas entradas, gerando mensagens a serem enviadas ao usuário. O notificador aplica a política de notificação na lista de mensagens.

### **3.7 Trabalhos Relacionados**

A Biblioteca Digital da ACM centraliza documentos científicos publicados pela ACM *Publishing* em periódicos, jornais e conferências da área de computação (ACM D.L., 2002).

A biblioteca oferece os três tipos de busca discutidos na seção 3.2: por texto completo, por campo bibliográfico e por navegação. A busca principal disponibilizada na página inicial da biblioteca é a busca por texto completo, que tem uma interface similar a ferramentas de busca na Web.

A busca por navegação pode ser feita por periódicos, por SIGs (*Subject Interested Group*) e por conferências. Na navegação por periódico, o usuário navega pelo nome do

periódico (ACM Transactions, Communications of ACM, etc.) e seleciona ano e edição de interesse. O sistema de busca retorna a lista de documentos publicados na edição.

A busca por campos específicos da catalogação é disponibilizado como busca avançada, onde o usuário define a consulta utilizando os campos título, resumo, autor, data de publicação, ISBN, entre outros. O usuário pode ainda definir a consulta através de uma expressão booleana.

O sistema de busca realiza a ordenação da lista de documentos retornados pela similaridade entre estes e a consulta. O sistema apresenta a porcentagem de quanto o documento satisfaz a especificação da consulta. O usuário pode mudar a ordenação do resultado por data de publicação. O resultado da busca apresenta os títulos dos documentos encontrados, que correspondem a ligações para as referências bibliográficas dos documentos. A cada documento que o usuário consulta, uma nova janela do navegador é aberta, o que pode tornar a interface bastante confusa.

A página de referência bibliográfica, ilustrada na **Figura 7**, apresenta: descrição do documento; acesso ao arquivo PDF; palavras-chaves utilizadas na classificação do documento e lista de documentos que outras pessoas viram ao consultar o documento selecionado. Este último item é similar ao “*Customers who bought*” da Amazon.com, descrito no capítulo dois. As palavras-chaves associadas ao documento são links para consultas na biblioteca. Ao selecionar a palavra-chave, é feita a busca de documentos classificados com a mesma. O acesso ao conteúdo do documento só é disponível para membros.

Outro serviço disponível, a partir da referência bibliográfica, é a consulta por documentos similares. A opção está ilustrada na **Figura 7**. O sistema procura por documentos similares baseado na classificação por palavra-chave. O resultado é uma lista de documentos ordenados por similaridade.

acm **PORTAL** > home > about > feedback > login  
CAPES

**Conference on Information and Knowledge Management** > archive  
Proceedings of the eleventh international conference on Information and knowledge management > toc 2002, McLean, Virginia, USA

**Meta-recommendation systems: user-controlled integration of diverse recommendations**

**Authors**  
J. Ben Schafer University of Northern Iowa, Cedar Falls, IA  
Joseph A. Konstan University of Minnesota, Minneapolis, MN  
John Riedl University of Minnesota, Minneapolis, MN

**Busca por documentos similares.**

> full text > abstract > references > index terms > peer to peer

> Discuss > **Similar** > Review this Article > BibTeX Format Save to Binder

↑ FULL TEXT: Access Rules pdf 429 KB Acesso ao arquivo.

↑ ABSTRACT  
In this paper, we address such systems and introduce a new class of recommender system called meta-recommenders. Meta-recommenders provide users with personalized control over the generation of a single recommendation list formed from a combination of rich data using multiple information sources and recommendation techniques. We discuss experiments conducted to aid in the design of interfaces for a meta-recommender in the domain of movies. We demonstrate that meta-recommendations fill a gap in the current design of recommender systems. Finally, we consider the challenges of building real-world, usable meta-recommenders across a variety of domains.

↑ INDEX TERMS  
**Primary Classification:**  
H. Information Systems  
    ↳ H.1 MODELS AND PRINCIPLES  
        ↳ H.1.2 User/Machine Systems  
            ↳ **Subjects:** Human factors

**Palavras-chaves utilizadas na classificação do documento.**

**Keywords:**  
collaborative filtering, information filtering, recommender systems

↑ Peer to Peer - Readers of this Article have also read: **Lista de documentos que outras pessoas também viram.**

◆ On the design of a learning crawler for topical resource discovery  
**ACM Transactions on Information Systems (TOIS)** 19, 3  
Charu C. Aggarwal, Fatima Al-Garawi, Philip S. Yu

**Figura 7 - Referência bibliográfica na Biblioteca Digital ACM.**

A biblioteca da ACM oferece o serviço personalizado chamado *binders*, onde o assinante define consultas que são aplicadas periodicamente a base da biblioteca. O serviço é semelhante a um serviço de alerta baseado em perfil do usuário. Quando novos documentos são inseridos ao catálogo, agentes inteligentes comparam estes documentos com consultas pré-definidas e indicam por correio eletrônico. O *binder* é composto por uma especificação da consulta, periodicidade e forma da entrega. A **Figura 8** mostra a tela de definição do *binder*. O usuário pode habilitar ou desabilitar o

*binder* por um período de tempo. Se o usuário quiser, este pode se tornar de acesso público. Dessa forma, outros usuários podem utilizar a definição do *binder* para submeter a consulta no catálogo.

**acm** **PORTAL**  
DIGITAL LIBRARY

**Binder "Recommender Systems"**

<b>Binder Definition</b>		[View Contents]
Name:	Recommender Systems	(xexeo)
Update:	<input type="radio"/> Manually <input checked="" type="radio"/> Automatically	Especificação da consulta.
<b>Saved Search:</b>		
Recommender <and> Systems		
Enter your search above.		Ativa / desativa o binder.
<b>Notification:</b> Automatic Binders On		
<input type="radio"/> None <input checked="" type="radio"/> Notify me: catarina@cos.ufrj.br		
<b>Sharing:</b>		
<input checked="" type="radio"/> Not Shared <input type="radio"/> Share as:		
<b>Users:</b> Subscribers can access this Binder.		
Enter the Username above of a Portal Subscriber and click ➔ to add them to the list or click ✖ to clear the list.		Compartilhamento do binder.
<b>Groups:</b> Active Members can access this Binder.		
Choose a Group ...		
Select a Group from the pull-down above to add it to the list or click ✖ to clear the list.		
Update Binder		Delete Binder

Figura 8 – Definição de um binder na Biblioteca Digital ACM.

Utilizando o modelo de perfil de usuário definido por AMATO & STTRACIA (1999), o binder representa informações das categorias interesse do usuário, entrega da informação e política de segurança. A **Tabela 7** mostra as informações representadas no binder segundo esse modelo.

**Tabela 7 - Definição de um binder na biblioteca da ACM.**

<b>Categoria</b>	<b>Sub-categoria</b>	<b>Informações associadas</b>
<b>Interesse do usuário</b>	Conteúdo do documento.	<ul style="list-style-type: none"> <li>• consulta booleana.</li> </ul>
	Estrutura do documento.	<ul style="list-style-type: none"> <li>• formato do documento (PDF);</li> <li>• tipo do documento (livro, artigo, etc.);</li> <li>• data de publicação.</li> </ul>
	Fontes de busca do documento.	<ul style="list-style-type: none"> <li>• periódicos da ACM;</li> <li>• <i>anais de conferências</i> selecionadas.</li> </ul>
<b>Entrega de informação</b>	Modo de entrega.	<ul style="list-style-type: none"> <li>• forma de entrega das informações por correio eletrônico ou no próprio <i>binder</i></li> </ul>
	Tempo de entrega.	<ul style="list-style-type: none"> <li>• o usuário pode consultar manualmente;</li> <li>• envio automático por correio eletrônico, a medida que é feita atualização do catálogo.</li> </ul>
<b>Política de segurança</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• compartilhamento do binder;</li> <li>• definição de uma lista de usuários que podem acessar o binder.</li> </ul>

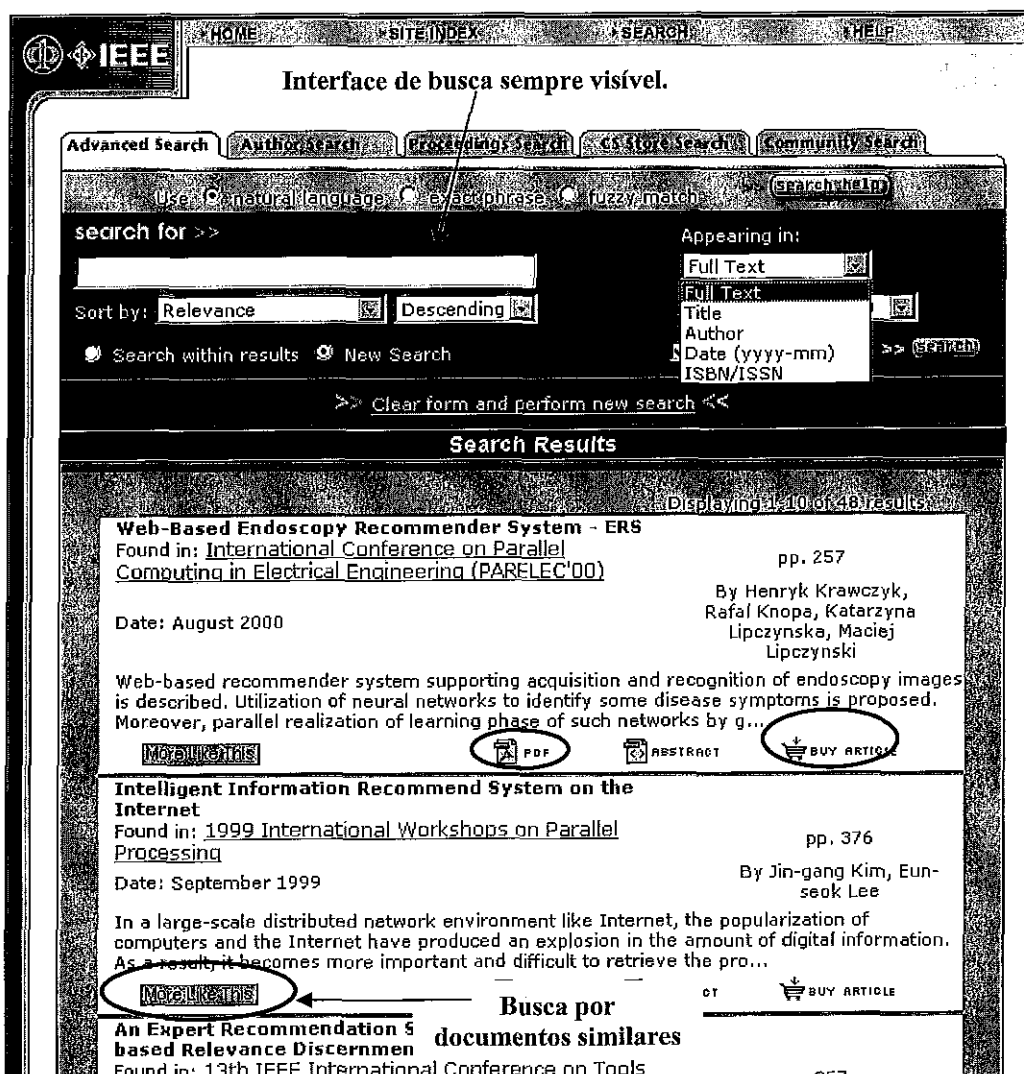
A Biblioteca Digital da IEEE oferece acesso a 20 periódicos (*Magazines e Transactions*) publicados pela *IEEE Computer Society* e a mais de 900 anais de conferências selecionadas (IEEE D.L., 2003).

A biblioteca oferece três tipos de busca: por texto completo, por campos da catalogação e por navegação. A busca por navegação é realizada através dos periódicos. A navegação apresenta a mesma estrutura utilizada na biblioteca ACM. A IEEE oferece uma opção a mais para os membros fazerem o download do PDF contendo toda a edição do periódico, ao invés de ter que fazer o download de cada um dos documentos.

A busca por texto completo e por campos bibliográficos são apresentadas em uma única interface. As opções são exclusivas, ou seja, não é possível combinar mais de duas opções na especificação da consulta. O usuário pode especificá-la através de uma consulta booleana ou utilizar uma frase exata com o uso de aspas. Além disso, o usuário pode restringir a busca selecionando um dos periódicos em uma lista de opções. O



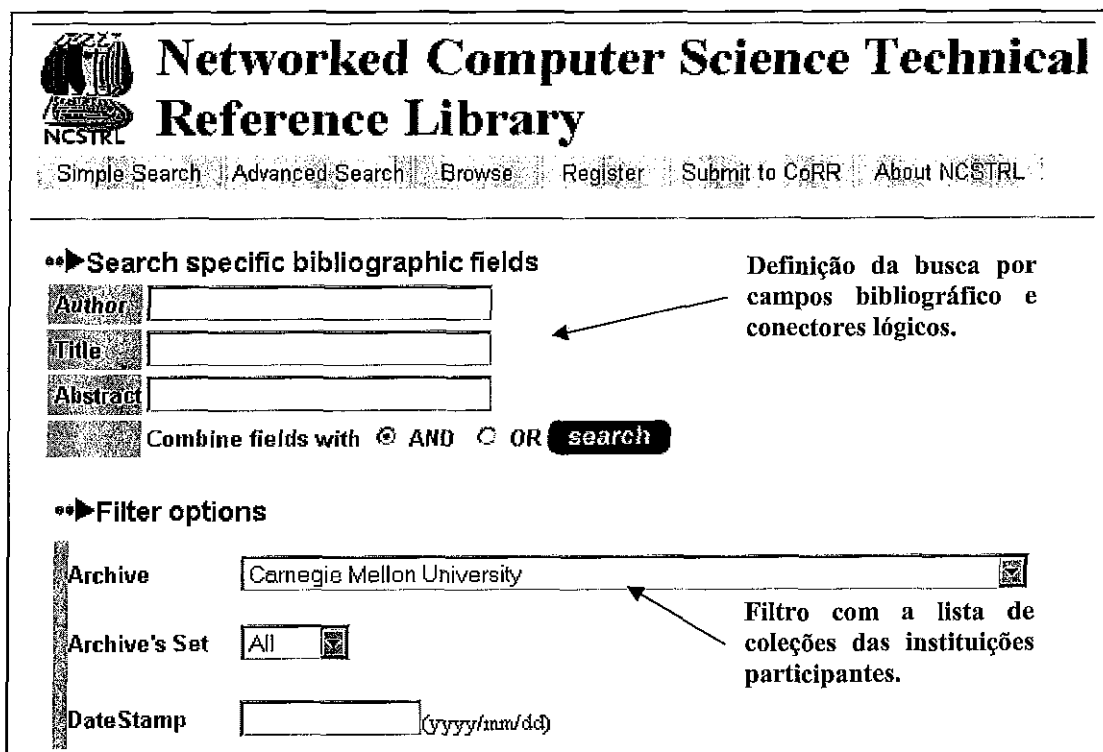
resultado da consulta apresenta dados da publicação, resumo, opção de acesso ao documento (para membros) e opção de compra (para não membros). O resultado da consulta é exibido na mesma página da busca, o que torna a interface simples e funcional, uma vez que as opções de busca estão sempre disponíveis. A **Figura 9** ilustra a interface de busca e o resultado de uma consulta. Também é possível buscar por documentos similares, como na ACM. Esta opção é chamada “*More like this*”, ilustrada na **Figura 9**.



**Figura 9 – Resultado da busca na biblioteca digital da IEEE.**

A biblioteca virtual **NCSTRL** (NCSTRL, 2003) centraliza coleções de universidades e instituições de pesquisa americanas (NASA, 2003, OLD DOMINION, 2003, VIRGINIA, 2003, VIRGINIA TECH, 2003).

A primeira opção da biblioteca é uma busca por campos bibliográficos. Na busca avançada, o usuário pode especificar a coleção de qual universidade deseja consultar, o ano de publicação, além dos campos bibliográficos combinados através de conectores lógicos. A **Figura 10** mostra a busca avançada, onde são disponibilizados três campos de texto para o usuário especificar autor, título e palavra-chave, podendo combiná-los com um conector AND ou OR. A figura também mostra os filtros disponíveis como: escolher o acervo de uma universidade específica em uma lista de opções e o ano de publicação.



**Networked Computer Science Technical Reference Library**

Simple Search | Advanced Search | Browse | Register | Submit to CoRR | About NCSTRL

►► Search specific bibliographic fields

Author

Title

Abstract

Combine fields with  AND  OR

►► Filter options

Archive

Archive's Set

DateStamp  (yyyy/mm/dd)

Definição da busca por campos bibliográfico e conectores lógicos.

Filtro com a lista de coleções das instituições participantes.

**Figura 10 - Busca avançada da NCSTRL.**

O resultado da busca é particionado em cada coleção participante. Ao selecionar um documento, a ficha bibliográfica e o resumo são exibidos. A **Figura 11** mostra o resultado de uma busca, onde é apresentado a lista das coleções e o número de documentos encontrados em cada uma delas, em um *frame* lateral esquerdo. A figura também mostra a descrição e o resumo de um documento selecionado, em um *frame* lateral direito.

A utilização de *frames* dificulta a visualização completa da tela. Outro aspecto a considerar, é a separação do resultado por coleção. O usuário teria que consultar coleção a coleção para ter uma visão completa dos documentos retornados. Seria interessante

oferecer a união dos resultados das diversas coleções, como funcionam os SMBs discutidos no capítulo dois.

**Networked Computer Science Technical Reference Library**  
 Ficha bibliográfica e resumo.

Simple Search | Advanced Search | Browse | Register | Submit to CoRR | About NCSTRL | OAI | Help

Lista de coleções.

archive	Hits
<u>Auburn-Eng</u>	1
<u>Brown Univ</u>	8
<u>Dartmouth</u>	4
<u>Discrete Mathematics and CS</u>	2
<u>ERCIMCWI</u>	4
<u>ERCIMSICS</u>	2
<u>FIT</u>	2
<u>IDSIA</u>	1
<u>Illinois at Urbana-Champaign</u>	2
<u>LTRS</u>	7
<u>MIT AI Lab</u>	4
<u>MTA SZTAKI</u>	3
<u>ODU</u>	3
<u>Oregon St</u>	1
<u>Oxford</u>	1
<u>RIACS</u>	1
<u>Stanford Univ</u>	2

**SEARCH RESULTS GROUPED BY ARCHIVE**

**ARCHIVE :Auburn-Eng**

<b>Title</b>	<i>Case-Based Reasoning: An Implemented Method</i>
<b>Authors</b>	Raman, Pradeep
<b>Abstract</b>	Case Based Reasoning (CBR) is the process by which system solves a given problem based on the knowled from solving precedents in the past. It is a popular AI the area of customer service or HelpDesks, whereby uses CBR as its problem solving approach is
<b>Archive</b>	Auburn-Eng
<b>Discovery Date</b>	1995-03-10
<b>Document ID</b>	oai:ncstrlh:auburn_eng:AUBURN_ENG//CSE95-04

Figura 11 - Resultado da busca na NCSTRL.

A biblioteca Springer (*Lecture Notes*) (SPRINGER, 2002) oferece uma combinação de busca por navegação e busca por referência bibliográfica. Primeiramente, o usuário navega por uma árvore contendo nomes de jornais e periódicos. Após selecionar um periódico, a biblioteca oferece a opção de buscar por campos bibliográficos com conectores lógicos. Outras opções de busca são disponibilizadas na biblioteca como busca avançada e busca no texto completo do documento. Estas opções ficam visíveis no menu principal da biblioteca. A Figura 12 exibe a busca avançada por referência bibliográfica. Esta figura apresenta dois campos de texto, listas de opções onde o usuário especifica campos da referência e conectores lógicos. Ainda na figura, são exibidas listas de opções onde o usuário especifica o período de publicação do documento e caixas de opções onde escolhe-se a língua para realizar fazer a consulta: francês, inglês, italiano e alemão.

A biblioteca possui um serviço de alerta onde o usuário escolhe um periódico no qual deseja receber atualizações. Essa abordagem é diferente da apresentada no sistema Hermes e na ACM, pois o usuário não precisa especificar uma consulta. Uma vantagem é a simplicidade de definição. Por outro lado, nem todos os documentos indicados podem ser de interesse, pois o usuário não informou nenhum critério de seleção, além

do nome do periódico. A Tabela 8 apresenta as informações utilizadas pelo serviço de alerta segundo o modelo definido em (AMATO & STRACCIA, 1999).

Forum

What's New Search Comment Helpdesk

Expert Search Bibliographic Expert Search Command Line Expert Search Fulltext Expert Search DOI LINK Site

**LINK Bibliographic Search**

Opções de busca disponíveis em um menu superior

Search for

within Abstract And

within Abstract And

case-sensitive CrossSearch in PubMed/Medline

add line start search clear form

**Command Line Search**

Published

Since Month Year Before Month Year

Languages

English German French Italian

start search clear form

Definição do período de publicação.

Figura 12 - Busca por campo bibliográfico na Springer.

Tabela 8 - Informações utilizadas no serviço de alerta da biblioteca Springer.

Categoria	Sub-categoria	Informações associadas
Interesse do usuário	Conteúdo do documento.	<ul style="list-style-type: none"> <li>nome do periódico.</li> </ul>
Entrega de informação	Modo de entrega.	<ul style="list-style-type: none"> <li>correio eletrônico.</li> </ul>
	Tempo de entrega.	<ul style="list-style-type: none"> <li>envio automático de correio eletrônico quando uma nova edição do periódico é publicada.</li> </ul>

A Biblioteca Digital da Nova Zelândia (WITTEN et al., 1996, NEW ZEALAND D.L., 2003), desenvolvida na University of Waikato, integra diversas coleções de documentos científicos. As coleções são separadas por áreas de interesse pré-determinadas. A principal coleção da biblioteca é a CSTR (*Computer Science Technical Reports*) que contém relatórios técnicos da área de ciência da computação. O gerenciamento da coleção é feito através de mecanismos automáticos que explora o conteúdo de documentos científicos em formato PS (*PostScript*) ou PDF. Nesta abordagem, o custo e o tempo de cadastro dos documentos é reduzido. A biblioteca oferece busca por texto completo (*full-text*). Foi desenvolvida uma ferramenta que cria um índice externo da coleção e realiza as buscas de forma mais rápida. Esta ferramenta é de domínio público e pode ser utilizada por outras bibliotecas.

O software CiteSeer (RUCKER & POLANCO, 1997, GILES et al., 1998) é um gerador automático de bibliotecas digitais, desenvolvido no NEC Research Center. CiteSeer utiliza ferramentas de busca na Web para procurar documentos científicos publicados em formato PS e PDF, fazendo download dos mesmos. Em seguida, é feito um processamento (*parsing*) do conteúdo do arquivo onde são extraídos os campos da referência bibliográfica e suas citações (GILES et al., 1998). O software Citesser e as coleções geradas estão disponíveis para uso, contanto que não seja para fins comerciais.

A biblioteca CiteSeer (*Scientific Literature Digital Library*) (LAWRENCE et al., 1999a, 1999b, BOLLACKER et al., 2000) é composta por documentos científicos encontrados na Web através do software Citeseer. A biblioteca possui um sistema de busca por texto completo no conteúdo dos artigos e nas referências. Quando o usuário seleciona um documento retornado na busca são exibidos: a descrição do documento, resumo, acesso ao arquivo PS ou PDF, lista de artigos citados no texto e uma lista de documentos similares. A **Figura 13** ilustra a página com a descrição do documento. O acesso ao arquivo é feito diretamente da URL de onde este foi publicado. A similaridade entre documentos é medida usando termos de maior frequência no texto e o inverso da frequência na coleção (GILES et al., 1998). O usuário pode consultar também uma lista de documentos vistos por outros usuários que também viram o documento selecionado. Esta opção é semelhante ao serviço “Customers who bought” da Amazon.com.

**Methods and Metrics for Cold-Start Recommendations (2002)** (Make Corrections) (3 citations)  
 Andrew J. Schein, Alexandrin Popescul, Lyle H. Ungar

View or download:  
[new.com/homepages/.../2002coldstart.ps](#)  
 Cached: [PS](#):[gz](#) [PS](#):[PDF](#) [DjVu](#) [image](#) [Update](#) [Help](#)

From: [hec.com/homepages/dipenrock/pap...](#) (more)  
 Homepages: [A. Popescul](#), [L. Ungar](#) [2]  
[HSRSearch](#) ([Update links](#))

**CiteSeer** Home Search Bookmark Context Related

**Avaliar ou comentar documentos.** →  1 2 3 4 5 (best)  
 Comment on this article

**Abstract:** We have developed a method for recommending items that combines content and collaborative data under a single probabilistic framework. We benchmark our algorithm against a naïve Bayes classifier on the cold-start problem, where we wish to recommend items that no one in the community has yet rated. We systematically explore three testing methodologies using a publicly available data set, and explain how these methods apply to specific real-world applications. We advocate heuristic... ([Update](#))

**Cited by: [More](#)**  
 Learning User Profiles for Content-Based - Filtering In Commerce (2002) ([Correct](#))  
 Notes on the CROC Curve - Andrew Schein, Department ([Correct](#))  
 REFEREE: An open framework for practical testing of... - Cosley, Lawrence.. (2002) ([Correct](#))

**Documentos similares ao selecionado.**

**Similar documents (at the sentence level):**  
**30.2%:** Methods and Metrics for Cold-Start Recommendations - Schein, Popescul, Ungar (2002) ([Correct](#))  
**5.8%:** Generative Models for Cold-Start Recommendations - Schein, Popescul, Ungar (2001) ([Correct](#))

**Active bibliography (related documents): [More](#) [All](#)**  
**0.5:** Probabilistic Models for Unified Collaborative and... - Popescul, Ungar.. (2001) ([Correct](#))  
**0.5:** Application of ART2 Networks and Self-Organizing Maps to... - Guntram Graef And (2001) ([Correct](#))  
**0.3:** Building Recommender Systems using a Knowledge Base of Product... - Ghani, Fano ([Correct](#))

**Similar documents based on text: [More](#) [All](#)**  
**0.5:** Clustering and Identifying Temporal Trends in... - Popescul, Flake.. (2000) ([Correct](#))  
**0.4:** Towards Structural Logistic Regression: Combining... - Popescul, Ungar.. (2002) ([Correct](#))  
**0.4:** Automatic Labeling of Document Clusters - Popescul, Ungar (2000) ([Correct](#))

**Citações (may not include all citations):**  
 102 Recommender systems (context) - Resnick, Varian - 1997  
 85 GroupLens: Applying collaborative filtering to Usenet news (context) - Nonstah, Miller et al. - 1997  
 79 Recommending and evaluating choices in a virtual community o... (context) - Hill, Stead et al. - 1995  
 55 Social information filtering: Algorithms for automating wor... - Shardanand, Maes - 1995  
 36 Recommendation as classification: Using social and content-b... - Basu, Hirsh et al. - 1998

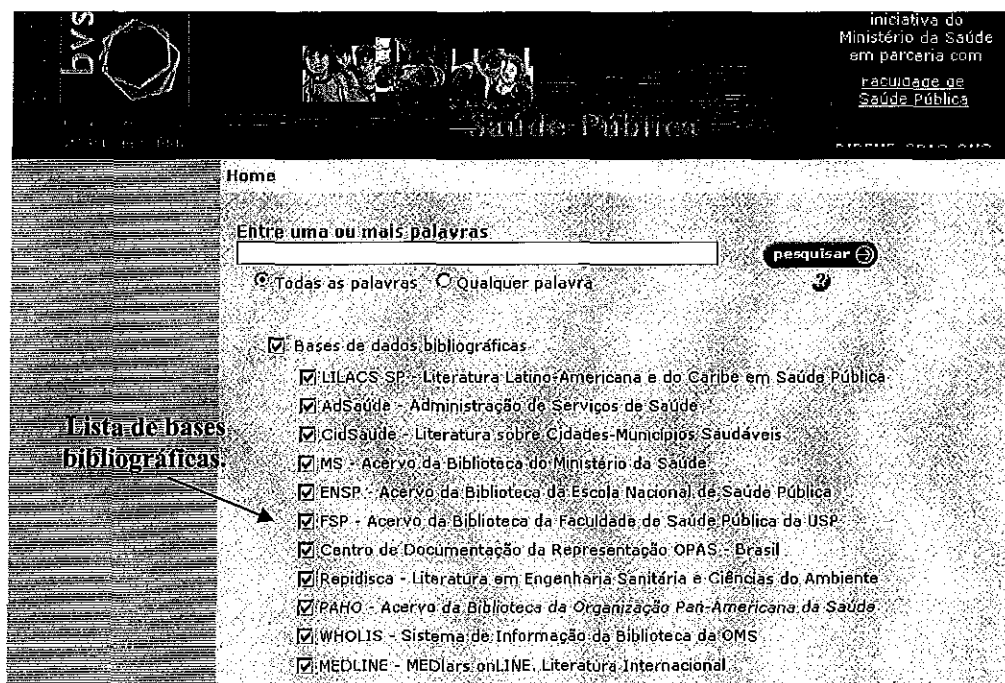
**Citações do documento.**

Figura 13 - Referência bibliográfica na Biblioteca Digital CiteSeer.

Na biblioteca CiteSeer, o usuário pode guardar as referências dos documentos em um perfil de usuário. O usuário avalia o documento de forma explícita com notas de 1 a 5. O sistema também observa ações do usuário e analisa o histórico de acesso do servidor Web para obter uma avaliação implícita do documento. Esta avaliação considera pesos para as ações do usuário. Documentos que o usuário guardou no perfil recebem um peso maior do que os documentos que ele somente acessou (BOLLACKER et al., 2000, LAWRENCE et al., 1999a, 1999b).

A interface da biblioteca apresenta diversas opções com cores diferentes na mesma página da descrição do documento, o que pode tornar a interface de difícil entendimento. Outro aspecto a considerar, é a opção de realizar a consulta no Google (GOOGLE, 2003), restringindo-a ao endereço (URL) da biblioteca. O resultado da consulta no Google apresenta acesso para a página de referência bibliográfica na biblioteca. Como a busca é feita na mesma janela da biblioteca, o usuário pode perder o objetivo da pesquisa.

A Biblioteca Virtual em Saúde (BVS, 2002) provê acesso a diversas bibliotecas virtuais e bases bibliográficas da área de saúde através de uma interface única de busca. A **Figura 14** mostra a interface de busca da biblioteca, composta por uma caixa de texto e a lista de bases de dados disponíveis para consulta. O resultado da consulta mostra quantos documentos foram encontrados por base. O nome da base é um acesso para a página com o resultado da consulta, aberta em uma nova janela do navegador.



**Figura 14 – Definição da busca na Biblioteca Virtual de Saúde (BVS).**

A vantagem dessa abordagem é a interface de consulta ser a mesma para qualquer base. Isto evita do usuário ter que preencher diferentes formulários de buscas das bibliotecas digitais participantes. O problema dessa abordagem é que o usuário precisa verificar o resultado da consulta nos sites das bibliotecas. O interessante seria prover transparência ao usuário de que base está sendo consultada. O sistema de busca poderia aplicar a consultas em diferentes bases de dados e retornar uma lista com a agregação dos resultados, como funcionam os sistemas de meta busca na Web, descritos no capítulo dois.

O comparativo entre características e serviços oferecidos nas bibliotecas discutidas estão descritos na **Tabela 9**.

**Tabela 9 - Comparativo entre as bibliotecas digitais.**

Bibliotecas Digitais	Serviços oferecidos	Características
ACM	<ul style="list-style-type: none"> <li>• busca por texto completo, por campos da catalogação e por navegação;</li> <li>• serviço de alerta (por correio eletrônico);</li> <li>• lista de documentos acessados por usuários que consultaram o documento selecionado;</li> <li>• busca por documentos similares.</li> </ul>	Base de artigos científicos da área de computação.
IEEE Computer Society	<ul style="list-style-type: none"> <li>• busca por texto completo, por campos da catalogação e por navegação.</li> </ul>	Base de artigos científicos da área de computação.
Springer – Lecture Notes	<ul style="list-style-type: none"> <li>• busca por navegação, campo bibliográfico e texto completo;</li> <li>• serviço de alerta por correio eletrônico.</li> </ul>	Base de artigos científicos área de computação. Oferece também de outras ciências como medicina e matemática.
Citeseer	<ul style="list-style-type: none"> <li>• perfil do usuário;</li> <li>• busca por texto completo;</li> <li>• navegação entre citações.</li> </ul>	Base de artigos científicos da área de computação.
Biblioteca Digital da Nova Zelândia	<ul style="list-style-type: none"> <li>• alimentação automática da coleção através de <i>crawlers</i> na Web;</li> <li>• busca por texto completo do documento.</li> </ul>	Base de artigos científicos da área de computação.



### **3.8 Considerações Finais**

Neste capítulo, foram apresentados conceitos de bibliotecas digitais e aplicações desenvolvidas na Web. Foram discutidas questões referentes a sistemas de busca, gerência da coleção e serviços personalizados. Um comparativo destas características foi apresentado para as bibliotecas analisadas. Foi apresentada uma modelagem de perfil de usuário. O modelo foi aplicado nas bibliotecas discutidas que oferecem serviços personalizados.

No próximo capítulo, será apresentada uma proposta de utilização de serviços personalizados e um sistema de recomendação para uma biblioteca digital na Web.

## Capítulo 4

# RecDoc: um sistema de recomendação para uma biblioteca digital virtual na Web

---

*Este capítulo apresenta uma biblioteca digital virtual na Web. A customização desta biblioteca para a Biblioteca Digital Virtual de Engenharia de Software é apresentada. Em seguida, é apresentado um sistema de recomendação, RecDoc, que sugere documentos de interesse aos leitores da biblioteca virtual. O sistema de recomendação proposto recomenda documentos ao usuário baseado na similaridade entre documentos ou entre usuários.*

### 4.1 Considerações Iniciais

Bibliotecas digitais, assim como outros provedores de informação, deveriam ajudar os leitores a encontrar documentos de interesse, através de recomendações personalizadas (MOONEY & ROY, 2002). Prover um sistema de busca personalizada e um serviço de entrega de informação parece ser um primeiro passo para disponibilizar informação relevante para as pessoas (AMATO & STRACCIA, 1999).

Sistemas de recomendação são serviços personalizados que permitem as pessoas compartilharem opiniões e se beneficiarem com a experiência alheia (TERVEEN & HILL, 2001).

Uma condição prévia para prover serviços personalizados é definir perfis de usuário. O perfil é uma representação de preferências do usuário. Baseado nesta representação, o sistema personalizado pode achar a informação pertinente.

Considerando estas questões, propomos a utilização de um sistema de recomendação RecDoc (ROCHA, 2002a) como um serviço de entrega de documentos aos usuários de uma biblioteca digital virtual na Web.

## 4.2 Construção de uma Biblioteca Digital Virtual na Web

Uma biblioteca digital virtual na Web foi desenvolvida com propósito genérico. Como exemplo de utilização, a biblioteca genérica foi configurada e customizada para a linha de pesquisa de Engenharia de Software da COPPE/UFRJ.

A biblioteca genérica foi desenvolvida em três etapas, seguindo o ciclo de vida de prototipação evolutiva (PRESSMAN, 2001). Em cada etapa, novas funcionalidades foram desenvolvidas e integradas.

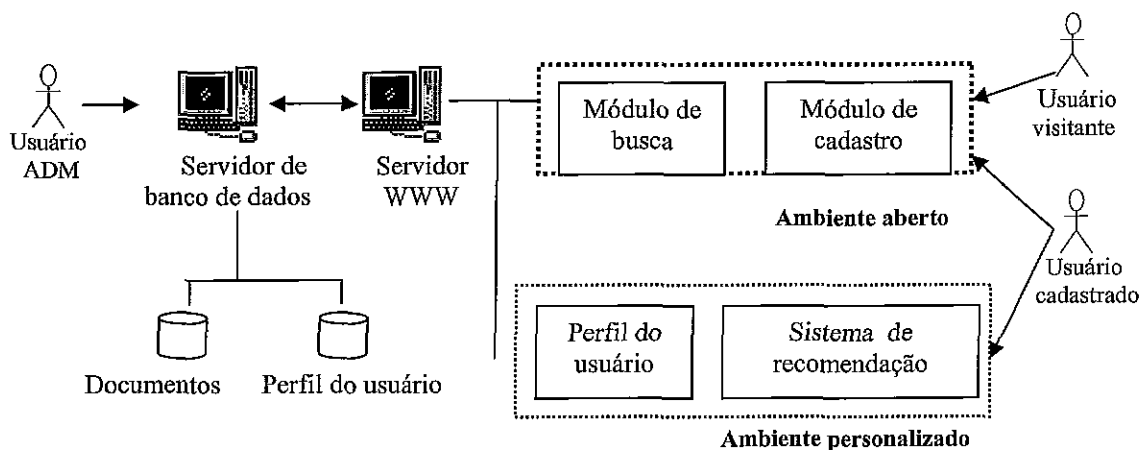
Na primeira etapa, foram especificados e implementados o **módulo de busca** e o **módulo de cadastro** da biblioteca. A especificação foi feita em casos de usos em UML, listados no ANEXO 1. Os dois módulos desenvolvidos fazem parte do que chamamos **ambiente aberto**, por ser acessível a todos os usuários visitantes da biblioteca. Este ambiente possibilita a busca de documentos e a gerência do catálogo.

Em uma segunda etapa, o perfil de usuário foi modelado para representar as necessidades de informação do usuário. As funcionalidades relacionadas ao perfil também foram especificadas em casos de uso em UML, implementadas em um módulo separado e integradas à biblioteca. Foram desenvolvidas funcionalidades que permitissem o compartilhamento de opiniões entre os usuários. O perfil e as *funcionalidades personalizadas* foram agrupados em um **ambiente personalizado**. Este ambiente está disponível para usuários cadastrados na biblioteca.

Na terceira etapa, foi desenvolvido o sistema de recomendação **RecDoc**. O sistema manipula o catálogo da biblioteca e as informações dos usuários registradas nos perfis.

A interação entre os ambientes e o sistema de recomendação está representada na arquitetura da biblioteca, ilustrada na **Figura 15**. Esta figura mostra os usuários da biblioteca, recursos utilizados e as bases de dados definidas. Os tipos de usuários são: visitante, cadastrado e administrador (ADM). O usuário ADM gerencia a base de dados, composta de uma base de documentos e uma de perfis de usuários. A biblioteca é executada no servidor Web com acesso ao servidor de banco de dados. A biblioteca foi desenvolvida em ASP (VBScript e JavaScript) e banco de dados SQL Server.





**Figura 15 - Arquitetura da biblioteca.**

As descrições dos módulos da biblioteca, assim como características referentes ao sistema de busca, gerenciamento da coleção e serviços personalizados serão discutidos na seção seguinte, utilizando o exemplo construído para Engenharia de Software.

### **4.3 Biblioteca Digital Virtual de Engenharia de Software**

A biblioteca genérica foi configurada e customizada para a linha de pesquisa de Engenharia de Software da COPPE/UFRJ. A biblioteca centraliza os documentos científicos produzidos pela linha de pesquisa, bem como referências para documentos produzidos por centros afins. Dessa forma, o acervo apóia os alunos nas suas atividades de pesquisa, oferecendo catálogo do material utilizado nos cursos, material relacionado aos projetos e conferências da área (ROCHA, 2002b). Além disso, a biblioteca está disponível on-line na Internet, possibilitando que os usuários consultem o acervo a qualquer hora e em qualquer lugar. Os usuários da biblioteca são professores, pesquisadores, estudantes de mestrado e doutorado.

Durante a configuração da Biblioteca de Engenharia de Software alguns ajustes foram necessários na interface e na base de dados da biblioteca genérica. Na customização, novas funcionalidades foram desenvolvidas e novas tabelas adicionadas à base de dados, para atender a demandas específicas do grupo de pesquisa.

### 4.3.1 Sistema de Busca da Biblioteca

O sistema de busca de documentos da biblioteca apresenta três opções de pesquisa: por palavra-chave, por título e por autor.

A busca por palavra-chave apresenta duas alternativas de busca. A primeira alternativa é a busca por texto onde pode ser digitado uma palavra-chave ou um conjunto de palavras. A segunda alternativa é uma busca por seleção, onde o usuário escolhe uma palavra-chave na lista de palavras da biblioteca. A **Figura 16** mostra as duas opções de busca por palavra-chave. A figura exibe uma caixa de texto para digitação e as letras iniciais das palavras-chaves, que ao serem selecionadas, atualizam a lista de opções. O menu principal da biblioteca com as opções de login e acesso ao perfil do usuário são exibidos na parte superior da figura. Na parte inferior, os acessos a busca por autor e por título são apresentados.

The screenshot shows the top navigation bar with links: 'Acervo Bibliográfico', 'Perfil do Usuário', 'Quem Somos?', 'Login', an email icon, a question mark icon, and 'Cadastre-se'. Below this is the header 'Engenharia de software' and 'Biblioteca Digital e Virtual'. The main search area is titled 'BUSCA POR PALAVRA-CHAVE' and contains two sections. The first section, 'busca por texto', has a text input field and a 'pesquisar' button. The second section, 'busca por seleção', has a dropdown menu with 'ADS' selected and a 'pesquisar' button. Below the search options is a section 'OUTRAS OPÇÕES DE BUSCA' with links for 'BUSCA POR AUTOR', 'BUSCA POR TÍTULO', 'BUSCA POR CURSO', and 'BUSCA POR CONFERÊNCIA'. A message above the second search option reads: 'Você pode selecionar uma palavra-chave em nossa lista, para isso basta selecionar uma letra do alfabeto abaixo' followed by the alphabet 'A B C D E F G H I J K L M N O P Q R S T U V X Z'.

**Figura 16 - Busca por palavra-chave.**

Durante a customização da biblioteca de Engenharia de Software, foram desenvolvidas três novas buscas: por curso, por conferência e por projeto. Estas buscas oferecem um acesso mais rápido à lista de documentos dos cursos da linha de pesquisa, à produção científica relacionada aos projetos do grupo e às referências das principais conferências da área. Um estudante que esteja procurando documentos sobre o projeto

de pesquisa Conhecer, por exemplo, não precisa fazer a busca por palavra-chave e depois filtrar o resultado manualmente. Para isso, basta que o estudante faça uma pesquisa por projeto e selecione a opção Conhecer. As três buscas foram implementadas utilizando o tipo de busca por seleção, onde é apresentada uma lista de opções para o usuário escolher e submeter a consulta. O acesso para essas três buscas está disponível a partir da tela inicial de busca, ilustrado na parte inferior da **Figura 16**.

O resultado de uma busca submetida na biblioteca apresenta o total de documentos encontrados, a lista dos documentos, o número da página de resultado atual e opções de navegação entre as demais páginas resultantes. Cada página de resultado apresenta dez documentos. A **Figura 17** exibe o resultado de uma busca. A figura mostra o título e os autores para cada documento retornado. Na parte inferior da figura, é ilustrada a navegação entre as páginas do resultado. O título do documento é um acesso para a página de referência bibliográfica. Para qualquer tipo de busca realizada, a interface do resultado é a mesma. Este modelo de interface é utilizado em sistema de busca na Web como Google e Yahoo (GOOGLE, 2003, YAHOO, 2003).

▼ Acervo Bibliográfico Perfil do Usuário Quem Somos? Login Cadastro de usuário

Engenharia de Software Biblioteca Digital e Virtual

◀ Volta

RESULTADO DA BUSCA : 51 documentos encontrados. **Página de 6**

▶Título: [A Case-Based Reasoning Approach for managing qualitative experience](#)  
 Autor(es): TAUTZ,C, Althoff,Klaus-Dieter, Nick,Markus.  
 Nota média:3. **Total de documentos e número da página atual.**

▶Título: [A Concept to Support Process Model Maintenance through Systematic Experi](#)  
 Autor(es): BECKER-KORNSTAEDT,U, REINERT,R.  
 Nota média:4.

▶Título: [A Hybrid Approach for Corporate memory management Systems in Software R&D Organizations](#)  
 Autor(es): VON WANGENHEIM,C., LICHTNOW,D., VON WANGENHEIM,A., **Acesso a referência bibliográfica.**

▶Título: [A Learning Component for Workflow Management Systems](#)  
 Autor(es): BERGER,M., ELLMER,E., MERKL,D.,

▶Título: [A Look at Nasa Goddard Space Flight Center's Knowledge Management Initiatives](#)  
 Autor(es): Liebowitz,J.,  
 Nota média:4.

Página de resultados: [1 2 3 4 5 6] **Navegação entre as páginas do resultado.**

**Figura 17 - Resultado da busca.**

### 4.3.2 Gerenciamento da Coleção

A biblioteca foi desenvolvida baseada no controle manual de gerenciamento do acervo e da classificação. A biblioteca disponibiliza um formulário de cadastro para registrar e classificar os documentos. O formulário é dividido em três partes: referência bibliográfica, classificação por palavra-chave e associação com cursos e projetos. A **Figura 18** ilustra o formulário de cadastro.

**Engenharia de Software** Biblioteca Digital e Virtual

**MÓDULO DE CADASTRO - REFERÊNCIA BIBLIOGRÁFICA**  
Os campos marcados com \* são de preenchimento obrigatório.

Titulo \*

Ano de publicação:  Local de publicação:

Idioma: \*  Tipo: \*   PDF

Editor:   Novo

Complemento:

**Referência bibliográfica.**

Autor \* (Novo)  
ROCHA, CATARINA  
ROCHA, ANA REGINA DA  
RODRIGUES, MARIA ANDRÉIA  
ROGERS, P. A.  
ROMBACH, H. D.

Autores do documento  
JACOBSON, IVAR  
BOOCH JAMES RUMBAUGH, GRADY  
RUMBAUGH, JAMES

**Classificação por palavra-chave.**

**CLASSIFICAÇÃO POR PALAVRA-CHAVE**

Palavras-chave (Nova)  
ONTOLOGIA  
PLANEJAMENTO  
POSTMORTEN  
PROCESSO  
QUALIDADE DE SOFTWARE

Palavras-chave do documento  
UML  
ORIENTAÇÃO A OBJETOS

**ASSOCIAR DOCUMENTO COM CURSO E/OU PROJETOS**

Curso:   Novo

Projeto (Novo)  
CardioEducar  
Conhecer  
TABA

Projetos associados ao documento

**Associação a um curso e a projetos.**

**Figura 18 - Cadastro de documento.**

O cadastro de novos documentos, assim como o registro de novos usuários, são realizados no módulo de cadastro. Ao inserir um documento, o usuário pode adicionar



novas palavras-chaves. Dessa forma, a lista de classificação da biblioteca é atualizada continuamente.

Os tipos de documentos que fazem parte da coleção da biblioteca de Engenharia de Software são: artigos de congresso, artigos de periódico, teses, relatórios técnicos, livros e normas técnicas. A catalogação dos documentos seguiu o padrão de referência bibliográfica definido pela norma NBR 6023 da ABNT. O módulo de cadastro pode ser customizado para permitir a inserção de outros tipos de documento. Para isso, é preciso identificar a existência de campos de catalogação específicos do novo tipo de publicação.

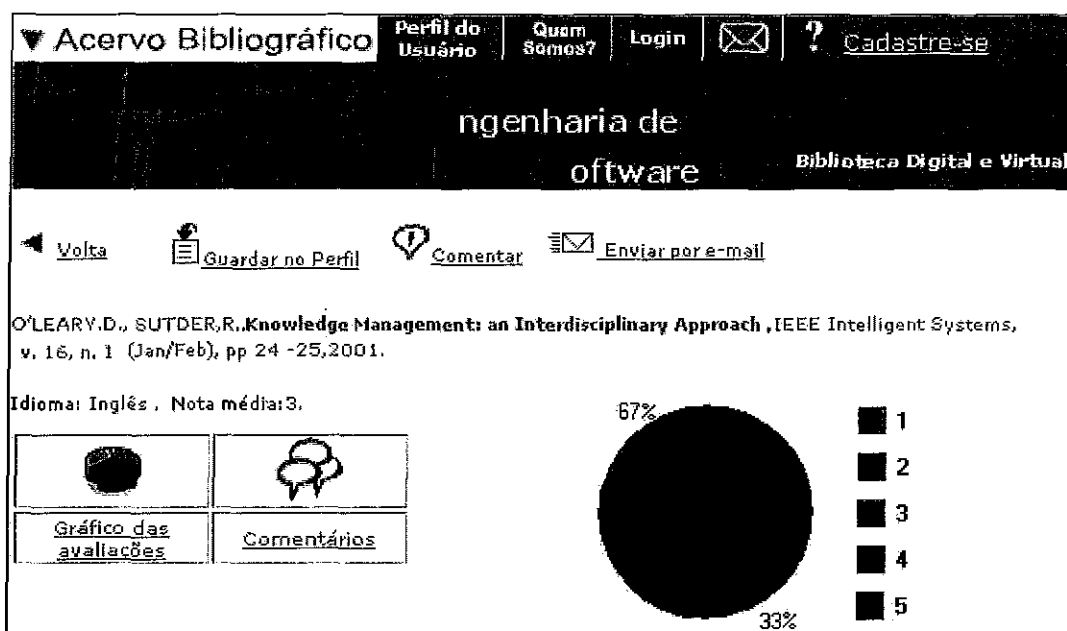
Durante a customização, o formulário de cadastro do documento foi alterado para possibilitar a associação do documento a um curso e a um projeto, ilustrado na parte inferior da **Figura 18**.

A biblioteca disponibiliza um formulário de gerenciamento da classificação onde as palavras-chaves associadas aos documentos podem ser revisadas. Somente o administrador da biblioteca tem acesso a essa funcionalidade.

#### **4.3.3 Serviços Personalizados**

A biblioteca oferece um ambiente personalizado onde os usuários armazenam as referências dos documentos lidos, compartilham opiniões sobre os mesmos e recebem recomendações personalizadas (ROCHA, 2002b).

Após efetuar login na biblioteca, as funcionalidades do ambiente são disponibilizadas ao usuário. A interface de consulta dos documentos apresenta opções personalizadas, não disponíveis no ambiente aberto. O usuário consulta o gráfico das avaliações e os comentários associados aos documentos. O usuário também avalia e comenta os documentos. Outra opção disponível é o envio de referências por correio eletrônico. A **Figura 19** mostra a referência bibliográfica de um documento e serviços personalizados do ambiente.



**Figura 19 - Referência bibliográfica e serviços personalizados.**

A avaliação dos documentos possibilita a troca de opiniões entre os usuários. Um exemplo disso é a utilização da nota média das avaliações como critério de seleção de documentos a serem lidos. Como a nota média é exibida na página de resultado da busca, ilustrado na **Figura 17**, o usuário pode priorizar a leitura dos documentos previamente bem avaliados, ou seja, os que obtiveram maior nota média. Outra vantagem, é oferecer ao professor de um curso a avaliação do material utilizado. O professor pode consultar documentos através da busca por curso e selecionar aqueles que os usuários mais gostaram. Estes documentos podem ser utilizados no próximo curso, ou podem ser indicados a alunos que façam pesquisa no assunto relacionado.

O principal serviço personalizado oferecido na biblioteca é o sistema de recomendação de documentos RecDoc (ROCHA, 2002a), que sugere documentos de interesse ao usuário baseado nas informações armazenadas no perfil do usuário. O sistema será detalhado na seção 4.4.

#### 4.3.4 Perfil do Usuário Definido na Biblioteca

Utilizou-se o modelo de perfil de usuário descrito no capítulo 3 para modelar e definir as informações dos usuários a serem representadas no perfil. O modelo é composto de 5 categorias (AMATO & STRACCIA, 1999): informações pessoais; interesse do usuário; modo de entrega das informações; ações do usuário e política de

segurança. Para cada categoria, foram associadas informações do usuário ou características da biblioteca. O esquema resultante é exibido na **Tabela 10**.

**Tabela 10 – Definição do perfil do usuário.**

<b>Categoria</b>	<b>Sub-categoria</b>	<b>Definições no perfil</b>
<b>Informações pessoais</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• nome, correio eletrônico, senha e ocupação (estudante de mestrado, doutorado, professor, pesquisador).</li> </ul>
<b>Interesse do usuário</b>	Conteúdo do documento.	<ul style="list-style-type: none"> <li>• uma lista de palavras-chaves que representa a área de interesse do usuário;</li> <li>• filtros personalizados.</li> </ul>
	Estrutura do documento.	<ul style="list-style-type: none"> <li>• os documentos digitais disponibilizados estão no formato PDF.</li> </ul>
	Fontes de busca do documento.	<ul style="list-style-type: none"> <li>• o usuário busca na base da biblioteca por palavra-chave, autor e título.</li> </ul>
<b>Entrega de informação</b>	Modo de entrega.	<ul style="list-style-type: none"> <li>• o usuário recebe um e-mail informando a existência de recomendações;</li> <li>• um aviso também é emitido no ambiente personalizado.</li> </ul>
	Tempo da entrega.	<ul style="list-style-type: none"> <li>• a recomendação por conteúdo pode ser requisitada no perfil do usuário.</li> <li>• a recomendação por colaboração é executada com uma periodicidade quinzenal.</li> </ul>
<b>Ações do usuário</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• armazenar documentos no perfil;</li> <li>• avaliar os documentos em notas de 1 a 5;</li> <li>• avaliar recomendações com uma função gostou/não gostou;</li> <li>• todas as avaliações são explícitas.</li> </ul>
<b>Política de segurança</b>	Não se aplica.	<ul style="list-style-type: none"> <li>• somente o usuário tem acesso ao seu próprio perfil;</li> <li>• as avaliações e comentários são exibidos na referência do documento, mas como usuário anônimo.</li> </ul>

Inicialmente, o perfil é criado quando o usuário submete o formulário de cadastro, disponível no menu principal da biblioteca. Neste formulário, o usuário preenche os dados pessoais, dados acadêmicos e seleciona um conjunto de palavras-chaves que representam a sua área de interesse. O perfil é continuamente atualizado à medida que o usuário interage com o sistema de busca da biblioteca, armazena referências dos documentos no perfil, submete avaliações e comentários sobre os mesmos. A **Figura 20**

mostra a tela de consulta do perfil do usuário. A figura mostra os títulos dos documentos armazenados e as notas associadas a cada documento. O título é um acesso a referência bibliográfica. O usuário avalia os documentos selecionando a opção Avaliar, ao lado do título. Para alterar a avaliação, o usuário seleciona a nota associada ao documento. Estas duas opções são acessos ao formulário de avaliação do documento.

Perfil do Usuário    Acervo Bibliográfico    Quem Somos?    Login    ?    Cadastro de usuário

Engenharia de Software    Biblioteca Digital e Virtual

[Voltar](#)   [Verificar Recomendações](#)   [Alterar dados pessoais](#)

PERFIL DE catirocha

Lista de documentos:

TÍTULO	NOTA
<a href="#">The Unified Software Development Process</a>	4
<a href="#">Knowledge Management: an Interdisciplinary Approach</a>	3
<a href="#">Knowledge Management Systems: Emerging Views and Practices from the Field</a>	4
<a href="#">Organizational Learning in Dynamic Domains</a>	4
<a href="#">Supporting Knowledge Management in University Software R&amp;D Groups</a>	<a href="#">Avaliar</a>

Figura 20 - Perfil do usuário.

O formulário de avaliação de um documento é exibido na Figura 21. Esta figura mostra opções com as notas de 1 a 5 e um campo de digitação de múltiplas linhas para preenchimento do comentário. Somente a nota é de preenchimento obrigatório.

Perfil do Usuário    Acervo Bibliográfico    Quem Somos?    Login    ?    Cadastro de usuário

Engenharia de Software    Biblioteca Digital e Virtual

COMENTAR DOCUMENTO

Título: [The Unified Software Development Process](#)

Nota:  1  2  3  4  5 - escala de ruim (1) a excelente (5)

Comentário:

Figura 21 - Formulário de avaliação do documento.

No perfil, o usuário também define filtros que são aplicados nas recomendações de documentos. O filtro contém, inicialmente, as palavras-chaves definidas como área de interesse. O usuário pode selecionar outros critérios para o filtro. Os critérios são divididos em três grupos: dados da referência bibliográfica, área de interesse e participação em projetos e cursos. A **Figura 22** ilustra o formulário de definição do filtro personalizado. Com relação a referência bibliográfica, o usuário escolhe que tipo de documento está interessado (artigo de periódico, tese, livro, etc.) e a partir de que ano de publicação os documentos devem ser selecionados. Na área de interesse, são utilizadas as palavras-chaves definidas no cadastro, mas estas podem ser editadas no filtro.

Engenharia de Software
Biblioteca Digital e Virtual

---

**FILTRO PERSONALIZADO**

REFERÊNCIA BIBLIOGRÁFICA

Documentos publicados a partir de:

Tipo:

---

ÁREA DE INTERESSE

Palavras-chave (Nova)

ADS

ADS CENTRADO EM PROCESSO

ADS ORIENTADO A DOMÍNIO

ADS ORIENTADO A ORGANIZAÇÃO

ANÁLISE DE PONTOS DE FUNÇÃO

ANÁLISE DE REQUISITOS

APRENDIZADO ORGANIZACIONAL

AQUISIÇÃO DO CONHECIMENTO

Palavras-chave selecionadas

BASE DE CONHECIMENTO

GESTÃO DO CONHECIMENTO

FÁBRICA DE EXPERIÊNCIAS

---

CURSO E/OU PROJETOS

Curso:

Projeto

CardioEducar

Conhecer

TABA

Projetos selecionados

Conhecer

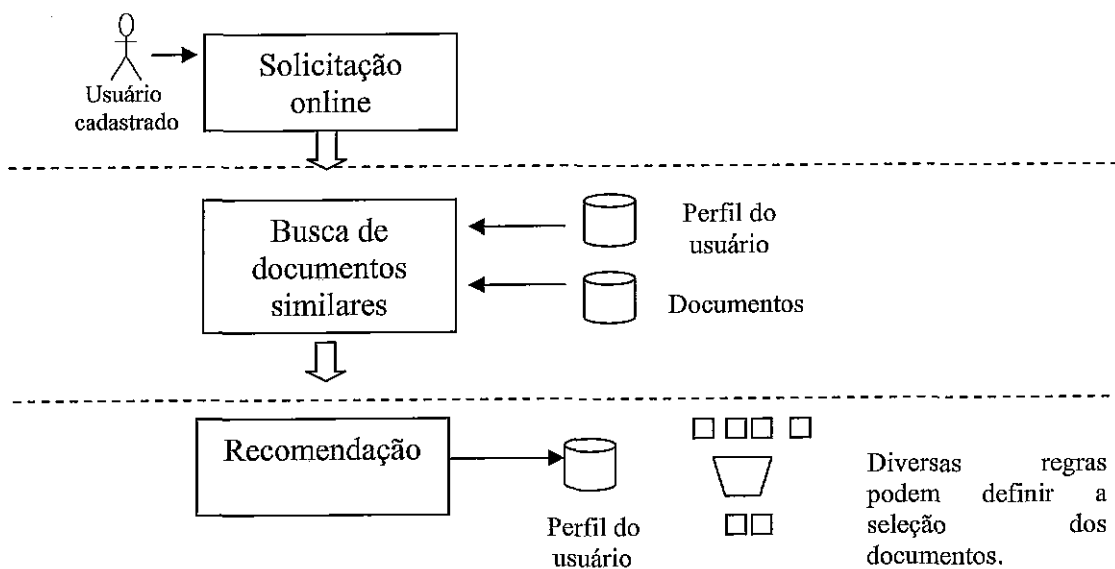
**Figura 22 – Definição do filtro personalizado.**

## 4.4 RecDoc: sistema de recomendação

RecDoc é um sistema de recomendação híbrido, de acordo com a classificação proposta por BALABANOVIC (1997) na seção especial sobre sistemas de recomendação da Communications of ACM. RecDoc é considerado híbrido pois implementa a abordagem baseada em conteúdo e a abordagem baseada em colaboração. As abordagens são discutidas a seguir.

### 4.4.1 Recomendação por Conteúdo

A abordagem recomendação por conteúdo é realizada em 3 etapas: solicitação de recomendação online, busca de documentos similares e recomendação. A **Figura 23** ilustra a seqüência das etapas e recursos utilizados em cada uma delas.



**Figura 23 - Esquema de recomendação por conteúdo.**

A **solicitação da recomendação por conteúdo** pode ser feita através de um link disponível no perfil do usuário. Para isso, basta que o usuário esteja cadastrado na biblioteca e que tenha definido uma lista de palavras-chaves no momento do cadastro.

Na **etapa de busca de documentos similares**, o sistema procura documentos similares ao que o usuário armazenou no perfil. A comparação entre documentos é feita com base na classificação por palavra-chave.

Na **etapa de recomendação**, são aplicados filtros ou regras sob a lista de documentos encontrados na etapa anterior. Os critérios para seleção dos documentos podem ser: recomendar documentos mais novos ou documentos bem avaliados (que obtiveram nota média acima de 3.5, por exemplo). Além disso, podem ser aplicados os filtros personalizados definidos no perfil.

O pseudo-código utilizado no esquema de recomendação por conteúdo está exemplificado na **Figura 24**. Este foi implementado em ASP (VbScript, JavaScript) e em *stored procedures* no SQL Server. Esta tecnologia foi utilizada no desenvolvimento da biblioteca.

```
Selecione o usuário logado no sistema
Selecione palavras-chaves dos documentos avaliados no perfil
acessando a base de perfis de usuário
Busque documentos classificados com as palavras-chaves, na base de
documentos
Aplique os filtros definidos o perfil.
Selecione, na lista de documentos retornados, a quantidade
correspondente ao limite de recomendação
Grave os documentos no perfil do usuário como recomendação por
conteúdo.
Exiba a página de recomendação por conteúdo com formulário de
avaliação da recomendação.
```

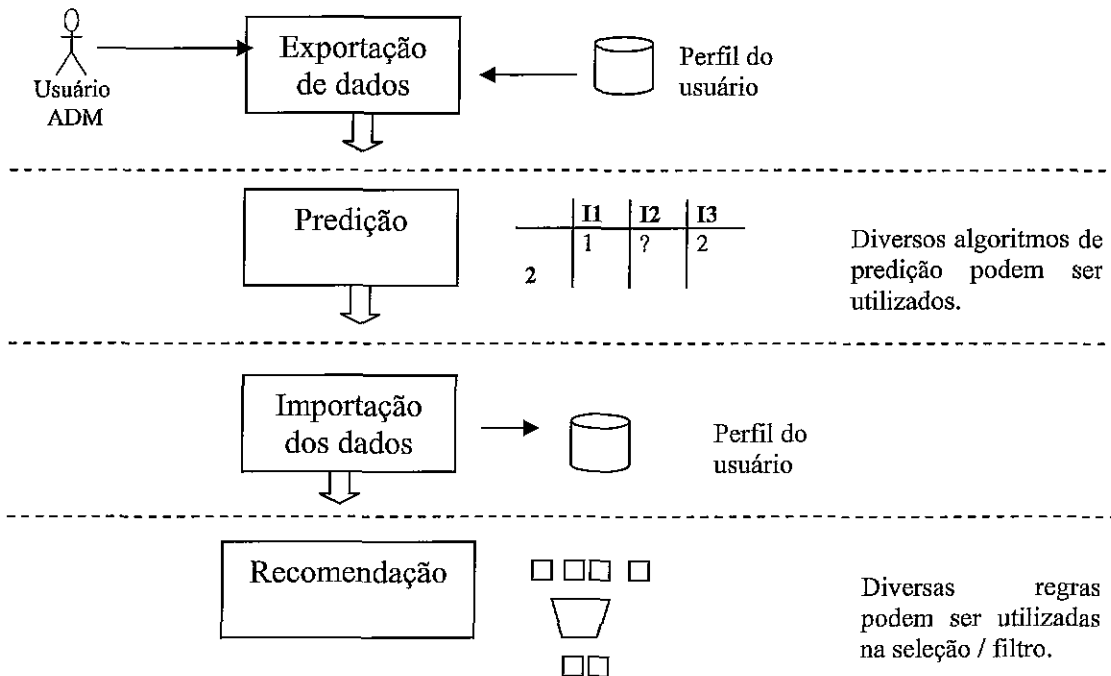
**Figura 24 - Pseudo-código da recomendação por conteúdo.**

Uma abordagem similar é utilizada na Biblioteca Digital da ACM, onde o usuário define consultas booleanas que são aplicadas periodicamente à base (ACM D.L., 2003). Uma lista de novos documentos é enviada por e-mail ao usuário. A principal diferença dessa abordagem para RecDoc, é que o usuário não precisa definir uma consulta para obter a indicação de documentos novos. Essa etapa é transparente ao usuário.

#### **4.4.2 Recomendação por Colaboração**

A abordagem recomendação por colaboração é executada periodicamente de forma off-line pelo usuário administrador do sistema. A recomendação é realizada em 4 etapas: exportação dos dados, predição das notas, importação dos dados e recomendação de documentos. Estas etapas foram definidas de acordo com processo de

filtragem colaborativa proposto em (SARWAR et al., 2001), descrito no capítulo dois, seção 2.4.2 - Processo de Filtragem Colaborativa. A **Figura 25** ilustra a seqüência e interação entre as etapas desse esquema de recomendação.



**Figura 25 - Esquema de recomendação por colaboração.**

Na **etapa de exportação**, o administrador da biblioteca (usuário ADM) exporta as avaliações dos documentos armazenadas no perfil do usuário para arquivos texto. O administrador utiliza ferramentas do SGBD (Sistema Gerenciador de Banco de Dados) para realizar transferências de dados entre a base e os arquivos. O arquivo texto contém uma matriz “usuário x documentos”.

A **etapa de predição** tem como entrada os arquivos gerados na exportação. Algoritmos de predição podem ser executados para calcular as notas dos documentos que o usuário não avaliou.

Na **etapa de importação**, o usuário ADM importa a matriz de predição para a base de dados.

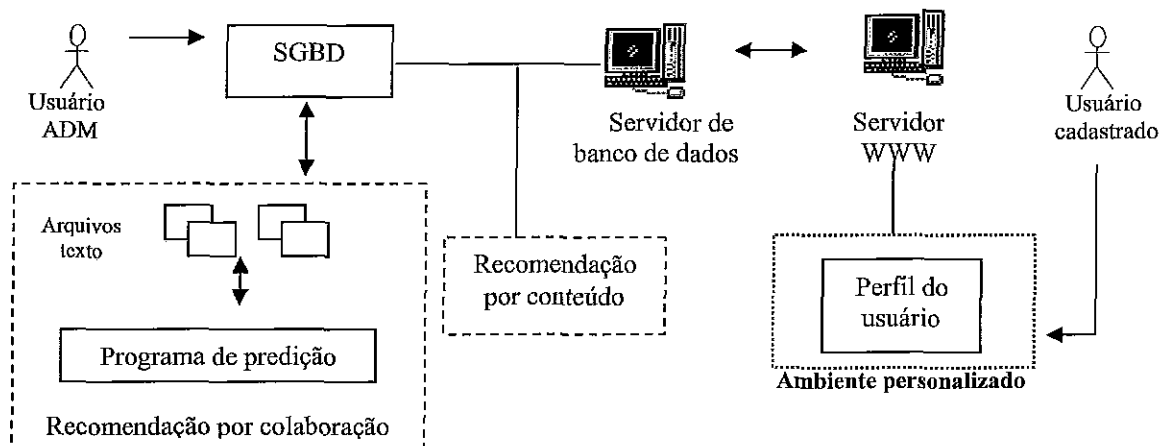
Na **etapa de recomendação**, são selecionados os documentos a serem indicados aos usuários. Um passo inicial é o corte a partir de um limite pré-definido. Os documentos que obtiveram notas superiores a esse limite são selecionados. Outras regras podem ser utilizadas para selecionar os documentos. Os filtros personalizados ou



a área de interesse podem ser aplicados. Esta etapa atualiza os perfis dos usuários com novos documentos a serem indicados.

#### 4.4.3 Arquitetura

A arquitetura do sistema de recomendação RecDoc representa características e recursos definidos nos dois esquemas de recomendação descritos nas seções anteriores. A arquitetura do sistema está ilustrada na **Figura 26**. O usuário cadastrado acessa as recomendações no ambiente personalizado da biblioteca. Este ambiente é executado no servidor Web. O usuário ADM utiliza ferramentas do SGBD para gerenciar a base. A abordagem por conteúdo manipula os dados diretamente na base de dados, enquanto que a abordagem por colaboração manipula arquivos texto exportados da base.



**Figura 26 – Arquitetura do sistema de recomendação RecDoc.**

A arquitetura definida possibilita que o programa de predição não seja restrito a biblioteca de Engenharia de Software. O programa pode ser utilizado em outros sistemas de recomendação, sendo necessário apenas que os arquivos contendo as avaliações estejam em um formato pré-definido.

Outro fator a considerar sobre a arquitetura, é o fato do programa de predição poder ser desenvolvido em qualquer linguagem de programação que manipule arquivos texto, não sendo restrito a linguagens que possuam diretivas de acesso ao SQL Server. Devido a esta característica, algoritmos de filtragem colaborativa desenvolvidos por terceiros podem ser integrados ao RecDoc.

#### 4.4.4 Algoritmo de filtragem colaborativa

No esquema de recomendação por colaboração, podem ser utilizados diversos algoritmos de predição. O único requisito necessário é que este manipule o arquivo contendo a matriz “usuário x itens” no formato no qual este é gerado.

Neste trabalho, optamos por um algoritmo **baseado em memória**, segundo classificação de (BREESE et al., 1998), para medir a correlação entre os usuários na etapa de predição. Segundo os autores, a escolha do algoritmo depende da natureza dos dados e da aplicação, tamanho da base de dados e do tempo de processamento. Considerando essas questões, o algoritmo baseado em memória se apresentou o mais apropriado para a biblioteca. Os algoritmos baseados em modelo apresentam um tempo de processamento menor para matrizes muito esparsas com milhares de documentos, mas esta questão não é considerada relevante para a biblioteca. Diversos estudos comparativos entre algoritmos de predição, tanto no grupo de algoritmos baseados em memória, quanto algoritmos baseados em modelo, apresentaram resultados similares, sendo o primeiro grupo o mais utilizado. Este foi mais um aspecto considerado.

Dentre os algoritmos de predição baseado em memória, foi escolhido o algoritmo utilizado no GroupLens (RESNICK et al., 1994) que usa o coeficiente Pearson para medir a correlação entre usuários.

A nota calculada na etapa de predição segue a fórmula definida na **Equação 1**.

$$p_{a,j} = \bar{n}_a + \frac{\sum_i w(a, i) \cdot (n_{i,j} - \bar{n}_i)}{\sum_i |w(a, i)|}$$

**Equação 1 – Fórmula de predição (BREESE et al., 1998).**

Considere:

$p_{a,j}$  é a nota predição do usuário a para o item j;

j corresponde ao item no qual se deseja calcular a avaliação;

$w(a, i)$  é a correlação entre o usuário a e o usuário i.

$\bar{n}_i$  é a média aritmética das avaliações do usuário i;

$n_{i,j}$  corresponde a nota do item j dada pelo usuário i;

A média aritmética das notas para um usuário  $i$  é calculada como na **Equação 2**. Considere  $n_i$  a média a ser calculada,  $I_i$  o total de documentos avaliados pelo usuário  $i$  e  $n_{i,j}$  a nota de cada item  $j$  avaliado pelo usuário.

$$\bar{n}_i = (1 / |I_i|) \cdot \sum_{j=1}^N (n_{i,j})$$

**Equação 2 - Média aritméticas das avaliações para o usuário  $i$ .**

Para o método de Pearson, o valor de  $w(a, i)$  é calculado com base na correlação dos usuários. Este método foi utilizado nos sistemas Ringo (SHARDANAD & MAES, 1995) e GroupLens (RESNICK et al., 1994). Em (SWAMI, 2000) são apresentados estudos experimentais de avaliação sobre o desempenho deste método. Outro estudo comparou o desempenho deste, com a média aritmética das avaliações (MILLER et al., 1997). Estes trabalhos foram discutidos no capítulo dois.

A correlação entre o usuário  $a$  e o usuário  $i$  é calculado como na **Equação 3**. Considere  $j$  os itens avaliados por ambos usuários.

$$w(a,i) = \frac{\sum_j (n_{a,j} - \bar{n}_a) \cdot (n_{i,j} - \bar{n}_i)}{\sqrt{\sum_j (n_{a,j} - \bar{n}_a)^2} \cdot \sqrt{\sum_j (n_{i,j} - \bar{n}_i)^2}}$$

**Equação 3 - Coeficiente de correlação Pearson.**

O coeficiente  $w(a,i)$  obtém valores no intervalo  $[-1, +1]$ . Quanto mais próximo de  $-1$ , menos relacionados são os usuários. Em contrapartida, quanto mais próximo de  $+1$ , mais similares eles são.

O pseudo-código do esquema de recomendação está ilustrado na **Figura 27**.

```

Ler arquivo com a matriz "usuário x itens"
Para cada usuário
    Definir o usuário como ativo ( $u_a$ )
    Calcular média aritmética das avaliações de  $u_a$ 
    Salvar em vetor "nota média"
    Verificar itens sem avaliar para  $u_a$ , correspondente a colunas em
branco na matriz
    Para cada item sem nota
        Selecionar usuários que avaliaram o item e colocar em uma lista
        "usuários a analisar"
        Se nenhum usuário avaliou o item
            Pegar o próximo item sem nota do usuário ativo
        Fim Se
        Para cada usuário da lista "usuários a analisar", digamos  $u_i$ 
            Verificar se já existe uma entrada no vetor "nota
            média" para o usuário  $u_i$ 
            Se não existe
                Calcular média aritmética das avaliações de  $u_i$ 
                Salvar em vetor "nota média"
            Fim Se
            Verificar se já existe uma entrada para o coeficiente
             $w(a,i)$  ou  $w(i,a)$  na matriz "correlação"
            Se não existe
                Calcular  $w(a,i)$ .
                Salvar em matriz "correlação"
            Fim Se
        Fim para
        Calcular nota  $p_{a,j}$ 
        Salvar nota em matriz "predição"
    Fim Para
Fim Para
Gravar arquivo "etapa_predição" com a matriz "predição"
Gravar arquivo "etapa_correlação" com a matriz "correlação"
Para cada coluna preenchida da matriz "predição"
    Se nota > nota mínima de recomendação
        num_avalições = num_avalições + 1
        Gravar nota na matriz "recomendação"
    Fim Se
Fim Para
Se num_avalições > limite de recomendações
    Aplicar filtro para selecionar documentos na matriz recomendação
Fim Se
Gravar arquivo com a matriz "recomendação"

```

**Figura 27 - Pseudo-código do algoritmo de recomendação.**

O algoritmo tem como entrada o arquivo de avaliações. Outros parâmetros de entrada fornecidos ao algoritmo são:

1. Nota mínima de recomendação: valor utilizado para descartar documentos baseado na nota obtida na predição. Por exemplo, somente devem ser recomendados itens com notas acima de 3.
2. Limite de recomendação: quantos itens a recomendar por vez.

O algoritmo de predição foi desenvolvido na linguagem C. Os filtros foram desenvolvidos como stored-procedures no banco de dados.

A simulação do funcionamento do algoritmo do método de Pearson, é apresentado a seguir. Considere uma matriz de itens avaliados como descrito na **Tabela 11** . A primeira linha da matriz corresponde aos itens I1, I2, .., I6. A primeira coluna refere-se aos usuários U1, U2, U3 e U4. Os itens avaliados pelo usuário são correspondentes às células preenchidas na matriz. Por exemplo, o usuário U1 avaliou o item I2 com nota 5. As células em branco correspondem aos itens cuja nota o algoritmo deve calcular.

**Tabela 11 - Matriz de avaliações – “Usuários x Itens”**

Usuário \ Item	I1	I2	I3	I4	I5	I6
U1	1	5		2	4	
U2	4	2		5	1	2
U3	2	4	3			5
U4	2	4		5	1	

Seguindo alguns passos do algoritmo, calcula-se a predição da nota referente ao item 6 para o usuário U1. O primeiro passo é calcular a média aritmética das avaliações dos usuários U2 e U3 que avaliaram o item I6. A **Tabela 12** apresenta as médias calculadas.

**Tabela 12 - Média aritmética das notas para os usuários**

Usuário	Média aritmética das notas
U1	3
U2	2,8
U3	3,5
U4	3

O segundo passo é calcular o coeficiente  $w(U1, U2)$  equivalente à correlação entre o usuário U1 e U2, e o coeficiente  $w(U1,U3)$  entre usuários U1 e U3. O coeficiente  $w(U1,U4)$  não precisa ser calculado pois o usuário U4 não avaliou o item I6. Os coeficientes resultantes são ilustrados na **Figura 28**.

$$w(U1, U2) = -2 -2 -2 -2 / (\sqrt{10} \cdot \sqrt{10}) = -0,8$$

$$w(U1, U3) = 1$$

**Figura 28 – Coeficientes de correlação entre usuários U1 e U2, U1 e U3.**

Do coeficiente de correlação, podemos observar que o usuário U1 tende a discordar do usuário U2, visto que o coeficiente obteve valor próximo de  $-1$ . Verificando na **Tabela 11**, os usuários costumavam realmente discordar das avaliações. Já o usuário U3 costuma concordar com U1 nas avaliações, obtendo coeficiente de correlação 1.

Com estes valores calculados, é possível prever a nota  $p_{U1, I6}$  para o item I6 com relação ao usuário U1, como mostra a **Figura 29**.

$$p_{U1, I6} = 3 + ((2 - (-0,8)) / (|1| + |-0,8|)) = 4,56$$

**Figura 29 – Predição da avaliação do usuário U1 para o item I6.**

Se for definido que a nota mínima de recomendação for 4, o item I6 será inserido na lista de itens a recomendar. Fazendo os mesmos cálculos para o usuário U4, em relação também ao item I6, é obtida uma predição 3,75. Se for usada a mesma nota limite, este item não será recomendado.

#### 4.4.5 Justificativa e Avaliação da Recomendação

A recomendação é apresentada ao usuário como uma lista de documentos no perfil do usuário. Não é feita a ordenação dos mesmos. A **Figura 30** ilustra a recomendação de documentos. O título do documento é um link para a referência bibliográfica. O usuário é sugerido a avaliar a recomendação em uma função gostou / não gostou. Na figura, a avaliação corresponde a ícones representando que o usuário está satisfeito ou insatisfeito com a recomendação. A avaliação é utilizada para medir o desempenho do sistema de recomendação e pode ser usada para melhorar o desempenho das próximas indicações.

Engenharia de Software		Biblioteca Digital e Virtual
<b>Perfil do usuário - Recomendação</b>		
Lista de documentos recomendados para o usuário <b>montoni</b>		
Título	Avaliação	
Representação da Distribuição de Conhecimento e Habilidades através da Estrutura Organizacional	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	
Knowledge Processes and Ontologies	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	
Better Knowledge Management through Knowledge Engineering	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	
Planejamento de Riscos em Ambientes de Desenvolvimento de Software Orientados à Organização	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	
Defining Knowledge: an Epistemological Foundation for Knowledge Management	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	
Knowledge-Management Systems: Converting and Connecting	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/>	

**Figura 30 – Recomendação de documentos no perfil do usuário.**

Para cada documento recomendado, o sistema apresenta a justificativa da recomendação, que corresponde a uma explicação do porquê o documento foi indicado. A justificativa se mostrou um fator importante na confiabilidade da recomendação e consequentemente na aceitação da mesma por parte do usuário.


Na recomendação por conteúdo, é exibida a classificação do documento que corresponde a palavras-chaves da área de interesse do usuário.

Em (HERLOCKER, 2001), foram apresentados diversos tipos de justificativa da recomendação por colaboração. Foi escolhida a alternativa de mostrar o desempenho do sistema nas recomendações anteriores e os usuários utilizados na etapa de predição.

Para apresentar os usuários utilizados em cada predição, foi necessário alterar o algoritmo de predição utilizado. O programa de predição gera como saída mais uma matriz contendo a predição, o número de usuários utilizados no cálculo e quais foram os usuários. Com esta alteração, foi possível montar um gráfico com a justificativa das recomendações, como ilustrado na **Figura 31**.

**Engenharia de software**

Biblioteca Digital e Virtual

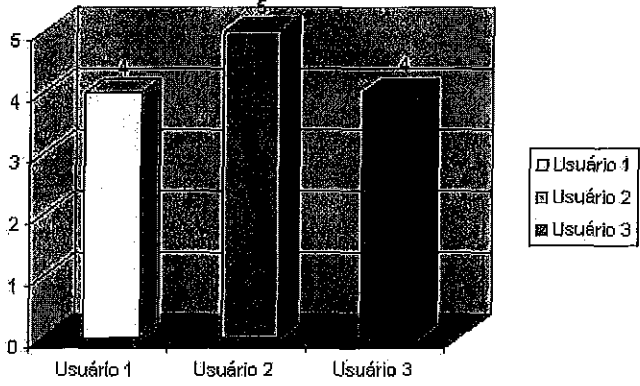
**JUSTIFICATIVA DA RECOMENDAÇÃO** 

▣ **Título:** Knowledge – Management Systems: Converting and Connecting

▣ **Percentual de acerto do sistema de recomendação:** 82%

▣ **Percentual de recomendações satisfatórias para o usuário: montoni:** 78%

▣ **Justificativa:** Este documento foi recomendado devido a avaliação de usuários semelhantes a você.  
Veja estas avaliações no gráfico abaixo.



Usuário	Avaliação
Usuário 1	4
Usuário 2	5
Usuário 3	4

**Figura 31 - Justificativa da recomendação colaborativa.**

## 4.5 Base de dados da biblioteca

Conceitualmente, a base de dados da biblioteca foi representada na arquitetura e nos esquemas de recomendação como sendo composta por uma base de documentos e uma base de perfis de usuários. Esta representação foi utilizada para facilitar o entendimento da arquitetura e do sistema de recomendação. Entretanto, fisicamente, essas bases foram unificadas para facilitar a implementação no SQL Server.

O modelo de dados lógico da biblioteca foi definido usando a ferramenta CASE ErWin versão 3.52. O modelo de dados está exibido na **Figura 32**.

A tabela **AcervoBibliografico** representa a referência bibliográfica de um documento, com os atributos comuns a qualquer publicação. As tabelas **Livro**, **RelatorioTecnico** e **ArtigoPeriodico** são especializações de **AcervoBibliografico**, representando atributos específicos para essas publicações, como o atributo ISBN da tabela **Livro**. O atributo `dt_cadastro_documento` foi adicionado a tabela



**AcervoBibliografico** com o objetivo de registrar a data de cadastro do documento na biblioteca. Este atributo é utilizado para reconhecer documentos novos a serem recomendados na abordagem por conteúdo, bem como pode ser usado como critério de seleção na abordagem colaborativa.

Quando o documento tem um arquivo digital associado, essa relação é armazenada na tabela **Localizacao**. A coluna `localizacao_arquivo` desta tabela contém um caminho no servidor Web onde o arquivo pode ser encontrado.

A tabela **Autor** contém as informações de autoria da publicação, como nome e sobrenome do autor. A tabela **AutorAcervo** relaciona os autores a um determinado documento cadastrado na biblioteca. A coluna `autor_ordem` identifica a ordem na qual os autores aparecem na publicação.

A tabela **PalavraChave** armazena a lista de palavras-chaves utilizadas para classificar os documentos. A classificação de cada documento fica relacionada na tabela **PalavraChaveDocumento**.

As tabelas **Projeto**, **Curso** e **Conferencia** representam, respectivamente, projetos de pesquisa, cursos ministrados na linha de pesquisa de ES e conferências da área. Essas tabelas foram adicionadas ao modelo no momento de configuração e customização da biblioteca de Engenharia de Software.

A tabela **Usuario** armazena os dados pessoais do usuário. A tabela **UsuarioAcervo** relaciona documentos a um determinado usuário. Estão representados nesta tabela documentos lidos, avaliações e comentários feitos pelo usuário.

A tabela **UsuarioPalavraChave** relaciona o usuário a uma lista de palavras-chave. Estes dados representam a área de interesse do usuário e são utilizados como filtros.

A tabela **UsuarioRecomendacao** guarda a lista de documentos recomendados. É representado, para cada usuário, o documento recomendado, a data e a avaliação da recomendação. Dessa forma, é possível manter um histórico das recomendações. Esses dados podem ser utilizados futuramente para eliminar documentos já recomendados, bem como eliminar itens similares a documentos mal recomendados.

A tabela **PredicaoUsuario** armazena todas as notas dos documentos geradas na etapa de predição.

A base atual contém 325 documentos, 15 usuários e 350 avaliações.

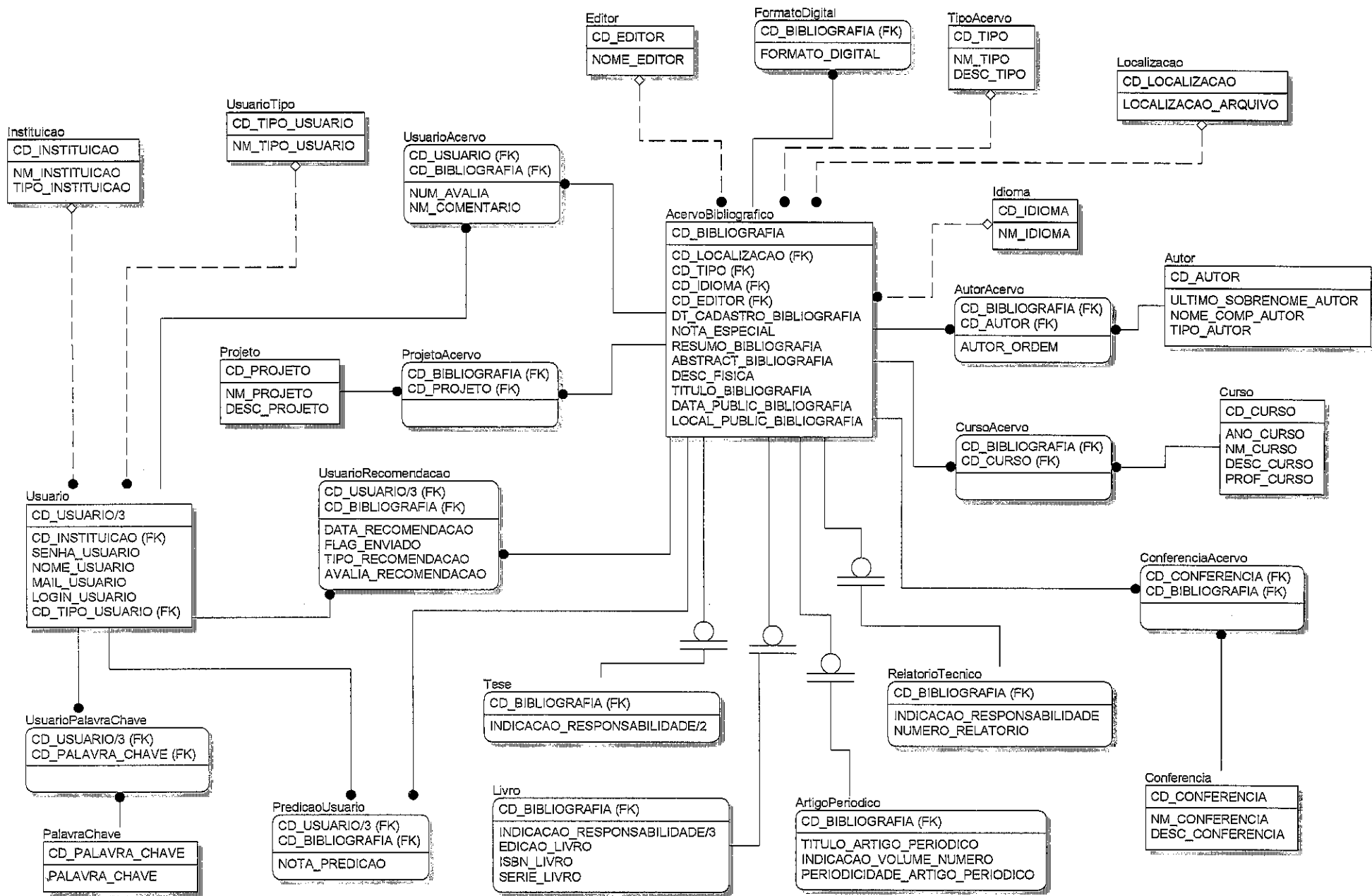


Figura 32 - Modelo de dados da biblioteca.

## 4.6 Comparativo com Trabalhos Relacionados

### 4.6.1 Comparativo entre a Biblioteca de Engenharia de Software e demais Bibliotecas

A biblioteca digital virtual de ES apresenta algumas das características e serviços oferecidos nas bibliotecas discutidas no capítulo três. A **Tabela 13** mostra um comparativo entre as bibliotecas utilizando os critérios: sistema de busca, serviço de alerta, perfil do usuário, avaliação e comentários dos documentos e sistema de recomendação.

A biblioteca de ES utiliza um sistema de busca por campos de catalogação, assim como a maioria das bibliotecas estudadas, por este tipo de busca ser similar a buscas oferecidas em bibliotecas reais. As bibliotecas da ACM e da IEEE apresentam inicialmente uma busca por texto completo do documento, mas oferecem opções para o usuário especificar a busca com campos da catalogação. No caso da ACM, esta opção está disponível na busca avançada.

No perfil do usuário da biblioteca de ES, o usuário avalia e comenta documentos, que ficam disponíveis para outros usuários consultarem no ambiente personalizado. Na biblioteca Citeseer, o usuário avalia o documento, mas não fica disponível para consulta posterior. O compartilhamento de opiniões é muito utilizado em sites de entretenimento, como o ViaRio, sites de comércio eletrônico como a Amazon.com e em guias de viagem como TripAdvisor.com (VIARIO, 2003, AMAZON, 2002, TRIPADVISOR, 2003). A biblioteca ACM oferece *binders* que podem ser considerados similares a perfis de usuário.

A biblioteca ACM oferece serviço de alerta através de agentes de filtragem para avisar os usuários do cadastro de novos documentos que satisfaçam uma consulta pré-estabelecida. A biblioteca de ES oferece um serviço similar através do sistema de recomendação RecDoc. Este sistema implementa uma combinação das abordagens por conteúdo e por colaboração, o que poucas bibliotecas possuem. A biblioteca QuickStep tem um sistema de recomendação por colaboração através de algoritmo de grafo (MIDDLETON et al., 2002).

**Tabela 13 - Biblioteca de Engenharia de Software x demais bibliotecas.**

Biblioteca Digital	Sistema de busca		Serviço de alerta	Perfil do usuário	Avaliação de documentos	Comentários do documento	Sistema de recomendação	
	Por campos da referência bibliográfica	Por texto completo do documento					Por conteúdo	Por colaboração
ACM	✓	✓	✓	✓				
IEEE	✓	✓						
Citeseer		✓		✓		✓	✓	
Springer	✓	✓	✓					
NCSTRL	✓							
QuickStep		✓						✓
Biblioteca Digital de ES	✓			✓	✓	✓	✓	✓

#### 4.6.2 Comparativo entre RecDoc e demais Sistemas de Recomendação

O sistema de recomendação RecDoc integra algumas propostas encontradas em trabalhos relacionados. A **Tabela 14** mostra um comparativo entre as características dos sistemas estudados e RecDoc. Os critérios de comparação foram: abordagem por conteúdo, por colaboração, justificativa da recomendação, perfil do usuário, avaliação explícitas dos itens, filtros personalizados e utilização de agentes inteligentes.

Em todos os sistemas de recomendação discutidos, as avaliações dos itens são explícitas. A escala de avaliação com notas de 1 a 5 é a mais utilizada, sendo aplicada no RecDoc, no GroupLens, no MovieLens e na Amazon.com (GROUPLENS, 2002, MOVIELENS, 2002, AMAZON, 2002). O sistema Ringo (SHARDANAD & MAES, 1995) utiliza uma escala de 1 a 7. Já os sistemas InfoFinder e NewsWeeder utilizam uma função gostou / não gostou para avaliar páginas Web.

Os sistemas InfoFinder e NewsWeeder utilizam agentes de filtragem na recomendação por conteúdo. RecDoc implementa a abordagem utilizando a similaridade entre documentos através da classificação por palavra-chave. Esta abordagem é utilizada também no sistema Fab (BALABANOVIC & SHOHAM, 1997).

RecDoc implementa a abordagem por colaboração utilizando um algoritmo baseado em memória, assim como o GroupLens e o MovieLens (GROUPLENS, 2002, MOVIELENS, 2002). Durante o processo de recomendação, é utilizado a classificação por palavra-chave na seleção dos documentos.

RecDoc apresenta a justificativa da recomendação realizada tanto na abordagem por conteúdo, quanto na colaborativa. A maioria dos sistemas estudados só apresentam a nota média das avaliações.

Na abordagem por conteúdo, o sistema RecDoc apresenta a justificativa da recomendação baseada na classificação por palavra-chave. A Amazon.com mostra o documento que o usuário avaliou, utilizado como base para encontrar os documentos similares recomendados (AMAZON, 2002).

Na abordagem colaborativa, o RecDoc apresenta como justificativa um gráfico com as avaliações dos vizinhos utilizados na predição. Este tipo de justificativa foi o mais votado no estudo realizado no sistema MovieLens (HERLOCKER, 2001). Apesar disso, MovieLens não disponibiliza esta justificativa no sistema on-line.

**Tabela 14 - RecDoc x sistemas de recomendação.**

Sistema de recomendação	Abordagem por conteúdo	Abordagem colaborativa	Justificativa da recomendação	Perfil do usuário	Avaliação explícita	Filtros personalizados	Agentes
Recdoc	✓	✓	✓	✓	✓	✓	
InfoFinder	✓						✓
NewsWeeder	✓						✓
Fab	✓	✓		✓			✓
GroupLens		✓		✓	✓		
Movielens		✓		✓	✓		
Amazon.com	✓	✓	✓	✓	✓		
Teamworks		✓		✓		✓	
QuickStep		✓					

## **4.7 Considerações Finais**

Este capítulo apresentou a proposta de utilização do sistema de recomendação RecDoc para uma biblioteca digital virtual na Web.

Inicialmente, foram mostrados as etapas de desenvolvimento de uma biblioteca Web genérica. Depois, foi apresentada a customização da biblioteca para o grupo de pesquisa de Engenharia de Software da COPPE/UFRJ. Utilizando este exemplo, as características e arquitetura da biblioteca foram discutidas.

Em seguida, o capítulo apresentou dois esquemas de recomendação utilizados no sistema de recomendação RecDoc. Estes esquemas são genéricos e podem ser utilizados com uma variedade de algoritmos. Um algoritmo baseado em memória foi escolhido e apresentado.

O capítulo seguinte apresenta estudos experimentais realizados para validar a utilização do sistema de recomendação, o desempenho das abordagens por conteúdo e colaborativa e o algoritmo de predição utilizado.

# Capítulo 5

## Estudos de Caso

---

*Este capítulo apresenta estudos de caso realizados para avaliar o desempenho do sistema de recomendação RecDoc proposto. Foram realizados testes tanto na abordagem por conteúdo, quanto na colaborativa. O capítulo apresenta também um estudo de caso que avalia a precisão do algoritmo de predição utilizado na abordagem colaborativa.*

### 5.1 Considerações Iniciais

Os estudos de caso descritos neste capítulo utilizaram a base de dados da Biblioteca de Engenharia de Software, construída como exemplo nesta dissertação. A biblioteca está em uso desde março de 2002. A base de dados contém atualmente 250 documentos e 310 avaliações. Para cada estudo de caso, foi utilizada uma sub-parte da base de dados.

Foram realizados 3 estudos de caso. O primeiro estudo avaliou a precisão do algoritmo de predição utilizado na abordagem colaborativa. O segundo estudo avaliou o desempenho desta abordagem. O terceiro estudo avaliou o desempenho do sistema de recomendação comparando resultados obtidos nas abordagens por conteúdo e por colaboração. Neste último estudo de caso, também é observado se o desempenho da abordagem colaborativa foi a mesma do estudo anterior, para uma nova amostragem com um número maior de usuários e avaliações.

### 5.2 Estudo de Caso: Precisão do Algoritmo de Predição

O objetivo deste primeiro estudo de caso é medir a precisão do algoritmo de predição proposto na recomendação colaborativa. O pré-requisito para a realização deste estudo é a seleção de um conjunto de documentos da coleção da biblioteca no qual todos eles tenham



sido avaliados por um grupo de usuários. É necessário que cada usuário deste grupo tenha avaliado todos os documentos.

O estudo é composto de três passos. O primeiro passo consiste em apagar aleatoriamente parte das avaliações. No passo seguinte, é executado o algoritmo proposto e obtido a predição para as notas apagadas. No terceiro e último passo é calculado a precisão do algoritmo através de métricas discutidas em (O'CONNOR & HERLOCKER, 1999).

Para este estudo de caso foram utilizados 14 documentos associados ao curso Aprendizado Organizacional na COPPE/UFRJ no segundo período de aulas de 2002. Cinco alunos participaram do curso, sendo 4 alunos de mestrado e 1 aluno de doutorado. Estes alunos utilizaram a biblioteca durante o curso para avaliar e comentar os documentos. Para incentivar os alunos a participarem, esta atividade fazia parte da avaliação do curso.

Foram obtidas 84 avaliações dos 6 usuários participantes (1 professora e 5 alunos). Os códigos de identificação e os títulos dos documentos do curso encontram-se listados na **Tabela 15**. As 84 avaliações estão representadas na Tabela 16. Por uma questão de simplificação, será utilizado o código do documento como identificação do mesmo nas tabelas seguintes deste estudo.

**Tabela 15 - Lista de documentos do curso Aprendizado Organizacional.**

<b>Código</b>	<b>Título do documento</b>
97	Organizational Learning in Dynamic Domains
98	The Organic and the Mechanistic Form of Managing Knowledge in Software Development
99	CORONET-Train: A Methodology for Web-based Colaborative Learning in Software organizations
100	Supporting Knowledge Management in University Software R&D Groups
101	Organizational Learning and Software Process Improvement: a Case Study
102	Lessons Learned about Structing and Describing Experience
103	Knowledge Elicitation through Web-Based Data Mining Services
104	A model for Analyzing Measurement Based Feedback Loops in Software Development Projects
105	Process-Integrated Learning: The ADVISOR Approach for Corporate Development
106	Task-Specific Knowledge Management in a Process-Centred SEE
107	Computer-Supported Reuse of Project Management Experience with an Experience Base
108	The Integrated Corporate Information Network iCoIN: A comprehensive, Web-Based Experience Factory
109	Reuse of Measurement Plans Based on Process and Quality Models
110	Engineering Experience Base Maintenance Knowledge

**Tabela 16 – Avaliações dos documentos do curso Aprendizado Organizacional.**

Documentos Usuários	97	98	99	100	101	102	103	104	105	106	107	108	109	110
Professora	4	3	5	5	4	4	3	4	5	5	5	5	5	5
Aluno 1	4	3	5	5	4	4	3	4	4	4	3	4	5	3
Aluno 2	3	4	4	5	2	5	2	3	5	5	3	5	2	2
Aluno 3	3	4	3	5	5	3		4	5	5	5	4	4	4
Aluno 4	4	5	4	5	4	5	3	4	5	5	3	5	5	3
Aluno 5	4	3	3	4	2	3	1	3	3	5	3	3	3	4
Média das avaliações	3,7	3,7	4	4,83	3,67	4	2,5	3,7	4,5	4,83	3,67	4,33	4	3,5

No primeiro passo do estudo, foram apagadas aleatoriamente 15 % das avaliações do curso. As notas apagadas estão marcadas na Tabela 16.

No passo seguinte, foram calculadas as predições das notas apagadas. Neste passo, foram geradas três matrizes durante a execução do algoritmo de predição: matriz de média aritmética das avaliações por usuário, matriz de correlação entre os usuários e matriz de predição.

A primeira matriz calculada no algoritmo é a média aritmética das avaliações por usuário, ilustrada na Tabela 17.

**Tabela 17 - Média aritmética das avaliações por usuário.**

Código do usuário	Média aritmética das avaliações
Professora	4,42
Aluno 1	4
Aluno 2	3,57
Aluno 3	4,08
Aluno 4	4,28
Aluno 5	3,14

A segunda matriz gerada contém as correlações entre os usuários, como mostra a Tabela 18. Vale ressaltar que a correlação entre os usuários é calculada com base nos documentos que ambos avaliaram. Por exemplo, no cálculo da correlação entre o aluno 2 e o 4 não foi considerado o documento de código 107, pois as avaliações de ambos foram apagadas neste estudo.

**Tabela 18 – Matriz de correlação entre os usuários.**

Usuário x usuário	Professora	Aluno 1	Aluno 2	Aluno 3	Aluno 4	Aluno 5
Professora	1	0,45	0,18	0,16	0,18	0,56
Aluno 1		1	0,31	-0,20	0,47	0,27
Aluno 2			1	-0	0,73	0,49
Aluno 3				1	0,28	-0,04
Aluno 4					1	0,39
Aluno 5						1

A matriz de correlação entre os usuários é simétrica. Correlações obtêm valores entre [-1, +1]. Quanto mais próximo de + 1, mais similares são os usuários na forma de avaliar, ou seja, gostam ou desgostam dos mesmos documentos. Quanto mais próximo de - 1, mais dissimilares são os usuários.

A terceira matriz calculada no algoritmo é equivalente a predição das notas que foram apagadas no estudo. O resultado obtido está ilustrado na Tabela 19.

**Tabela 19 - Matriz de predição para as notas apagadas no estudo**

Documentos \ Usuários	97	100	101	103	104	105	106	107
Professora		5					5	
Aluno 1			2,98			4,37		
Aluno 2					3,38			3,46
Aluno 3				3,61			4,18	
Aluno 4			3,50					4,35
Aluno 5	2,74				2,82			

O último passo do estudo consiste em medir a precisão do algoritmo de predição. Inicialmente, é necessário obter o erro de cada predição, comparando as notas calculadas pelo algoritmo (nota predição) com a avaliação feita pelo usuário (nota real). A Tabela 20 lista cada nota de predição calculada, cada nota real dada pelo usuário e o erro calculado entre as duas. Quanto mais próximo de zero for o erro, maior a precisão do algoritmo.

**Tabela 20 – Notas reais e notas obtidas na predição.**

Usuário	Código do documento	Nota predição	Nota real	Erro
Professora	100	5	5	0
	106	5	5	0
Aluno 1	101	2,99	4	-1,01
	105	4,37	4	0,37
Aluno 2	104	3,39	3	0,39
	107	3,46	3	0,46
Aluno 3	103	3,61	3	0,61
	106	4,19	5	-0,81
Aluno 4	101	3,50	4	-0,50
	107	4,35	3	1,35
Aluno 5	97	2,74	4	-1,26
	104	2,82	3	-0,18

Em (O'CONNOR & HERLOCKER, 1999) o autor propôs avaliar os algoritmos de filtragem colaborativa com base nas medidas MAE (*Mean Absolute Error*), desvio padrão do erro e coeficiente Pearson entre as notas obtidas pelo algoritmo e as avaliações dos usuários. O índice MAE é calculado de acordo com a **Equação 4**. Nesta equação,  $E_i$  é o erro de cada predição e  $N$  o número de predições realizadas. Quanto mais próximo de 0 for o valor do MAE, mais preciso é o algoritmo.

$$MAE = \sum |E_i| / N$$

**Equação 4 – MAE.**

Foram calculadas as três medidas, utilizando os dados coletados neste estudo. Os valores obtidos estão listados na **Tabela 21**. A tabela apresenta um comparativo dos resultados obtidos no estudo realizado por MILLER et al. (1997) para o sistema GroupLens com os fóruns de discussão “Rec.humor” e “Comp.os.linux.development”; no estudo (SHARDANAD & MAES, 1995) para o sistema de recomendação Ringo e nos experimentos executados com o Movielens (O'CONNOR & HERLOCKER, 1999). Os dois primeiros estudos utilizaram algoritmos baseados em memória, enquanto o último utilizou algoritmos baseados em modelo. O resultado obtido no algoritmo utilizado no RecDoc foi similar ao do GroupLens, apresentando um erro inferior a este. A obtenção de um melhor resultado pode ser atribuído a amostragem utilizada e o número de avaliações do estudo experimental.

**Tabela 21 – Comparativo entre resultados do algoritmos de filtragem colaborativa.**

Sistema de recomendação	Algoritmo de predição	Tipo do algoritmo	MAE	Desvio padrão do erro	Correlação Pearson
<b>Grouplens – fórum Rec.humor</b>	Correlação de Pearson	baseado em memória	0,67	0,68	0,62
<b>Grouplens – fórum Comp.os.linux.development</b>	Correlação de Pearson	baseado em memória	0,71	0,64	0,55
<b>Ringo</b>	Correlação de Pearson	baseado em memória	1,1	1,4	-
	Mean square difference	baseado em memória	1	1,3	-
<b>Movielens</b>	Average link k = 50	baseado em modelo	0,77	-	-
<b>Movielens</b>	KMetis k = 5	baseado em modelo	0,8	-	-
<b>RecDoc</b>	Correlação de Pearson	baseado em memória	0,63	0,31	0,57

O valor desejável para a métrica MAE é o mais próximo de zero. Quanto menor o valor, maior a precisão do algoritmo.

Obs: o sistema Ringo utiliza uma escala de avaliação de 1 a 7.

### 5.3 Estudo de Caso: Desempenho da Abordagem Colaborativa

O objetivo deste segundo estudo de caso é avaliar a abordagem colaborativa proposta. Esta abordagem utiliza o algoritmo de predição analisado no estudo de caso anterior.

O estudo é composto de dois passos. No primeiro passo são geradas recomendações para os usuários da biblioteca. O segundo passo do estudo consiste em avaliar o resultado das recomendações. Cada documento recomendado foi avaliado pelos usuários como sendo de interesse ou não. O desempenho da recomendação foi calculado como a percentagem de recomendações positivas (*de interesse*) e negativas (*não interessantes ao usuário*).

A amostra utilizada continha dados de 5 usuários e 200 avaliações. Os usuários participantes realizaram no mínimo 15 avaliações.

O primeiro passo do estudo foi executado conforme o esquema de recomendação colaborativa proposto no capítulo quatro. Este esquema é composto de 4 etapas: exportação, predição, importação e recomendação. Cada etapa é descrita a seguir:

- a) etapa exportação: as avaliações foram exportadas da base de dados da biblioteca para um arquivo texto (input.txt). Foi utilizada a ferramenta DTS (“Data Transfer Service”) do SQL Server que faz transferência de dados entre os arquivos e a base de dados.
- b) etapa predição: o algoritmo de predição foi executado tendo como entrada o arquivo “input.txt”. Esta etapa gera uma matriz de correlação entre usuários e uma matriz de predição.
- c) etapa importação: as matrizes geradas na etapa anterior foram importadas para a base de dados. Foi utilizada a ferramenta DTS do SQL Server.
- d) etapa recomendação: 3 filtros foram aplicados no arquivo que contém a matriz de predição, para eliminar documentos que não são de interesse do usuário.

Na etapa de recomendação foram aplicados os filtros de nota mínima na predição, de palavra-chave e de nota média do documento. No filtro de nota mínima na predição, foram eliminados os documentos que obtiveram nota de predição inferior a 3,5. Em seguida foi aplicado o filtro de palavra-chave, no conjunto de documentos resultantes da filtragem anterior. Foram eliminados os documentos que não estavam classificados por palavras-

chaves definidas como área de interesse do usuário. Por fim, foram eliminados os documentos que tinham nota média inferior a 4.

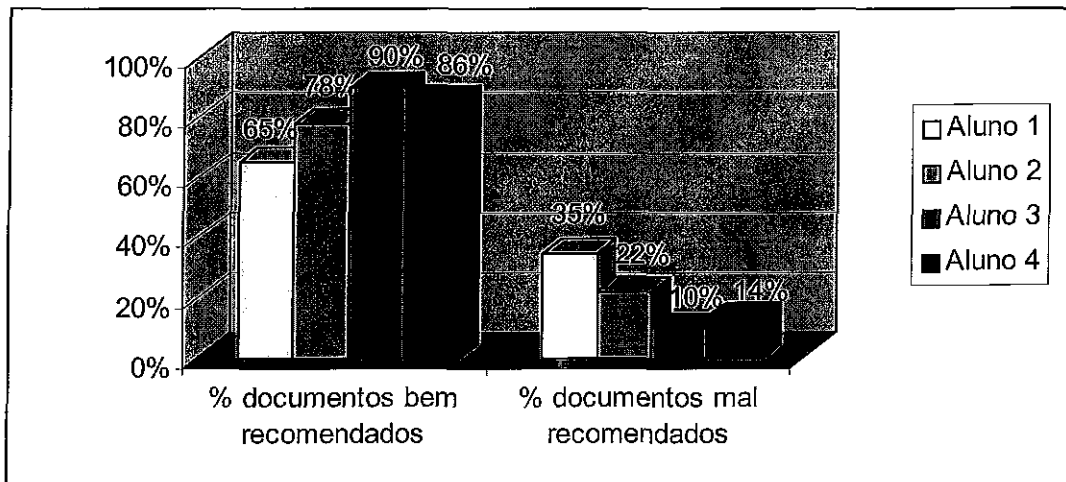
No segundo passo do estudo, o desempenho do sistema foi avaliado. A etapa de avaliação foi realizada em duas fases. Na primeira fase, os usuários avaliaram a lista de documentos recomendados. O usuário preencheu um formulário indicando se os documentos eram de interesse ou não. Com as avaliações dos usuários, foram calculadas as percentagens de recomendações positivas e negativas. Na segunda fase, os usuários avaliaram a lista de documentos eliminados da recomendação devido aos filtros aplicados. Os usuários informaram se os documentos eliminados eram de interesse ou não.

Em média 77% das recomendações foram positivas. Dentre os documentos eliminados na filtragem, 80% realmente não interessavam ao usuário, por serem documentos não relacionados com sua área de interesse. A **Tabela 22** apresenta o número de documentos recomendados por usuário, número de documentos filtrados e a percentagem das recomendações positivas e negativas por usuário. O **Gráfico 1** apresenta os resultados obtidos na recomendação. O **Gráfico 2** apresenta o desempenho da filtragem.

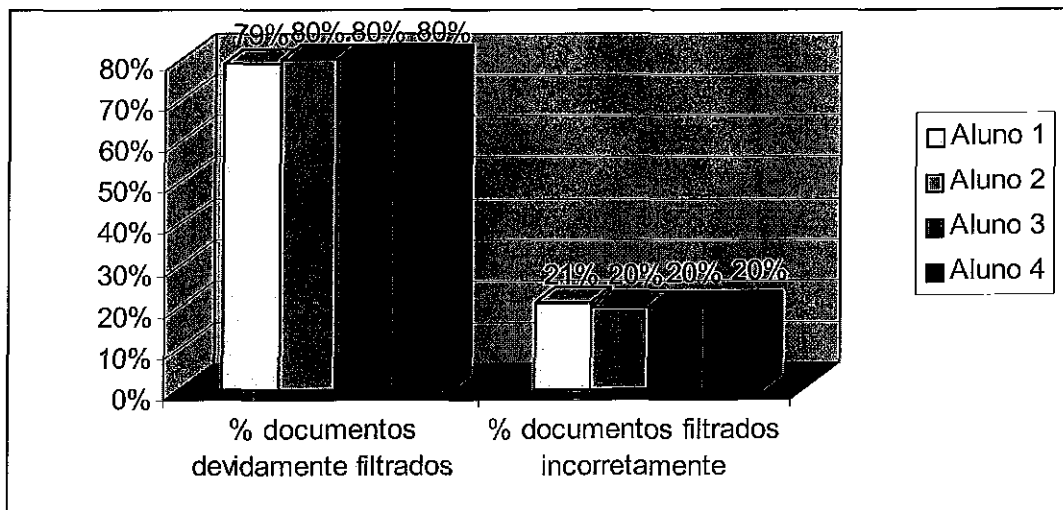
**Tabela 22 - Resultados obtidos na recomendação colaborativa.**

<b>Usuário</b>	<b>Número de documentos recomendados</b>	<b>% documentos bem recomendados</b>	<b>% documentos mal recomendados</b>	<b>Número de documentos eliminados</b>	<b>% documentos devidamente filtrados</b>	<b>% documentos filtrados incorretamente</b>
<b>Aluno 1</b>	40	65%	35%	23	79%	21%
<b>Aluno 2</b>	9	78%	22%	5	80%	20%
<b>Aluno 3</b>	21	90%	10%	5	80%	20%
<b>Aluno 4</b>	14	86%	14%	5	80%	20%
<b>Média</b>	21	77%	23%	9,5	80%	20%





**Gráfico 1 - Desempenho da recomendação colaborativa em %.**



**Gráfico 2 - Desempenho da filtragem em %.**

Foi observado que o desempenho do sistema foi maior para os usuários que avaliaram mais documentos. Outro fator detectado foi que os documentos melhor recomendados tiveram mais avaliações também. Foi observado que o número de avaliações por usuário e por documento influem diretamente no resultado da recomendação.

Outras considerações foram obtidas com uma análise mais detalhada das avaliações. Foi feito um levantamento de justificativas do porquê documentos irrelevantes foram recomendados e os relevantes eliminados na filtragem. Utilizou-se como exemplo o aluno 1 que obteve um índice de recomendações positivas de 65%, inferior aos demais usuários.

Observou-se que os documentos mal recomendados ao aluno 1 não foram eliminados na filtragem pois eram classificados com pelo menos uma palavra-chave definida como área de interesse no seu perfil do usuário. Por exemplo, o usuário definiu no seu perfil que estava interessado em documentos sobre gerência do conhecimento. Dentre os documentos mal recomendados, 6 estavam classificados com esta palavra-chave. Entretanto, os documentos eram classificados também com as palavras-chaves informática médica, medicina, ou biomédica, que não fazem parte da sua área de interesse. Deste fato, podemos verificar que seria interessante manter também uma lista de palavras-chaves na qual o usuário não tenha interesse. Isto poderia ser feito no momento do cadastro do usuário ou então permitir que o próprio usuário definisse os filtros a serem aplicados na etapa de recomendação.

Para complementar o estudo, foi verificado se os documentos eliminados no filtro eram de interesse deste aluno. Dentre os 23 documentos eliminados, 18 realmente não interessavam (79%) e 5 documentos interessavam ao usuário (21%). Foi constatado que 2 documentos de interesse foram eliminados no filtro palavra-chave. Estes documentos eram sobre ontologia, mas o usuário não havia selecionado essa palavra-chave como área de interesse. A palavra-chave foi acrescentada no perfil e estes documentos devem ser indicados em uma próxima recomendação. Outros 2 documentos foram eliminados no filtro de nota média pois os documentos obtiveram nota média 1. Entretanto, a nota obtida na predição foi de 3,5. Deste fato observamos que as pessoas que avaliaram com nota 1, deveriam ter coeficiente de correlação baixo em relação ao aluno. Neste caso, estas pessoas realmente discordam de opinião. O aluno já tinha lido os documentos e os avaliariam com nota 4. Talvez aplicar o filtro de nota média não seja necessário, pois na predição é considerada a relação dos usuários para determinar a nota, geralmente uma estimativa melhor do que a média aritmética das notas.

## **5.4 Estudo de Caso: Desempenho do Sistema de Recomendação**

O objetivo deste terceiro estudo de caso é avaliar o desempenho do sistema de recomendação proposto. Foi utilizada uma amostra de 11 usuários e 310 avaliações.

O estudo é composto de três passos. Nos dois primeiros passos, foram executadas as abordagens de recomendação por conteúdo e colaborativa. No terceiro passo, foram

coletadas avaliações dos usuários. Foi medido o desempenho das abordagens e feito um comparativo entre elas.

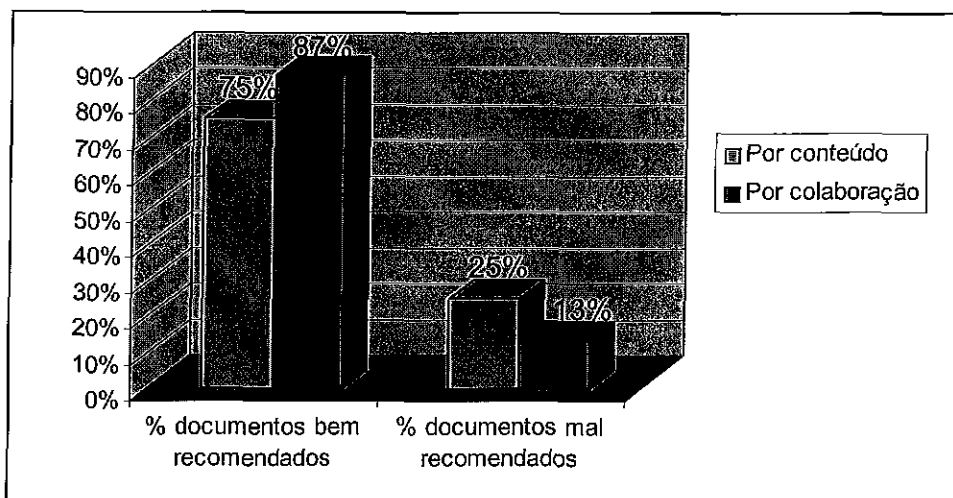
No passos 1 e 2, foram aplicados o filtro de palavras-chaves e mais um filtro personalizado. Não foi aplicado o filtro de nota média como no estudo anterior. No filtro personalizado, o usuário informou se participa de algum projeto de pesquisa, tipos de documento que tem mais interesse e a partir de qual ano de publicação deseja filtrar os documentos. A definição do filtro personalizado foi descrito no capítulo quatro.

No passo 3, cada documento recomendado foi avaliado pelos usuários como sendo de interesse ou não. O desempenho da recomendação foi calculado como a percentagem de recomendações positivas (de interesse) e negativas (não interessantes ao usuário), da mesma forma que no estudo anterior.

Na recomendação por conteúdo, foram geradas em média 12 recomendações por usuário. Já na colaborativa, a média foi de 8 documentos. Os resultados obtidos na abordagem por conteúdo foram: 75% de recomendação positivas e 25% negativas. Na abordagem colaborativa, a recomendação positiva obteve um valor de 87% das indicações. Apenas 13% dos documentos não interessavam ao usuário. A **Tabela 23** mostra os resultados obtidos. O **Gráfico 3** apresenta um comparativo do desempenho da abordagem por conteúdo e por colaboração.

**Tabela 23 – Desempenho do sistema de recomendação.**

<b>Abordagem</b>	<b>Número médio de documentos recomendados</b>	<b>% documentos bem recomendados</b>	<b>% documentos mal recomendados</b>
<b>Por conteúdo</b>	12	75%	25%
<b>Por colaboração</b>	8	87%	13%



**Gráfico 3 – Desempenho da abordagem por conteúdo x por colaboração em %.**

Este estudo foi o que obteve um resultado mais positivo dos usuários. Após receberem a lista de documentos recomendados, alguns enviaram mensagens perguntando “Onde posso encontrar esse documento?” ou “A seleção de documentos vai ser bastante útil para iniciar a pesquisa no tema.”. Isso mostra que as pessoas mesmo convivendo em um mesmo ambiente de pesquisa, ou trabalho, não têm a oportunidade de compartilhar documentos e experiências.

Os estudantes que estão iniciando a pesquisa de tese foram os mais receptivos aos documentos recomendados. Neste período, os estudantes estão constantemente realizando busca de novas referências bibliográficas. Neste caso, receber uma lista de materiais pré-selecionados é um ponto de partida mais eficiente do que iniciar buscas em ferramentas ou bibliotecas na Web. Para os alunos que já tinham realizado a revisão bibliográfica da tese, foi mais comum o sistema recomendar documentos que o usuário já conhecia.

Ocorreu um caso onde a abordagem por conteúdo apresentou resultado insatisfatório (< 60%). Parte dos documentos mal recomendados eram relacionados a área de interesse do usuário, mas este já conhecia o documento e o avaliou segundo a qualidade.

A utilização de um sistema de recomendação se mostrou uma abordagem interessante para a troca de documentos entre os estudantes desse grupo de pesquisa.

# Capítulo 6

## Conclusões e Perspectivas Futuras

---

*Neste capítulo são apresentadas as principais conclusões e as perspectivas futuras para a continuidade deste trabalho.*

### 6.1 Conclusões

A importância de disponibilizar serviços personalizados em provedores de informação na Web, incluindo bibliotecas digitais, foram discutidos em diversos trabalhos estudados (AMATO & STRACCIA, 1999, LAWRENCE, 2000, MANBER & ROBISON, 2000). Esta é uma tendência no desenvolvimento de novas ferramentas de busca na Web (BAEZA-YATES & RIBEIRO-NETO, 1999).

Recentemente, os sistemas de recomendação têm sido um dos serviços personalizados mais utilizados em diversas aplicações. Estes sistemas podem ser construídos com base na similaridade entre documentos ou na colaboração entre os usuários. Em ambas abordagens, geralmente as informações dos usuários são estruturadas em perfis de usuário.

Muitos trabalhos foram realizados para recomendação de mensagens em fóruns de discussão, USENET e recomendação de filmes. Entretanto, a utilização destes sistemas em sites de comércio eletrônico e bibliotecas digitais é recente. Já foram feitos estudos comparativos sobre desempenho e precisão dos diversos algoritmos utilizados na abordagem por colaboração. Novos enfoques foram direcionados a pesquisa como o problema da partida a frio, que aborda questões como indicar itens quando o sistema ainda não sabe informação suficiente sobre o usuário. Outra questão, é a apresentação da justificativa da recomendação, que aumenta a confiabilidade das indicações realizadas pelo sistema.

Baseado nessas questões, esta dissertação propôs a utilização de um sistema de recomendação híbrido para uma biblioteca digital virtual na Web. O sistema utiliza a própria experiência do usuário (abordagem por conteúdo) e a colaboração entre usuários para indicar documentos aos leitores da biblioteca (abordagem por colaboração).

A combinação de sistemas de recomendação, serviços e filtros personalizados parece ser um caminho para disponibilizar informação relevante os usuários em bibliotecas digitais.

Dentre as principais contribuições deste trabalho temos:

- (1) Definição, especificação e construção de uma biblioteca genérica Web configurável e customizável. A biblioteca possibilita a interação entre usuários e compartilhamento de avaliações e comentários sobre os documentos. Foram apresentadas a arquitetura e etapas de desenvolvimento desta biblioteca;
- (2) Representação do perfil para as bibliotecas discutidas, baseada no modelo proposto por (AMATO & STRACCIA, 1999);
- (3) Definição de dois esquemas de recomendação genéricos: um esquema para a abordagem por conteúdo e outro para abordagem por colaboração, baseada no processo de filtragem colaborativa proposto por SARWAR et al. (2001);
- (4) Definição e implementação do sistema de recomendação RecDoc.
- (5) Definição de filtros personalizados para a biblioteca digital genérica. Os filtros são utilizados na seleção de documentos a serem recomendados pelo RecDoc, bem como uma alternativa para solucionar o problema da partida a frio.
- (6) Desenvolvimento de justificativa de recomendação baseado nos usuários que avaliaram os documentos e no desempenho anterior do sistema. Esse tipo de justificativa foi proposto em (HERLOCKER, 2001) e não havia sido implementado nos sistemas estudados.
- (7) Validação do sistema de recomendação proposto através de estudo de casos. Estudo da precisão do algoritmo de filtragem colaborativa utilizado.

## **6.2 Perspectivas Futuras**

Buscando melhorar e expandir a biblioteca desenvolvida e o sistema de recomendação proposto, algumas perspectivas de trabalhos futuros são destacadas.

Em relação à biblioteca genérica desenvolvida, um primeiro passo seria a definição e implementação de uma busca avançada onde fossem combinados os tipos de busca desenvolvidos. Em um único formulário de pesquisa o usuário poderia escolher

palavras-chaves, especificar um autor e selecionar um curso, por exemplo. Uma outra funcionalidade interessante a ser adicionada ao sistema de busca, é permitir que o usuário salve a busca no perfil do usuário para aplicá-la posteriormente.

Um fator importante é utilizar um padrão de metadados para descrever a referência bibliográfica, como o padrão MARC (MARC, 2003) ou o padrão proposto por MOURA et al. (2001). Com a utilização de metadados, será possível integrar a biblioteca desenvolvida com demais bibliotecas existentes.

Um outro aspecto interessante é desenvolver um módulo de configuração automática da biblioteca, de forma que a implantação da biblioteca genérica permita a escolha de características da interface como fonte e cor do texto. Como a biblioteca foi desenvolvida de forma modular, o usuário poderia também selecionar funcionalidades a serem utilizadas.

Em relação ao sistema de recomendação desenvolvido, a abordagem por conteúdo poderia incluir análise de campos de catalogação para medir a similaridade entre documentos, como complemento a comparação realizada por palavra-chave.

Um trabalho interessante a ser feito na recomendação colaborativa é utilizar um outro algoritmo de predição. Isto validaria a possibilidade de integração proposta na arquitetura. Neste caso, poderia ser feito um estudo comparativo do desempenho deste novo algoritmo e o desenvolvido neste trabalho.

Uma melhoria a ser considerada no esquema de recomendação por colaboração é possibilitar a requisição on-line da recomendação. Rotinas de exportação e importação automática dos dados poderiam ser desenvolvidas e agentes inteligentes poderiam agir como um serviço de alerta automático.

Outra possibilidade, é realizar os estudos de casos com uma outra instância da biblioteca genérica. A biblioteca está sendo configurada para a Fundação Bahiana de Cardiologia (FBC) para centralizar a produção científica da instituição e referências produzidas por centros afins. A FBC realiza diversas atividades de ensino e pesquisas na área de cardiologia, sendo de extrema importância a utilização da biblioteca para apoiar essas atividades. A biblioteca e o sistema proposto serão validados em uma outra instituição utilizando uma amostragem de documentos e usuários diferentes da realizada nesse trabalho.

## Referências Bibliográficas

---

- ACHEI, 2003, Sistema de busca Achei, <http://www.achei.com.br>, acessado em 28/02/2003.
- ACM D.L., 2003, Biblioteca Digital da ACM, <http://www.acm.org/dl/>, acessado em 28/02/2003.
- AOL, 2003, Sistema de busca AOL, <http://search.aol.com>, acessado em 10/03/2003.
- ALTAVISTA, 2003, Sistema de busca AltaVista, <http://www.altavista.com>, acessado em 20/12/2002.
- AMATO, G & STRACCIA, U., 1999, "User Profile Modeling and Applications to Digital Libraries". In: Proceeding of ECDL European Conference on Digital Libraries, pp. 184-197, Paris, France.
- AMAZON, 2002, Amazon homepage, <http://www.amazon.com>, acessado em 20/12/2002.
- BAEZA-YATES, R. & RIBEIRO-NETO, B., 1999, Modern Information Retrieval. Addison-Wesley-Longman Publishing co.
- BAEZA-YATES, R., 2000, "Desenredando La Madeja". In: Anais da VII Jornada IberoAmericana de Informática, Cartagena, Colômbia.
- BALABANOVIC, M., SHOHAM, Y., YUN, Y., 1995, "An Adaptive Agent for Automated Web Browsing", Stanford University Digital Library Project, Working Paper SIDL-WP-1995-0023.
- BALABANOVIC, M., & SHOHAM, Y., 1997, "Fab: Content-Based, Collaborative Recommendation", *Communications of the ACM*, v. 40, n. 3, pp. 66-72.
- BELKIN, N. J. & CROFT, W. B., 1992, "Information filtering and information retrieval: Two sides of the same coin ?", *Communications of the ACM*, v. 35, n. 12, pp. 29-38.
- BILLSUS, D. & PAZZANI, M., 1998, "Learning Collaborative Information Filters". In: Proceedings of the Fifteenth International Conference Machine Learning, pp. 46-54, Madison, WI, USA.
- BOLLACKER, K., LAWRENCE, S., GILES, C., 2000, "Discovering Relevant Scientific Literature on the Web", *IEEE Intelligent Systems*, v. 15, n. 2, pp. 42-47.



- BORGMAN, C., 1999, "What are digital libraries? Competing visions", *Information Processing and Management: an International Journal*, v. 35, n. 5, pp. 227–243.
- BREESE, J. S., HECKERMAN, D., KADIE, C., 1998, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering". In: *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 43-52, Madison, WI, USA.
- BUCKLEY, C., 1995, "Optimization of relevance feedback weights". In: *Proceeding of 1995 ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 351 - 357, Seattle, Washington, USA.
- BVS, 2002, Biblioteca Virtual de Saúde, [www.saude.gov.br/bvs/](http://www.saude.gov.br/bvs/), acessado 20/12/2002.
- CADE, 2003, Ferramenta de busca CADE, <http://www.cade.com.br>, acessado em 28/03/2003.
- CDNOW, 2002, CDNow homepage, <http://www.cdnw.com/>, acessado em 20/12/2002.
- CITeseer, 2002, Biblioteca Digital CiteSeer, <http://citeseer.nj.nec.com/cs>, acessado em 20/12/2002.
- DAHLEN, B. J., KONSTAN, J. A., HERLOCKER, J. L., et al, 1998, "Jump-starting Movielens: User benefits os starting a collaborative system dead data", University of Minnesota, TR 98-017.
- DEERWESTER, D., DUMAIS, S. T., LANDAUER, T. K., et al, 1990, "Indexing by Latente Semantic Analysis", *Journal of the Society of Information Science*, v.41, n. 6, pp. 391-407.
- DELOS, 2003, Delos Digital Library Network of Excellence, <http://delos-noe.iei.pi.cnr.it/>, acessado em 20/02/2003.
- DRUGSTORE, 2002, Drugstore homepage, <http://www.drugstore.com/>, acessado em 10/12/2002.
- EACHMOVIE, 2002, Eachmovie dataset, Compaq Research Center, <http://www.research.compaq.com/SRC/eachmovie/>, acessado em 28/12/2003.
- ENTLICH, R., OLSEN, J., GARSON, L, et al, 1997, "Making a digital library: the contents of the CORE project", *Communications of ACM*, v.15 , n. 2.
- FAENSEN, D., FAULSTICH, L., SCHWEPPE, H., et al., 2001, "Hermes A Notification Service for Digital Libraries". In: *Proceedings of ACM/IEEE Joint Conference on Digital Libraries*, pp. 373-380, Roanoke, Virginia, USA.
- FISHER, D., HILDRUM, K., NEWMAN, M., et al., 2000, "SWAMI: A Framework for Collaborative Filtering development and evaluation". In: *Proceedings of the 2000*

- Conference on Research and Development in Information Retrieval, pp. 366-368, Athens, Greece.
- FOLTZ, P. & DUMAIS, S., 1992, "Personalized information delivery: an analysis of information filtering methods", *Communications of ACM*, v. 32, n. 12, pp. 51-60.
- FOX, E., AKSCYN, R., FURUTA, R., et al., 1995, "Digital Libraries", *Communications of ACM*, v.38, n.4, pp.23-28.
- FOX, E. & MARCHIONINI, G., 1998, "Toward a WorldWide Digital Library", *Communications of ACM*, v. 41, n.4, pp.29-32.
- GEY, F., 1996, "Models in Information Retrieval". In: *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, Zurich, Switzerland.
- GILES, C. L., BOLLACKER, K. D., LAWRENCE, S., 1998, "CiteSeer: An Automatic Citation Indexing System". In: *Proceeding of The 3rd ACM Conference on Digital Libraries*, pp. 89-98, Pittsburgh, Pennsylvania, USA.
- GLOVER, E., LAWRENCE, S., GORDON, M., et al., 2001, "Web search – your way", *Communications of the ACM*, v. 44, n. 12, pp. 97–102.
- GOLDBERG, N., NICHOLS, D., OKI, B. M., et al., 1992, "Using collaborative filtering to weave information Tapestry", *Communications of the ACM*, v. 35, n. 12, pp. 61-70.
- GOOGLE, 2003, Ferramenta de busca Google, <http://www.google.com>, acessado em 20/02/2003.
- GROUPLENS, 2002, GroupLens homepage, <http://www.grouplens.org>, acessado em 20/12/2002.
- HERLOCKER, J., 2001, "Understanding and Improving Automated Collaborative Filtering Systems". Dissertação de doutorado, University of Minnesota, USA.
- HERLOCKER, J., KONSTAN, J., BORCHERS, J., et al., 1999, "An algorithmic framework for performing collaborative filtering". In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, p.230-237, August, Berkeley, California, USA.
- HOASHI, K. et al., 2000, "Document Filtering Method Using Non\_Relevant Information Profile". In: *Proceedings of the 2000 Conference on Research and Development in Information Retrieval*, pp. 313-317, Athens, Greece.

- HUANG, Z., CHUNG, W., ONG, T., et al., 2002, "A graph-based recommender system for digital library". In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries), pp.65-73, Portland, Oregon, USA.
- IEEE D.L., 2003, IEEE Computer Society Digital Library, <http://www.computer.org/publications/dlib/>, acessado em 15/02/2003.
- IMDB, 2002, Internet Movie Database, <http://us.imdb.com>, acessado em 20/12/2002.
- INFOSEEK, 2003, Sistema de busca Infoseek, <http://infoseek.go.com/>, acessado em 28/02/2003.
- KAPUR, K. & ZHANG, J., 2000, "Searching the Web Using Synonyms and Senses", WebNet Journal, July-september.
- KIM, Y., HAHN, S.Y., ZHANG, B.T., 2000, "Text Filtering by Boosting Naive Bayes Classifiers". In: Proceedings of the 2000 Conference on Research and Development in Information Retrieval, pp. 168-175, Athens, Greece.
- KONSTAN, J., MILLER, V, MALTZ, D., et al, 1997, "GroupLens: applying collaborative filtering to Usenet news", Communications of the ACM, v. 40 , n. 3, pp. 77-87.
- KRULWICH, B. & BURKEY, C., 1997, "The Infofinder Agent: Learning User Interests Through Heuristic Phrase Extraction", IEEE Expert Intelligent Systems & their applications.
- KUFLIK, T. & SHOVAL, P., 2000, "Generation of user profiles for information filtering - research agenda". In: Proceedings of the 2000 Conference on Research and Development in Information Retrieval, pp. 313 - 315 , Athens, Greece.
- LANG, K., 1995, "NewsWeeder: Learning to Filter NetNews". In: Proceeding of the 12<sup>th</sup> International Conference on Machine Learning, California, USA.
- LAWRENCE, S., GILES, C. L., BOLLACKER, K., 1999a, "Digital Libraries and Autonomous Citation Indexing", IEEE Computer, v. 32, n. 6, pp. 67-71.
- LAWRENCE, S., BOLLACKER, K., GILES, C. L., 1999b, "Indexing and Retrieval of Scientific Literature". In: Proceedings of the Eighth International Conference on Information and Knowledge Management (CIKM 99), pp.139-146, Kansas City, Missouri, USA.
- LAWRENCE, S., 2000, "Context in Web Search", IEEE Data Engineering Bulletin, v. 23, n. 3, pp. 25-32.
- LOOKSMART, 2003, Sistema de busca LookSmart, <http://www.looksmart.com>, acessado em 10/03/2003.

- MAES, P., 1994, "Social interface agents: acquiring competence by learning from users and other agents". In: Working Notes of the AAAI Spring Symposium on Software Agents, pp. 71-78, Stanford, California.
- MALTZ, D. & EHRLICH, K., 1995, "Pointing the way: active collaborative filtering". In: Proceedings of Conference on Human Factors in Computing Systems, pp. 202-209.
- MANBER, U. & ROBISON, J., 2000, "Experience with personalization of Yahoo!". Communications of the ACM, v. 43 , n. 8, pp. 35-39.
- MARC, 2003, MARC Standards, <http://www.loc.gov/marc/>, acessado em 02/04/2003.
- METAMINER, 2003, Sistema de meta-busca MetaMiner, <http://miner.bol.com.br/index.html>, acessado em 15/02/2003.
- MIDDLETON, S.E., ALANI, H., SHADBOLT, N.R., et al., 2002, "Exploiting Synergy Between Ontologies and Recommender Systems". In: proceedings of the Eleventh International World Wide Web Conference (WWW2002), Semantic Web Workshop 2002, pp. 41-50, Hawaii, USA.
- MILLER, B. N., RIEDL, J. T., KONSTAN, J. A., 1997, "Experiences with GroupLens: Making Usenet useful again". In: Proceedings of the Usenix Winter Technical Conference, Anaheim, California, USA.
- MIZZARO, S., 1997, "Relevance: the whole history", Journal of American Society for Information Science, v. 48, n.9, pp. 810-832.
- MOONEY, R. & ROY, L., 2000, "Content-based Book Recommending Using Learning for Text Categorization". In: Proceedings of the fifth ACM conference on Digital libraries, pp.195-204, San Antonio, Texas, USA.
- MOTTA, C. L. R., 1999, "Um Ambiente de Recomendação e Filtragem Cooperativas para Apoio a Equipe de Trabalhos". Dissertação de doutorado, Universidade Federal do Rio de Janeiro/UFRJ, Rio de Janeiro, Brasil.
- MOUKAS, A. M. C., 1997, "Amathea: Information Filtering and discovery Using a Multiagent Evolving System". Dissertação de mestrado, MIT, USA.
- MOURA, A. M. C., SILVA, L. A. E., CAMPOS, M. L. M., 2001, "A Metadata Approach for Designing Configurable Interfaces in Digital Libraries". In: 12th International Workshop on Databases and Expert Systems Applications, DEXA 2001, Munique. Los Alamitos - California: IEEE Computer Society Press, 2001, v. 1, pp. 942 – 947.

- MOURA, A., 2002, "The semantic Web: fundamentals, technologies, trends". In: Anais do XVII Simpósio Brasileiro de Banco de Dados, Gramado, Brasil, Outubro.
- MOVIELENS, 2002, MovieLens homepage, <http://movielens.umn.edu/>, acessado em 20/12/2002.
- MOVIEFINDER, 2002, MovieFinder homepage, <http://www.moviefinder.com>, acessado em 20/12/2002.
- NASA, 2003, National Aeronautics and Space Administration, <http://www.nasa.gov/>, acessado em 20/02/2003.
- NCSTRL, 2003, Networked Computer Science Technical Reports Library, <http://www.ncstrl.org>, acessado em 20/02/2003.
- NETPERCEPTIONS, 2002, Net Perceptions homepage, <http://www.netperceptions.com/>, acessado em 20/12/2002.
- NEW ZEALAND D.L., 2003, New Zealand Digital Library, <http://www.nzdl.org/>, acessado em 15/02/2003.
- NSF, 2003, National Science Foundation, <http://www.nsf.gov>, acessado em 15/02/2003.
- O'CONNOR, M. & HERLOCKER, J., 1999, "Clustering Items for Collaborative Filtering". In: Workshop on Recommender Systems: Algorithms and Evaluation, 1999 Conference on Research and Development in Information Retrieval, California, USA.
- OLD DOMINION, 2003, Old Dominion University, <http://web.odu.edu/>, acessado em 15/02/2003.
- PAEPCKE, A., 1996, "Digital Libraries: Searching is not Enough", D-Lib Magazine, May, ISSN 1082-9873, <http://dlib.org/dlib/may96/stanford/05paepcke.html>, acessado em 28/02/2003.
- PAEPCKE, A. & GARCIA-MOLINA, H., 1999, "Beyond Document Similarity: Understanding Value-based Search and Browsing Technologies", Stanford University, <http://www.stanford-db.edu/>, acessado em 25/03/2002.
- PRESSMAN, R. S., 2001, "*Software Engineering: A Practitioner's Approach*", McGraw-Hill International Editions, 5th ed.
- PROSSIGA, 2002, projeto PROSSIGA de apoio a pesquisa e desenvolvimento na área de ciência e tecnologia do CNPq, <http://www.prossiga.br>, acessado em 20/12/2002.
- RADAR UOL, 2003, Sistema de busca Radar UOL, <http://www.radaruol.com.br>, acessado em 10/03/2003.

- RASHID, A. M., ALBERT, I. A., COSLEY, D., et al., 2002, "Getting to know you: Learning new user preferences in recommender Systems". In: Proceedings of the 2002 International Conference on Intelligent User Interfaces, pp. 127-134, San Francisco, CA.
- RESNICK, P., IACOVU, N., SUSHAK, M., et al, 1994, "GroupLens: An open architecture for collaborative filtering of netnews". In: Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 175 – 186, Chapel Hill, North Carolina, USA.
- RESNICK, P. & VARIAN, H., 1997, "Recommender Systems", Communications of the ACM, v. 40 n.3, pp. 56-58.
- ROCHA, C., XEXEO, G. B., ROCHA, A. R. C., 2002a, "RecDoc: a recommender system to a web digital library". In: Proceedings of EdMedia Conference, Denver, USA.
- ROCHA, C., XEXEO, G. B., ROCHA, A. R. C., 2002b, "MyLibrary: a web personalized digital library". In: Proceedings of E-learn Conference, pp. 824-829, Montreal, Canada.
- RUCKER, J. & POLANCO, M., 1997, "Citeseer: personalized navigation for the Web", Communications of the ACM, v. 40, n. 3, pp. 73–76.
- SARWAR, B., KARYPIS, G., KONSTAN, J., REIDL, J., 2001, "Item-based Collaborative Filtering Recommendation Algorithms". In: Proceedings of the tenth international conference on World Wide Web, Hong Kong, pp.73-76.
- SCHAFFER, J. B., KONSTAN, J. A., RIEDL, J., 2001, "E-Commerce Recommendation Applications". In: proceedings of Data Mining and Knowledge Discovery, pp.115-153.
- SCHITIR, B., et al, 1998, "Digital Library Information Appliances". In: Proceedings of ACM Digital Library 98, Pitsburg, USA.
- SHARDANAD, U. & MAES, P., 1995, "Social information filtering: Algorithms for automating "word of mouth"". In: Proceedings of the Conference on Human Factors in Computing Systems – CHI 95, Denver, pp. 210-127.
- SILVA, I., RIBEIRO NETO, B., 1998, *Modelagem em Sistemas de Recuperação de Informação*, Relatório técnico, UFMG, Minas Gerais.
- SOBOROFF, I., NICHOLAS, C., PAZZANI, M., 1999, Workshop on Recommender Systems: Algorithms and Evaluation., 1999 International Conference on Research and Development in Information Retrieval, pp. 36-43, California, USA.

- SPRINGER, 2002, Biblioteca Digital Springer, <http://link.springer.de/search.htm>, acessado em 20/12/2002.
- SWAMI, 2002, SWAMI homepage, <http://guir.cs.berkeley.edu/projects/swami/>, acessado em 20/12/2002.
- TERVEEN, L., HILL, W., MCDONALD, D., et al, 1997, "Phoaks: a system for sharing recommendations", *Communications of the ACM*, v. 40, n. 3, pp. 59-62.
- TERVEEN, L. & HILL, W., 2001, "Beyond Recommender System: Helping People to Find Each Other". In: *Proceedings of HCI in the New Millennium*, Jack Carroll, ed., Addison-Wesley.
- TRIPADVISOR, 2003, Trip Advisor Homepage, <http://www.tripadvisor.com>, acessado em 28/02/2003.
- VIARIO, 2003, ViaRio Homepage, <http://www.viario.com.br>, acessado em 28/02/2003.
- VIRGINIA, 2003, University of Virginia, <http://www.virginia.edu/>, acessado em 15/02/2003.
- VIRGINIA TECH, 2003, Virginia Tech University, <http://www.vt.edu/>, acessado em 15/02/2003.
- WEBCRAWLER, 2003, WebCrawler homepage, <http://www.webcrawler.com/>, acessado em 28/02/2003.
- WEBLEARNER, 2002, <http://www.ics.uci.edu/~pazzani/>, acessado em 20/12/2002.
- WITTEN, I. H, NEVILL-MANNING, C., CUNNINGHAM, S., 1996, "Digital Libraries Based on Full-Text Retrieval". In: *Proceedings of WebNet 96*.
- WOLFRAM, D. & XIE, H., 2002, Traditional IR for web users: a context for general audience digital libraries. *Information Processing & Management*, v. 38, n. 5, pp. 627-648.
- WORDNET, 2002, Cognitive Science Laboratory at Princeton University homepage, <http://www.cogsci.princeton.edu/~wn/>, acessado em 20/02/2002.
- WRS, 1999, Workshop on Recommender Systems: Algorithms and Evaluation, <http://www.csee.umbc.edu/~ian/sigir99-rec/>, acessado em 20/12/2002.
- WRS, 2001, Workshop on Recommender Systems, <http://www.cs.orst.edu/~herlock/rsw2001/>, acessado em 20/02/2003.
- YAHOO, 2003, Ferramenta de busca Yahoo, <http://www.yahoo.com>, acessado em 28/02/2003.

# Anexo 1

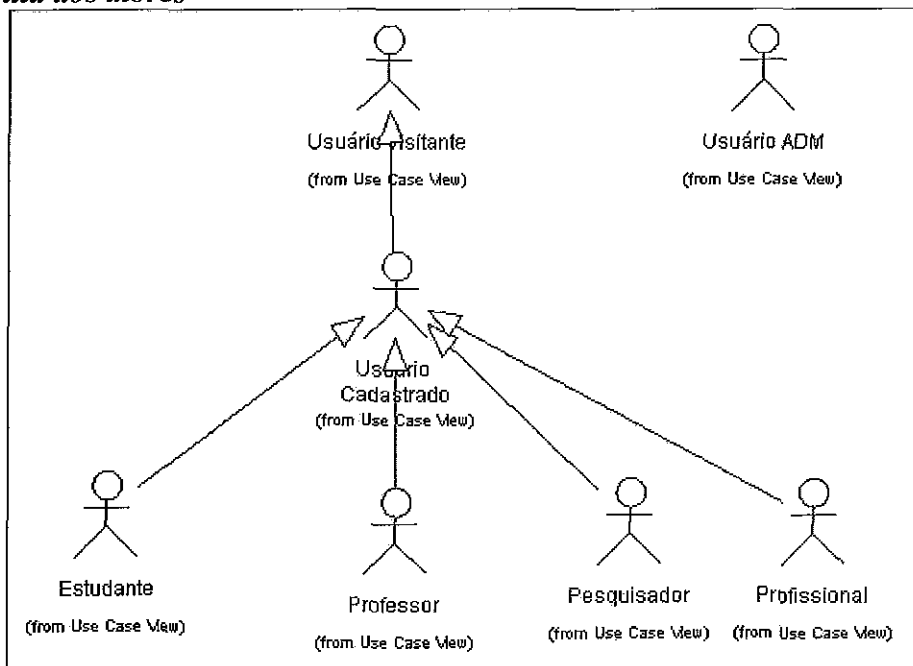
## Casos de Uso da Biblioteca Digital de Engenharia de Software

*Este anexo apresenta atores e casos de uso definidos para a Biblioteca de Engenharia de Software. O anexo também apresenta o diagrama dos casos de uso que ilustra a interação entre eles.*

### Atores

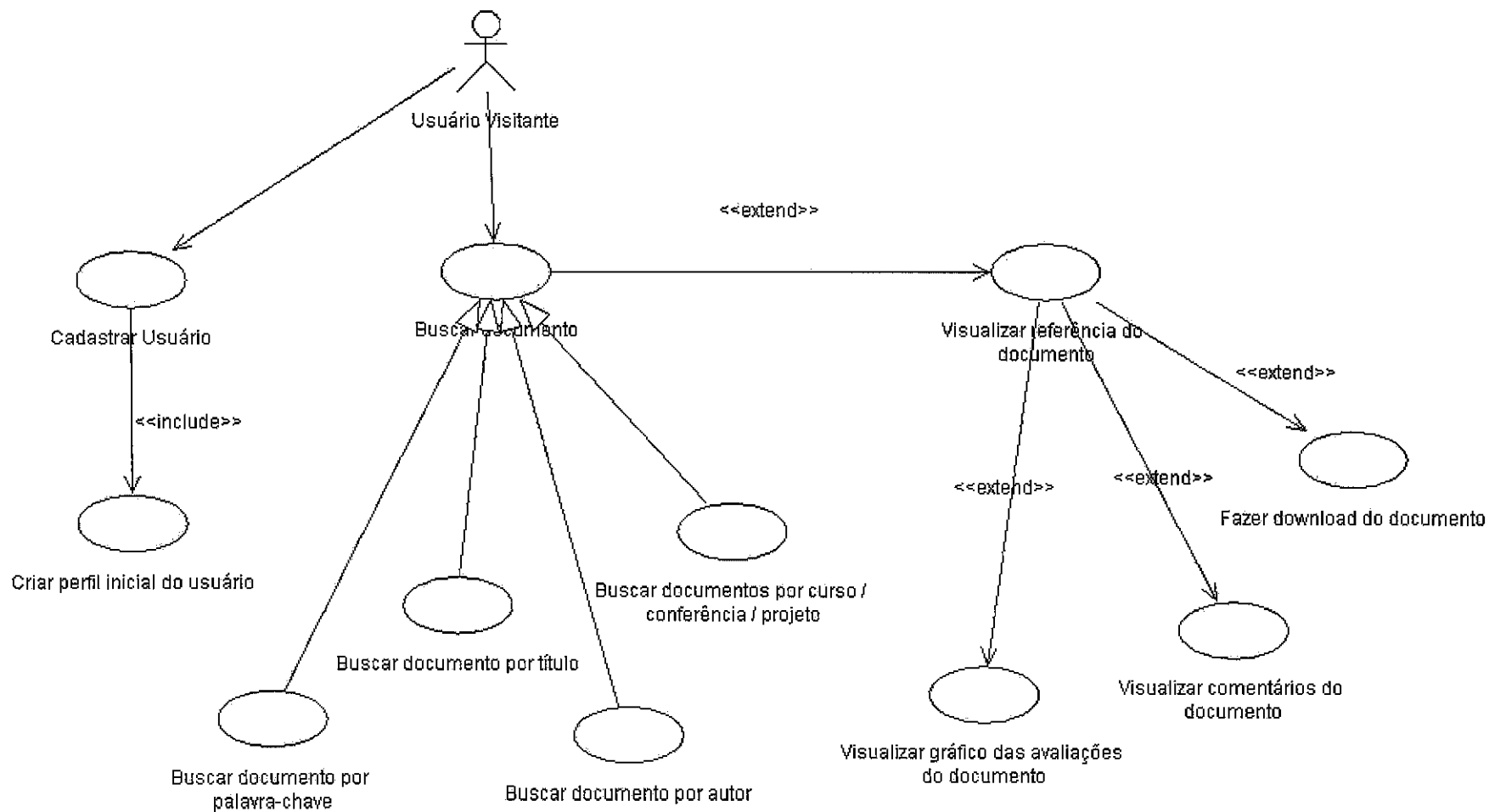
<b>Ator</b>	<b>Descrição</b>	<b>Especialização</b>
Usuário visitante	Qualquer pessoa que utiliza o ambiente aberto da biblioteca para consultar o catálogo.	Usuário cadastrado
Usuário cadastrado	Usuário que se cadastra na biblioteca, tem acesso ao ambiente personalizado e ao sistema de recomendação.	Estudante
		Professor
		Profissional
		Pesquisador
Usuário ADM	Realiza tarefas de manutenção da base de dados.	

### Hierarquia dos atores

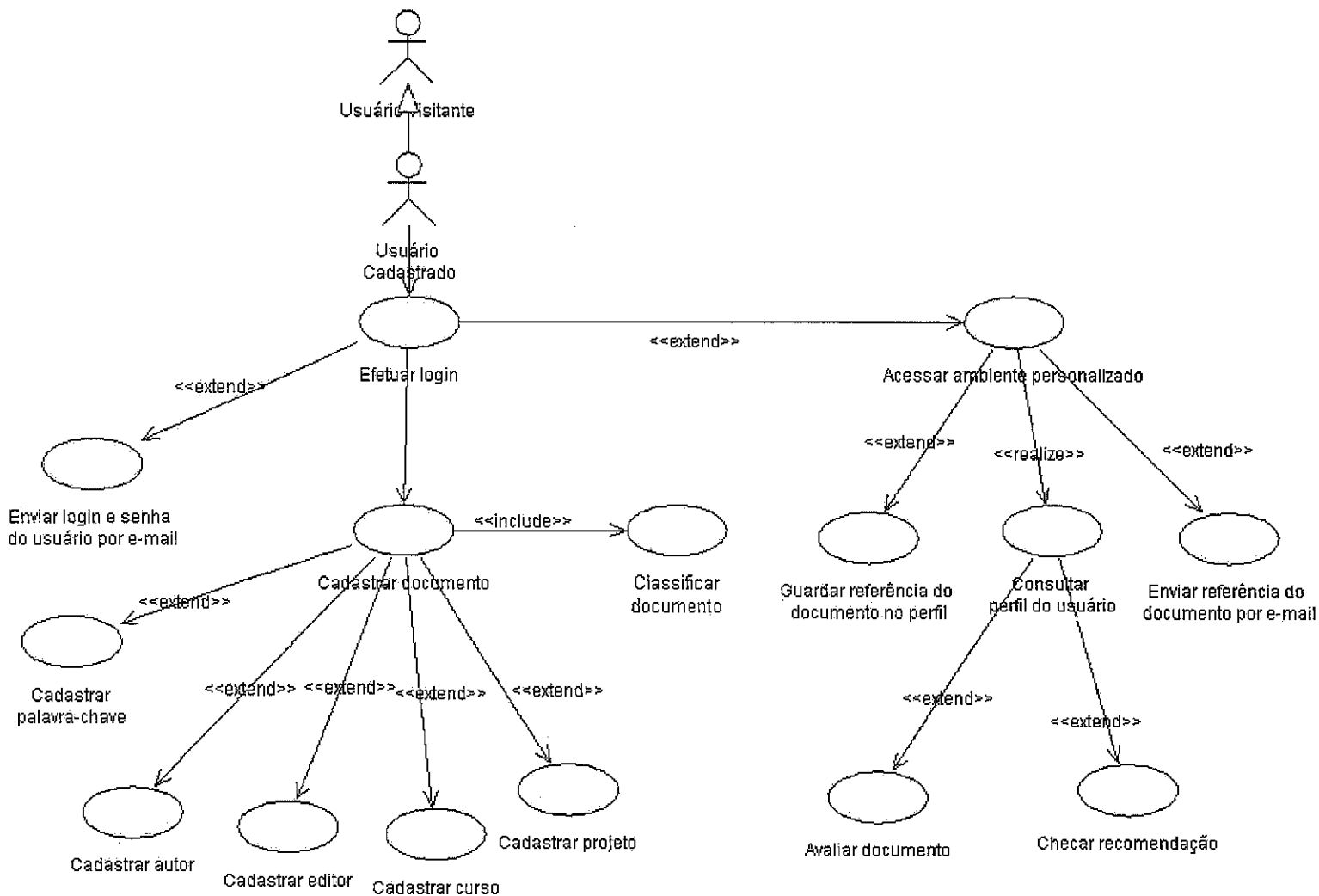




**Diagrama dos casos de uso para o Usuário visitante.**



*Diagrama dos casos de uso para o Usuário cadastrado.*



## *Descrição dos Casos de Uso*

<b>Nome do caso de uso</b>
Buscar documento.
<b>Descrição do caso de uso</b>
Caso de uso abstrato que representa qualquer tipo de busca de documentos realizada na biblioteca.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
<b>Pontos de extensão</b>
Buscar documento por palavra-chave. Buscar documento por título. Buscar documento por autor. Buscar documentos por curso / conferência / projeto.

<b>Nome do caso de uso</b>
Buscar documento por palavra-chave.
<b>Descrição do caso de uso</b>
O usuário especifica uma consulta digitando uma palavra-chave ou um grupo de palavras na caixa de texto.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
1. Usuário solicita busca no acervo bibliográfico. 2. Sistema exibe tela de busca principal. 3. Usuário digita uma ou mais palavras-chave na caixa de texto e pressiona o botão "pesquisar". 4. Sistema realiza a busca de documentos na base de dados. 5. Sistema retorna uma lista de documentos equivalentes a consulta do usuário.
<b>Fluxo alternativo</b>
No passo 3, o usuário pode selecionar uma palavra-chave pré-definida na lista da biblioteca.
<b>Pontos de extensão</b>
Buscar documento por título. Buscar documento por autor. Buscar documentos por curso / conferência / projeto. Visualizar referência do documento.

<b>Nome do caso de uso</b>
Buscar documento por título
<b>Descrição do caso de uso</b>
O usuário especifica a busca de documentos por título.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita busca por título no acervo bibliográfico.</li> <li>2. Sistema exibe tela de busca por título, com campo de texto para digitação e opções para o usuário marcar “começa com”, “contêm o texto” e “exatamente igual”.</li> <li>3. Usuário digita parte do título do documento a ser procurado e pressiona o botão “pesquisar”.</li> <li>4. Sistema realiza a busca de documentos na base de dados.</li> <li>5. Sistema retorna uma lista de documentos equivalentes a consulta do usuário.</li> </ol>
<b>Fluxo alternativo</b>
No passo 3, o usuário pode selecionar se o texto especificado na consulta é referente ao início do título ou o título completo.
<b>Pontos de extensão</b>
Visualizar referência do documento.

<b>Nome do caso de uso</b>
Buscar documento por autor.
<b>Descrição do caso de uso</b>
O usuário especifica a busca de documentos por autor.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita busca por autor no acervo bibliográfico.</li> <li>2. Sistema exibe tela de busca por autor, com uma caixa de texto de digitação e opções para o usuário marcar se a busca deve ser feita por nome ou sobrenome do autor.</li> <li>3. Usuário digita sobrenome do autor a ser procurado e pressiona o botão “pesquisar”.</li> <li>4. Sistema realiza a busca de documentos na base de dados.</li> <li>5. Sistema retorna uma lista de documentos equivalentes a consulta do usuário.</li> </ol>
<b>Fluxo alternativo</b>
No passo 3, o usuário pode selecionar “buscar por nome do autor”, ao invés de último sobrenome.
<b>Pontos de extensão</b>
Visualizar referência do documento
<b>Casos de uso incluídos</b>

<b>Nome do caso de uso</b>
Buscar documento por curso / conferência / projeto.
<b>Descrição do caso de uso</b>
O usuário deseja buscar documentos associados a um curso / conferência / projeto.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita busca por curso / conferência / projeto no acervo bibliográfico.</li> <li>2. Sistema exibe tela de busca com uma lista de opções para o usuário selecionar.</li> <li>3. Usuário seleciona um curso / conferência / projeto na lista e pressiona o botão “pesquisar”.</li> <li>4. Sistema realiza a busca de documentos na base de dados.</li> <li>5. Sistema retorna uma lista de documentos equivalentes a consulta do usuário.</li> </ol>
<b>Fluxo alternativo</b>
<b>Pontos de extensão</b>
Visualizar referência do documento.

<b>Nome do caso de uso</b>
Visualizar referência do documento.
<b>Descrição do caso de uso</b>
Após selecionar um documento, o usuário deseja ver a referência bibliográfica completa.
<b>Atores</b>
Usuário visitante e Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita ver a referência bibliográfica do documento, selecionando o título na lista de documentos retornados na busca.</li> <li>2. Sistema exibe tela de referência bibliográfica do documento.</li> </ol>
<b>Fluxo alternativo</b>
<p>No passo 2, se o documento possuir um arquivo digital, o usuário pode requisitar o download do documento. Neste caso, chamar caso de uso “Fazer download do documento”.</p> <p>No passo 2, se o usuário estiver logado na biblioteca, chamar caso de uso Exibir serviços personalizados.</p>
<b>Pontos de extensão</b>
<p>Fazer download do documento.</p> <p>Visualizar gráfico das avaliações do documento.</p> <p>Visualizar comentários do documento.</p> <p>Guardar documento no perfil.</p> <p>Comentar documento.</p> <p>Enviar documento por e-mail.</p>

<b>Nome do caso de uso</b>
Fazer download do documento.
<b>Descrição do caso de uso</b>
Usuário faz download do PDF referente ao documento selecionado.
<b>Atores</b>
Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário requisita download do documento, selecionando a figura “download” na tela de referência bibliográfica.</li> <li>2. O sistema abre janela do navegador de download de arquivo.</li> <li>3. Usuário seleciona diretório onde o arquivo deve ser salvo e aperta o botão “OK”.</li> <li>4. Sistema confirma final de download do arquivo.</li> </ol>
<b>Fluxo alternativo</b>

<b>Nome do caso de uso</b>
Visualizar comentários do documento
<b>Descrição do caso de uso</b>
Após selecionar um documento, o usuário solicita visualizar comentários feito por outros usuários que já leram o documento.
<b>Atores</b>
Usuário Cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita visualizar os comentários do documento, através do acesso “Comentários” na tela de referência bibliográfica.</li> <li>2. O sistema abre uma nova janela do navegador com a lista de comentários associados ao documento.</li> <li>3. Usuário fecha janela com os comentários.</li> </ol>

<b>Nome do caso de uso</b>
Visualizar gráfico das avaliações do documento
<b>Descrição do caso de uso</b>
Após selecionar um documento, o usuário solicita visualizar avaliações feitas por outros usuários.
<b>Atores</b>
Usuário Cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita visualizar as avaliações relacionadas ao documento, selecionando o acesso “Gráfico das avaliações” na tela de referência bibliográfica.</li> <li>2. O sistema abre uma nova janela do navegador com o gráfico da distribuição percentual das notas dadas ao documento.</li> <li>3. Usuário fecha janela com o gráfico.</li> </ol>

<b>Nome do caso de uso</b>
Cadastrar usuário.
<b>Descrição do caso de uso</b>
O usuário se cadastra na biblioteca para ter acesso aos serviços personalizados.
<b>Atores</b>
Usuário Visitante.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita cadastro na biblioteca, clicando na opção “Cadastro de Usuário” no menu principal da biblioteca.</li> <li>2. Sistema abre nova janela com formulário de cadastro, composto por campos para preenchimento de dados pessoais e uma lista de palavras para seleção de área de interesse.</li> <li>3. Usuário preenche os dados no formulário e clica no botão “enviar”.</li> <li>4. Sistema verifica preenchimento de dados obrigatórios e cadastra as informações na base de dados.</li> <li>5. Sistema envia e-mail de confirmação com login e senha do usuário.</li> <li>6. Sistema exibe tela de confirmação de cadastro.</li> </ol>
<b>Fluxo alternativo</b>
No passo 4, se o usuário não fornecer algum dado de preenchimento obrigatório, o sistema exibe uma mensagem de alerta em janela pop-up.
<b>Ponto de extensão</b>
Consultar perfil do usuário. Buscar documento no acervo.
<b>Casos de uso incluídos</b>
Caso de uso “Criar perfil inicial do usuário”.

<b>Nome do caso de uso</b>
Criar perfil inicial do usuário
<b>Descrição do caso de uso</b>
Definir um perfil inicial com dados preenchidos no cadastro do usuário.
<b>Atores</b>
Usuário visitante.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Sistema insere dados pessoais na tabela Usuário na base de dados.</li> <li>2. Sistema verifica se foram selecionadas palavras-chaves no cadastro do usuários.</li> <li>3. Sistema atualiza a tabela UsuarioPalavraChave.</li> <li>4. Sistema verifica se usuário selecionou um projeto de pesquisa.</li> <li>5. Sistema atualiza a tabela UsuarioProjeto.</li> </ol>
<b>Pontos de extensão</b>

<b>Nome do caso de uso</b>
Efetuar login na biblioteca.
<b>Descrição do caso de uso</b>
O usuário faz o login na biblioteca para acessar o perfil ou acessar os serviços personalizados.
<b>Atores</b>
Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita login clicando na opção “Login” no menu principal da biblioteca.</li> <li>2. Sistema exibe tela de login, com campos para preenchimento de login e senha.</li> <li>3. Usuário preenche dados do formulário e clica no botão “OK”.</li> <li>4. Sistema emite aviso de sucesso e exibe opções de navegação para o acervo bibliográfico ou perfil do usuário.</li> </ol>
<b>Fluxo alternativo</b>
No passo 2, se o usuário tiver esquecido o login ou senha, pode solicitar o envio desses dados por e-mail. Neste caso, chamar “Enviar login e senha do usuário por e-mail”.
No passo 3, se o usuário preencher login ou senha inválidos, o sistema emitido aviso de erro.
<b>Pontos de extensão</b>
<p>Buscar documento.</p> <p>Consultar perfil do usuário.</p>

<b>Nome do caso de uso</b>
Enviar login e senha do usuário por e-mail.
<b>Descrição do caso de uso</b>
O usuário requisita envio de e-mail com dados de login e senha.
<b>Atores</b>
Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita dados de login e senha de seu cadastro na biblioteca.</li> <li>2. Sistema exibe tela com formulário para preenchimento de e-mail.</li> <li>3. Usuário informa o endereço de e-mail e clica no botão “enviar”.</li> <li>4. Sistema envia e-mail com dados de login.</li> <li>5. Sistema emite aviso de envio de e-mail com sucesso.</li> </ol>
<b>Fluxo alternativo</b>
No passo 3, se o usuário preencher um e-mail inválido, o sistema emite aviso de erro.
<b>Pontos de extensão</b>



<b>Nome do caso de uso</b>
Guardar referência do documento no perfil.
<b>Descrição do caso de uso</b>
Após ver a referência bibliográfica do documento, o usuário solicita que a referência seja armazenada no seu perfil.
<b>Atores</b>
Usuário cadastrado
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita armazenamento da referência do documento em seu perfil. O usuário seleciona a figura “Guardar documento” na página de referência bibliográfica.</li> <li>2. Sistema atualiza a tabela UsuarioAcervo onde fica armazenado as referências dos documentos associados ao usuário.</li> <li>3. Sistema emite aviso de sucesso.</li> </ol>
<b>Casos de uso incluídos</b>

<b>Nome do caso de uso</b>
Avaliar documento.
<b>Descrição do caso de uso</b>
O usuário avalia o documento em escala de 1 (ruim) a 5 (excelente) e faz comentários sobre o mesmo.
<b>Atores</b>
Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita avaliar documento na página de referência bibliográfica. O usuário seleciona a opção “Comentar Documento”.</li> <li>2. Sistema exibe formulário de avaliação de documento, com opções mutuamente exclusivas com notas de 1 a 5 e um campo de digitação de múltiplas linhas para o comentário.</li> <li>3. Usuário preenche o formulário de avaliação e clica no botão “enviar”.</li> <li>4. Sistema verifica o preenchimento do campo de avaliação.</li> <li>5. Sistema atualiza a tabela UsuarioAcervo.</li> <li>6. Sistema exibe tela e confirmação de sucesso na avaliação.</li> </ol>
<b>Fluxo alternativo</b>
No passo 4, se o usuário não selecionou uma nota para avaliação, o sistema emite aviso de que o campo é de preenchimento obrigatório.
<b>Casos de uso incluídos</b>
Guardar documento no perfil.

<b>Nome do caso de uso</b>
Checar recomendação.
<b>Descrição do caso de uso</b>
Usuário verifica recomendações de documentos no perfil do usuário.
<b>Atores</b>
Usuário cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. O usuário seleciona a opção “Checar recomendações” no perfil do usuário.</li> <li>2. Sistema exibe tela com lista de documentos. O título é um acesso a tela de referência bibliográfica. Ao lado de cada documento tem opções para avaliação da recomendação como satisfatória ou insatisfatória.</li> <li>3. Usuário avalia as recomendações e clica no botão “enviar”.</li> <li>4. Sistema exibe tela do perfil do usuário.</li> </ol>
<b>Pontos de extensão</b>
Caso de uso Visualizar referência bibliográfica.

<b>Nome do caso de uso</b>
Cadastrar documento.
<b>Descrição do caso de uso</b>
Usuário cadastro um novo documento no acervo da biblioteca.
<b>Atores</b>
Usuário Cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita cadastro de documento, selecionando a opção “Cadastro de documento” no menu principal da biblioteca.</li> <li>2. Sistema abre nova janela do navegador com formulário de cadastro de documento, composto por campos de preenchimento da referência bibliográfica, opção para classificação por palavra-chave e opções para associar documento a um curso e/ou projeto.</li> <li>3. Usuário preenche formulário de cadastro e aperta o botão “enviar”.</li> <li>4. Sistema verifica preenchimento de campos obrigatórios para o cadastro.</li> <li>5. Sistema emite aviso de sucesso no cadastro do documento.</li> </ol>
<b>Fluxo alternativo</b>
Se no passo 4, o usuário não tiver preenchido campos obrigatórios, o sistema emite um aviso em janela pop-up.
<b>Pontos de extensão</b>
Cadastrar palavra-chave.
Cadastrar autor / editor.
Cadastrar curso / projeto.

<b>Nome do caso de uso</b>
Cadastrar palavra-chave
<b>Descrição do caso de uso</b>
Usuário cadastra uma nova palavra na lista de termos da biblioteca.
<b>Atores</b>
Usuário Cadastrado
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita cadastro de palavra-chave, selecionando a opção “Novo” no formulário de cadastro de documento.</li> <li>2. Sistema abre nova janela do navegador com formulário de cadastro da palavra-chave.</li> <li>3. Usuário preenche formulário e aperta o botão “OK”.</li> <li>4. Sistema fecha a janela aberta e retorna para o formulário de cadastro do documento.</li> <li>5. Sistema atualiza a lista de palavras-chaves no formulário.</li> </ol>
<b>Fluxo alternativo</b>
Se no passo 3, o usuário não tiver digitado uma palavra-chave, o sistema emite um aviso na mesma janela.
<b>Pontos de extensão</b>

<b>Nome do caso de uso</b>
Cadastrar autor / editor.
<b>Descrição do caso de uso</b>
Usuário cadastra um novo autor na biblioteca.
<b>Atores</b>
Usuário Cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita cadastro de autor / editor, selecionando a opção “Novo” no formulário de cadastro de documento.</li> <li>2. Sistema abre nova janela do navegador com formulário de cadastro de autor / editor.</li> <li>3. Usuário preenche formulário e aperta o botão “OK”.</li> <li>4. Sistema fecha a janela aberta e retorna para o formulário de cadastro do documento.</li> <li>5. Sistema atualiza a lista de autores / editores no formulário.</li> </ol>
<b>Fluxo alternativo</b>
Se no passo 3 o usuário não tiver digitado nome e sobrenome do autor / nome do editor, o sistema emite um aviso na mesma janela.
<b>Pontos de extensão</b>

<b>Nome do caso de uso</b>
Cadastrar curso / projeto.
<b>Descrição do caso de uso</b>
Usuário cadastra um novo curso / projeto na biblioteca.
<b>Atores</b>
Usuário Cadastrado.
<b>Fluxo principal</b>
<ol style="list-style-type: none"> <li>1. Usuário solicita cadastro de curso / projeto, clicando no link “Novo” no formulário de cadastro de documento.</li> <li>2. Sistema abre nova janela do navegador com formulário de cadastro de curso / projeto.</li> <li>3. Usuário preenche formulário e aperta o botão “OK”.</li> <li>4. Sistema fecha a janela aberta e retorna para o formulário de cadastro do documento.</li> </ol>
<b>Fluxo alternativo</b>
Se no passo 3 o usuário não tiver digitado os dados obrigatórios do curso/projeto no formulário, o sistema emite um aviso na mesma janela.
<b>Pontos de extensão</b>
<b>Casos de uso incluídos</b>