



**UNIVERSIDADE DE BRASÍLIA**  
**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E**  
**CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO**  
**DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO**

**ABORDAGENS DE RECOMENDAÇÃO PARA A RECUPERAÇÃO DE**  
**PERFIS: UMA PROPOSTA DE MODELO**

**TIAGO MIRANDA MARQUES**

Brasília  
2007

**ABORDAGENS DE RECOMENDAÇÃO PARA A RECUPERAÇÃO DE  
PERFIS: UMA PROPOSTA DE MODELO**

TIAGO MIRANDA MARQUES

Dissertação apresentada à banca examinadora como requisito parcial à obtenção do Título de Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília.

Orientadora: Doutora Marisa Bräscher  
Basílio Medeiros

BRASÍLIA

2007

## FOLHA DE APROVAÇÃO

**Título:** *“Abordagens de Recomendação para a Recuperação de Perfis: uma proposta de modelo”*

**Autor:** Tiago Miranda Marques

**Área de concentração:** Transferência da Informação

**Linha de Pesquisa:** Arquitetura da Informação

Dissertação submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação e Documentação, da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Dissertação aprovada em: 30 de maio de 2007

Aprovado por:

---

Profa. Dra. MARISA BRÄSCHER BASÍLIO MEDEIROS  
Presidente - Orientador

---

Profa. Dra. SUZANA PINHEIRO MULLER  
Membro interno – (UnB/PPGCinf)

---

Prof. Dr. MARCELO LADEIRA  
Membro externo – (UnB/CIC)

---

Prof. Dr. EDGARD COSTA OLIVEIRA  
Membro suplente – (UnB/CID)

Dedico esta obra a Allysson Leite.

## AGRADECIMENTOS

Primeiramente, agradeço a meu antigo gerente Jair de Menezes. Seu apoio inicial foi fundamental para garantir minha participação neste programa de mestrado. Agradeço a meu amigo e antigo colega de trabalho, Osmar Ribeiro Torres, por toda colaboração e apoio durante as quedas do Control-M nas madrugadas, e a entrada da GSM que tanto nos tomou o sono e a saúde — sem sua participativa colaboração não teria conseguido seguir em frente. Agradeço a meu segundo antigo gerente, Luiz Sávio, por permitir a continuidade do cronograma elaborado pela gerência anterior. E agradeço à minha atual coordenadora, Sônia Cazzola, por entender como ninguém esses difíceis momentos finais, e por todo o auxílio prestado, tanto neste projeto como em minha vida particular, incentivando meu progresso e me consolando em minhas derrotas.

Um agradecimento especial é carinhosamente dedicado à minha orientadora, Marisa Bräscher, não só pela competência, dedicação e seriedade, mas por harmonizar todas estas qualidades com amor e respeito, encorajando-me desde a fase de aluno especial, engrandecendo meu projeto e acreditando no meu trabalho. Deixo registrado os meus sinceros reconhecimento e gratidão.

Acima de tudo, agradeço a Deus por cuidar de mim enquanto eu ando distraído. E a meus intercessores, Belchior, Baltazar e Gaspar.

“Não importa o que tiraram de você. O que importa é o que você vai fazer com o que sobrou”.

*Autor Anônimo*

## Resumo

Este trabalho propõe analisar o uso de recomendações na recuperação de informação no domínio de perfis de usuário. Espera-se demonstrar que o uso de abordagens de recomendação é um critério essencial para a identificação de documentos relevantes em vários cenários de recuperação neste domínio. As abordagens de recomendação surgiram da necessidade de se prever os documentos mais relevantes para o usuário, problema central da recuperação de informação. As recomendações servem para auxiliar no processo de receber ou fornecer indicações, e um sistema de recomendação é um sistema de informação que auxilia o usuário a recuperar informação através da previsão de seus interesses. Nesta pesquisa exploratória, foi realizado um estudo do sistema Currículo Lattes identificando casos onde o sistema não oferece respostas com a recuperação de informação tradicional, e que poderiam ser solucionados utilizando as abordagens de recomendação. Generalizando essas dificuldades, foi criado um modelo geral de recuperação de perfis de usuário, que pode ser aplicado na recuperação de perfis em qualquer contexto, não só na recuperação de currículos. Um sistema de recuperação de perfis de usuário em *sites* de relacionamento foi desenvolvido com base no modelo geral, a fim de validar esse modelo, que depois foi transposto para o Currículo Lattes. Com os resultados alcançados, espera-se não só contribuir com os conhecimentos da academia na arte da recuperação da informação, mas ampliar os recursos para satisfação da necessidade do usuário com o estudo de novos procedimentos, e, também, consolidar a recomendação como estratégia importante da recuperação de informação, com vistas a influenciar o desenvolvimento, sob a ótica dessa estratégia, de novas ferramentas de maior qualidade.

Palavras-chave: Abordagens de recomendação; sistemas de recomendação; perfil de usuário; recuperação de informação; recuperação de perfis de usuário; modelo de recomendação de perfis de usuário.

## **Abstract**

In this paper the use of recommendation for information retrieval in user profiles domain will be analyzed. The recommendation approaches are expected to be an essential criterion to the identification of relevant documents in several retrieval scenarios in this domain. The recommendation approaches emerged from the need for predicting the most relevant documents to the user, the main issue of information retrieval. Recommendation is useful in the process of receiving or providing advice, and a recommendation system is an information system that helps users retrieve information through the precognition of their interests. In this research, Curriculum Lattes' retrieval system has been analyzed and, in many cases of its traditional information retrieval, it has been identified as not being able to answer some questions that would be solved using recommendation approaches. These examples were generalized in order to create a general model of user profiles retrieval that was applied at profile recovering in any context, not only at curriculum retrieval. A profile retrieval system based on the general model was developed and applied to a dating *website* to validate the model and then the system's behavior will be transposed to Curriculum Lattes' system. The results achieved are expected to help magnify the academy's knowledge on the information retrieval art, increase resources for the satisfaction of user needs through the research of new procedures, and also consolidate recommendation as an important tactic of information retrieval, influencing the development of new tools with higher quality under these approach goals.

**Key Words:** Recommendation approaches; recommender systems; user profile; information retrieval; user profiles retrieval; user profiles recommendation model.

# Sumário

Resumo .....	7
Abstract .....	8
Lista de Siglas e Abreviaturas.....	11
Lista de Tabelas .....	12
Lista de Figuras.....	13
Introdução .....	15
1. Especificação .....	18
1.1. Objetivo Geral.....	18
1.1.1. Objetivos Específicos.....	18
1.2. Problema .....	18
1.2.1. Definição do Problema .....	18
1.2.2. Escopo do Problema .....	19
1.3. Justificativa .....	19
2. Sistemas de Recomendação .....	24
2.1. Recuperação de Informação.....	24
2.1.1. Modelos de Recuperação da Informação .....	27
2.1.2. Processo de Recuperação da Informação.....	30
2.2. Recomendações.....	36
2.2.1. Abordagens e Técnicas .....	37
2.2.2. Relevância e <i>Feedback</i> do Usuário .....	52
2.3. Redes Sociais .....	53
2.3.1. Importância das Redes Sociais.....	55
3. Plataforma Lattes .....	56
3.1. Histórico.....	56
3.2. Funcionalidades .....	59
3.2.1. Recuperação de Currículos .....	60
3.2.2. Outras Funcionalidades.....	62
3.3. Carências do Sistema Currículo Lattes .....	63
4. Metodologia da Pesquisa .....	65
5. Desenvolvimento da Pesquisa.....	69
5.1. Modelo Geral de Recomendação de Perfis de Usuário.....	73
5.1.1. Domínio e Indicadores .....	77
5.1.2. Modelo do Sistema Proposto .....	96
5.1.3. Origem dos Dados.....	97
5.1.4. Interface de Consulta .....	97
5.2. Aplicando o Modelo Geral em um <i>Site</i> de Relacionamentos .....	97
5.2.1. Domínio e Indicadores .....	98
5.2.2. Sistema Proposto.....	101
5.2.3. Origem dos Dados.....	103
5.2.4. Interface de Consulta .....	103
5.2.5. Implementação .....	103
5.2.6. Execução da Pesquisa Experimental.....	116
6. Análise dos Dados.....	119
6.1. Análise dos Módulos.....	119
6.2. Análise dos Resultados .....	132
6.3. Transposição do Sistema Implementado para o Currículo Lattes.....	133
6.3.1. Identificação dos Indicadores .....	133

6.3.2.	Telas .....	135
6.3.3.	Resultados Esperados.....	146
6.4.	Considerações Finais .....	146
7.	Conclusão.....	148
	Referências.....	151
	Anexo I – Estrutura do Perfil .....	154
	Anexo II – Termos de Uso do Sistema .....	166
	Anexo III – Sobre.....	170
	Anexo IV – Código-fonte do Sistema 2coracoes.....	171
	Anexo V – Estrutura do Banco de Dados .....	188
	Anexo VI – Recomendação do <i>Site</i> .....	192
	Glossário .....	193

## **Lista de Siglas e Abreviaturas**

AVG – Avaliação geral

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico.

FC – Filtragem Colaborativa

RI – Recuperação da informação

RR – *Ranking* de Recomendação

SGBD – Sistema Gerenciador de Banco de Dados

SQL – Linguagem de Consulta Estruturada (*Structured Query Language*)

SRI – Sistema de recuperação de informação

UF – Unidade Federativa

USR9 – Usuário 9

## Lista de Tabelas

Tabela 1 - Comparação entre recuperação de dados e recuperação de informação segundo Rijsbergen .....	25
Tabela 2- Variáveis qualitativas extraídas do perfil do usuário .....	78
Tabela 3 - Pesos associados aos valores qualitativos de nível acadêmico .....	79
Tabela 4 - Pesos associados aos valores qualitativos de nível profissional .....	80
Tabela 5 - Variáveis quantitativas extraídas do perfil do usuário .....	81
Tabela 6 - Valores de currículos acadêmicos (não normalizados).....	82
Tabela 7 - Variáveis normalizadas para calcular o RR.....	83
Tabela 8 - Resultado do Ranking de Recomendação.....	84
Tabela 9 - Áreas de interesse de currículos da Ciência da Computação.....	85
Tabela 10 - Usuários e suas áreas de interesse.....	86
Tabela 11 - Resultado do Ranking de Recomendação por área de interesse .....	88
Tabela 12 - Identificação dos usuários com opinião mais relevante.....	89
Tabela 13 - Dados de $RR_{\text{área}}$ do usuário USR9 .....	90
Tabela 14 - Grupos de opinião para o sistema de recomendação de currículos .....	91
Tabela 15 - Avaliação do USR9 pelos usuários do sistema.....	91
Tabela 16 - Agrupamento dos usuários de acordo com o grupo de opinião .....	92
Tabela 17 - Resultado da avaliação dos usuários utilizando FC pura.....	92
Tabela 18 - Resultado da avaliação dos usuários.....	93
Tabela 19 – $RR_{\text{total}}$ de cada usuário .....	94
Tabela 20 - Pontuação dos usuários do grupo GRP1 .....	94
Tabela 21 - Pontuação normalizada dos usuários do grupo GRP1 .....	94
Tabela 22 - Variáveis qualitativas.....	98
Tabela 23 - Variáveis quantitativas.....	100
Tabela 24 - Valores não normalizados do cálculo de $RR_{\text{htfcopas}}$ .....	113
Tabela 25 – Cálculo do AVG para futura normalização de $RR_{\text{htfcopas}}$ .....	113
Tabela 26 – Valores normalizados para o cálculo de $RR_{\text{htfcopas}}$ .....	115

## Lista de Figuras

Figura 1 - Arquitetura Básica de um Sistema de Recomendação .....	24
Figura 2 - Classificação de técnicas de recuperação de informação .....	29
Figura 3 - Representação de um conjunto de documentos.....	30
Figura 4 - Representação de estratégias de busca com três descritores diferentes .....	32
Figura 5 - Precisão e revocação da informação recuperada, e registros adjacentes....	33
Figura 6 - Relação perda/ruído.....	34
Figura 7 - Relação perda/ruído com recomendações .....	35
Figura 8 - Google Toolbar, com propriedades avançadas ativas .....	38
Figura 9 - Abordagem baseada em conhecimento (regra do sistema) .....	40
Figura 10 - Abordagem baseada em conhecimento ( <i>data mining</i> ) .....	40
Figura 11 - Resultado da busca por "recuperação da informação" no Vivísimo .....	42
Figura 12 - Resultado da busca por " <i>information retrieval</i> " no Vivísimo .....	42
Figura 13 - Resultado da busca por " <i>information retrieval</i> " no Google .....	43
Figura 14 - Convite para se cadastrar e receber recomendações na Amazon.com ....	44
Figura 15 - Pesquisa por "buscape" no Google.....	46
Figura 16 - Implantação da abordagem de recomendação social no site Buscapé .....	48
Figura 17 - Alguns dos principais clientes do Buscapé .....	49
Figura 18 - Página inicial do site Buscapé.....	50
Figura 19 - Filtros disponíveis para a busca avançada do Currículo Lattes .....	61
Figura 20 - Depoimentos de empresas sobre o Buscapé.....	72
Figura 21 - Áreas de atuação de um currículo do sistema Currículo Lattes .....	74
Figura 22 - Equação para normalizar variáveis no RR .....	76
Figura 23 - Equação do Ranking de Recomendação (RR) .....	76
Figura 24 - Representação do cálculo do $RR_{\text{área}}$ .....	87
Figura 25 - Equação do cálculo da pontuação .....	93
Figura 26 - Interface de consulta "busca pré-definida".....	106
Figura 27 - Interface de consulta "busca rápida" .....	107
Figura 28 - Interface de consulta "busca por apelido" .....	107
Figura 29 - Tela inicial de acesso ao sistema 2coracoes.....	109
Figura 30 - Fluxo de execução da pesquisa experimental .....	117
Figura 31 - Cadastro (página 1 de 3) .....	120
Figura 32 - Observação do campo "nome" .....	121
Figura 33 - Funcionalidade do campo "localização" .....	123
Figura 34 - Seleção de "Exterior" na localização .....	123
Figura 35 - Cadastro (página 2 de 3) .....	124
Figura 36 - Cadastro (página 3 de 3) - parte I.....	126
Figura 37 - Cadastro (página 3 de 3) - parte II .....	127
Figura 38 - Tela de seleção de interesses.....	128
Figura 39 - Tela principal do sistema 2coracoes .....	130
Figura 40 - Resultado apresentado pela abordagem <i>Cross-Sell</i> .....	131
Figura 41 - Busca pelo assunto "inteligência artificial".....	137
Figura 42 - Redução do resultado na mesma consulta por assunto .....	138
Figura 43 - Opções de visualização .....	139
Figura 44 - Mudança nos resultados da busca por assunto .....	140
Figura 45 - Resultados da consulta por "recuperação da informação" .....	141
Figura 46 - Primeiro currículo da consulta por "recuperação da informação" .....	142
Figura 47 - Uso da abordagem <i>Cross-Sell</i> no Currículo Lattes .....	143

Figura 48 - Funcionalidade para avaliação de currículos semelhantes.....	145
Figura 49 - Tela inicial após autenticação no Currículo Lattes (parte superior).....	146
Figura 50 - Tabelas do banco de dados.....	188
Figura 51 - Campos das tabelas - parte I.....	189
Figura 52 - Campos das tabelas - parte II .....	190
Figura 53 - Links do menu superior.....	192
Figura 54 - Indicador de recomendação do <i>site</i> .....	192

## Introdução

A recuperação de informação é fortemente voltada para algoritmos que trabalham o conteúdo dos documentos. Para alcançar melhores resultados, esforços vão do *Data Mining* ao *Text Mining*, dos Tesouros às Ontologias. O objetivo é fazer o computador se tornar o que o cérebro é, com todas as suas relações semânticas e com toda a sua subjetividade. Essa realidade, porém, está longe de ser alcançada.

Para fugir da dificuldade da implementação de soluções no domínio da Inteligência Artificial, a Ciência da Informação faz uso da recomendação de informação, uma idéia que não é nova, mas que não foi bem explorada no contexto da Tecnologia da Informação.

No meio acadêmico, por exemplo, as recomendações estão presentes em várias situações: para realizar uma escolha de um item em uma biblioteca há recomendações pessoais, críticas de livros e listas de leitura. Recomendações implícitas também podem ser consideradas na avaliação de determinado item pelo usuário, como a procedência desse item ou mesmo sua presença em uma biblioteca especializada (NICHOLS et al., 2004).

No cotidiano, lidamos com informações que podem ou não ter relevância, e, diante disso, recomendações das mais diversas são realizadas. É bastante comum recomendar um filme ou livro a um amigo. A recomendação de informação já vem sendo praticada muito antes de surgir o tratamento automático da informação.

Em sistemas computacionais, a existência de muitas informações ao dispor do usuário acaba ocasionando uma sobrecarga de informação (FISCHER e WIVES, 2005). Os sistemas de recomendação surgiram em uma tentativa de contornar esse problema. São programas que tentam prever itens de informação — sejam eles filmes, músicas, livros, notícias, *sítes*, imagens, etc. — em que um usuário possa estar interessado, normalmente fazendo uso do perfil do usuário ou de um grupo de usuários.

Com isso, os sistemas de recomendação auxiliam no processo de realizar e receber recomendações de outras pessoas.

Podemos definir que um sistema de recomendação é um sistema de informação que auxilia o usuário a recuperar informação através da previsão de seus interesses.

Os benefícios dos sistemas de recomendação podem ser conferidos em vários ambientes, tais como sistemas P2P<sup>1</sup>, em *sites* de rádios, fóruns eletrônicos, sistemas de informações de livros e filmes, sistemas bancários, *sites* de comércio eletrônico, e, ainda pouco difundidos, há também sistemas eletrônicos de guias de TV e sistemas para a recomendação de notícias. Não disponível para o público em geral, foi construído em meio acadêmico um sistema para a recuperação de artigos científicos (TORRES JR., 2004) que se beneficia ativamente das abordagens de recomendação. Sempre que o universo é grande — ou desconhecido — e quando há o registro do comportamento dos usuários, as vantagens no uso de sistemas de recomendação são fortemente percebidas.

Ainda que vagarosamente, os sistemas de recomendação estão ganhando espaço e sendo implementados em outros domínios. Nos domínios pouco explorados com as técnicas desses sistemas, podemos citar a recuperação de imagens e a recuperação de perfis de usuário.

Boa parte dos sistemas de recuperação de imagens é voltada para descritores da imagem e seus metadados (por exemplo, no padrão MPEG-7); esses sistemas podem usar ontologias, e há outros mais voltados para a decomposição de padrões da imagem — histograma, composição de cores, limites e bordas. Mas não há um sistema que utilize a abordagem de recomendação social<sup>2</sup> para a recuperação de imagens.

Sistemas que armazenam perfis de usuário se apóiam predominantemente no modelo booleano de recuperação de informação,

---

<sup>1</sup> P2P ou *Peer-to-Peer* (ponto-a-ponto) é uma tecnologia que possibilita a distribuição de arquivos em rede e que tem como característica permitir o acesso de qualquer usuário dessa rede a um nó (peer), ou seja, a outro usuário de forma direta.

<sup>2</sup> Veremos adiante quais são as abordagens de recomendação, e detalharemos cada uma delas.

onde é possível selecionar, por exemplo, usuários de determinada região que tenham uma específica formação acadêmica. Quando do retorno dos resultados, há uma avalanche de perfis que nem sempre deixa claro qual o critério de ordenação dos resultados. Neste contexto as abordagens de recomendação não são aproveitadas.

A idéia desse projeto é analisar os benefícios das abordagens de recomendação na recuperação de informação em um domínio onde ela tenha sido pouco explorada. Dentre os domínios citados, recuperação de imagens e recuperação de perfis de usuário, a segunda alternativa foi a escolhida para o projeto pela maior facilidade de generalização — a recuperação de perfis de usuário pode ser utilizada em diferentes sistemas de informação de recursos humanos. Pretende-se, como produto deste trabalho, propor um modelo genérico de recomendação de perfis de usuário.

O modelo proposto utilizará estratégias de recomendação. Espera-se demonstrar que os benefícios da recomendação de informação para a recuperação de informação podem ser aproveitados em um novo domínio, e, com isso, encorajar pesquisadores e arquitetos de sistema a utilizarem as abordagens de recomendação na construção de novos modelos de sistemas de informação que melhor atendam às necessidades do usuário, seja em domínios onde a utilização já mostrou vantagens, seja em domínios onde as abordagens de recomendação ainda não foram aplicadas.

# **1. Especificação**

## **1.1. Objetivo Geral**

Propor um modelo de recuperação de perfis de usuário<sup>3</sup> baseado em abordagens de recomendação.

### **1.1.1. Objetivos Específicos**

Analisar o sistema Currículo Lattes, identificando necessidades na funcionalidade de recuperação de informação que possam ser tratadas com a utilização de estratégias de recomendação. Criar um modelo geral de recomendação de perfis de usuário a partir dessas necessidades. Adequar o modelo geral a um domínio de recuperação de perfis de usuário, diferente da recuperação de currículos. Transpor o modelo de sistema de recomendação para o ambiente Lattes.

Através desse estudo, espera-se ter como produto um modelo geral de recomendação de perfis de usuário.

Espera-se obter um levantamento sistemático sobre as abordagens de recomendação em recuperação da informação que reúna um referencial teórico e metodológico no qual essas abordagens se apóiam.

Com esse projeto, espera-se que os benefícios da utilização de recomendações demonstrem eficiência e robustez, o suficiente para influenciar desenvolvedores a aplicar essa abordagem em domínios ainda não explorados.

## **1.2. Problema**

### **1.2.1. Definição do Problema**

O problema desta pesquisa é mostrar que a busca de perfis de usuário em sistemas de informação apresenta perguntas que não podem ser

---

<sup>3</sup> Perfil de usuário, no contexto deste trabalho, refere-se ao registro dos dados sobre cada usuário em um sistema de informações. Cada registro é considerado um documento, que deve ser recuperado de acordo com a necessidade do usuário.

respondidas simplesmente com o uso do modelo lógico de recuperação da informação baseado em álgebra booleana, senão com a utilização de abordagens de recomendação; e, partindo das carências identificadas em um sistema real e amplamente conhecido, propor um modelo geral baseado em recomendações de recuperação de perfis de usuário, que possa ser aplicado na recuperação de perfis de usuário em domínios distintos desse sistema real e nele, inclusive.

## **1.2.2. Escopo do Problema**

O presente projeto se restringirá à criação do modelo geral baseado em recomendações de recuperação de perfis de usuário, à validação de tal modelo na recuperação de perfis de usuário em um domínio diferente da recuperação de currículos e à transposição dessa aplicação para o sistema original de recuperação de currículos. Embora haja diversos domínios onde o modelo geral aqui proposto possa ser aplicado, não há como validar todos, não sendo esta atividade, portanto, objeto de análise no presente trabalho.

## **1.3. Justificativa**

Apesar de não apresentar respostas a diversas perguntas na recuperação de informação, o modelo de busca baseada em álgebra booleana sem dúvida apresenta um resultado melhor do que o apresentado pela busca manual. Por isso, e por sua simplicidade e facilidade de implementação, é um modelo popular e utilizado na maioria dos sistemas comerciais.

Um dos objetivos da recuperação de informação é atender à necessidade dos usuários. Certo de que os usuários são fábricas de sonhos<sup>4</sup>, nem todos esses desejos podem ser atendidos. Mas há determinadas situações que, conforme apresentadas, podem ser identificadas e solucionadas com estratégias de recuperação que estão em

---

<sup>4</sup> "Usuários são fábricas de sonhos", divertida expressão citada por Vitor Almeida, coordenador-geral de Tecnologia e Segurança da Informação da Receita Federal do Brasil (2006).

constante estudo, aperfeiçoamento e evolução pelos profissionais das Ciências da Informação e Computação.

Podemos acompanhar no mercado a crescente demanda por sistemas de apoio à decisão e a carência de profissionais com conhecimentos em *Data Warehouse*, *Balanced Scorecards* ou outras soluções que estão se tornando cada vez mais populares. A Inteligência Artificial está sendo constantemente utilizada para identificar tendências, padrões e apontar soluções. O foco das pesquisas é habitualmente voltado para o produto final: o resultado das buscas, a tendência do mercado, a solução para a sua empresa. Mas qual era mesmo a origem do problema?

O problema em recuperação de informação nunca foge à regra: a necessidade do usuário. Várias alternativas estão disponíveis para solucionar os vários problemas apresentados. E sempre uma delas parece resolver melhor determinada necessidade. Mas como a fábrica de sonhos nunca pára, outros problemas surgem e soluções antes adotadas não são agora capazes de alcançar o desejo do usuário.

Então se percebe que uma parte do que eles sonham tem solução com uma atividade simples: poderia ser resolvido no dia-a-dia com a recomendação de informação. O problema é social. O foco é no usuário e no contexto em que ele está inserido. Quem nunca leu a coluna do jornal para escolher o cinema no fim-de-semana?

As abordagens de recomendação em recuperação de informação — campo de estudo cujo interesse pela comunidade de Ciência da Informação está em ascensão — surgem, enfim, como mais uma estratégia para satisfazer as necessidades do usuário. Também na Ciência da Computação este campo de estudo está sendo pesquisado como parte da Inteligência Artificial, sendo seus algoritmos e técnicas conhecidos como filtragem colaborativa e filtragem baseada no conteúdo.

Alguns domínios, citados na introdução deste projeto, já demonstram o benefício do uso de recomendações para a recuperação de

informação. Há diversos outros onde o uso de recomendações poderia apresentar ganhos perceptíveis, e um domínio no qual os benefícios podem ser fortemente conferidos é na recuperação de perfis de usuário, talvez principalmente pela forte característica social associada à recuperação de informação nesse domínio.

Sistemas de perfis de usuário se apresentam de diferentes formas, por exemplo, no cadastro de profissionais — como nos *sites* Catho Online<sup>5</sup> e apinfo.com<sup>6</sup> —, nos recursos humanos das empresas, em *sites* de relacionamento, no cadastro de clientes, em sistemas curriculares, em sistemas escolares, dentre outros. Cada uma das formas possui necessidades específicas de recuperação de informação que o uso de recomendações pode auxiliar.

No cadastro de profissionais, as recomendações auxiliam as empresas a encontrarem o profissional mais preparado para determinada atividade, e auxiliam o profissional a identificar empresas onde é mais provável que ele seja chamado para uma entrevista. Na área de recursos humanos, as recomendações identificam talentos, auxiliando o processo de promoções e premiações interno. Em *sites* de relacionamento, as recomendações ajudam o sistema a viabilizar os resultados mais prováveis para a formação de pares, para ambos os envolvidos. No cadastro de clientes, as recomendações ajudam a identificar os produtos, planos ou pacotes a serem indicados aos clientes, de forma a economizar recursos de divulgação e promoção, como também o tempo dos funcionários. Em sistemas curriculares, as recomendações permitem a formação de grupos de usuário em diferentes classificações, de forma que, ao realizar uma pesquisa, seja possível encontrar currículos mais relevantes aproveitando-se das avaliações realizadas por outros usuários com perfis semelhantes. Em sistemas escolares, as recomendações sugerem atenção diferenciada aos alunos que dela necessitem, seja em reforço escolar ou ainda na indicação de acompanhamento social de uma psicopedagoga. Estes são exemplos

---

<sup>5</sup> <http://www.catho.com.br>

<sup>6</sup> <http://www.apinfo.com>

singulares, há muitas outras indicações para o uso de recomendações que podem surgir no estudo minucioso de cada uma das diferentes aplicações no domínio de perfis de usuário.

Lendo o parágrafo anterior, podemos visualizar o componente social de cada aplicação com facilidade: a empresa quer indicações para saber quem contratar, o profissional não quer perder seu tempo visitando as empresas indiscriminadamente, o mercado quer saber o que oferecer e para quem oferecer... Enfim, percebe-se que a recomendação é uma necessidade e está fortemente presente no domínio de perfis de usuário.

O componente social não quer dizer que a recomendação precisa necessariamente do fator humano. A identificação automática de padrões atende, muitas vezes, e muito bem, às perguntas para as quais foi implementada. Conforme será visto ao decorrer dessa dissertação, muitas recomendações são geradas com base no conteúdo dos documentos. No caso, o componente social aqui descrito refere-se à necessidade do usuário do sistema. É como aquela indicação que ele recebe para tomar determinado caminho em uma viagem: ela pode ter sido feita por um amigo, ou ele pode ter lido em um guia de turismo, ou consultado um mapa pela Internet. Sua necessidade de informação foi satisfeita, e não precisa necessariamente ser alcançada pela consulta direta a outra pessoa.

Voltando ao objetivo desta seção, o estudo das recomendações para propor um modelo geral de sistema de recuperação de informações no domínio de perfis de usuário justifica-se pelas necessidades identificadas e que estão carentes de solução. Os problemas encontrados e sugestões apontadas podem servir não só ao domínio estudado, mas servirão de exemplo e ponto de partida para influenciar os pesquisadores a adotarem as abordagens de recomendações em domínios pouco ou não explorados. Espera-se que a flexibilidade dos indicadores<sup>7</sup> — que será a base de generalização do modelo de recomendação de perfis de usuário — torne

---

<sup>7</sup> Neste trabalho, o termo indicador refere-se a rótulos de dados, variáveis que podem ser trabalhadas quantitativamente e qualitativamente pelo sistema.

este modelo, produto deste trabalho, facilmente convertível e prontamente reutilizável.

Por fim, o presente trabalho é de interesse acadêmico, pois a pesquisa do tema traz visões inovadoras para o desenvolvimento da recuperação de informação. A ousadia do trabalho verificou a falta de pesquisas científicas que validem as recomendações na satisfação das necessidades dos usuários. Há um longo caminho pela frente. Pode-se supor que o modelo de recomendação de perfis de usuário será de grande utilidade para a comunidade científica.

## 2. Sistemas de Recomendação

Um sistema de recomendação é um sistema de informação que auxilia o usuário a recuperar informação através da previsão de seus interesses, informando-lhe conteúdo, fontes de consulta ou outras informações.

De acordo com Torres Jr. (2004), um sistema de recomendação geralmente possui: (i) dados prévios: dados armazenados pelo sistema antes do início do processo de recomendação; (ii) dados de entrada: dados que precisam ser inseridos pelo usuário a fim de receber recomendações; (iii) um algoritmo que combine os dados de entrada com os dados prévios para produzir recomendações. Em um sistema real, os dados prévios são obtidos do perfil ou histórico dos usuários, e os dados de entrada são obtidos através das ações que o usuário realiza para obter recomendações. Este processo é mostrado na Figura 1.

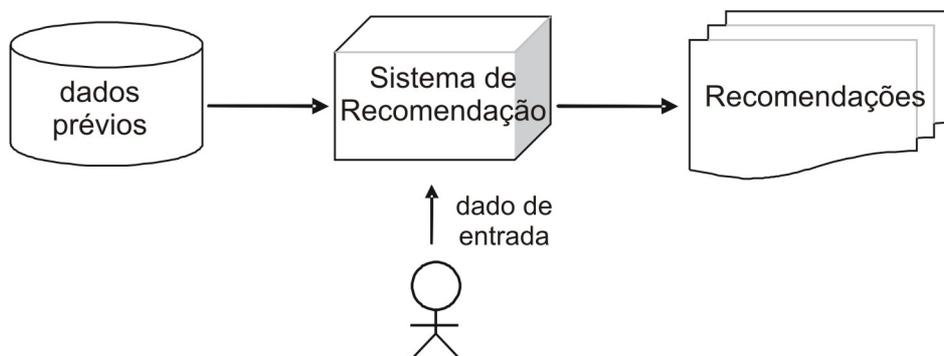


Figura 1 - Arquitetura Básica de um Sistema de Recomendação

Fonte: Torres Jr., 2004.

### 2.1. Recuperação de Informação

Lancaster (apud RAMOS, 1999, p.67) considera o problema da recuperação de informação como sendo o de procurar cotejar aproximações de necessidades de informação com aproximações de mensagens.

O objetivo da recuperação, dada uma pergunta formalizada por descritores que a definem corretamente, é que o sistema de informação providencie a comparação desses com aqueles que descrevem o documento

e obtenha as referências bibliográficas que atendem à pergunta em questão (ROBREDO, 2005).

O termo “recuperação da informação” foi primeiro utilizado por Mooers, em 1952, e depois popularizado pela comunidade de pesquisa à emergente área de recuperação da informação por meio dos escritos de Faithorne, em 1961 (RAMOS, 1999).

Com o crescimento no final da década de 50 da atividade científica e técnica e o conseqüente aumento da literatura, as bibliotecas viram-se obrigadas a preocupar-se mais com a especificidade dos assuntos. Ao mesmo tempo, começavam com Luhn as primeiras tentativas de desenvolvimento de métodos automatizados para armazenamento e recuperação da informação (LUHN, 1961).

O termo “recuperação de informação” tem sido caracterizado de forma diferenciada de recuperação de dados, pois segundo Rijsbergen (1998), alguma confusão tem se formado no meio científico. Procurando elucidar suas proposições, ele apresenta esta tabela:

*Tabela 1 - Comparação entre recuperação de dados e recuperação de informação segundo Rijsbergen*

	Recuperação de Dados	Recuperação de Informação
Combinação	<i>Combinação exata</i>	<i>Combinação parcial</i>
Inferência	<i>Dedução</i>	<i>Indução</i>
Modelo	<i>Determinístico</i>	<i>Probabilístico</i>
Especificação da consulta	<i>Completa</i>	<i>Incompleta</i>
Itens procurados	<i>Combinados</i>	<i>Relevantes</i>
Resposta a erros	<i>Sensitivo</i>	<i>Insensitivo</i>

Embora há mais de meio século o assunto esteja em evidência nas pesquisas, não há uma solução completa para a recuperação de documentos que satisfaça aos critérios de eficiência e rapidez. As

ferramentas de pesquisa apresentam muito ruído no resultado, em outras palavras, grande quantidade de documentos recuperados, mas apenas uma pequena parte é relevante para o usuário.

Percebendo a importância do tema, o Departamento de Defesa dos Estados Unidos, em conjunto com o Instituto Nacional de Padrões e Tecnologia (NIST), do mesmo país, patrocinou, em 1992, a *Text Retrieval Conference*<sup>8</sup> (TREC) — Conferência de Recuperação de Textos — como parte do programa TIPSTER<sup>9</sup>. O objetivo era observar a evolução das pesquisas na área de recuperação de informação, com o apoio de uma infraestrutura de suporte para a gigantesca avaliação das metodologias de recuperação de textos. Este evento consolidou a importância da pesquisa em recuperação de informação.

Como ponto de partida para a recuperação de informação, a “consulta” é, tradicionalmente, como o usuário comunica sua necessidade de informação a um sistema de informação. A necessidade de informação, segundo Mizaro (1996), é a falta de conhecimento do usuário para realizar determinada atividade. Para solucionar um determinado problema, o usuário necessita de um determinado conhecimento. Se ele não possui tal conhecimento, ele passa a ter uma necessidade de informação. Essa necessidade deve ser satisfeita, geralmente através de uma consulta a um sistema de recuperação de informação.

Normalmente a necessidade de informação é especificada na interface do mecanismo de recuperação em palavras-chave, e não em linguagem natural (BAEZA-YATES e RIBEIRO-NETO, 1999). Os dados são enviados para consulta em um formulário eletrônico. Após a introdução desses dados, o sistema dispara os processos de busca que vão localizar documentos que possam ser relevantes para o usuário.

---

<sup>8</sup> Disponível em <<http://trec.nist.gov>>. A TREC de mera observadora tornou-se referência mundial. A conferência TREC ocorre anualmente, e influencia o foco das linhas de pesquisa e as tecnologias do futuro em recuperação de informação.

<sup>9</sup> O programa TIPSTER (Profeta, em português) foi um projeto de pesquisa avançada do Depto. de Defesa Americano em tecnologias de processamento de texto, através da cooperação do Governo, indústria e academia (NIST, 2006).

Um sistema de recuperação de informação não tem a obrigação de responder precisamente a uma consulta, mas sim de identificar documentos que permitam satisfazer a necessidade de informação do usuário.

Não importa o quão bom seja o mecanismo de pesquisa e o algoritmo de classificação, ele não garante que os documentos retornados com alto grau de relevância — determinado pelo sistema — tenham a relevância necessária para o usuário (YUWONO et al, 1995). De acordo com a quantidade e qualidade dos argumentos de pesquisa, a única solução é uma interação<sup>10</sup> entre o usuário e o sistema, buscando um contínuo refinamento da consulta.

A recuperação de informação segue em constante aprimoramento, haja vista que a necessidade do usuário demanda desafios que extrapolam os limites das estratégias disponíveis e consolidadas. Os modelos consolidados de recuperação de informação são o ponto de partida para qualquer estudo sério sobre este assunto, portanto tais modelos serão abordados a seguir.

### **2.1.1. Modelos de Recuperação da Informação**

Os modelos de recuperação se subdividem em duas grandes áreas: técnicas de combinação exata e técnicas de combinação parcial (RAMOS, 1999). As técnicas de combinação exata requerem que o conteúdo das respostas obtidas represente precisamente a expressão de busca apresentada. São implementadas por meio de expressões de busca booleanas, de textos completos ou de varreduras em cadeias de caracteres (*strings*), e representam o conjunto de técnicas mais utilizado nos sistemas de recuperação da informação.

---

<sup>10</sup> Como só há coleta de recomendações sociais com a interação do usuário, esta é uma indicação de que os sistemas de recomendação promovem resultados mais satisfatórios do que sistemas que não utilizam o feedback do usuário.

As técnicas de combinação parcial surgiram para minimizar falhas evidentes da combinação exata, como a perda de muitos documentos relevantes que não podem ser recuperados pelas expressões de busca apresentadas ou a impossibilidade de priorizar os documentos recuperados. A combinação parcial é subdividida em técnicas individuais e técnicas em rede.

As técnicas individuais buscam comparar expressões de busca com documentos representados por termos índice e as técnicas em rede trabalham com conjuntos de documentos que possuem interligações em comum, e se dividem em: agrupamento, navegação e *spreading activation*<sup>11</sup>.

A técnica de agrupamento é baseada no agrupamento de documentos com conteúdos semelhantes; a técnica de navegação baseia-se na representação dos documentos por meio de uma rede de nós e conexões; e, por último, a técnica de *spreading activation* busca ativar, utilizando a expressão de busca, partes da rede que descrevem os conteúdos dos documentos e como eles se inter-relacionam.

Ramos (1999) representou a classificação das técnicas de recuperação conforme a Figura 2. Esta representação está bastante atual, e serve como base para este trabalho.

---

<sup>11</sup> Bastante utilizado em modelos de inteligência artificial de buscas em redes semânticas.

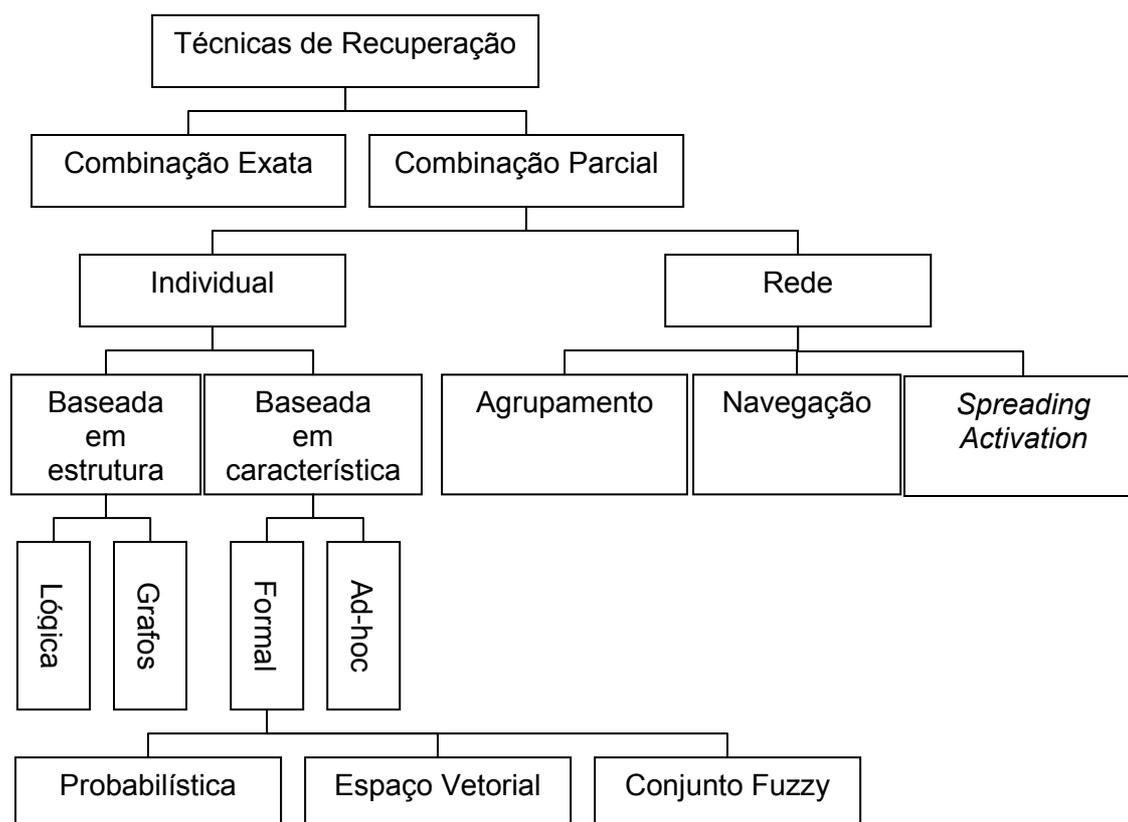


Figura 2 - Classificação de técnicas de recuperação de informação

Fonte: adaptação de Ramos (1999).

As abordagens de recomendação, embora à primeira vista pareçam fazer parte das técnicas em rede, são uma estratégia à parte. As recomendações, para funcionarem em um sistema de informação, podem ter como base quaisquer das técnicas de recuperação existentes. Não estão representadas acima porque se tratam de um modelo ainda em estudo, não consolidado<sup>12</sup> — se pudessem ser representadas, deveriam se apresentar disseminadas por todo o fluxo<sup>13</sup>.

Com os conceitos estabelecidos, fica a pergunta: como se dá a recuperação de informação com recomendações? Para entender esta atividade, é preciso entender como funciona o processo de recuperação da informação.

<sup>12</sup> As técnicas de recomendação citadas neste trabalho — filtragem colaborativa e filtragem baseada no conteúdo — se referem a algoritmos de programação da Ciência da Computação, e não aos métodos de recuperação de informação estudados pelos pesquisadores em Ciência da Informação.

<sup>13</sup> As abordagens de recomendação podem ser ditas como estratégias que melhoram os resultados alcançados através dos métodos de recuperação da informação.

## 2.1.2. Processo de Recuperação da Informação

Aproveitando os modelos expostos por Robredo (2005, p.197), vamos representar a recuperação de informação de um esquema de busca simplificado em diagramas de Venn, para então envolver a estratégia de recomendação.

Digamos um volume de informação em um conjunto delimitado, representado pela superfície ( $Z$ ), como na Figura 3, lado (a). Dentro deste conjunto há documentos que contêm o descritor  $A$  e outros documentos, ou suas referências, que contêm o descritor  $B$ .

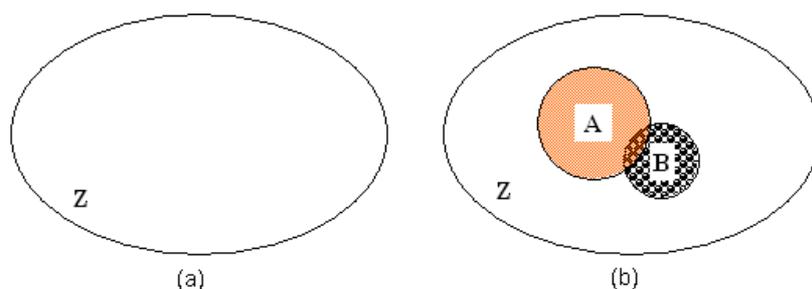


Figura 3 - Representação de um conjunto de documentos. a) Conjunto total; b) Dois subconjuntos inclusos no conjunto total

Fonte: adaptação de Robredo (2005).

Digamos que há outros descritores<sup>14</sup>  $C, D, E \dots$ . O grupo de cada um desses descritores forma um subconjunto dentro de  $Z$ . O conjunto  $Z$  é simbolizado por  $(A, B, C \dots)$  e cada um desses elementos é um elemento de  $Z$ . Isto pode ser representado como  $A \subset Z$  ( $A$  está contido ou incluído em  $Z$ ) ou  $Z \supset A$  ( $Z$  contém ou inclui  $A$ ).

Se considerarmos uma pergunta que fica definida por ambos os descritores  $A$  e  $B$ , teremos obtido todos os documentos que contêm o descritor  $A$  mais todos os documentos que contêm o descritor  $B$ . O número de documentos obtidos é menor que a soma dos documentos que contêm cada um dos descritores, já que alguns dos documentos contêm os dois

<sup>14</sup> Para deixar claro, descritor é qualquer informação de propriedade representativa do documento. São exemplos de descritores: autor, organizador, título do trabalho, palavras-chave, etc.

descritores simultaneamente. Não se trata de uma soma aritmética, e sim de uma soma lógica, representada por  $A \cup B$  que representa um novo subconjunto formado pela união dos elementos de  $A$  e  $B$ .

Se a pergunta for definida pelos descritores  $A$  e  $B$  simultaneamente, o resultado é diferente. Haverá documentos que contêm o descritor  $A$  mas não o  $B$ , então estes não respondem às exigências da pergunta. São boas (relevantes) as referências que contêm ambos os descritores. O resultado da recuperação é um subconjunto formado pelo produto lógico de  $A$  e  $B$ . Esta interseção é simbolizada por  $A \cap B$ .

Podemos representar o resultado das buscas com operadores booleanos. Em alguns casos é preciso fazer uso de parênteses para evitar ambigüidades na formulação das perguntas. Tais estratégias de recuperação estão representadas na Figura 4, onde:

- a) A e B e C, ou  $A \cap B \cap C$
- b) (A e B) ou C, ou  $(A \cap B) \cup C$
- c) A e (B ou C), ou  $A \cap (B \cup C)$
- d) (A ou B) exceto C, ou  $(A \cup B) - C$
- e) (A e B) exceto C, ou  $(A \cap B) - C$
- f) (A e C) ou (B e C), ou  $(A \cap C) \cup (B \cap C)$

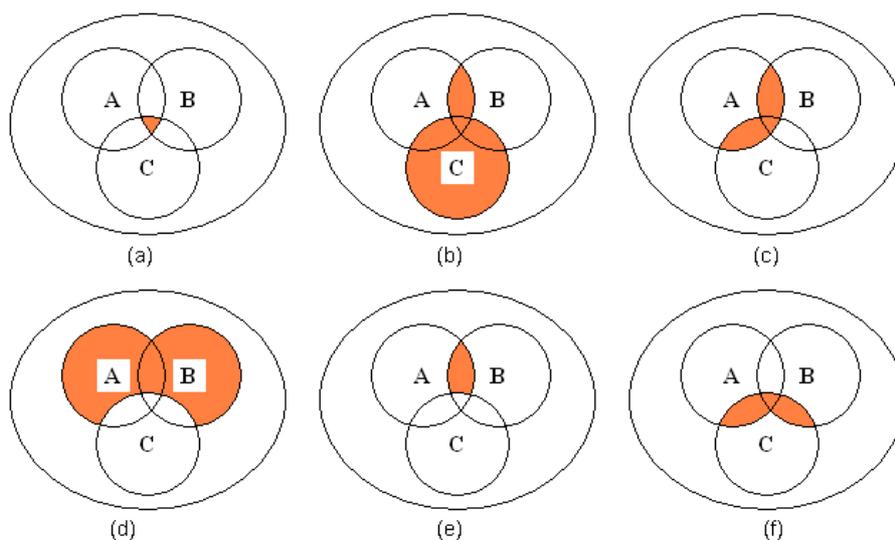


Figura 4 - Representação de algumas estratégias de busca com três descritores diferentes

Fonte: Robredo, 2005.

No resultado da recuperação, apenas uma fração do total das referências recuperadas corresponde à necessidade do usuário. Robredo (2005, p.200) nos relembra alguns conceitos bem conhecidos dos técnicos da documentação, e como tais conceitos são peças fundamentais no estudo de Recuperação da Informação, serão aqui também representados:

a) Índice de revocação ou exaustividade (*recall*):

$$R = \frac{\text{registros\_relevantes\_recuperados}}{\text{total\_de\_registros\_relevantes}} = \frac{Q}{F}$$

b) Índice de precisão ou pertinência (*relevance*):

$$P = \frac{\text{registros\_relevantes\_recuperados}}{\text{total\_de\_registros\_recuperados}} = \frac{Q}{A}$$

c) Índice de perda de informação ou razão de exaustividade (*information loss*):

$$L = \frac{F - Q}{F}$$

d) Índice de ruído de informação (*information noise*):

$$N = \frac{A - Q}{A}$$

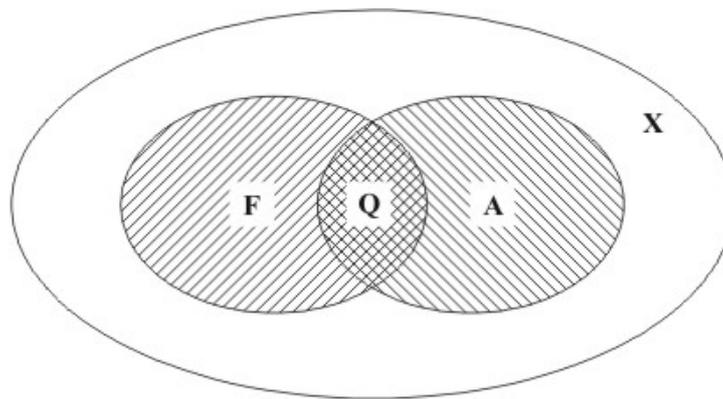


Figura 5 - Precisão e revocação da informação recuperada, e registros adjacentes

Fonte: Robredo, 2005.

Para entender bem o processo de recuperação da informação é preciso atenção ao significado dos símbolos:

- $X$  , todos os registros do banco
- $F$  , total de registros relevantes no banco (círculo completo)
- $A$  , total de registros recuperados (círculo completo)
- $Q$  , total de registros relevantes recuperados ( $F \cap A$ )
- $F - Q$  , perda da informação
- $A - Q$  , ruído da informação
- $R = Q/F$  , índice de revocação ou exaustividade
- $P = Q/A$  , índice de precisão ou pertinência<sup>15</sup>
- $L = (F - Q)/F$  , índice de perda de informação
- $N = (A - Q)/A$  , índice de ruído de informação

Apresentados os símbolos, vamos analisar o processo tradicional de recuperação de informação.

<sup>15</sup> Para o contexto da recuperação de informação, pertinência e relevância significam o mesmo.

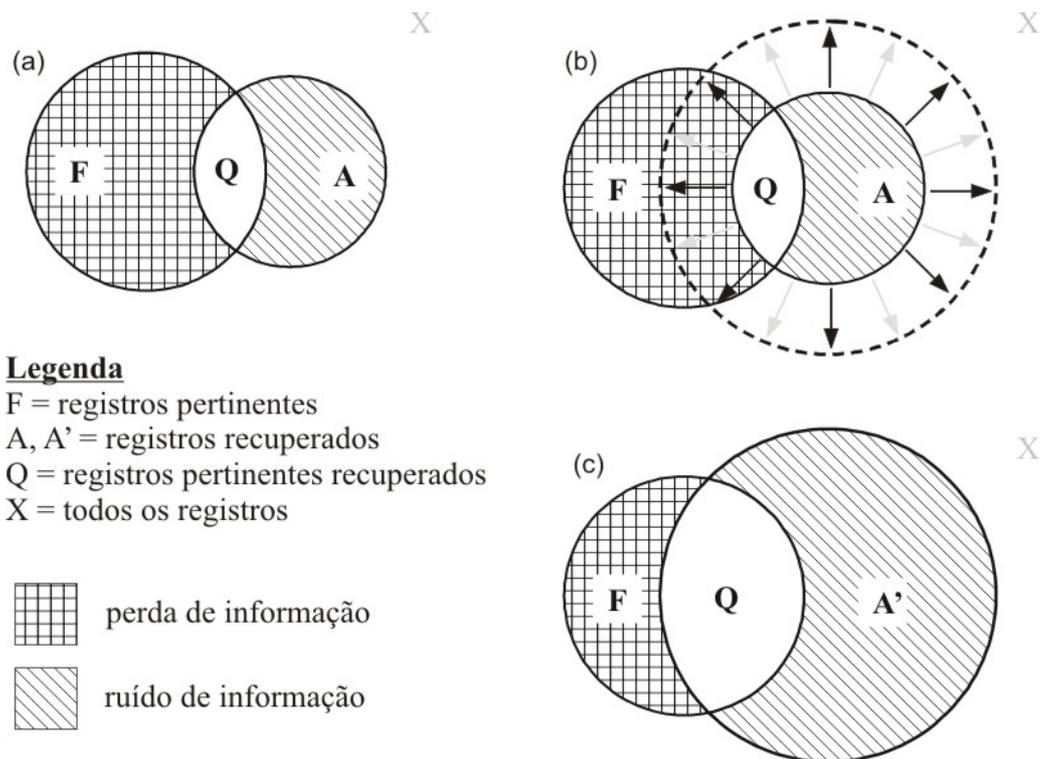


Figura 6 - Relação perda/ruído

Fonte: adaptação de Robredo (2005).

No esquema *a* da Figura 6, foi representado um processo de recuperação de informação com alta precisão, porém baixa exaustividade (pouco ruído, perda elevada de resultados relevantes). Se o campo da pergunta for expandido com, por exemplo, a utilização de termos mais genéricos — esquema *b* — aumenta-se a revocação, mas a precisão é reduzida (perda baixa, ruído elevado). Tal resultado foi representado no esquema *c*.

Como auxílio no processo de recuperação da informação, o usuário de um sistema de recomendação se beneficia do julgamento prévio realizado por outros que tiveram acesso à informação. Ele aproveita as associações realizadas por usuários com interesses semelhantes ao seu, que o remetem a documentos que são mais relevantes para o grupo<sup>16</sup>. Como o resultado da busca não apresenta documentos fora da área *A* (onde se encontram os documentos recuperados, sejam eles relevantes — área *Q* —

<sup>16</sup> Documentos mais relevantes são indicados e documentos de menor relevância tendem a ser desprezados. Um sistema de recomendação só é eficiente com a definição bem elaborada dos grupos de usuários com perfis semelhantes, para que haja real relevância dos indicadores coletados.

ou ruído de informação), a diferença da recuperação tradicional pode ser percebida no esquema da Figura 7.

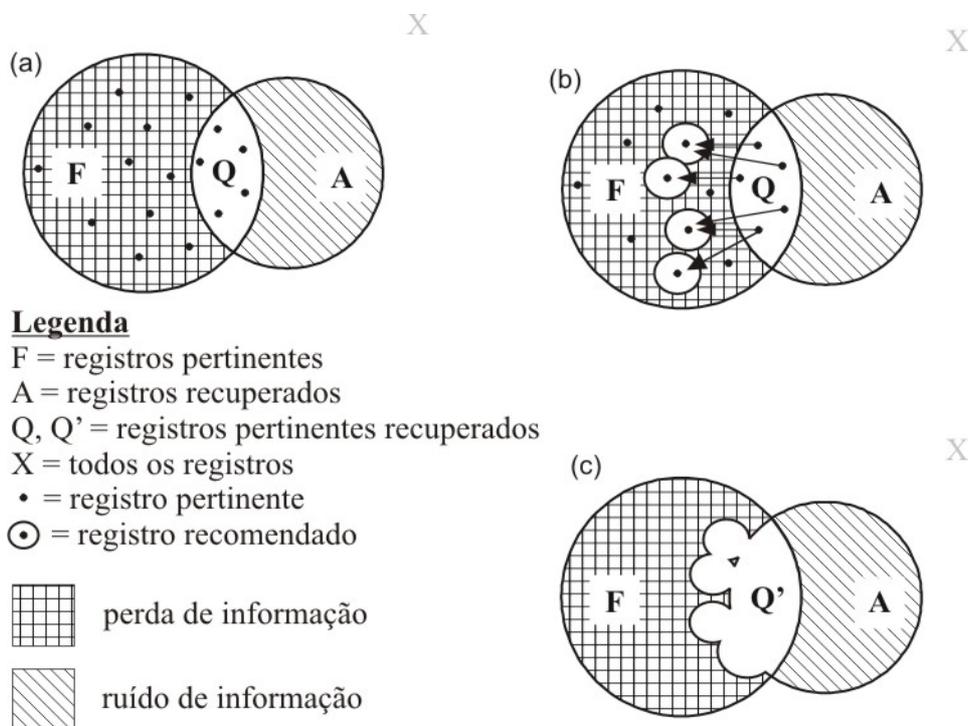


Figura 7 - Relação perda/ruído com recomendações

Fonte: elaboração própria.

Resultados relevantes fazem parte de  $F$  (círculo completo), e os documentos relevantes presentes na área  $Q$  apontam para outros documentos relevantes, presente na área de  $F$  não alcançada por  $Q$  (esquema  $b$ ). O uso de recomendações permite ao usuário alcançar resultados que ele não seria capaz na busca tradicional — exceto com modificações na busca original gerando mais ruído no resultado. O sistema de recomendação apresenta uma busca mais precisa sem aumentar o ruído de informação (esquema  $c$ ). Matematicamente isto é comprovado com o aumento dos índices de precisão  $P$  e revocação  $R$ , calculados por  $P = Q/A$  e  $R = Q/F$ , pois  $Q'$  (resultado no sistema de recomendação) é sempre maior que  $Q$  (resultado no sistema tradicional).

Como o assunto é de extrema importância para o entendimento deste projeto, precisamos falar das recomendações em uma seção à parte.

Aqui foi evidenciado que a utilização de abordagens de recomendações auxilia as técnicas de combinação parcial nos modelos já consolidados de recuperação da informação.

## **2.2. Recomendações**

Recomendações são importantes porque constituem um comportamento natural para se lidar com recuperação de informação. Elas possibilitam que sistemas de recuperação de informação (SRIs) confrontem uma volumosa quantidade de dados para localizar o material desejado. As recomendações expressam a experiência e opinião de um público, e são o meio pelo qual indivíduos podem se beneficiar da experiência prévia de outros (GROSSER, 1991).

Na economia, uma larga indústria de recomendações, tais como guias de viagem, avisos financeiros e corretores de ações, tem crescido para guiar os consumidores através das mais diferentes opiniões promovidas pelo mercado varejista e pelos anunciantes (NICHOLS et al., 2004). No comércio eletrônico, sistemas de recomendação são utilizados nos *sites* para auxiliar o cliente a localizar produtos, bem como oferecer outros produtos que também sejam do seu interesse (SCHAFER et al., 1999). Nas instituições financeiras é interessante a utilização de tais sistemas para o apoio à decisão. Em situações como a análise de crédito, por exemplo, o sistema de recomendação verifica os padrões de transações e faz previsões futuras. O sistema pode apresentar, através do histórico de seus clientes, os riscos da transação, bem como os fatores críticos encontrados. As recomendações podem indicar clientes que merecem atendimento especial, além de outras recomendações cruciais para o sucesso da corporação, como serviços que são oferecidos, mas que nunca são utilizados por determinado grupo de clientes (SCHAFER et al., 2001). Em fóruns eletrônicos, as recomendações podem indicar mensagens ao usuário, descobrindo seus interesses através da análise do conteúdo das mensagens postadas no fórum (FISCHER e WIVES, 2005).

No momento em que passam a ser utilizados como uma nova tendência, os sistemas de recomendação tornam-se estratégicos e passam a fazer parte do negócio (SCHAFER et al., 2001). Há recomendações para todos os tipos de bens e serviços.

Se não há imparcialidade na origem da recomendação, um risco que pode ocorrer é a sua distorção — por exemplo, alguém informar que determinado produto é melhor por interesse próprio ou duvidoso. Para evitar esta tendência, é importante conhecer a reputação<sup>17</sup> de quem recomenda (TORRES JR., 2004).

A coleta de recomendações é importante e não é uma tarefa fácil. Pode ser explícita (crítica ou *rating*), implícita (comportamento do usuário, através, por exemplo, da navegação, itens comprados, vistos, clicados, rejeitados) ou por inferência (aprendizado do perfil ou comportamento do usuário). No contexto de recuperação da informação, as recomendações se mostram úteis na utilização das técnicas<sup>18</sup> de filtragem colaborativa e filtragem baseada no conteúdo, técnicas de programação que precisam ser mais exploradas, especialmente em domínios nos quais possuem benefícios em potencial.

### **2.2.1. Abordagens e Técnicas**

As abordagens de recomendação surgiram da necessidade de se prever os documentos mais relevantes para o usuário. A primeira estratégia foi a abordagem de recomendação social, que faz uso de relevância em grupo — critérios de relevância estabelecidos por um vasto número de usuários diferentes, que classificam ou categorizam informações de interesse geral (BAEZA-YATES e RIBEIRO-NETO, 1999). Os sistemas que fazem uso da abordagem de recomendação social são conhecidos como sistemas de recomendação de informação social (MUKAYAMA, 1997) ou, mais comumente, como sistemas de recomendação colaborativa (POPESCU et

---

<sup>17</sup> Reputação, nesta situação, pode ser substituída por “a fonte que recomenda” ou “a origem da recomendação”. O status de quem recomenda influencia o valor da recomendação.

<sup>18</sup> Não confundir as técnicas de programação (Ciência da Computação) com as abordagens de recomendação (Ciência da Informação). Os conceitos estão detalhados na seção “Abordagens e técnicas”.

al., 2001). Na Ciência da Computação, esta abordagem é implementada através da técnica de filtragem colaborativa (*collaborative filtering*).

O termo “filtragem colaborativa” surgiu em 1992 no trabalho de Goldberg, Nichols e Terry (GOLDBERG et al., 1992). No sistema por eles proposto, o usuário avaliava artigos distribuídos em uma empresa, baseando-se em anotações. Os usuários podiam receber recomendações através das palavras-chave contidas no artigo ou através das anotações deixadas por outros. O sistema era útil para grupos pequenos, onde os usuários se conheciam para avaliar a recomendação.

Em 1994, foi publicado o estudo que desenvolveu o sistema GroupLens (RESNICK et. al., 1994), que se tornou o principal estudo sobre filtragem colaborativa (RESNICK e VARIAN, 1997; TORRES, 2004). O sistema GroupLens indicava notícias para os usuários utilizando avaliações feitas por outros, em pontos de 1 a 5. Com as pontuações determinadas, era possível identificar usuários que avaliavam as notícias de forma semelhante, e agrupá-los em uma "vizinhança". Notícias que recebiam altas notas por vizinhos e não lidas pelo usuário ativo eram recomendadas.

Pode-se dizer que a abordagem de recomendação social é baseada na idéia de que melhores recomendações são as realizadas por usuários com gostos semelhantes, e que a filtragem colaborativa depende de que usuários avaliem alguma coisa e dêem suas opiniões ou notas para que essa coisa possa ser recomendada.

Um exemplo prático do uso da abordagem de recomendação social é a barra de ferramentas Google:



Figura 8 - Google Toolbar, com propriedades avançadas ativas

Fonte: Google Toolbar para o Internet Explorer, disponível em <<http://toolbar.google.com>>

Os botões com os ícones feliz e triste — nesta ordem, situados à esquerda do botão *PageRank* — permitem a qualquer usuário no mundo expressar sua opinião acerca de uma página *web* (STEVEN, 2004). No

entanto, não é explícita a forma como trabalha esse algoritmo de recomendação de informação (FÓRUM DE DISCUSSÃO GOOGLE, 2004).

De maneira geral, a utilização da técnica de filtragem colaborativa tem a vantagem de que nenhuma análise da informação é necessária. A desvantagem<sup>19</sup> é que uma nova informação que ninguém classificou não será recomendada.

Para contornar a desvantagem da filtragem colaborativa é possível combiná-la com outra técnica, conhecida por filtragem baseada no conteúdo (*content-based filtering*). A utilização de filtragem baseada no conteúdo implementa diversas abordagens de recomendação (MUKAYAMA, 1997; POPESCU et al., 2001; MOONEY e ROY, 2000), dentre elas:

- Abordagem baseada em conhecimento: indica itens com base em regras editadas pelo administrador do sistema ou extraídas por mineração de dados. Exemplo: ao realizar uma busca no Google sobre “recuperação de dados” aparecem os *links* patrocinados (regra do sistema), ou na busca sobre “leite moça” uma sugestão para pesquisar o termo “nestlé” (*data mining*).

---

<sup>19</sup> Há sistemas que usam apenas similaridade entre critérios de relevância de pessoas com mesmo interesse, ignorando representações julgadas por pessoas em geral, criando assim mais uma desvantagem. Há diversas

Google Web Imagens Grupos Notícias mais »

recuperação de dados Pesquisar Pesquisa avançada Preferências

Pesquisar:  a web  páginas em português  páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 5.820.000 para recuperação de dados (0,06 segundos)

**Recuperação de Dados** Links Patrocinados  
[www.datarecover.com.br](http://www.datarecover.com.br) HD, RAID, SCSI, DVD, Memory Card, Fotos, Arquivos. DDG: 0800 702-9022

**Recuperação de Dados - HD** Links Patrocinados  
[www.DoctorByte.us](http://www.DoctorByte.us) Recuperamos **Dados** em HDs, RAID, CDs, Pen Drive, Memory Card, etc

**Recuperação de Dados** Links Patrocinados  
[www.RecoveryLab.com.br](http://www.RecoveryLab.com.br) Análise grátis na hora, Ligue ! Grátis 0800 771 7242 24H

**Recuperação de Dados | Clube do Hardware** Links Patrocinados  
 Para isso você precisará usar um programa de **recuperação de dados**. ... Tenho algumas dúvidas em relação ao software de **recuperação de dados** Goback. ...  
[www.clubedohardware.com.br/duvidas/30](http://www.clubedohardware.com.br/duvidas/30) - 64k - 5 out. 2006 -  
 Em cache - Páginas Semelhantes

**Data Disk - 0800 770 3282** Links Patrocinados  
 Laboratório Avançado de **Recuperação de Dados** (11) 6497-2000  
[www.datadisk.com.br](http://www.datadisk.com.br)

**Hospital do Hd 24 horas** Links Patrocinados  
**Recuperação dados** no mesmo dia sem taxas. S.P 11-7721-6708  
[www.hospitaldohd.com.br](http://www.hospitaldohd.com.br)

**Recuperar dados em HD Mac** Links Patrocinados  
 Conferencia remota de arquivos do Mac. Mesmo em HD danificado ligue.  
[www.InstitutoDoHD.com](http://www.InstitutoDoHD.com)

Figura 9 - Abordagem de recomendação baseada em conhecimento (regra do sistema)

Fonte: elaboração própria, a partir da consulta ao Google

Google Web Imagens Grupos Notícias mais »

leite moça Pesquisar Pesquisa avançada Preferências

Pesquisar:  a web  páginas em português  páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 423.000 para leite moça (0,16 segundos)

**Receita de Pudim de leite moça - Tudo Gostoso**  
 3 ovos; 1 lata de **leite moça**; 1 medida da lata de **leite moça** de **leite** ... bata no liquidificador os ovos, o **leite moça** eo **leite** por 3 minutos. ...  
[tudogostoso.uol.com.br/receita/1045-pudim-de-leite-moca.html](http://tudogostoso.uol.com.br/receita/1045-pudim-de-leite-moca.html) - 18k -  
 Em cache - Páginas Semelhantes

Ver resultados para: [nestle](http://www.nestle.com.br/)

**NESTLÉ**  
[www.nestle.com.br/](http://www.nestle.com.br/)

Figura 10 - Abordagem de recomendação baseada em conhecimento (data mining)

Fonte: elaboração própria, a partir da consulta ao Google

- Abordagens baseadas no conteúdo: indica itens de acordo com o perfil do usuário (ex: livros na mesma área de interesse) ou baseando-se no histórico do usuário (ex: filmes já assistidos, CDs já comprados).

Métodos de filtragem de informação baseada no conteúdo criam recomendações através da combinação da consulta ou de outra informação do usuário com informações de propriedades representativas do documento (MOONEY e ROY, 2000), como, por exemplo, índice ou palavras-chave.

Sistemas puramente implementados com essa técnica não são capazes de realizar uma recomendação social: um filtro de informação poderia recomendar o filme “Sexo, Amor e Traição” para um usuário que gosta de Murilo Benício e Malu Mader, embora o filme tenha desagradado muitas pessoas que apreciam esses atores.

Um bom exemplo de sistema que utiliza recomendações baseadas no conteúdo, e que não poderia deixar de ser citado, é o *site* de buscas Vivísimo<sup>20</sup>. A grande maioria dos *sites* de busca baseia-se nas *tags* para identificar os resultados, o Vivísimo apostou em uma outra corrente: o agrupamento de termos semelhantes (VIVISIMO, 2006). Cruzando algoritmos estatísticos com um conhecimento lingüístico profundo, os pesquisadores da Carnegie Mellon University — desenvolvedores da ferramenta — foram capazes de identificar relacionamentos entre os termos de busca. Trabalhando dessa forma, os resultados das consultas realizadas no Vivísimo não são disponibilizados em categorias pré-definidas: o programa determina as categorias aleatoriamente, dependendo do resultado da busca. A classificação é feita por meio de uma combinação da análise lingüística e estatística, que embora esteja preparada para trabalhar com outros idiomas — é sabido funcionar bem no espanhol —, não apresentou resultados satisfatórios em buscas simples realizadas em português.

Para validar o propósito da ferramenta, foi realizada uma consulta por "recuperação da informação" tanto no Vivísimo quanto no Google, a fim de comparar os resultados dos dois motores de busca. O Vivísimo não retornou sequer um resultado, e perguntou se a busca não seria sobre "recreação da informação" (Figura 11). A busca pelo termo em inglês "*information retrieval*" retornou 27 resultados (Figura 12), enquanto que no Google nos perderíamos nos 27.700.000 resultados (Figura 13).

---

<sup>20</sup> Disponível em <<http://vivisimo.com/>>

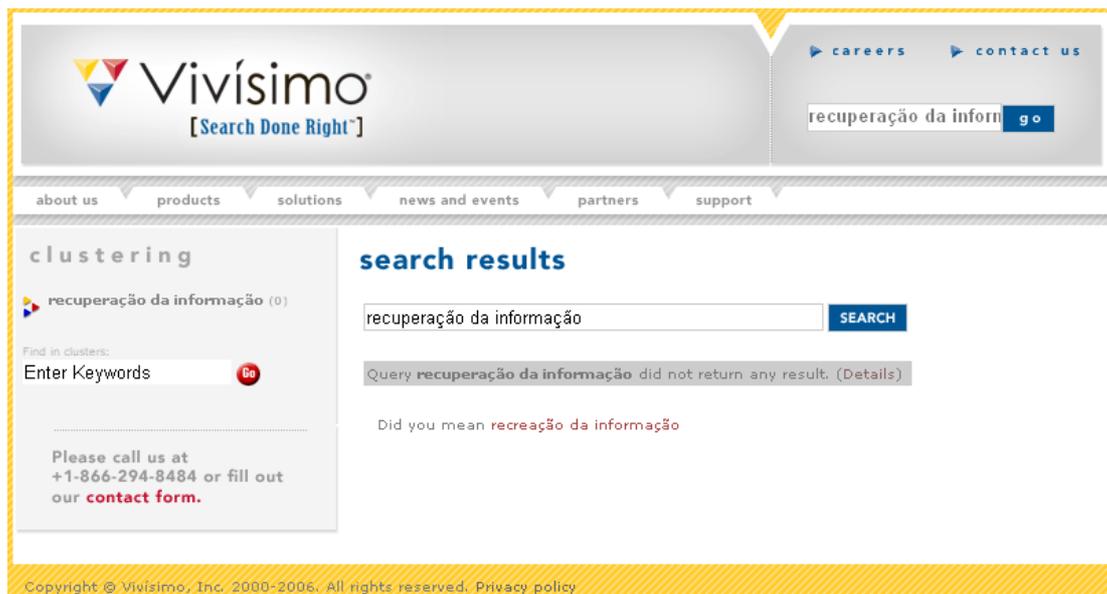


Figura 11 - Resultado da busca por "recuperação da informação" no Vivísimo

Fonte: elaboração própria, a partir da consulta ao Vivísimo

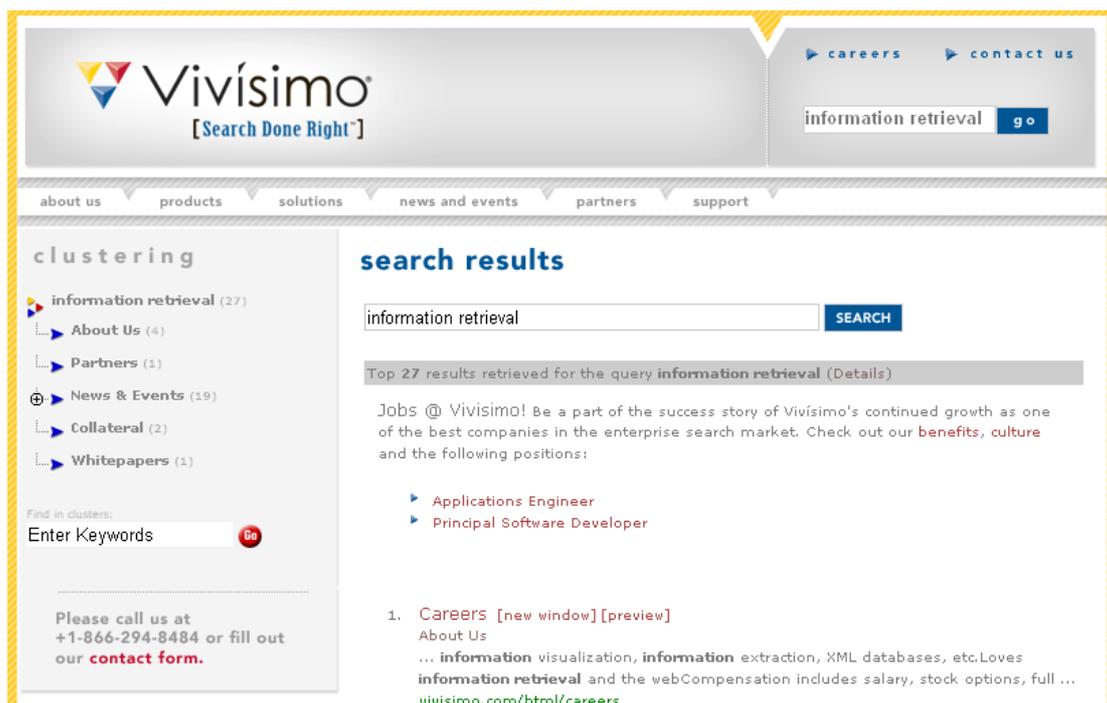


Figura 12 - Resultado da busca por "information retrieval" no Vivísimo

Fonte: elaboração própria, a partir da consulta ao Vivísimo

The screenshot shows the Google search interface. At the top left is the Google logo. To its right are navigation links: Web, Imagens, Grupos, Notícias, and mais ». Below these is a search bar containing the text 'information retrieval'. To the right of the search bar is a 'Pesquisar' button. Further right are links for 'Pesquisa avançada' and 'Preferências'. Below the search bar, there are radio buttons for 'a web' (selected), 'páginas em português', and 'páginas do Brasil'. A blue banner below the search bar indicates 'Web Resultados 1 - 10 de aproximadamente 27.700.000 para information retrieval (0,21 segundos)'. The main results area shows a list of academic articles under the heading 'Artigos acadêmicos sobre information retrieval'. The first result is 'Automatic text processing: the transformation, analysis, ... - Salton - Citado por 2612'. Below it are two more results: 'Introduction to Modern Information Retrieval - Salton - Citado por 4019' and 'Information Retrieval - Van Rijsbergen - Citado por 2438'. To the right of the main results is a 'Links Patrocinados' section featuring 'Enterprise Information' with the text 'Access your business data without switching applications.' and the URL 'www.babyon-enterprise.com'. Below the main results, there is a link for 'Information Retrieval - [ Traduzir esta página ]' followed by the text 'An online book by CJ van Rijsbergen, University of Glasgow.' and the URL 'www.dcs.gla.ac.uk/Keith/Preface.html - 7k - Em cache - Páginas Semelhantes'.

Figura 13 - Resultado da busca por "information retrieval" no Google

Fonte: elaboração própria, a partir de consulta ao Google

Grande parte das representações deste trabalho foi feita no Google por duas razões: primeiro, o Google é bastante conhecido e facilmente acessível; segundo porque não foi localizado outro trabalho acadêmico que ilustrasse as abordagens de recomendação da ferramenta Google. A maioria dos trabalhos internacionais destaca as abordagens de recomendação da empresa Amazon.com<sup>21</sup>. Seu sucesso deve-se principalmente à história da empresa.

A Amazon.com abriu suas portas para a *World Wide Web* em 1995, foi uma das pioneiras na venda de livros pela Internet, e hoje tem a maior seleção de livros do planeta. Desde cedo percebeu que a personalização da interface na venda pela Internet aumentaria as vendas, da mesma forma que um vendedor conhecendo determinado cliente, poderia aumentar as vendas indicando a chegada de um novo livro do gênero ou autor que o cliente goste (EADS, 2000).

Nas palavras de Jeff Bezos, diretor-executivo da Amazon.com: "Eu quero transportar a venda de livros *on-line* para o tempo das pequenas livrarias, que conheciam você muito bem e poderia dizer coisas como 'eu sei que você gosta de John Irving'<sup>22</sup>, e veja só este novo autor, seu trabalho é muito parecido com o do John Irving". É claro que no caso da Amazon.com é

<sup>21</sup> <http://www.amazon.com>

<sup>22</sup> Autor do *bestseller* "The world according to Garp" (O mundo segundo Garp, tradução de Luiz Corção). O livro rendeu um filme de mesmo título, estrelado por Robin Williams (1982 - 131m).

a tecnologia, e não o vendedor da loja, que lembra do que o cliente gosta ou não. Jeff Bezos também é autor do famoso discurso "Nosso objetivo é abastecer nossos 20 milhões de usuários com 20 milhões de lojas diferentes. Somos a maior companhia do mundo com foco total no cliente" (EADS, 2000).

O sistema de recomendação da Amazon.com grava o que um cliente adquiriu, avaliou ou viu no *website*, e então, através de uma abordagem híbrida<sup>23</sup> de recomendação colaborativa e recomendação baseada no conteúdo, combina esses dados com as informações agregadas de outros usuários que adquiriram produtos semelhantes. Pela comparação realizada, a ferramenta pode oferecer recomendações objetivas, relevantes e interessantes para seus clientes. Exemplos de funcionalidades que usam esta tecnologia no *website* podem ser conferidos em *New for You* (novidades para você), *Instant Recommendations* (recomendações instantâneas), *Gift Wizard* ("wizard" é um termo utilizado na informática para designar um programa que assiste ao usuário executar uma função da melhor forma, assim *Gift Wizard* seria algo como "mago dos presentes", um programa para ajudar o usuário a presentear alguém), *Similarity Explorer* (explorador de similaridades), *Recommendation Wizard* (mago das recomendações), "*Customers who bought...*" (Quem comprou isto, comprou também...), e sempre na apresentação personalizada de novos lançamentos. Já na página principal do *website*, há um convite para se cadastrar e começar a receber recomendações, conforme mostrado na Figura 14:

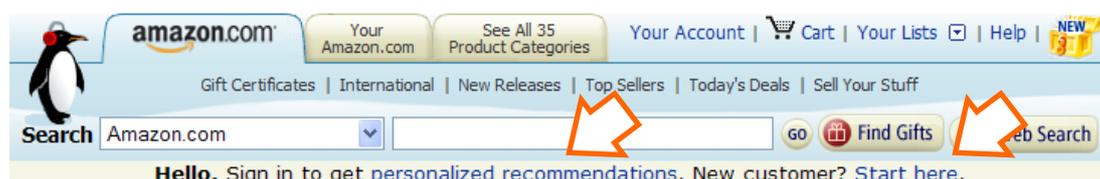


Figura 14 - Convite para se cadastrar e receber recomendações na Amazon.com

Fonte: elaboração própria, a partir de consulta ao amazon.com

Sistemas híbridos, como o *website* Amazon.com, atenuam as desvantagens das técnicas de filtragem colaborativa e filtragem baseada no

<sup>23</sup> Sistemas híbridos estão explicados logo a seguir.

conteúdo, e estão sendo investigados por diversos pesquisadores (MOONEY e ROY, 2000). Acrescentando as vantagens das técnicas de *data mining*, onde os sistemas de recomendação “aprendem” com o comportamento dos usuários no sistema, é possível adotar as mais variadas abordagens de recomendação, como por exemplo:

- Abordagem *Top-N*: indica os itens mais bem classificados por um grupo ou indivíduo. O método de coleta de recomendações pode ser implícito ou explícito, através de *ratings* ou estatística. Exemplo: Livros mais lidos, CDs mais vendidos, filmes mais vistos, artigos mais citados.
- Abordagem *Cross-Sell*: indica itens com base na seleção “casada” feita por outros usuários. Exemplo: quem comprou X, também comprou Y (quem visitou o *site* X, quem leu o artigo X). Utiliza recursos de mineração de dados.

Suponhamos um contexto de pesquisa de preços em supermercados<sup>24</sup>. Na interface de busca é apresentada, na primeira consulta a ser realizada por cada usuário cadastrado, uma interface para deixar registrada sua preferência por produtos a serem pesquisados, com uma descrição como esta: “Este é um sistema de busca e recomendações inteligente, para que ele funcione é necessário que você defina uma busca de acordo com o seu interesse. Preencha os campos abaixo”. Na seção de produtos de feira, diferentes tipos de fruta que o sistema trabalha na pesquisa de preços são apresentados e o usuário seleciona as frutas que deseja. Toda vez que se autenticar no sistema, poderá haver recomendações de preço de banana — se for este o caso de ele ter selecionado “banana” — indicando ao usuário os supermercados onde pode encontrar os melhores preços.

Esse tipo de busca por menor preço existe no *site* Buscapé<sup>25</sup>, e facilita bastante a vida de quem precisa comprar qualquer produto. De um

---

<sup>24</sup> Exemplo fictício.

<sup>25</sup> <http://www.buscape.com.br>

chocalho ao preço da gasolina, é difícil haver um produto que não esteja presente no Buscapé.



Figura 15 - Pesquisa por "buscape" no Google

Fonte: elaboração própria, a partir de consulta ao Google

O Buscapé é um exemplo brasileiro de sistema de recomendações que deu certo. Criado em São Paulo em 1999, vasculhava na Internet o preço dos produtos aplicados pelas lojas virtuais e fornecia a pesquisa gratuitamente a quem consultasse. Essa facilidade atraiu muitos usuários, e também outras empresas que gostariam de ter seus produtos aparecendo no *site* porque aplicavam preços mais baixos e aumentariam suas vendas. O Buscapé passou a cobrar para mostrar o cadastro de pequenas empresas virtuais. As grandes empresas que já tinham preços mais vantajosos e disponibilizavam o preço de seus produtos virtualmente viram que as consultas realizadas a partir dos robôs do *site* Buscapé foram crescendo exponencialmente, e então o Buscapé começou a cobrar pela publicidade. O objetivo principal do *site* era a variedade de produtos para pesquisa.

Em 2000, surge o concorrente carioca BondFaro, também para oferecer pesquisa de preços, mas focando seu diferencial na variedade, na consulta a empresas virtuais. Então havia dois sistemas de recomendação de preços concorrentes, o que não era bom nem para o consumidor, porque tinha que realizar sua busca duas vezes, nem para o Buscapé e BondFaro, porque tinham que disputar por publicidade e visibilidade.

Talvez porque os sistemas de recomendação passaram a ter maior destaque depois da TREC de 2002 — quando as pesquisas tiveram maior enfoque em filtragem de informação, recuperação na *Web* e *Question Answering*<sup>26</sup> (VOORHEERS, 2002) —, os programadores começaram a perceber que havia uma série de vantagens na implementação dos sistemas de recomendação híbridos. Em novembro de 2005, o Buscapé lança sua versão híbrida de sistema de recomendação. Como antes estava baseado na filtragem baseada no conteúdo, informações que ou eram consultadas nas vitrines virtuais ou fornecidas diretamente pelas empresas, agora seria preciso do apoio dos usuários para avaliar os produtos, aprimorando o sistema de recomendação. O Buscapé envia, então, convites por *email* aos usuários.

---

<sup>26</sup> *Question Answering* (QA) é um tipo de recuperação de informação. Dada uma coleção de documentos, o sistema deve ser capaz de fornecer respostas a questões elaboradas em linguagem natural. QA requer técnicas de processamento de linguagem natural mais complexas que outros tipos de recuperação de informação e é considerada como o próximo passo no desenvolvimento de motores de busca.

## Opinião de Produtos

Como você avalia sua experiência com relação a um Produto ou Serviço adquirido, ele atendeu às suas expectativas? É resistente? De fácil manuseio? A bateria dura bastante?

**Auxilie milhões de pessoas a comprar melhor!**  
Para opinar sobre um produto, basta encontrar o objeto de seu relato\* e clicar em "Escreva uma opinião"

\*(apenas para Eletrônicos, Informática, Eletrodoméstico e Telefonia)

---

**Samsung X480 GSM Tim / Oi / Claro** (9 ofertas)

votar para > Início > Telefonia > Celular  
 imprimir esta página    envie para um amigo

Veja também Celular em:  
 > Usados e Semi-novos  
 > Atacado e Distribuição

Já possui este produto?  
 Escreva uma opinião.

Já possui este produto?  
 Escreva uma opinião.

Avaliação: ★★★★★ 700 avaliações  
 Faixa de preço: R\$ 480,00 a R\$ 550,00  
 Alerta de Preço

> Clique e refine as 9 ofertas encontradas em 6 empresas

**Busque por um produto para opinar agora mesmo!**

Figura 16 - Implantação da abordagem de recomendação social no site Buscapé

Fonte: elaboração própria, a partir de email-convite enviado pelo site Buscapé

O Buscapé contava com a tecnologia de ponta em recomendação de informação. Com a avaliação de produtos, a indústria também se tornou cliente do *site*, para promover seus produtos nas pesquisas baseadas em regras e *data mining* e também comprando espaço publicitário. Ganhando o mercado, em maio de 2006 o Buscapé compra seu concorrente BondFaro, que não se preocupou em aprimorar suas estratégias de recuperação de informação. O BondFaro já tinha espalhado seu domínio para Argentina, Chile e México, e era uma marca forte. Talvez por isso o Buscapé tenha optado por manter a marca.

A fusão do BuscaPé com o Bondfaro resultou em uma plataforma integrada para a publicidade *on-line* e o comércio eletrônico com 40 mil lojas listadas, entre lojas virtuais e do mundo real, sendo que o BuscaPé agregou 230 novas lojas *on-line* anteriormente presentes apenas no Bondfaro e o

Bondfaro, por sua vez, acrescenta outras 2,2 mil lojas *on-line* em seu portfólio (BARBOSA, 2006).



Figura 17 - Alguns dos principais clientes do Buscapé

Fonte: elaboração própria, a partir de consulta ao Buscapé

A empresa se consolidou como a líder em comparação de preços na América Latina. A página inicial do *site* nos mostra seu escopo geográfico (Figura 18), e a página da Central de Negócios<sup>27</sup> ilustra alguns de seus principais clientes (Figura 17).

<sup>27</sup> <http://www.buscape.com.br/negocios>

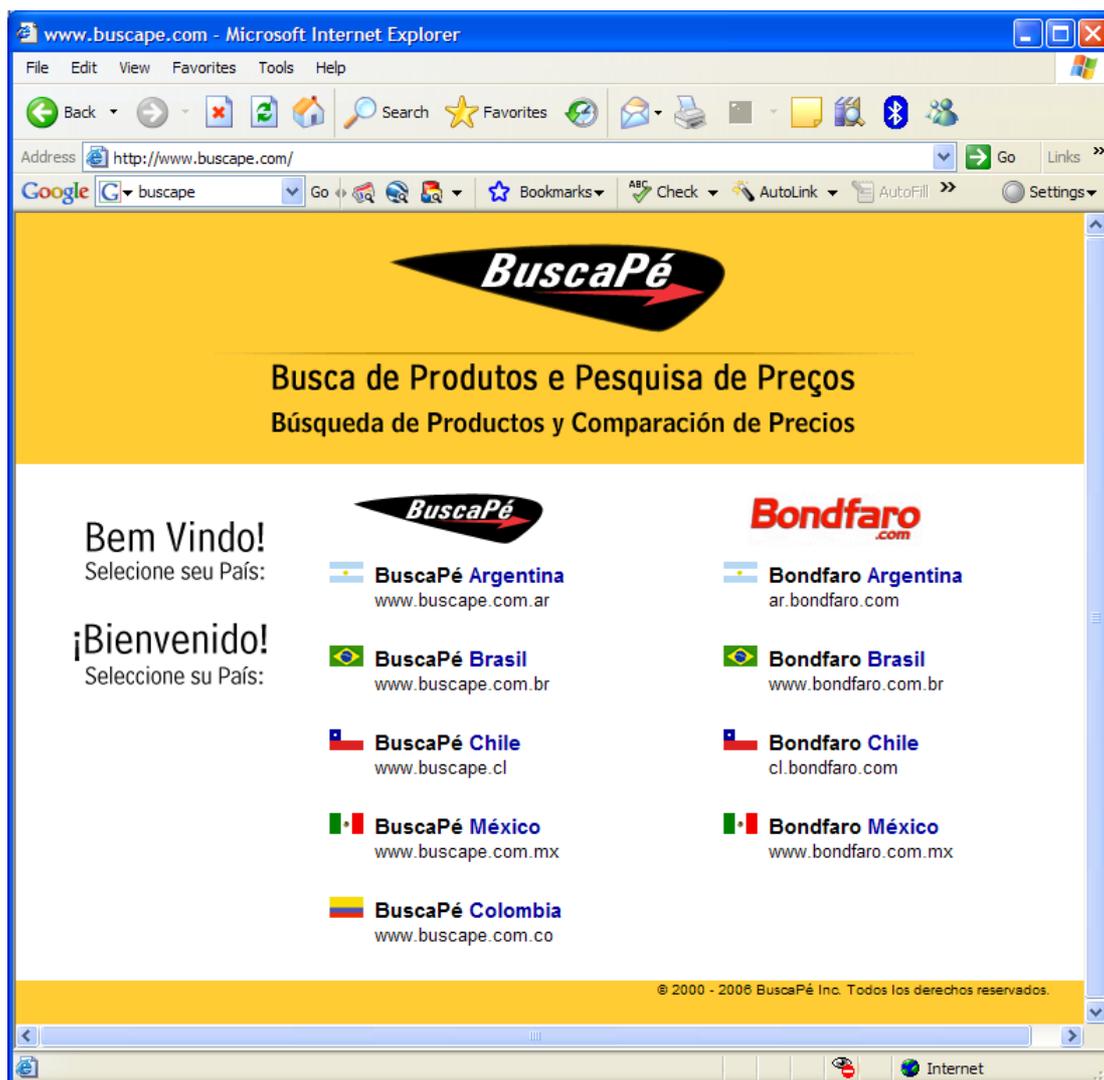


Figura 18 - Página inicial do site Buscapé

Fonte: disponível em <<http://www.buscape.com>>

De acordo com Schafer et al. (2001), em seus estudos sobre comércio eletrônico, os sistemas de recomendação podem aumentar as vendas de três maneiras: convertendo “curiosos” em compradores, realizando vendas “casadas” e, por fim, fidelizando o consumidor com a personalização. As recomendações se tornam uma ferramenta essencial para o comércio eletrônico, a ponto de se dizer que são necessárias para definir o sucesso ou fracasso de uma empresa.

Parece novidade quando dado o foco computacional, porém dentro da Ciência da Informação nunca foi novidade o uso de recomendações para a realização de pesquisas nos centros de

documentação. Quem já estudou as técnicas de bibliometria certamente se lembrou da análise de co-citações ao ler esta dissertação.

A análise de co-citações é um método, desenvolvido em meados da década de 70 por SMALL e GRIFFITH (1974), que mede o interesse intelectual em comum existente em um par de documentos. Esta medida, definida como força de co-citação (*cocitation strength*), é dada pelo número de documentos que foram citados por ambos. No ambiente *web*, a análise de co-citações se traduz na noção de que, quando um documento contém *links* referenciando outros documentos, por exemplo, documento A e documento B, estes documentos estão inter-relacionados de alguma maneira. Neste exemplo, os documentos A e B são ditos co-citados (ONODA, 2006). Em sua essência, os indicadores baseados em citações procuram refletir o impacto e a visibilidade das publicações citadas junto à comunidade científica. A citação é uma variável complexa, que não pode ser correlacionada à qualidade científica de um trabalho. O uso de recomendações baseadas em citações deve ser criterioso, tendo em vista as fragilidades e limitações inerentes à sua construção e seu uso.

Uma das grandes dificuldades dos sistemas tradicionais de recuperação de informação é a decisão quanto à relevância de um item de informação em resposta a uma busca (MARQUES e LADEIRA, 2005). Lancaster (2004) ressalta que a medida da relevância só pode ser atribuída pelo julgamento do usuário. A maioria dos sistemas de recuperação de informação disponíveis para consulta foi implementada em algoritmos de pura base estatística.

Os sistemas de recuperação de informação que aplicam as abordagens de recomendação, isto é, os sistemas de recomendação, levam em conta exatamente o julgamento do usuário. Nesse sentido, as abordagens de recomendação surgem como uma alternativa para o futuro dos sistemas de recuperação de informação. Não está claro, neste momento, se essas abordagens vão promover o desenvolvimento de novos sistemas com menor esforço.

## 2.2.2. Relevância e *Feedback* do Usuário

Segundo Le Coadic (2004, p.62), o conceito de relevância é importante, mas não é operacional. Nos sistemas de informação, o objetivo principal é fornecer ao usuário informação relevante. A relevância mede a correspondência entre documento e consulta. Esse conceito é básico na avaliação de desempenho dos sistemas de recuperação: vincula a necessidade do usuário a documentos, tendo a ver, portanto, com a satisfação dele.

A relevância automática, utilizada pelos sistemas, não costuma atender à necessidade do usuário. O usuário espera interagir com o sistema, de forma que possa influenciar na seleção dos documentos relevantes, isto é, o usuário deseja o *feedback* de relevância.

Baeza-Yates e Ribeiro-Neto (1999, p.303) explicam sua importância:

Uma parte importante no processo de acesso à informação é a formulação da consulta, e uma técnica eficaz comprovada para a formulação da consulta é o *feedback* de relevância. Originalmente, o *feedback* de relevância se refere a um ciclo de interação no qual o usuário seleciona um conjunto de documentos que parecem ser relevantes para a consulta, então o sistema utiliza características resultantes dos documentos relevantes selecionados para aprimorar a consulta original. Esta consulta melhorada é então executada e como resultado há um novo conjunto de documentos. Documentos do conjunto original podem aparecer na nova lista de resultados, provavelmente em uma diferente ordem de classificação.

Não é fácil criar esse *feedback*. A dificuldade se encontra em diversos pontos do sistema, seja encapsulada na inteligência dos algoritmos, seja exposta nas interfaces de usuário. A Ciência da Computação ainda não consegue cobrir falhas evidentes e os estudos de Ciência da Informação, que evoluíram para chegar ao encontro das necessidades do usuário, estão

sendo adaptados para os sistemas, nesse tempo em que a sociedade exige sistemas que não recuperem apenas dados e sim informação útil e relevante.

Utilizando *feedback* de relevância, os critérios de relevância do usuário podem melhorar a busca, atendendo, de acordo com a necessidade desse usuário, aos critérios de precisão e revocação.

Mas, se grande parte do interesse por *feedback* de relevância justifica-se pelo desejo do usuário em interagir com o sistema (CROFT, 1995), sabemos que esse desejo não é alcançado somente após a realização da consulta. As abordagens de recomendação nos mostram que é possível haver interação do usuário com o sistema antes mesmo de que a consulta seja realizada — após preencher seu cadastro em um sistema de perfis de usuário, por exemplo, o usuário antes de realizar uma primeira busca por algum perfil pode receber recomendações de pesquisas realizadas por outros que compartilham de sua afinidade.

Sendo assim, os sistemas de recomendação além de agradarem aos usuários por permitirem o *feedback* de relevância, ajudam na argumentação e ainda dão confiança ao usuário para aceitar as informações recuperadas.

### **2.3. Redes Sociais**

Rede Social é uma das formas de representação dos relacionamentos dos seres humanos entre si ou entre seus agrupamentos de interesses mútuos. Segundo Fritjot Capra (1997), "redes sociais são redes de comunicação que envolvem a linguagem simbólica, os limites culturais e as relações de poder". São também consideradas como uma medida de política social que reconhece e incentiva a atuação das redes de solidariedade local, como no combate à pobreza e à exclusão social e na promoção de desenvolvimento regional. As redes sociais são capazes de expressar idéias políticas e econômicas inovadoras com o surgimento de novos valores, pensamentos e atitudes. Esse segmento que proporciona a ampla informação a ser compartilhada por todos, sem canais reservados e

fornecendo a formação de uma cultura de participação é possível graças ao desenvolvimento das tecnologias de informação e comunicação. As redes unem os indivíduos, organizando-os de forma igualitária e democrática e, ainda, em relação aos objetivos que compartilham.

A rede é responsável pela difusão de idéias entre pessoas que possuem interesses, objetivos e valores a serem compartilhados. Assim, um grupo de discussão é composto por indivíduos que possuem identidades semelhantes. Muitas redes sociais estão presentes na Internet pelo fato desta possibilitar velocidade e variedade para a divulgação das idéias e absorção de conteúdo.

As redes sociais são inerentes às atividades humanas. O dia-a-dia, as relações que sustentam nossas rotinas, é um conjunto de redes. Como a teia de relações que cada pessoa tece em sua vida escolar: professores, colegas, aquela pessoa que compartilha o lugar no ônibus ou metrô, o cobrador, a servente da escola; a rede de relações que cada pessoa estabelece para abastecer a casa, comprar roupas, em sua vida profissional; a rede de afetos: as pessoas que ama. Todas essas atividades originam redes sociais. São redes espontâneas, que derivam da sociabilidade humana. Estão presentes o tempo inteiro, apenas não costumamos focar nosso olhar sobre elas, vendo-as como um sistema vivo e dinâmico.

As redes sociais são importantes para a difusão da informação. Um referendo, uma receita de bolo, qualquer seja o domínio de interesse, as pessoas consultam pessoas a fim de satisfazer suas necessidades de informação. As redes sociais são o ponto de partida para um sistema de recomendação que utiliza abordagem social.

Em qualquer rede social, em qualquer domínio de atuação dos sistemas de recomendação, o usuário se beneficia do julgamento feito por outros que já consultaram determinada informação; portanto, trabalhando com agrupamentos de usuário é mais provável que uma informação adequada, mais relevante, chegue a seu conhecimento. Com auxílio das

técnicas de mineração de dados, a eficiência das recomendações pode ser percebida estatisticamente (SCHAFER et al., 2001).

### **2.3.1. Importância das Redes Sociais**

Mesmo antes de a Internet se tornar uma ferramenta para a recuperação de informação, as redes sociais já atuavam como formadoras de opinião. E ainda hoje, essa rede de informações está disponível e é muito consultada. Como foi dito, na economia as redes sociais guiam os consumidores através de opiniões e existem para todos os tipos de bens e serviços. Somos seres sociais e antes de termos uma opinião formada sobre qualquer assunto precisamos nos informar, e a consulta à opinião de outros, principalmente se a origem tem forte reputação, facilita o entendimento sobre o assunto de interesse.

Com o apoio da tecnologia da informação, as redes sociais estão aumentando seu poder de atuação através dos sistemas de recomendação. Não só é possível consultar o seu vizinho ou o professor na sala de aula, não só é possível consultar o jornal ou guias de uma banca de revistas, mas também repositórios de informação armazenados em *sites* espalhados pela Internet. Por isso as redes sociais estão inseridas no domínio da tecnologia da informação — o orkut<sup>28</sup>, o candango<sup>29</sup>, o correioweb<sup>30</sup> ou qualquer outro sistema de recomendação. As redes sociais permitem que as pessoas encontrem com mais facilidade e mais rapidamente a informação desejada.

Na crescente demanda por informação, onde somos atacados diariamente com dados de todo tipo, é muito difícil selecionar o que é relevante. Nesse contexto, as recomendações se tornam essenciais. Para filtrar o que é importante, nada como uma boa indicação. E isso é o ponto forte dos sistemas de recomendação: onde atuam, auxiliando as pessoas a identificar o que elas realmente precisam.

---

<sup>28</sup> Disponível em <<http://www.orkut.com>>

<sup>29</sup> Disponível em <<http://www.candango.com.br>>

<sup>30</sup> Disponível em <<http://www.correieweb.com.br>>

### **3. Plataforma Lattes**

A Plataforma Lattes é a base de dados de currículos e instituições das áreas de Ciência e Tecnologia (CNPQ, 2005). Para acessar essa base, utilizamos o Sistema Currículo Lattes<sup>31</sup>. O Currículo Lattes é o componente da Plataforma Lattes desenvolvido para o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) como sistema de informação curricular.

Os usuários desse sistema são pesquisadores, estudantes, gestores, profissionais e demais atores no cenário de Ciência e Tecnologia. No CNPq, suas informações são aplicadas na avaliação da competência de candidatos à obtenção de bolsas e auxílios, na seleção de consultores, membros de comitês e grupos assessores e, por fim, no subsídio à avaliação da pesquisa e da pós-graduação brasileiras.

Para que os objetivos possam ser alcançados, o CNPq decidiu que todos os bolsistas (de pesquisa, mestrado, doutorado e iniciação científica), orientadores credenciados e outros clientes tenham um currículo Lattes cadastrado no CNPq. A inexistência do currículo impedirá pagamentos e renovações. O currículo é também obrigatório para todos os pesquisadores e estudantes que participam do Diretório de Grupos de Pesquisa no Brasil. Os sistemas ou formulários eletrônicos do currículo Lattes estão disponíveis no *site* permanentemente, podendo os usuários atualizá-los a qualquer momento.

#### **3.1. Histórico<sup>32</sup>**

De 1993 a 1999, o CNPq utilizou formulários em papel, sistema em ambiente DOS e sistema de currículos específicos para credenciamento de orientadores. Nesse período, a Agência acumulou cerca de 35 mil registros curriculares da atividade de Ciência e Tecnologia do País. Embora os instrumentos tenham viabilizado a operação de fomento da Agência, a natureza das informações dificultava a completa utilização dessa operação

---

<sup>31</sup> O sistema pode ser acessado em <<http://buscatextual.cnpq.br/buscatextual>>

<sup>32</sup> fonte: Histórico da Plataforma Lattes, disponível em <[http://lattes.cnpq.br/conheca/con\\_hist.htm](http://lattes.cnpq.br/conheca/con_hist.htm)>

em outros processos de gestão em Ciência e Tecnologia (por exemplo, não era possível separar co-autores ou mesmo contabilizar índices de co-autoria nos currículos).

Entre 1998 e 1999, o CNPq realizou levantamento visando estabelecer um modelo de currículo que atendesse tanto às suas necessidades de operação de fomento como de planejamento e gestão em Ciência e Tecnologia. Além disso, o grupo de desenvolvimento Grupo Stela da UFSC incluiu no formulário eletrônico diversas funcionalidades há muito solicitadas pela comunidade científica, tais como relatórios configuráveis, saída para outras fontes, indicadores de produção, dicionários individualizados, importação dos dados preenchidos em outros sistemas de currículos, dentre outros.

Entre março e abril de 1999, 140 dos 400 consultores que responderam à pesquisa avaliaram o primeiro protótipo do currículo Lattes. A avaliação geral alcançou 4.5 em escala de 0 - péssimo a 5 - excelente.

Em maio de 1999, CNPq e CAPES<sup>33</sup> acordaram completa compatibilização do novo currículo do CNPq com os dados de pós-graduação, sob a ótica dos indivíduos de um Programa (pesquisadores, docentes ou discentes). O encontro entre as agências resultou na modificação do protótipo, que se transformou no Sistema de Currículos Lattes e foi lançado a 16 de agosto de 1999.

Nos dois primeiros anos do Sistema de Currículos Lattes, a cobertura de currículos ligados a Ciência e Tecnologia aumentou em mais de 300%, com a base anterior de cerca de 35.000 registros sendo incrementada para mais de 100 mil currículos.

Em julho de 2000, a Coordenação Geral de Informática do CNPq iniciou um trabalho de intercâmbio com outras instituições ligadas a Ciência e Tecnologia no País. O resultado foi a ligação dinâmica dos currículos Lattes do CNPq com referência ao mesmo pesquisador em outras bases de dados.

---

<sup>33</sup> Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Consulte o glossário para mais informações.

Ao mesmo tempo que construiu o formulário *off-line*, a Coordenação Geral de Informática do CNPq também trabalhou na ferramenta *on-line*, que funciona sobre uma plataforma *Web* e permite que os pesquisadores atualizem os seus currículos diretamente na base do CNPq.

Nesse trabalho de intercâmbio, o CNPq vinculou os currículos Lattes com: INPI, para apresentação dinâmica das patentes de registro dos pesquisadores; com SCIELO<sup>34</sup>, LILACS<sup>35</sup>, MEDLINE<sup>36</sup> (fruto de acordo com BIREME<sup>37</sup>), para leitura dos textos completos publicados pelos pesquisadores (e para vínculo com os currículos dos co-autores); com as universidades, para vínculo com bases institucionais desses pesquisadores.

No ano de 2000, as Instituições Federais de Ensino Superior reuniram suas equipes de informática no Workshop de Sistemas de Informações das IFES (UFOP - Ouro Preto) e convidaram as agências federais para construção de um modelo único de informação, visando racionalizar o processo de captura de dados no Sistema Federal de Educação em Ciência e Tecnologia.

Na ocasião, o CNPq prontificou-se a construir projeto específico para atender a essa demanda, mas salientou a necessidade de manter a confiabilidade das informações (e a Plataforma operacional) dos pesquisadores, dado que estas são o principal subsídio ao processo de fomento.

Em fevereiro de 2001, UFSC, UNICAMP, UFRJ, USP, UFRGS, UFBA e UFRN, universidades que haviam procurado o CNPq solicitando abertura tecnológica de sua plataforma, participaram de workshop na Agência, visando à construção da Linguagem de Marcação da Plataforma Lattes (LMPL), sob coordenação da CGINF/CNPQ, sendo os trabalhos de desenvolvimento conduzidos pelo Grupo Stela da UFSC.

---

<sup>34</sup> *Scientific Electronic Library*, projeto de biblioteca eletrônica da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

<sup>35</sup> Literatura Latino-americana e do Caribe em Ciências da Saúde

<sup>36</sup> Base de dados da U.S. *National Library of Medicine*

<sup>37</sup> Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde

Desse encontro, resultou a formação da Comunidade Virtual LMPL, que definiu o modelo DTD (*Data Type Definition*) XML do Currículo Lattes. Com esse modelo, as universidades brasileiras podem extrair informações do currículo Lattes e/ou gerar informações para o mesmo a partir dos seus sistemas corporativos. O projeto viabilizou a abertura da Plataforma Lattes, do ponto de vista de conteúdo dos dados, e manteve inalterado o acesso técnico às informações, preservando a segurança dos pesquisadores.

### **3.2. Funcionalidades**

Com o propósito de manter conhecimento atualizado da realidade de seu campo de trabalho, o CNPq se esforça para integrar suas bases de informações. Essa integração tem como fonte primária de coleta de dados quatro projetos distintos, porém integrados (CNPQ, 2005).

O primeiro e principal deles, e objeto de estudo desta dissertação, se refere a um sistema eletrônico de currículos, o Currículo Lattes. O registro da vida pregressa e atual dos pesquisadores é elemento fundamental para a análise de seu mérito e competência. Nesta base constam mais de 200 mil currículos atualizados (CNPQ, 2005).

O segundo sistema é o Diretório dos Grupos de Pesquisa no Brasil. O Diretório é uma base de dados que registra todos os grupos de pesquisa em atividade no País (CNPQ, 2005). As informações constantes na base de dados dizem respeito aos recursos humanos engajados no grupo, às linhas de pesquisa em andamento, às especialidades do conhecimento, aos setores de aplicação, aos cursos de mestrado e doutorado com os quais os grupos interagem e à produção científica e tecnológica captada a partir do sistema eletrônico de currículos.

O terceiro sistema é o Diretório de Instituições (CNPQ, 2005). O registro acurado das instituições é fundamental para a disponibilização de um mapa preciso da distribuição de recursos e da localização da competência de pesquisa e desenvolvimento no País e no exterior.

O quarto sistema chama-se Sistema Gerencial de Fomento. Este sistema é imprescindível para uma gestão estratégica e para dar mais qualidade às atividades de fomento do CNPq (CNPQ, 2005).

Esses quatro sistemas de informação integrados, articulados com outras bases de dados localizadas fora da Agência, constituem a Plataforma Lattes.

### **3.2.1. Recuperação de Currículos**

O Sistema Currículo Lattes disponibiliza duas formas de busca: a busca simples e a busca avançada.

A busca simples permite que a pesquisa seja realizada de três maneiras: na base de dados de Doutores, na base de dados dos demais pesquisadores ou selecionando ambas opções. Essa busca pode ser feita pelo nome do pesquisador e/ou por assunto, onde assunto refere-se a título e palavras-chave da produção científica.

A busca avançada permite que a pesquisa seja realizada nas três maneiras descritas na busca simples, porém quando vamos preencher nome e/ou assunto, são disponibilizados campos para que a consulta seja construída com “todas as palavras”, com “esta frase exata”, com “qualquer uma dessas palavras” e “nenhuma dessas palavras”. Se o usuário possui conhecimento de álgebra booleana, há também disponível a construção da pesquisa com “esta expressão booleana” (e um aviso informando que expressão booleana usa termos tais como AND, OR e NOT).

Nos resultados da busca avançada podem ser aplicados os seguintes filtros, mostrados na Figura 19:

**Aplicar filtro aos resultados por:**



Figura 19 - Filtros disponíveis para a busca avançada do Currículo Lattes

Fonte: Currículo Lattes < <http://buscatextual.cnpq.br/buscatextual>>, 2005.

Cada um dos filtros permite configurações próprias, mostradas abaixo:

- Filtrar por pesquisador do CNPq: este filtro restringe a busca ao universo de Pesquisadores detentores de Bolsa de Produtividade em Pesquisa concedida pelo CNPq e atualmente vigente. Caso deseje selecionar uma Categoria / Nível de bolsa específica, desmarque as demais.
- Filtros relativos à formação: este filtro permite restringir a busca a currículos que possuam uma formação acadêmica específica, por nível, país, cidade, região ou UF<sup>38</sup> da formação<sup>39</sup>.
- Filtros relativos à área de atuação<sup>40</sup>: este filtro permite restringir a busca a currículos que possuam uma área ou grande área específica informada no módulo de áreas do conhecimento.
- Filtrar por idioma: este filtro permite restringir a busca a currículos que possuam determinada proficiência em idiomas.

<sup>38</sup> Unidade Federativa. Ex: Goiás, Distrito Federal, Pará, Pernambuco.

<sup>39</sup> Permite também filtrar por nome da Instituição, embora não descrito na ajuda do sistema.

<sup>40</sup> O Sistema precisa ser revisado, o texto informativo do CV Lattes é "filtros relativos ao idioma e à área de atuação", mas idioma possui um filtro próprio.

- Filtrar por atividade profissional: este filtro permite restringir a busca a currículos cujo profissional tenha atuado em uma instituição, UF, região ou país em determinado tipo<sup>41</sup> de atividade.
- Filtros relativos aos bolsistas do CNPq: restringe a busca a bolsistas do CNPq de uma modalidade específica.
- Nível do curso de pós-graduação onde é docente: este filtro permite restringir a busca a docentes atuando em cursos avaliados pela CAPES com um determinado conceito<sup>42</sup>.
- Filtrar por atividade de orientação: este filtro permite restringir a busca a orientadores de bolsistas do CNPq em diferentes níveis.
- Filtros relativos à produção C,T&I: permite restringir a busca a pesquisadores que tenham publicado ou produzido em determinada área, grande área ou setor de aplicação.
- Filtrar por presença no diretório de pesquisa: permite restringir a busca a pesquisadores que estejam registrados na base corrente do diretório de grupos de pesquisa, em grupo devidamente reconhecido pelo dirigente institucional competente.

E, por fim, a busca avançada permite selecionar a ordenação dos resultados, que pode ser por “nome” ou por “score”.

### **3.2.2. Outras Funcionalidades**

Há outras funcionalidades além da busca de currículos, tais como atualizar currículo, cadastrar-se na Plataforma Lattes, buscar instituição, atualizar instituição e cadastrar instituição. Nenhuma dessas funcionalidades

---

<sup>41</sup> O tipo ou natureza da atividade pode ser “conselhos, comissões e consultorias”, “direção e administração”, “ensino”, entre outros. Os termos estão disponíveis para a realização da busca.

<sup>42</sup> Os campos disponíveis para seleção de conceito variam de 1 a 7.

serve ao propósito desta dissertação, por isso não serão descritas nem estudadas aqui.

### **3.3. Carências do Sistema Currículo Lattes**

Embora as funcionalidades de recuperação de informação do Currículo Lattes sejam úteis e vão ao encontro da proposta para a qual foram desenhadas, o sistema não oferece respostas eficientes às necessidades de recuperação de informação que, com o uso de sistemas de recomendação, poderiam atender à necessidade do usuário.

Uma dessas deficiências pode ser descrita na situação a seguir: ao encontrar o currículo de um profissional que seja de interesse, o sistema não indica outros usuários com potencial semelhante, seja estudando na mesma linha de pesquisa, seja com produção semelhante, ou outras inúmeras possibilidades. Não há um agrupamento de usuários. Toda busca é única e só é possível haver um “agrupamento” se realizado através de filtros da busca avançada, o que não traz nenhuma vantagem para o sistema, apenas serve para apresentação dos resultados.

Outra deficiência é a impossibilidade de avaliação dos currículos pelos usuários. Sem avaliações, a situação hipotética é de difícil solução: localizando três perfis de potencial semelhante, qual desses profissionais seria o mais indicado a convidar para participar de um projeto?

Como não é possível haver *feedback* do usuário, não há possibilidade de um tratamento diferenciado dos resultados apresentados, nem é possível aprimorar o sistema para consultas futuras. O sistema tem uma estrutura rígida e não adaptável de recuperação de informação.

A busca do Currículo Lattes também não privilegia pesquisadores tradicionais, nem possibilita o inverso, que seria privilegiar pesquisadores de ponta quando se trata da busca por profissionais que trabalham com uma tecnologia nova. No caso da tecnologia nova, por exemplo, seria preciso avaliar dentre os currículos mais atuais os que trabalham com tal tecnologia

e que receberam melhores recomendações na área de interesse depois do surgimento de tal tecnologia.

Outro problema que seria amenizado com o agrupamento de usuários seria a classificação dos níveis de usuário para, dessa forma, gerar recomendações de acordo com o interesse do grupo. Essa classificação para gerar recomendações poderia ser diferenciada por nível de escolaridade (graduação, mestrado, doutorado...), por nível profissional (diretoria, gerência, sênior, pleno, júnior, *trainee*...), ou por qualquer outra forma de classificação — a flexibilidade é uma das características que tornam interessantes os sistemas de apoio à decisão.

## **4. Metodologia da Pesquisa**

De acordo com a classificação adotada por Silva e Menezes (2001, p.20-22) a pesquisa realizada é exploratória, pois visou proporcionar maior familiaridade com o problema com vistas a torná-lo explícito, e envolveu a apresentação de exemplos para estimular a compreensão. Do ponto de vista dos procedimentos técnicos, podemos classificá-la como estudo de caso — houve um estudo profundo de alguns objetos de forma a detalhar o conhecimento. A pesquisa gerou conhecimentos para aplicação prática dirigidos à solução de problemas específicos e, como o processo e seu significado são os focos principais de abordagem, esta pesquisa foi predominantemente qualitativa, muito embora haja variáveis quantitativas trabalhadas no modelo geral.

O trabalho consistiu na elaboração de um modelo geral de sistema de recomendação de perfis de usuário. Primeiramente foi realizada uma pesquisa em recuperação de informação, com ênfase nas abordagens de recomendação, identificando conceitos importantes e técnicas que se aplicassem ao projeto. O levantamento bibliográfico resultou na identificação de pesquisas sobre o tema e de ferramentas que utilizam recomendações para a recuperação de informação.

Dentre as técnicas estudadas, foi definido que o melhor algoritmo de construção de um sistema de recuperação de informações baseado em recomendações deveria adotar uma abordagem híbrida de filtragem baseada no conteúdo e filtragem colaborativa. Com a utilização conjunta de ambas as técnicas de programação, diversas abordagens de recomendação poderiam ser aplicadas e foram escolhidas as que melhor se adaptavam ao domínio de recuperação de informação de perfis de usuário.

Para propor o modelo, objetivo deste projeto, estudou-se um sistema que trabalha com perfis de usuário que é amplamente conhecido no

meio acadêmico, o sistema Currículo Lattes, um dos componentes da Plataforma Lattes<sup>43</sup>.

Foram identificadas muitas situações onde os ganhos com o uso de recomendações seriam perceptíveis a ponto de justificar a pesquisa. Além dessas, há outras situações no contexto de recuperação de perfis de usuário que nos fornecem mais material de trabalho: a necessidade de localização de perfis para montar grupos de projeto para a concorrência em editais, como no caso dos editais do Ministério da Ciência e Tecnologia; ou ainda o caso de um profissional de educação física que precise localizar outros profissionais de perfis semelhantes para convidar a montar uma sociedade em uma academia de ginástica.

Outros cenários foram identificados na recuperação de informações para o modelo genérico, que poderiam ser solucionados através do uso de recomendações. Tais cenários forneceram uma base para a melhoria do modelo tradicional baseado em busca booleana. A nova proposta é flexível o suficiente para ser aplicada em qualquer sistema de recuperação de perfis de usuário, utilizando indicadores que não sejam aplicados somente na recuperação de currículos.

Por isso, a fim de verificar se a proposta foi alcançada, o modelo geral foi aplicado em um domínio diferente, o da recuperação de perfis de usuário em um *site* de relacionamentos. Os indicadores identificados no levantamento de requisitos para o sistema genérico — sistema de recomendação de perfis de usuário — foram correlacionados com os indicadores da recuperação de currículos. A implementação e funcionamento do modelo de recomendação proposto em um *site* de relacionamentos teve por objetivo demonstrar a sua aplicação real. Como os indicadores são diferentes em cada subdomínio, foram estabelecidos alguns rótulos possíveis para o modelo genérico.

Um indicador genérico poderia ter o rótulo “indicador de eficiência”. Na recuperação de currículos acadêmicos, poderíamos utilizar

---

<sup>43</sup> Disponível em <<http://lattes.cnpq.br/>>

pontuações atribuídas ao número de publicações para formalizar o indicador de eficiência nesse contexto — o sistema entenderia que quanto mais o profissional publica, mais interessado ele está em realizar pesquisas. Na recuperação de perfis em um *site* de relacionamentos, o mesmo indicador poderia ser pontuado pela estatística de uso do sistema pelos usuários — o sistema entenderia que uma maior frequência de uso da ferramenta implicaria em maior interesse por encontrar um relacionamento. Estes dois exemplos são apenas para mostrar como um indicador genérico pode ser aplicado em diferentes casos de uso.

A flexibilidade na utilização dessas variáveis — os indicadores — é que tornou o modelo genérico reutilizável, quer dizer, o modelo poderá ser usado em diferentes situações, segundo as especificações de cada sistema de perfis de usuário, ou mesmo em outros sistemas de recomendação de informação, desde que bem determinadas as variáveis correlacionadas.

Com a aplicação real em funcionamento, a tarefa seguinte foi transpor tal aplicação para o modelo genérico e, em seguida, transpor o modelo para o Currículo Lattes, demonstrando que a generalização do modelo é realmente funcional e adaptável.

O modelo geral é o resultado do trabalho e deve servir de base para o desenvolvimento de novos sistemas.

Foi escolhido o *site* de relacionamentos pela impossibilidade de validar a aplicação baseada na recomendação de currículos junto ao CNPq, conforme será explicado na Seção 6.3.1 (p.133).

Como se trata de uma aplicação baseada na *Web*, optou-se pela linguagem de programação PHP com banco de dados MySQL. A escolha de PHP deve-se ao fato de que se trata de uma linguagem extremamente modularizada — o que a torna ideal para instalação e uso em servidores *web* —, livre<sup>44</sup>, pode ser trabalhada com recursos de orientação a objetos e

---

<sup>44</sup> Software livre é qualquer programa de computador que pode ser usado, copiado, estudado, modificado e redistribuído com algumas restrições.

possui forte suporte a várias bases de dados, o que torna simples a construção de páginas dinâmicas.

A escolha de MySQL deve-se primeiramente ao fato da facilidade de integração com PHP e da grande quantidade de tutoriais grátis disponíveis para programação. Fora isso, é um banco portátil, possui excelente estabilidade, exige poucos recursos de hardware, é reconhecido pelo seu desempenho e robustez e também por ser multitarefa e multiusuário. É possível utilizá-lo em sistemas de produção de alta exigência e em aplicações sofisticadas.

Para finalizar, este trabalho confirmou que as recomendações são úteis na recuperação de perfis de usuário, através da demonstração de que um modelo genérico de recomendação de perfis de usuário pode encontrar resultados que não seriam alcançados por um modelo tradicional de sistema de recuperação de informação.

## 5. Desenvolvimento da Pesquisa

Para solucionar as situações descritas na Seção 3.3, relativas às carências do sistema Currículo Lattes, é necessário um modelo de sistema de recuperação de informação que utilize estratégias de recomendação. A idéia é coletar, medir e avaliar indicadores de interesse, implícitos e explícitos.

Na busca de currículos, cada resultado selecionado segue com uma coluna lateral que apresenta currículos semelhantes ao apresentado. Desta forma, o sistema estaria se realimentando para apresentar novos resultados relevantes baseando-se na pesquisa original e no currículo selecionado. Exigindo que o usuário se autentique no sistema para realizar a busca, pode-se, então, aprimorar a consulta para a busca de profissionais na mesma linha de pesquisa ou com produção semelhante à sua. Essa busca por afinidade é completamente diferente da busca original realizada pelo sistema tradicional.

Havendo a funcionalidade que permita ao usuário avaliar os resultados de sua busca, o sistema pode ser melhorado para apresentar resultados mais relevantes para consultas futuras deste usuário. As opiniões acerca dos currículos podem ser explícitas — um campo de agradecimentos e elogios a determinados currículos, assinalar currículos de usuários que estejam relacionados à minha área de interesse — ou implícitas — currículos mais clicados, currículos mais buscados. O risco de distorção<sup>45</sup> da avaliação invisível é baixo no Currículo Lattes, pois acredita-se que esta ferramenta deve contar com um número reduzido de perfis falsos ou duvidosos, uma vez que é gerida por um órgão do governo e destinada a uma comunidade de profissionais e pesquisadores da área de ciência e tecnologia e, além disto, a ampla visibilidade dadas às informações contribui para inibir informações falsas.

---

<sup>45</sup> A fim de garantir a imparcialidade das recomendações, seria disponibilizada uma forma de avisar sobre perfis falsos ou duvidosos — por exemplo, um botão de denúncia no topo do currículo.

Com a coleta das informações explícitas sobre currículos que tenham a ver com a área de interesse, seria possível criar um *ranking* de recomendação de currículos nas diversas áreas, e as informações implícitas ajudariam a balancear esse *ranking*, de forma que o sistema possa realizar recomendações mais eficientes. Os agradecimentos e elogios mostrados<sup>46</sup> nos currículos ajudariam quem realizasse a busca a conhecer uma faceta social do resultado apresentado, podendo assim inferir: se esse profissional trabalhou com o outro renomado, e o renomado demonstrou explicitamente que o profissional realizou um bom trabalho, o usuário tem agora mais confiança em convidar esse profissional a participar de seu projeto.

Embora muitas das recomendações possam ser realizadas automaticamente, permitir que o usuário interaja com os critérios de recuperação é fundamental para que ele confie nas indicações apresentadas, trazendo maior satisfação e aceitação dos resultados alcançados. Assim, explicar porque um currículo recebeu melhor avaliação que outro não seria responsabilidade do sistema, mas dos critérios de relevância elaborados pelo próprio usuário. Muitas recomendações são elaboradas por coletas anônimas das consultas realizadas, e informar a origem extrapolaria os limites de privacidade<sup>47</sup>. Ainda mais porque o sistema poderia aproveitar para dar mais importância às pesquisas realizadas por currículos de profissionais de reputação<sup>48</sup> mais forte.

Deve ficar sempre estabelecida a diferença entre uma busca casual, para uma demanda de momento, ou aquela que o usuário pode não estar pesquisando para interesse próprio, e uma busca realizada com interesse objetivo, como um projeto no qual esteja trabalhando ou aquelas diretamente relacionadas à sua linha de pesquisa. Para assegurar que a coleta implícita armazenará resultados válidos, o sistema deve diferenciar a interface de busca em dois segmentos. O primeiro é para a busca casual, onde os resultados apresentados não serão aproveitados de forma implícita

---

<sup>46</sup> Quando o usuário X escreve um agradecimento ou elogio para o usuário Y, Y pode aceitar ou recusar este depoimento. O depoimento só será disponibilizado após o aceite de Y.

<sup>47</sup> Uma nova política de privacidade precisa ser elaborada para informar ao usuário que informações anônimas de suas consultas ao sistema podem ser coletadas a fim de aprimorar o funcionamento do sistema de recomendação.

<sup>48</sup> Os critérios de reputação de um pesquisador são baseados em indicadores flexíveis, como o número de artigos publicados, se lidera grupos de pesquisa, dentre outros.

para aprimoramento do sistema de recomendação. O segundo poderia ser chamado de “busca pré-definida”, onde os termos de consulta utilizados pelo usuário ficam armazenados em banco permitindo que recomendações possam ser realizadas mesmo que o usuário não esteja usando o sistema. Suponha que determinado perfil de profissional seja desejado, porém nada que esteja cadastrado no Currículo Lattes atende ao usuário no momento da consulta. Com os termos da busca pré-definida, a consulta pode ser realizada em um intervalo de tempo regular por um robô e, se o currículo de novos usuários atendem ao perfil que estava sendo procurado, recomendações podem ser enviadas por *email* a quem procurava ou apresentadas a este usuário na próxima vez que se autentica no sistema — poderíamos chamar de “busca *off-line*”.

Além de permitir a funcionalidade de busca *off-line*, o armazenamento dos termos de consulta da busca pré-definida permitirá distinguir grupos de usuários, tornando viáveis as recomendações por afinidade.

Aproveitando o exemplo da pesquisa de preços em supermercados (p.45), vamos demonstrar a vantagem da busca pré-definida. Digamos que queremos separar os clientes que gostam de banana dos que gostam de limão para recomendar o supermercado que esteja vendendo banana mais barato a quem goste de banana. Agora temos os clientes que gostam de frutas e mais especificamente os que querem banana e os que querem limão. Através de análise às consultas do grupo que gosta de banana, técnicas de *data mining* identificam que este grupo sempre pesquisa também por “farinha láctea”. O sistema poderá recomendar aos novos usuários que queiram bananas o produto farinha láctea, antes mesmo que esse usuário realize qualquer consulta ao sistema. Basta se autenticar e preencher suas preferências na busca pré-definida.

As recomendações explícitas, como quando um usuário avalia se um currículo é ou não relevante para sua linha de pesquisa, são coletadas sempre. Os elogios, expressos em forma de depoimentos que identificam a

fonte, como na Figura 20, ficam sempre disponíveis para acesso. Um cuidado que deve ser tomado é que depoimentos sempre são tendenciosos e que as pessoas tendem a não permitir um depoimento negativo. Porém, um depoimento serve para destacar determinado produto quando a fonte tem boa reputação. É um indicador qualitativo e não quantitativo.



The screenshot shows the 'Central de Negócios: A melhor maneira de aumentar suas vendas!' header with the BuscaPé logo and website URL. Below the header, there are navigation links: 'Voltar para: Home > Depoimento de empresas' and 'Já é cadastrado? Acesse sua conta'. The main heading is 'Depoimentos de empresas sobre o BuscaPé'. On the left, a sidebar menu lists: 'Faça um tour', 'Resultados de busca', 'Como participar?', 'FAQ', 'Depoimentos de Empresas', 'Cadastre-se agora', 'O BuscaPé Indica', and 'Fale conosco'. The main content area features two testimonials. The first is from Silvana Dias de Andrade, Gerente Marketing e atendimento at 'Fábrica de Emoções', with a quote praising BuscaPé's clarity, service, and professionalism. The second is from Walter Rodrigues Torres, Diretor Comercial at 'Acessórios Car', with a quote expressing enthusiasm for the sales results and the team's performance.

Figura 20 - Depoimentos de empresas sobre o Buscapé

Fonte: site Buscapé <[https://cn.buscape.com.br/cn\\_br\\_home/Default\\_DepoimentoEmpresas.aspx](https://cn.buscape.com.br/cn_br_home/Default_DepoimentoEmpresas.aspx)>, 2006.

Comparar currículos que recebem muitas indicações com outro que recebe poucas é um dos cuidados que o administrador do sistema deve tomar. De forma a minimizar os prejuízos que a relevância automática pode causar ao lidar com essa situação, o sistema deve deixar para o usuário a responsabilidade sobre a temporalidade dos resultados. Basta permitir que ele decida se deseja ou não receber indicações pelas avaliações mais recentes, e assim ordenar os resultados de acordo com o percentual das pontuações no caso de decidir que deseja as avaliações mais recentes. Um outro critério importante de temporalidade é definir como apresentar primeiro

os currículos que foram atualizados recentemente, em vez dos que estão há tempos sem receber atualização.

É importante evidenciar como uma quebra de paradigma pode influenciar a decisão do usuário: se o desejo é recuperar um perfil mais tradicional, então é melhor receber resultados que mostrem todas as avaliações que os currículos receberam, isto é, o critério de temporalidade deve cobrir tanto avaliações antigas quanto as recentes. Se o desejo é recuperar perfis mais recentes, o que ocorre com frequência quando há uma quebra de paradigma em tecnologia<sup>49</sup> e precisamos dos currículos mais atuais de profissionais de informática, o critério de temporalidade deve cobrir apenas as avaliações mais recentes, haja vista que, quando as avaliações mais antigas foram realizadas, os requisitos que deveriam estar presentes nos currículos dos profissionais sequer existiam.

De acordo com o perfil do usuário, poderemos criar classes variadas de níveis, de forma a selecionar a recomendação de e para cada nível nas afinidades dos grupos identificados a partir das informações coletadas do cadastro, das preferências da busca pré-definida e do comportamento aprendido no uso do sistema.

## **5.1. Modelo Geral de Recomendação de Perfis de Usuário**

Para a construção do modelo geral de recomendações de perfis de usuário, os pontos identificados através do estudo das carências do sistema Currículo Lattes precisam ser generalizados, criando os indicadores genéricos. O que chamamos de indicadores genéricos são representações de variáveis que podem assumir diferentes rótulos nos diferentes domínios de perfil de usuário — seriam os identificadores flexíveis que permitirão que este modelo geral possa ser aplicado em qualquer domínio de sistema de recuperação de perfis de usuário, não só na recuperação de currículos.

---

<sup>49</sup> Por exemplo, na busca por programadores que saibam trabalhar com Java para tecnologias móveis, avaliações mais antigas na área de informática obviamente não cobririam esses requisitos porque eles não existiam antes de 2001 (SUN MICROSYSTEMS, 2006).

Por exemplo, digamos que na recuperação de currículos o usuário pode avaliar de forma binária — sim ou não — se o currículo de um outro usuário apresentado no resultado de busca é ou não afim a sua área de atuação. Em um outro domínio, como em um *site* de relacionamentos, o usuário pode avaliar da mesma maneira binária se o perfil de uma outra pessoa, apresentado em um resultado de busca, atende ou não à sua necessidade. Embora seja totalmente diferente o objetivo da avaliação nos cenários apresentados, para um modelo de sistema este cenário pode ser representado de forma única: a variável em questão é um artefato que será avaliado de forma binária, e o resultado das avaliações dos usuários será armazenado para aprimorar o conhecimento do sistema acerca de cada um dos grupos de usuário.

No Currículo Lattes, duas variáveis disponíveis poderiam ser trabalhadas para criar os indicadores de recomendação:

- áreas de atuação: lista de áreas de conhecimento cujo pesquisador possui interesse e atua (veja Figura 21).
- indicadores de produção: conjunto de informações que representam as diversas atividades acadêmicas, tais como: produção científica, número de orientações concluídas, e participações em eventos.

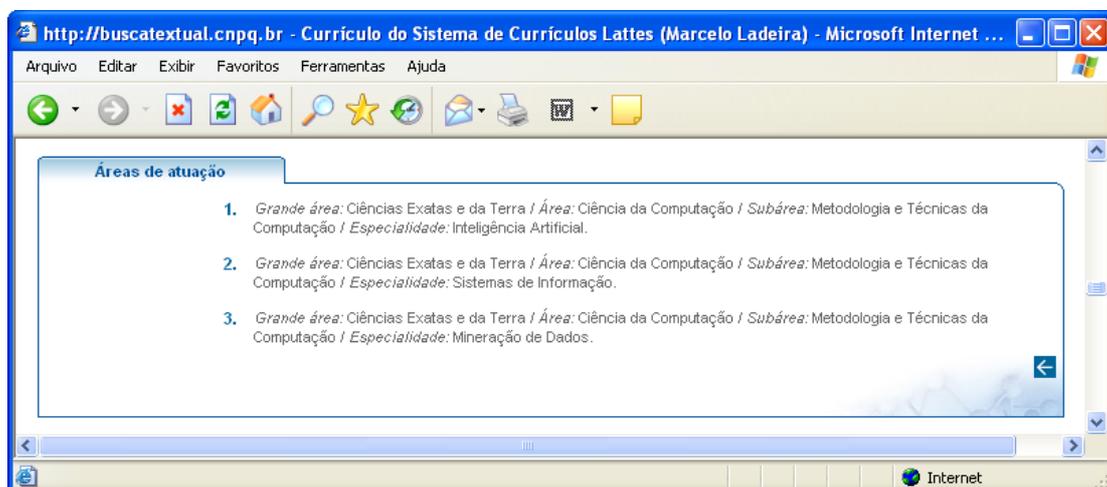


Figura 21 - Áreas de atuação de um currículo do sistema Currículo Lattes

Fonte: Consulta ao sistema Currículo Lattes < <http://buscatextual.cnpq.br/buscatextual> >, 2006.

Para utilizar os indicadores de forma quantitativa, isto é, de maneira que o sistema de recomendação possa trabalhar matematicamente, precisamos normalizar os valores expressos pelas variáveis. Nesta direção, faremos uso do estudo realizado por Cazella e Álvares (2005), onde desenvolveram uma métrica para representar a relevância da opinião do usuário. Esta métrica, chamada *Ranking* de Recomendação (RR), pretende mostrar quanto um usuário conhece sobre um item específico, em uma escala de 0 a 10.

Segundo Cazella e Álvares (2005, p.3), pessoas que ocupam uma posição mais alta em uma carreira específica geralmente têm opiniões mais bem consideradas pela comunidade. No exemplo, eles citam um maestro de uma orquestra de renome. Esta pessoa está no topo e não há dúvidas sobre sua competência devido a sua capacidade e conhecimento sobre sua área de atuação. Desta forma, a opinião dele é consagrada sobre coisas de sua área, e ainda mais relevante que a opinião de pessoas que não estão em seu campo de atuação. Se o maestro recomenda um conteúdo a alguém, esse alguém deveria dar mais importância a essa recomendação.

Equivalentemente, o exemplo pode ser aplicado no domínio de currículos de usuários e suas linhas de pesquisa. Um sistema pode prever o interesse de um usuário sobre um artigo baseando-se na opinião de outros, mas se esses outros não têm muita experiência ou conhecimento sobre a área que o artigo foi publicado. Pode ser que a opinião destas pessoas não seja tão relevante para o usuário, ainda que esses outros tenham avaliado itens da mesma forma que o usuário — usuários com idéias afins.

Recomendações de usuários com opiniões mais relevantes em determinada área são mais interessantes. Se o sistema detém conhecimento sobre a relevância das opiniões de quem recomenda, ele pode melhor avaliar e trabalhar as recomendações. A fim de representar a relevância da opinião dos usuários, Cazella e Álvares (2005) desenvolveram o *Ranking* de Recomendação.

Os critérios para medir a importância de uma recomendação variam de acordo com o domínio. Dessa forma, Cazella e Álvares (2005) desenvolveram um modelo geral baseado em variáveis relevantes, cada uma normalizada com um peso. Tais variáveis são selecionadas, de acordo com o domínio, de forma a representar um usuário. Como as variáveis podem ter diferentes escalas de valores, são normalizadas (transformação linear) conforme a equação abaixo:

$$an = \text{MinMax}(a) = \frac{a - \min A}{\max A - \min A} x(\text{new max } A - \text{new min } A) + \text{new min } A$$

Figura 22 - Equação para normalizar variáveis no RR

Fonte: Cazella e Álvares., 2005

Se  $\min A$  e  $\max A$  são os valores mínimo e máximo que uma variável pode assumir, a normalização  $\text{MinMax}(a)$  mapeia o valor  $a$  da variável  $A$  em um valor  $an$  na escala  $[\text{newmin}A, \text{newmax}A]$ . Essa normalização é essencial para manter a consistência de escala entre todas as variáveis.

Para calcular o *Ranking* de Recomendação (RR), Cazella e Álvares (2005, p.5) fizeram uso da equação abaixo:

$$RR = \frac{\sum_{i=1}^n an_i p_i}{\sum_{i=1}^n p_i}$$

Figura 23 - Equação do Ranking de Recomendação (RR)

Fonte: Cazella e Álvares., 2005, p.5

Esta equação é a média ponderada de todas as variáveis selecionadas para o *ranking* de recomendação — representadas por  $(an_1, an_2, \dots, an_n)$ . A equação converte toda a informação em um valor

quantitativo único em uma escala específica. O resultado é a relevância geral da opinião do usuário<sup>50</sup> (Cazella e Álvares, 2005, p.5).

### **5.1.1. Domínio e Indicadores**

Cada domínio possui as variáveis necessárias para a criação dos indicadores de recomendação. Na construção de um sistema de recomendação de perfis de usuário, precisamos identificar as variáveis qualitativas e quantitativas disponíveis, para podermos trabalhá-las nas equações demonstradas na seção anterior.

O primeiro passo é identificar se o domínio será trabalhado de forma integral ou será dividido em subdomínios. Se o domínio não é subdividido, o cálculo do RR é único e esse representa a relevância da opinião do usuário (autoridade). Se o domínio é subdividido, temos um  $RR_{\text{área}}$  para cada área/subdomínio e o  $RR_{\text{total}}$ , que representa o RR do domínio, será dado pela média dos  $RR_{\text{área}}$ . Um  $RR_{\text{área}}$  igual a zero indica que o usuário possui pouca experiência na área em questão. Um índice igual a, digamos, sete garante que o usuário possui experiência na área e um índice igual a dez indica profundo conhecimento da área.

---

<sup>50</sup> Domínios muito grandes podem ser subdivididos criando um  $RR_{\text{área}}$  para cada área específica. Isso ajuda o sistema de recomendação a trabalhar mais apropriadamente os recursos de relevância de opinião, fornecendo recomendações adequadas em cada subdomínio.

Tabela 2- Variáveis qualitativas extraídas do perfil do usuário

Variável	Valor
Nível ( $Nx_{USR}$ )	Para cada valor de $Nx_{USR}$ , há um peso específico associado
Áreas de Interesse ( $AI_{USR}$ )	Lista das áreas de interesse apontadas no perfil do usuário (se o domínio não é subdividido, a área de interesse é única)
Tempo de atuação na Área de Interesse ( $TAI_{USR}$ )	Variável medida em anos. Para cada área $AI_{USR}$ há um $TAI_{USR}$ correspondente.

Um domínio específico de perfil de usuário pode apresentar outras variáveis qualitativas próprias desse domínio. Para um modelo genérico, precisaremos sempre distinguir as áreas de interesse do usuário e seu nível, como representado na Tabela 2.

No domínio de recuperação de currículos, o nível do usuário pode estar tanto associado a sua experiência acadêmica quanto a sua experiência profissional, ou uma combinação de ambas. A forma de classificação é inerente ao sistema. Para exemplificar, foram criadas abaixo duas tabelas de valores de  $Nx_{USR}$  para a recuperação de currículos:

*Tabela 3 - Pesos associados aos valores qualitativos de nível acadêmico*

<b>NA<sub>USR</sub> (Nível Acadêmico)</b>	<b>Peso</b>
Estudante de graduação	1
Graduação	2
Especialista	3
Estudante de Mestrado	4
Mestrado	5
Estudante de Doutorado	6
Doutorado	9
Pós-doutorado	10

Tabela 4 - Pesos associados aos valores qualitativos de nível profissional

<b>NP<sub>USR</sub> (Nível Profissional)</b>	<b>Peso</b>
Auxiliar/Operacional	1
Técnico	2
Estágio	3
Júnior/ <i>Trainee</i>	4
Pleno	5
Sênior	6
Supervisão/Coordenação	7
Gerência	8
Diretoria	9
Presidência	10

Enquanto que para os níveis acadêmico e profissional apenas um valor pode ser obtido, as áreas de interesse do currículo de um usuário podem ser várias. Tomando como exemplo a Figura 21 (p.74), onde poderíamos trabalhar na recuperação de currículos da Ciência da Computação, teríamos três áreas de interesse para o usuário: Inteligência Artificial, Sistemas de Informação e Mineração de dados. Assim o sistema de recomendação colocaria este currículo como mais relevante nesses três subdomínios, ordenando os resultados segundo o  $RR_{\text{área}}$  de cada currículo.

Para o cálculo do  $RR_{\text{área}}$  é preciso trabalhar as variáveis  $AI_{\text{USR}}$  e  $TAI_{\text{USR}}$ , faremos isso logo a frente.

*Tabela 5 - Variáveis quantitativas extraídas do perfil do usuário*

Variável	Valor
Publicações em revistas ou Jornais ( $PRJ_{\text{USR}}$ )	Valor medido em unidades $PRJ_{\text{USR}}$
Artigos publicados em eventos ( $APE_{\text{USR}}$ )	Valor medido em unidades $APE_{\text{USR}}$
Livros e capítulos publicados ( $LCP_{\text{USR}}$ )	Valor medido em unidades $LCP_{\text{USR}}$
Supervisão de estudantes ( $SE_{\text{USR}}$ )	Valor medido em unidades $SE_{\text{USR}}$
Participação em bancas ( $PB_{\text{USR}}$ )	Valor medido em unidades $PB_{\text{USR}}$

Segundo Cazella e Álvares (2005, p.6), todas as métricas quantitativas devem ser normalizadas em uma escala específica, aplicando a equação para normalizar variáveis no RR (Figura 22, p.76). Os valores máximo e mínimo de cada variável devem ser obtidos da amostra.

Para ilustrar o uso, o modelo será aplicado em uma amostra fictícia de valores para currículos acadêmicos, conforme a Tabela 6. No cálculo da equação do RR (Figura 23, p.76), os pesos empiricamente atribuídos a nível acadêmico ( $NA_{\text{USR}}$  — Tabela 3, p.79) e para cada métrica de produção (Tabela 5, p.81) foram:  $NA_{\text{USR}}=2.0$ ,  $PRJ_{\text{USR}}=2.5$ ,  $APE_{\text{USR}}=1.5$ ,  $LCP_{\text{USR}}=2.5$ ,  $SE_{\text{USR}}=1.0$ ,  $PB_{\text{USR}}=0.5$ . Note que a soma dos pesos<sup>51</sup> é igual a 10. O valor mínimo atribuído a cada variável foi zero.

<sup>51</sup> Os pesos são de caráter ilustrativo. Os valores foram escolhidos arbitrariamente para demonstrar o exemplo.

Tabela 6 - Valores de currículos acadêmicos (não normalizados)

<i>Usuário</i>	$NA_{USR}$	$PRJ_{USR}$	$APE_{USR}$	$LCP_{USR}$	$SE_{USR}$	$PB_{USR}$
USR1	9	10	49	4	18	10
USR2	10	35	3	3	11	0
USR3	4	3	3	1	2	0
USR4	9	4	61	5	30	16
USR5	6	15	23	0	23	0
USR6	9	14	123	10	36	21
USR7	5	20	15	1	5	0
USR8	5	11	8	0	10	0

Na linha 3 está o usuário USR3, que é um estudante de mestrado, com 3 publicações em revistas ou jornais, 3 artigos apresentados em eventos, lançou 1 livro, supervisionou 2 estudantes e não participou de nenhuma banca. Após aplicar a equação para normalizar o RR, a tabela assume os valores a seguir.

Tabela 7 - Variáveis normalizadas para calcular o RR

<i>Usuário</i>	$NA_{USR}$	$PRJ_{USR}$	$APE_{USR}$	$LCP_{USR}$	$SE_{USR}$	$PB_{USR}$
USR1	9	2,9	4	4	5	4,8
USR2	10	10	0,2	3	3,1	0
USR3	4	0,9	0,2	1	0,6	0
USR4	9	1,1	5	5	8,3	7,6
USR5	6	4,3	1,9	0	6,4	0
USR6	9	4	10	10	10	10
USR7	5	5,7	1,2	1	1,4	0
USR8	5	3,1	0,7	0	2,8	0

Aplicando a equação do RR (Figura 23), com os pesos  $NA_{USR}=2.0$ ,  $PRJ_{USR}=2.5$ ,  $APE_{USR}=1.5$ ,  $LCP_{USR}=2.5$ ,  $SE_{USR}=1.0$ ,  $PB_{USR}=0.5$  citados anteriormente, chegamos ao resultado da Tabela 8:

Tabela 8 - Resultado do Ranking de Recomendação

<i>Usuário</i>	<i>RR<sub>resultado</sub></i>
USR1	4,8
USR2	5,6
USR3	1,4
USR4	5,3
USR5	3,2
USR6	8,3
USR7	3
USR8	2,2

O resultado apresentado na Tabela 8 já seria válido para a ordenação de currículos de Ciência da Computação em um sistema de recomendação baseado no conteúdo. Já é possível afirmar que, dentre todos os usuários deste domínio, a opinião de USR6 é a mais relevante. No entanto, ainda não estaríamos diferenciando os currículos por área de interesse. Suponhamos que os usuários tenham interesses descritos conforme a Tabela 9.

*Tabela 9 - Áreas de interesse de currículos da Ciência da Computação*

<b>Sigla</b>	<b>Área de Interesse</b>
A1	Inteligência Artificial
A2	Sistemas de Informação
A3	Mineração de Dados

Digamos que nossos usuários representados na Tabela 6 (p.82) tenham interesses nestas áreas, conforme Tabela 10:

Tabela 10 - Usuários e suas áreas de interesse

<i>Usuário</i>	<i>Área de Interesse (<math>AI_{USR}</math>)</i>	<i>Tempo de Atuação na Área (<math>TAI_{USR}</math>, em anos)</i>
USR1	A1	3
	A2	4,5
USR2	A1	7
USR3	A2	2
	A3	1
USR4	A2	2,5
USR5	A1	1
USR6	A1	6
	A2	1,5
	A3	3
USR7	A2	4
USR8	A1	4
	A3	0,5

Para o cálculo do  $RR_{\text{área}}$ , precisamos realizar uma média ponderada entre o  $RR_{\text{resultado}}$  obtido e o  $TAI_{USR}$ . Isso nada mais é do que passar os valores da Tabela 10 novamente pela equação do RR, associando

um peso ao  $RR_{\text{resultado}}$  e um peso ao  $TAI_{\text{USR}}$ . Para ficar mais claro, vamos representar esta equação:

$$RR_{\text{área}} = \frac{(RR_{\text{resultado}} * p_1) + (TAI_{\text{USR}} * p_2)}{\sum_{i=1}^n p_i}$$

*Figura 24 - Representação do cálculo do  $RR_{\text{área}}$*

*Fonte: elaboração própria*

Empiricamente, digamos que o  $RR_{\text{resultado}}$  deva ser associado ao peso  $p_1=6$  e que o tempo de estudo na área de interesse deva ser associado ao peso  $p_2=4$ . Após a normalização dos valores de  $TAI_{\text{USR}}$ , o  $RR_{\text{área}}$  para cada um dos usuários apresentados na Tabela 10 obtem-se o resultado da Tabela 11:

Tabela 11 - Resultado do Ranking de Recomendação por área de interesse

<i>Usuário</i>	$RR_{\text{resultado}}$	$TAI_{\text{USR}}$	$RR_{\text{área}}$
USR1 (A1)	4,8	4,3	4,6
USR1 (A2)		6,4	5,4
USR2 (A1)	5,6	10	7,4
USR3 (A2)	1,4	2,9	2
USR3 (A3)		1,4	1,4
USR4 (A2)	5,3	3,6	4,6
USR5 (A1)	3,2	1,4	2,5
USR6 (A1)	8,3	8,6	8,4
USR6 (A2)		2,1	5,8
USR6 (A3)		4,3	6,7
USR7 (A2)	3	5,7	4,1
USR8 (A1)	2,2	5,7	3,6
USR8 (A3)		0,7	1,6

Para facilitar a interpretação dos valores da Tabela 11, vamos representá-la segundo as áreas de interesse na Tabela 12:

Tabela 12 - Identificação dos usuários com opinião mais relevante

<i>Usuário</i>	<i>A1</i>	<i>A2</i>	<i>A3</i>
USR1	4,6	5,4	0
<b>USR2</b>	<b>7,4</b>	<b>0</b>	<b>0</b>
USR3	0	2	1,4
<u>USR4</u>	<u>0</u>	<u>4,6</u>	<u>0</u>
<b>USR5</b>	<b>2,5</b>	<b>0</b>	<b>0</b>
USR6	8,4	5,8	6,7
<u>USR7</u>	<u>0</u>	<u>4,1</u>	<u>0</u>
USR8	3,6	0	1,6

Podemos tirar as seguintes conclusões:

- O usuário com opinião mais relevante nas áreas A1, A2 e A3 é o USR6 (maior  $RR_{\text{área}}$  nas três áreas de interesse);
- Os usuários USR2 e USR5 pertencem a um mesmo grupo de opinião (só atuam na área A1), e dentre os dois, a opinião de USR2 é a mais relevante;
- Os usuários USR4 e USR7 pertencem a um mesmo grupo de opinião (só atuam na área A2), e dentre os dois, a opinião de USR4 é a mais relevante.

A relevância da opinião dos usuários pode, agora, ser utilizada para auxiliar a predição de perfis semelhantes na recuperação de perfis de usuário.

No sistema de recomendação, para captura explícita da opinião do usuário, é possível disponibilizar um meio para o usuário avaliar se determinado currículo está ou não relacionado à sua área de atuação. Digamos que os usuários — de USR1 a USR8 — podem entrar nesse sistema para avaliar o novo currículo USR9, apresentado na Tabela 13:

Tabela 13 - Dados de  $RR_{\text{área}}$  do usuário USR9

Usuário	A1	A2	A3
USR9	5,4	0	0

Imediatamente, podemos imaginar os melhores “vizinhos” para avaliar o currículo do usuário USR9: são os usuários USR2 e USR5, os únicos que participam do mesmo grupo de opinião que USR9, pois todos só atuam na área A1. No entanto, ao realizar uma busca por currículos relevantes para pesquisadores de determinada área, há aspectos subjetivos que não são representados formalmente em um sistema. Portanto, é necessário um critério na recomendação de currículos para um grupo de opinião.

Se o sistema possui três áreas de interesse, há então uma combinação de três elementos que representa os grupos de opinião possíveis para a recomendação de um currículo. Dado que  $C_{n,p}$  é um número binomial representado por  $\frac{n!}{p!(n-p)!}$ , ou seja, a combinação de  $n$  elementos distintos tomados  $p$  a  $p$ , o cálculo do número de grupos de opinião pode ser dado pela soma dos subconjuntos  $C_{3,1} + C_{3,2} + C_{3,3}$ , o que seria  $\frac{3!}{1!(3-1)!} + \frac{3!}{2!(3-2)!} + \frac{3!}{3!(3-3)!}$ , isto é, resultando em sete grupos de opinião.  $C_{3,0}$  deveria ser levado em conta se o sistema permitisse o cadastro de usuários sem nenhuma área de interesse — o *ranking* de recomendação desse usuário seria utilizado apenas pelas buscas gerais,  $RR_{\text{resultado}}$ . Caso o sistema hipotético que estamos apresentando fosse construído permitindo esta situação, teríamos que acrescentar  $C_{3,0}$  à soma anterior, isto é,  $\frac{3!}{0!(3-0)!} = \frac{3!}{3!} = 1$ , um subconjunto de um elemento, o conjunto vazio. Teríamos, para o caso, oito grupos de opinião.

A Tabela 14 mostra os sete grupos de opinião possíveis na situação que estamos estudando. Quanto mais áreas de interesse, maior o número possível de grupos de opinião.

*Tabela 14 - Grupos de opinião para o sistema de recomendação de currículos*

<i>Grupo de Opinião</i>	<i>A1</i>	<i>A2</i>	<i>A3</i>
GRP1	X		
GRP2		X	
GRP3			X
GRP4	X	X	
GRP5	X		X
GRP6		X	X
GRP7	X	X	X

Digamos que o usuário USR9 foi avaliado conforme a Tabela 15 pelos usuários do sistema:

*Tabela 15 - Avaliação do USR9 pelos usuários do sistema*

<i>Usuário</i>	<i>USR9 está relacionado à sua área de atuação?</i>
USR1	Não
USR2	Sim
USR3	Sim
USR4	Não
USR5	Não
USR6	—
USR7	Não
USR8	Sim

Dividindo os usuários segundo o grupo ao qual pertencem, chegamos à Tabela 16:

Tabela 16 - Agrupamento dos usuários de acordo com o grupo de opinião

<i>Grupo de Opinião</i>	<i>Membros</i>
GRP1	USR2, USR5
GRP2	USR4, USR7
GRP3	—
GRP4	USR1
GRP5	USR8
GRP6	USR3
GRP7	USR6

Se uma avaliação negativa pode ser representada por um valor +1 e uma avaliação positiva pode ser representada por um valor +2, o resultado das avaliações realizadas pelos usuários do sistema, considerando-se puramente a filtragem colaborativa, seria como o exposto na Tabela 17:

Tabela 17 - Resultado da avaliação dos usuários utilizando FC pura

<i>Grupo de Opinião</i>	<i>Pontuação do currículo USR9</i>
GRP1	3
GRP2	2
GRP3	0
GRP4	1
GRP5	2
GRP6	2
GRP7	0

Este resultado foi criado a título meramente ilustrativo. No caso do GRP1, por exemplo, estaríamos considerando que as opiniões de USR2 e USR5 valem o mesmo, sendo que o  $RR_{\text{área}}$  deles é bastante diferente — USR2 tem  $RR_{\text{área}}$  de 7,4 e o  $RR_{\text{área}}$  de USR5 é 2,5. As avaliações dos usuários com maior autoridade na área de interesse devem ter maior peso no cálculo da recomendação. É proposta a fórmula a seguir para o cálculo da pontuação do item para cada grupo de opinião:

$$p_i = \sum_{u=1}^n (p_{u,i} \times \frac{RR_{\text{área}_u}}{10})$$

Figura 25 - Equação do cálculo da pontuação

Fonte: elaboração própria

O símbolo  $p_i$  representa a pontuação final do item, que é dada pelo somatório da avaliação  $p_{u,i}$  realizada por cada usuário, adequada ao fator de ponderação — o  $RR_{\text{área}}$  de cada usuário.

No exemplo anterior, como  $USR2$  possui  $RR_{\text{área}}=7,4$  e  $USR5$  possui  $RR_{\text{área}}=2,5$  — e os votos foram sim e não para a avaliação do currículo  $USR9$ , isto é, pontuação de valores 2 e 1 respectivamente —, o valor de pontuação que  $GRP1$  deve assumir para o currículo  $USR9$  deve ser:

$$p_{USR9} = 2 \cdot 0,74 + 1 \cdot 0,25 = 1,73$$

Assim chegamos à Tabela 18:

Tabela 18 - Resultado da avaliação dos usuários

Grupo de Opinião	Pontuação do currículo $USR9$
GRP1	1,73
GRP2	0,87
GRP3	0
GRP4	0,50
GRP5	0,52
GRP6	0,34
GRP7	0

Para o cálculo da Tabela 18, os dados de  $RR_{\text{área}}$  dos usuários foram calculados realizando-se uma média dos  $RR_{\text{área}}$  disponíveis (diferentes de zero), o que chamaremos de  $RR_{\text{total}}$ . Os valores utilizados estão mostrados abaixo. O  $RR_{\text{total}}$  só faz sentido se utilizado em cálculos para recomendações dentro de cada grupo de opinião. Para o cálculo de *ranking* de recomendação em buscas gerais de currículos, sem considerar a área de interesse, deve ser utilizado o  $RR_{\text{resultado}}$ .

Tabela 19 –  $RR_{total}$  de cada usuário

Usuário	A1	A2	A3	$RR_{total}$
USR1	4,6	5,4	0	5
USR2	7,4	0	0	7,4
USR3	0	2	1,4	1,7
USR4	0	4,6	0	4,6
USR5	2,5	0	0	2,5
USR6	8,4	5,8	6,7	7
USR7	0	4,1	0	4,1
USR8	3,6	0	1,6	2,6
USR9	5,4	0	0	5,4

Os valores da Tabela 18 não devem ser normalizados porque cada métrica está representando a avaliação de diferentes grupos de opinião. Vamos nos concentrar no GRP1, onde temos três representantes: USR2, USR5 e USR9. Suponhamos que a pontuação de cada um deles seja dada conforme as duas tabelas seguintes:

Tabela 20 - Pontuação dos usuários do grupo GRP1

Usuários de GRP1	Pontuação do currículo
USR2	0,5
USR5	0,74
USR9	1,73

Tabela 21 - Pontuação normalizada dos usuários do grupo GRP1

Usuários de GRP1	Pontuação do currículo
USR2	2,9
USR5	4,3
USR9	10

Admitindo-se que USR2 avaliou USR5 como não-relevante, USR5 avaliou USR2 como relevante, e USR9 não avaliou nenhum dos dois currículos, o sistema de recomendação baseando-se apenas na avaliação colaborativa dos usuários recomendaria nesta ordem: USR9, USR5 e USR2.

Para amostras extremamente pequenas, tira-se a conclusão que o *ranking* baseado apenas em filtragem colaborativa, devido a sua subjetividade intrínseca, pode apresentar resultados que nem sempre eram esperados — USR2, que tem o melhor  $RR_{total}$ , ficou no fim da recomendação. No entanto, fica claro que a opinião de USR2 foi levada em conta e que isso influenciou o resultado do *ranking*, apresentando USR9 na frente de USR5. Em bases maiores haveria outros usuários com  $RR$  maior que poderiam ter avaliado USR2 diferentemente e o resultado final poderia ter sido<sup>52</sup> outro.

Para que o usuário possa trabalhar as respostas fornecidas pelo sistema e avaliar que resultados são mais relevantes para sua necessidade, o sistema a ser desenvolvido com base nesse modelo deve tratar os resultados de busca de forma diferenciada: aqueles onde se considera apenas a filtragem baseada no conteúdo para a recomendação de informação — isto é, resultados ordenados segundo o  $RR_{resultado}$  ou segundo  $RR_{área}$ , se realizada uma busca por área — e aqueles nos quais é levada em conta a avaliação dos usuários. Os resultados devem ser apresentados em uma mesma tela, com indicações claras ao usuário das listas de resultado, sem deixar dúvidas sobre como cada *ranking* foi realizado, de forma que o usuário estabeleça qual o critério de ordenação mais relevante para a sua necessidade naquele momento.

Uma abordagem híbrida poderia ser realizada colocando em um *ranking* único os resultados com base no  $RR_{total}$  e pontuação dos usuários, e só os usuários seriam capazes de avaliar se a relevância automática que o sistema forneceria seria eficaz no atendimento às suas necessidades. Esta abordagem híbrida não vai ser adotada no sistema desenvolvido neste projeto, porém fica como alternativa para a tela única que apresenta os dois *rankings*, conforme explicado no parágrafo anterior.

O sistema de recomendação de perfis de usuário que utiliza o modelo proposto neste projeto é um sistema dinâmico, atemporal. A relevância da opinião dos usuários varia conforme o tempo, e isto influencia a

---

<sup>52</sup> Não podemos dizer “seria outro”, pois a avaliação dos usuários é algo subjetivo.

forma como suas informações implícitas e explícitas são trabalhadas. Com a evolução do perfil, varia seu RR, e, com a normalização dos RR dos demais usuários, quem não evolui seu perfil acaba ficando com menor pontuação no *ranking* normalizado. Menor e maior RR causam influência direta na coleta de informação explícita — isso se aplicaria ao Currículo Lattes, se o usuário pudesse avaliar se determinado currículo está ou não relacionado à sua área de atuação. Um RR menor fará com que o voto do usuário receba um valor de menor importância no cálculo da pontuação de um item por ele avaliado.

Um sistema no domínio da recuperação de currículos que utilize o modelo deste projeto seria útil para a identificação de especialistas em áreas de pesquisa, consultores ad-hoc, editores de revista que necessitem de peritos em assuntos, contatos para eventos.

### **5.1.2. Modelo do Sistema Proposto**

Os módulos básicos para o sistema de pesquisa do sistema de recomendação de perfis de usuário são: Módulo de cadastro, Módulo de indexação, Módulo de consulta, Módulo de recuperação e Módulo de recomendação.

- Módulo de cadastro: é a porta de entrada do usuário no sistema de informações proposto neste projeto. Um cadastro bem feito facilita a recuperação de informação. A tarefa deste módulo é obter e armazenar as informações de cadastro do perfil em um banco de dados.
- Módulo de indexação: efetua a indexação dos itens relativos aos perfis. Pode ser feita de forma automática pelo SGBD (Sistema Gerenciador de Banco de Dados).
- Módulo de consulta: faz parte da camada de persistência, isto é, fica na ligação entre o ambiente de execução e o banco de dados. O módulo de consulta é a tarefa de interagir com o usuário para obter os dados que possam ser enviados ao sistema para recuperar informação.

- Módulo de recuperação: executa a tarefa de recuperar e apresentar informações através dos dados capturados pelo módulo de consulta. Também faz parte da camada de persistência.
- Módulo de recomendação: é responsável por manipular os dados dos perfis de usuário e as avaliações dos itens — notas dadas por usuários do sistema aos itens do domínio —, de forma a criar regras e gerar informações que permitam ao sistema gerar listas de recomendações — seleção de itens de melhor probabilidade de atender à necessidade do usuário.

### **5.1.3. Origem dos Dados**

Em cada domínio de um sistema de informações de perfis de usuários, os dados podem se originar de forma distinta. Podem ser preenchidos por uma autoridade do sistema, podem ser coletados por repositórios de informações em diferentes bancos de dados e ainda podem ser preenchidos diretamente pelos usuários. O sistema deve trabalhar com uma base de dados independente da origem que a alimentou.

### **5.1.4. Interface de Consulta**

O sistema deve contemplar interfaces de consulta que permitam o uso de termos e descritores, isto é, o sistema deve permitir que o usuário realize busca tanto por termos que ele mesmo insira quanto por descritores do sistema — pela seleção de termos fixados pelo sistema.

Concluída a Seção 5.1, temos um modelo geral que pode ser aplicado a qualquer domínio de sistema de recomendações de perfis de usuário. Para validar esta proposta, o modelo será aplicado em um *site* de relacionamentos, conforme explicado na Metodologia da Pesquisa.

## **5.2. Aplicando o Modelo Geral em um *Site* de Relacionamentos**

Obedecendo aos passos do modelo geral, nesta seção serão definidos os indicadores do domínio da recuperação de perfis em um *site* de relacionamentos. Em seguida será apresentado o modelo do sistema de pesquisa, com seus respectivos módulos, a origem dos dados e o protótipo de interface de consulta.

### 5.2.1. Domínio e Indicadores

O domínio da recuperação de perfis de usuário no qual o modelo foi aplicado é um *site* de relacionamentos. Os indicadores para este domínio estão representados nas duas tabelas a seguir.

*Tabela 22 - Variáveis qualitativas*

Variável	Valor
Nível (N)	Não se aplica
Áreas de Interesse (AI)	São cinco: <ul style="list-style-type: none"> <li>• A1 - Romance</li> <li>• A2 - Amizade</li> <li>• A3 - Atividades a dois</li> <li>• A4 - Atividades em grupo / Outras atividades</li> <li>• A5 - Mais de um ou todos os anteriores</li> </ul>
Tempo de atuação na Área de Interesse (TAI)	Não se aplica

Em um primeiro momento, pensou-se em colocar o nível profissional como um classificador de nível de usuário para o domínio. No entanto, nível profissional é um indicador qualitativo de sistemas que trabalham recursos humanos e, para o domínio de relacionamentos, isso não se aplica. O nível profissional da pessoa, que deverá ser colhido no cadastro, servirá apenas para filtro na recomendação de informação, e não como uma

variável que forneça mais ou menos peso para itens (perfis) que serão recomendados.

A área de interesse A5 foi criada com o objetivo de limitar a uma, e apenas uma, a opção de área de interesse do usuário ao se cadastrar no sistema deste domínio. Foi uma alternativa para diminuir a complexidade na criação de listas de recomendação — complexidade que foi observada no cálculo da combinação dos elementos de áreas de interesse no exemplo de recuperação de currículos de Ciência da Computação.

O tempo de atuação na área de interesse também não é um bom indicador para a criação de recomendações em um *site* de relacionamentos, nada garante que se alguém está interessado em romance há mais de três anos é mais relevante do que alguém que decidiu por isso há uma semana. Portanto, tempo de atuação também não se aplica neste contexto. Os indicadores considerados relevantes como variáveis quantitativas estão representados na Tabela 23.

Tabela 23 - Variáveis quantitativas

Variável	Valor
Visitantes do perfil (VPF)	Valor medido em unidades VPF
Contatos primários recebidos (CPR)	Valor medido em unidades CPR
Contatos secundários recebidos (CSR)	Valor medido em unidades CSR
Avaliações Positivas (AVP)	Valor medido em unidades AVP
Avaliações Negativas (AVN)	Valor medido em unidades AVN
Bloqueios (BLQ)	Valor medido em unidades BLQ

A variável VPF armazena a quantidade de visitantes que um perfil recebe. Perfis mais visitados provavelmente são mais relevantes que perfis menos visitados.

No sistema a ser desenvolvido, a interação inicial, isto é, a forma que um usuário pode a princípio contatar outro, será através do utilitário “piscada”. Perfis de usuários que recebem mais contatos primários certamente devem ser considerados mais relevantes que perfis que recebem menos contatos. A variável quantitativa que armazenará esta informação é a CPR.

Após um contato inicial, os usuários têm à sua disposição contato secundário, que é um utilitário para o envio de endereço de *email* ou número de telefone. Espera-se que perfis que recebam mais contatos secundários sejam mais relevantes. A variável que armazena quantitativamente esta informação é CSR.

Cada perfil pode ser avaliado positivamente ou negativamente por um usuário, ou ainda, um usuário pode bloquear outro para que não entre em

contato com ele. Todas estas informações são armazenadas em AVP, AVN e BLQ, respectivamente.

### **5.2.2. Sistema Proposto**

O sistema foi desenvolvido para a *Web*, utilizando PHP e MySQL. Abaixo segue uma explicação sobre cada um dos módulos do sistema de pesquisa, conforme o modelo definido na Seção 5.1:

- **Módulo de cadastro:** o módulo de cadastro foi dividido em três etapas, para não tornar cansativo o preenchimento pelo usuário. A primeira etapa, chamada de geral, possui as informações principais da identificação do perfil do usuário. Nesta etapa muitos dos campos são obrigatórios, como poderá ser visto na definição dos campos na implementação do sistema. A segunda etapa é chamada de social, onde o usuário colocará dados menos relevantes para a identificação de seu perfil, porém que serão de grande valia na geração de recomendações. A terceira e última etapa, chamada de pessoal, é a qual o usuário tem livre espaço para criar a descrição de seu perfil. Os campos não são obrigatórios. Nesta seção também há a possibilidade de seleção de interesses, como arte, culinária ou viagens. Esses campos de interesse, não obrigatórios, poderão ser úteis na geração de recomendações no futuro, descobrindo padrões de gosto semelhantes e agrupamentos de usuários — neste projeto, estes campos não serão utilizados para a geração de recomendações, ficando esta tarefa para posterior evolução do sistema.
- **Módulo de indexação:** os registros são ordenados por identificadores únicos, gerenciados automaticamente pelo SGBD.

- Módulo de consulta: é subdividido em três interfaces de busca: pré-definida, rápida e por apelido. Quando o usuário realiza sua primeira consulta na interface de busca pré-definida, o sistema coleta essas informações para gerar recomendações futuras com base no interesse explicitado pelo usuário. É permitido ao usuário alterar as definições de busca pré-definida a cada nova consulta realizada nesta interface. Os resultados da busca pré-definida sempre são apresentados em tempo real. A busca rápida se resume em uma consulta diferenciada, onde os gostos do usuário não são levados em conta para a apresentação dos resultados. Nesta interface de consulta, o usuário tem maior visibilidade dos registros do banco de dados. A única filtragem estabelecida é a de orientação sexual. Na interface de busca rápida, serão disponibilizados filtros para região geográfica e consulta apenas dos usuários *on-line*. A última interface é a busca por apelido. O usuário é conhecido pelo sistema através do identificador único que é gerenciado pelo SGBD, mas é conhecido por outros usuários através de seu apelido, que também deve ser único. A interface de busca por apelido permite uma consulta pelo termo digitado, seja ele o apelido por completo ou parte dele.
- Módulo de recuperação: Os algoritmos de recuperação de informação poderão ser conferidos no código-fonte do sistema. São baseados em consultas SQL simples.
- Módulo de recomendação: Os indicadores qualitativos e quantitativos levantados na seção anterior foram utilizados para criar recomendações em diversas abordagens. Como as recomendações são o cerne deste sistema, o funcionamento desde módulo será mais detalhado nas seções seguintes.

### 5.2.3. Origem dos Dados

*Sites* de relacionamento possuem uma característica em comum: os dados com as informações dos usuários são preenchidos por eles mesmos<sup>53</sup>. Baseando-se nesta premissa, é muito importante que os campos que serão preenchidos pelos usuários tenham alguma validação automática. Por exemplo, que a data de nascimento seja validada para impedir de se preencher o mês com dígitos inválidos, dentre outros.

É preciso deixar claro que “*site* de relacionamentos” ou “*website* de relacionamentos” não significa o mesmo que um sistema de relacionamento que pode ser utilizado por uma empresa que promove encontros. Embora este sistema também trate da recuperação de perfis de usuários com objetivos semelhantes aos do *site* de relacionamentos, os dados do sistema da empresa são preenchidos por uma autoridade do sistema, e não diretamente pelo usuário.

### 5.2.4. Interface de Consulta

Obedecendo às orientações, a interface deve permitir o uso de termos e descritores nas consultas aos perfis dos usuários. A coleta de informações explícitas deve ser realizada através da interação do usuário com o sistema, por meio de utilitários amigáveis, que façam uso, se possível, de ícones e símbolos de interpretação consagrada pela população à qual se refere.

Os detalhes da interface para o protótipo do sistema poderão ser visualizados na próxima seção.

### 5.2.5. Implementação

A preocupação principal da arquitetura da informação em um *website*, na prática, refere-se à organização da estrutura deste e seu conteúdo, rotulagem e categorização da informação e o projeto dos sistemas de navegação e de busca.

---

<sup>53</sup> Neste caso, o mesmo ocorre no sistema Currículo Lattes.

Para a implementação do sistema, foi fechado um contrato com uma empresa desenvolvedora de sistemas para Internet, que criou o esqueleto de um *website* com estruturas padrão de consultas SQL a registros de um banco de dados. Não era interesse nem foco da pesquisa a criação do *site* de relacionamentos, mas sim do modelo de recomendações.

No início de qualquer projeto é difícil explicar a outras pessoas como será a interação entre usuários e artefatos, pois a interação acontece no espaço-tempo. As relações entre as formas do artefato (espaço) e as seqüências de transformações (tempo) podem estar claras na mente, porém ao apresentar ou assimilar uma idéia, a tradução é uma atividade extremamente complexa.

Mapear um modelo conceitual, entretanto, pode ajudar a perceber que alguns dos conceitos pré-estabelecidos não se adequam à realidade. Nessa etapa é possível considerar alternativas, supondo que há um bom raciocínio para sustentar esta teoria. Após estabelecer qual é o artefato a ser modelado, ele é colocado em um contexto e cenários são imaginados, antes de surgirem os casos de uso desse artefato.

O modelo conceitual do artefato não é igual ao modelo mental do usuário. Isso é impossível de acontecer, pois cada pessoa interpreta o mundo de forma diferente e o artefato é sempre o mesmo. Segundo o conceito de modelo mental, o usuário já possui um esquema prévio criado em sua mente que é basicamente "recuperado" quando está diante do artefato, bastando mapeá-lo para definir o artefato. Mas, na prática, as pessoas só começam a formar idéias sobre os artefatos a partir do primeiro contato com tal artefato, ou quando há a necessidade de construir um artefato imaginário. Então é a interação com o artefato que constrói o modelo mental e não os conceitos prévios à sua utilização. O modelo mental é algo dinâmico, está em constante transformação, e não pode ser capturado de forma estática.

A fim de contornar essas dificuldades, próprias do estudo da usabilidade, que não é o foco deste projeto, a metodologia de

desenvolvimento de software adotada foi a *eXtreme Programming*, ou simplesmente XP, uma metodologia ágil para equipes pequenas e médias e que pretendem desenvolver software com requisitos vagos e em constante mudança. Para isso, a metodologia prega a estratégia de constante acompanhamento e realização de vários pequenos ajustes durante o desenvolvimento de software.

A interface do sistema, por exemplo, foi definida em diversas reuniões entre o representante da empresa e o representante deste projeto, sendo que cada novo detalhe ou mudança de requisito implicava em um novo ciclo de implementação. A empresa dividiu a responsabilidade da organização do conteúdo, rotulagem e categorização da informação.

Após montada a estrutura inicial, o foco foi voltado para a implementação dos módulos do sistema de pesquisa. Primeiramente tratou-se do módulo de cadastro, definindo os campos que estão no Anexo I “Estrutura do Perfil”.

O módulo de cadastro foi especificado e os campos para sua validação foram tratados com rigor, a fim de evitar erros cadastrais por parte do usuário. Durante o tratamento do módulo de cadastro, algumas questões de privacidade foram levantadas e outros requisitos foram registrados. Dentre eles:

- Só é possível realizar busca após autenticar-se no *site*;
- Usuários só vêem perfis de usuários de sua mesma orientação sexual;
- Não há mural de recados. Não é possível enviar mensagem pelo *site*.

Não é necessário discorrer sobre a indexação dos registros, pois é gerenciada automaticamente pelo MySQL.

A interface de consulta adotou a orientação de uso de termos e descritores e foi dividida em três telas, conforme as figuras seguintes.

Busca pré-definida
Busca rápida
Por apelido

**Altura:**

- Anão a 1,40m
- 1,41m a 1,65m
- 1,66m a 1,70m
- 1,70m a 1,72m
- 1,73m a 1,75m
- 1,76m a 1,78m
- 1,79m a 1,81m
- 1,82m a 1,84m
- 1,82m a 1,83m
- 1,84m a 1,85m
- 1,86m a 1,87m
- 1,88m a Gigante

**Biotipo:**

- Esquelético
- Magro/Pequeno
- Magro a Médio
- Médio
- Médio a Definido
- Médio a Gordinho

**Localização:**

<input type="checkbox"/> AL	<input type="checkbox"/> PB
<input type="checkbox"/> AM	<input type="checkbox"/> PE
<input type="checkbox"/> AP	<input type="checkbox"/> PI
<input type="checkbox"/> BA	<input type="checkbox"/> PR
<input type="checkbox"/> CE	<input type="checkbox"/> RJ
<input checked="" type="checkbox"/> DF	<input type="checkbox"/> RN
<input type="checkbox"/> ES	<input type="checkbox"/> RO
<input checked="" type="checkbox"/> GO	<input type="checkbox"/> RR
<input type="checkbox"/> MA	<input type="checkbox"/> RS
<input type="checkbox"/> MG	<input type="checkbox"/> SC
<input type="checkbox"/> MS	<input type="checkbox"/> SE
<input type="checkbox"/> MT	<input type="checkbox"/> SP
<input type="checkbox"/> PA	<input type="checkbox"/> TO

**Pelo corporal:**

- Nenhum
- Pouco
- Médio
- Muito

Idade entre:  e

Em ordem de:

Salvar alterações da busca pré-definida

Figura 26 - Interface de consulta "busca pré-definida"<sup>54</sup>

Fonte: elaboração própria

<sup>54</sup> Há mais campos descritores na busca pré-definida do sistema implementado. Eles foram reduzidos nesta tela para que a figura fosse apresentada em uma só página.

Figura 27 - Interface de consulta "busca rápida"

Fonte: elaboração própria

Figura 28 - Interface de consulta "busca por apelido"

Fonte: elaboração própria

Já foi explicada anteriormente nesse projeto a necessidade das três telas de interface de consulta, mas vale ressaltar algumas características oriundas da implementação do sistema.

Os campos descritores existentes na busca pré-definida auxiliarão o sistema na identificação de perfis que agradem ao usuário que está realizando a consulta. E de maneira inversa, o sistema poderá se utilizar dessas informações para evitar recomendar usuários a outros que já não passam na pré-definição de quem seria recomendado. Em outras palavras, se o usuário X já definiu que procura por perfis de DF e GO, embora ele seja de SP, o sistema não recomendará X ao usuário Y que está procurando por perfis de SP, pois o sistema já entende que Y não é uma boa recomendação para X, evitando que no futuro X recuse os contatos que Y possa realizar. Esta parece uma atividade própria da recomendação de perfis no contexto de *sites* de relacionamento.

A busca pré-definida limita os resultados apresentados na consulta por duas razões. A primeira limitação, que é clara, é causada pela filtragem realizada segundo os campos descritores selecionados pelo usuário. A segunda limitação é realizada automaticamente pelo sistema, que vai apresentar como resultados apenas perfis de usuário associados à área de interesse de quem está realizando a consulta.

Em uma analogia aos naipes do baralho, as cinco áreas de interesse que um usuário pode selecionar foram representadas da seguinte forma:

- A1 – Perfil Copas (romance)
- A2 – Perfil Ouros (amizade)
- A3 – Perfil Espadas (atividades a dois)
- A4 – Perfil Paus (atividades em grupo ou outras atividades)
- A5 – Perfil Coringa (mais de um ou todos os anteriores)

Esta analogia originou-se da necessidade de dar um nome para o sistema, que precisava ter um domínio de fácil dedução. O sistema adotou o nome 2coracoes<sup>55</sup>, em uma referência à carta “dois de copas”, e o slogan adotado foi “O que você está esperando para encontrar alguém especial? Jogue suas cartas nesse jogo de sedução”. Atrair usuários para utilizarem este sistema não é um objetivo do projeto, porém como era preciso garantir um sistema amigável e talvez definitivo para o *website* que vai fornecer a base de usuários, estratégias de *marketing* foram testadas e avaliadas no decorrer do projeto.

A Figura 29 mostra a tela inicial de acesso ao sistema.

---

<sup>55</sup> O domínio registrado foi [www.2coracoes.com](http://www.2coracoes.com) (não foi possível adotar [www.doiscoracoes.com](http://www.doiscoracoes.com) porque já estava registrado para uma padaria).

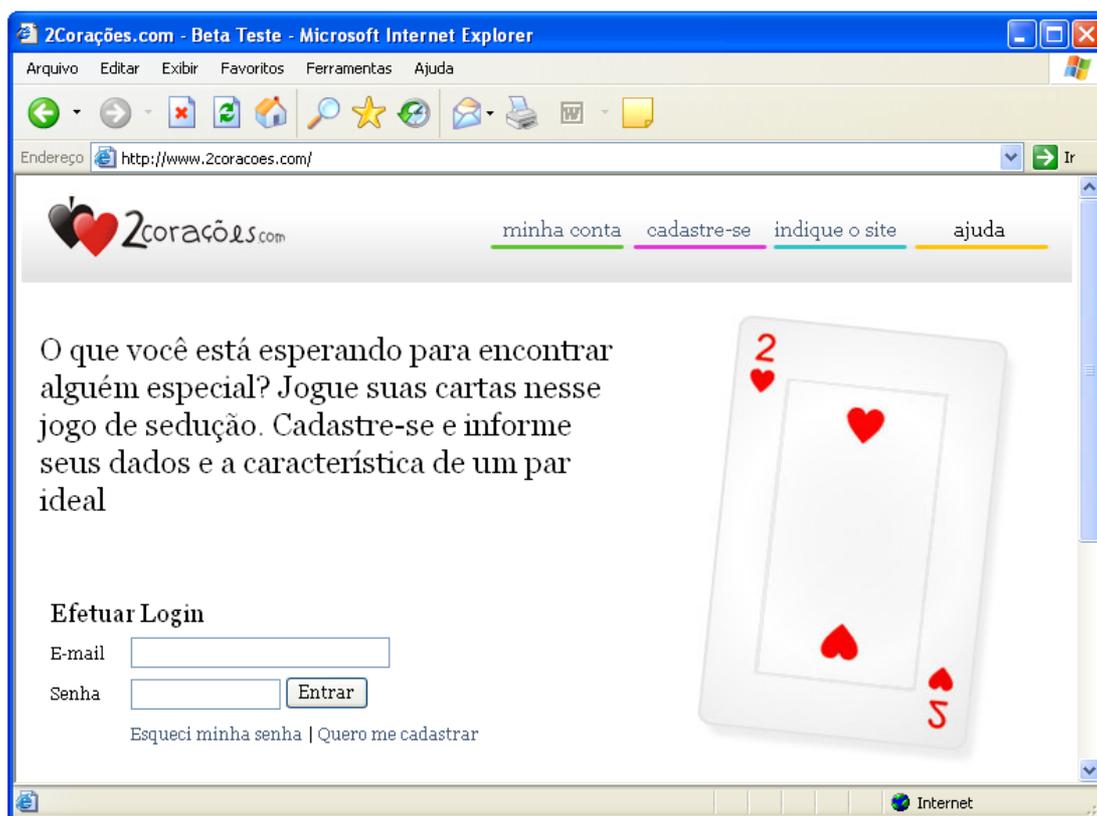


Figura 29 - Tela inicial de acesso ao sistema 2coracoes

Fonte: elaboração própria, disponível em <<http://www.2coracoes.com>>

Explicada a origem dos naipes — a representação das áreas de interesse — podemos falar sobre a busca rápida. Como a busca pré-definida não permite ao usuário um acesso total aos perfis armazenados em banco, a busca rápida permite a consulta a toda a base<sup>56</sup>, sendo permitido realizar filtragem por naipe ou estado (UF), e escolher a ordem de apresentação dos resultados — por idade ou estado (UF). Ainda há a possibilidade de filtrar somente os usuários que estão *on-line*, isto é, outros usuários que no momento da consulta estão também autenticados no sistema.

Por fim, a busca por apelido facilita o acesso direto ao perfil de um usuário, haja vista que não é necessário consultar toda uma base se já sabemos o perfil que queremos localizar.

Vale observar que a busca pré-definida é o principal utilitário do sistema de recomendações. Sem essa ferramenta conseguiríamos alcançar

<sup>56</sup> Há uma regra básica que não permite a apresentação de resultados de perfis com orientação sexual diferente da orientação do usuário que realiza a consulta. Essa regra é para garantir privacidade aos usuários do sistema.

apenas um sistema de recomendação de informações mais próximo da recuperação tradicional com uso de filtros. As informações coletadas da seleção dos campos pelos usuários permitem melhor trabalhar a geração de recomendações para os grupos de usuários com base nas avaliações dos usuários de interesse semelhante.

No desenho das interfaces de consulta, vários algoritmos foram rascunhados para se trabalhar a recuperação dos dados. O módulo de recuperação não foi construído à parte, seus algoritmos foram desenhados como parte da estrutura nos módulos das interfaces de consulta. Os algoritmos baseiam-se em consultas SQL simples, como podem ser conferidos no Anexo IV.

A parte fundamental do projeto, que diferencia o sistema 2coracoes de um sistema de informações baseado em recuperação tradicional, é o módulo de recomendação.

Os *sites* de relacionamento disponíveis no mercado trabalham algumas das abordagens de recomendação estudadas. Foram identificadas as seguintes utilizações:

- Abordagem baseada no conteúdo — na recuperação de perfis de outros usuários com base na filtragem de informações do perfil do usuário cadastrado, isto é, do usuário que realiza a busca. Ex: Msn paquera<sup>57</sup>, Par perfeito<sup>58</sup>.
- Abordagem *Top-N* — nas estatísticas sobre interatividade com os perfis de usuários (páginas mais visitadas, perfis mais bem avaliados, dentre outros). Ex: Disponível.com<sup>59</sup>, Love Happens<sup>60</sup>, Manhunt<sup>61</sup>.

---

<sup>57</sup> Disponível em <<http://match.br.msn.com/index.aspx>>

<sup>58</sup> Disponível em <<http://www.parperfeito.com.br>>

<sup>59</sup> Disponível em <<http://www.disponivel.com>>

<sup>60</sup> Disponível em <<http://www.lovehappens.com>>

<sup>61</sup> Disponível em <<http://www.manhunt.net>>

Embora a abordagem baseada no conteúdo tenha sido usada, a estrutura de recomendação de informação não foi bem aproveitada. A recomendação só é apresentada após a realização de consultas. O sistema poderia, por exemplo, recomendar perfis assim que o usuário se autenticasse no sistema. Essa funcionalidade foi implementada no sistema 2coracoes.

A utilização da abordagem *Top-N* é trivial e necessária. A interação do usuário com o sistema permite o armazenamento de valores matemáticos, que podem ser bem aproveitados na geração de *rankings* de recomendação. Esses valores matemáticos, que chamaremos de “pontuação”, serão utilizados para a promoção de outras abordagens de recomendação no sistema 2coracoes, a saber:

- Abordagem baseada em conhecimento — no sistema foram criadas regras para a recomendação de perfis de usuário.

Embora não tenham sido aplicadas técnicas eficientes de mineração de dados para a realização de recomendações, com a utilização das pontuações capturadas pela interação do usuário com o sistema, foram criadas regras para sugerir perfis que possam atender a necessidade do usuário. Com apoio das pontuações dos perfis, *rankings* de recomendação são gerados apoiando-se nessas regras<sup>62</sup>.

- Abordagem *Cross-Sell* — embora seja mais utilizada por *sites* de comércio eletrônico, a abordagem *Cross-Sell* foi adaptada neste projeto para uso no contexto de *sites* de relacionamento. Aqui, no caso, quando o usuário visita um perfil que foi apresentado em um resultado de busca, o sistema mostra imediatamente ao lado do perfil que se está consultando uma barra vertical contendo perfis semelhantes ao que se está visualizando. Esses perfis semelhantes são mostrados cruzando-se informações do perfil consultado com informações extraídas da busca pré-definida.

---

<sup>62</sup> Apenas na interface de busca pré-definida.

O uso de todas essas abordagens seria já um avanço considerável se comparamos as estratégias utilizadas pelos sistemas tradicionais, baseados em consulta eletrônica sem uma maior preocupação com a necessidade do usuário. Mas para garantir uma ordenação dos resultados que venha ao encontro do real interesse extraído implícita e explicitamente da interação do usuário com o sistema, precisamos trabalhar o *Ranking* de Recomendação, conceito tratado na Seção 5.1.

Se no sistema 2coracoes há cinco áreas de interesse, devemos ter, então, cinco  $RR_{\text{área}}$ , cada um associado a uma área de interesse: copas, ouros, espadas, paus e coringa. Como as avaliações em um *site* de relacionamentos estão diretamente associadas aos grupos de interesse, temos que criar mais diferenças para os  $RR_{\text{área}}$  — separar os *rankings* por sexo e orientação sexual. Ao todo seriam vinte *Rankings* de Recomendação, categorizados segundo sexo, orientação sexual e área de interesse.

$RR_{\text{htfx}}$  e  $RR_{\text{htmx}}$  representam os *rankings* de recomendação dos usuários feminino e masculino de orientação heterossexual, e  $RR_{\text{hofx}}$  e  $RR_{\text{homx}}$ , dos usuários feminino e masculino de orientação homossexual, sendo  $x$  a área de interesse. O índice  $RR_{\text{htmouros}}$ , por exemplo, representa o *ranking* de recomendação dos usuários masculinos de orientação heterossexual interessados em amizade (área de interesse “ouros”).

No sistema foi implementado conforme a teoria. Digamos que a Tabela 24 apresenta todos os usuários heterossexuais do sexo feminino no sistema 2coracoes, com interesse A1 — copas, romance: Maria, Joana, Natália, Taíssa e Rosa. E ao lado o número de visitas que o perfil recebeu (VPF), quantidade de contatos primários (CPR), quantidade de contatos secundários (CSR), avaliações positivas (AVP), avaliações negativas (AVN) e bloqueios (BLQ).

Tabela 24 - Valores não normalizados do cálculo de  $RR_{htfcopas}$

Usuário	VPF	CPR	CSR	AVP	AVN	BLQ
Maria	9	2	1	4	1	0
Joana	1	0	0	3	0	1
Natália	4	2	1	2	0	0
Taíssa	7	3	3	5	0	0
Rosa	12	1	0	3	2	1

Para simplificar os cálculos de normalização, vamos diminuir o número de avaliações negativas e bloqueios do número de avaliações positivas, ou seja,  $AVP - (AVN + BLQ)$ . Esse resultado será chamado de avaliação geral (AVG). Se AVG for menor que zero, recebe valor zero.

Tabela 25 – Cálculo do AVG para futura normalização de  $RR_{htfcopas}$

Usuário	VPF	CPR	CSR	AVG
Maria	9	2	1	3
Joana	1	0	0	2
Natália	4	2	1	2
Taíssa	7	3	3	5
Rosa	12	1	0	0

Observe que a avaliação geral da Rosa deu zero, mas se tivesse dado valor negativo, Rosa receberia  $AVG=0$ .

O VPF de Maria normalizado é  $\left(\frac{9-1}{12-1}\right) \times 10 = \left(\frac{8}{11}\right) \times 10 = 7,3$

Fazendo o mesmo para os demais usuários teremos:

$$\text{Joana: } \left(\frac{1-1}{12-1}\right) \times 10 = \left(\frac{0}{11}\right) \times 10 = 0$$

$$\text{Natália: } \left(\frac{4-1}{12-1}\right) \times 10 = \left(\frac{3}{11}\right) \times 10 = 2,7$$

$$\text{Taíssa: } \left(\frac{7-1}{12-1}\right) \times 10 = \left(\frac{6}{11}\right) \times 10 = 5,5$$

$$\text{Rosa: } \left(\frac{12-1}{12-1}\right) \times 10 = \left(\frac{11}{11}\right) \times 10 = 10$$

Então Rosa que é o perfil mais visitado tem valor 10 e Joana tem valor zero. Os demais usuários recebem valores intermediários.

Abaixo segue demonstração do cálculo apenas para o usuário Maria:

$$\text{CPR de Maria: } \left(\frac{2-0}{3-0}\right) \times 10 = 6,7$$

$$\text{CSR de Maria: } \left(\frac{1-0}{3-0}\right) \times 10 = 3,3$$

$$\text{AVG de Maria: } \left(\frac{3-0}{5-0}\right) \times 10 = 6,0$$

Tabela 26 – Valores normalizados para o cálculo de  $RR_{htfcopas}$

Usuário	VPF	CPR	CSR	AVG
Maria	7,3	6,7	3,3	6,0
Joana	0	0	0	4,0
Natália	2,7	6,7	3,3	4,0
Taíssa	5,5	10	10	10
Rosa	10	3,3	0	0

Os dados da Tabela 26 são usados no cálculo de  $RR_{htfcopas}$ . Foi definido empiricamente que a avaliação geral é mais importante do que o número de visitantes do perfil, e que o contato primário e contato secundário têm a mesma importância. Os seguintes pesos foram atribuídos para cada uma das variáveis:  $AVG=3.0$ ,  $VPF=2.0$ ,  $CPR=2.5$  e  $CSR=2.5$ . Vejamos o cálculo de  $RR_{htfcopas}$  da Maria:

$$\text{Maria: } \frac{(2 \times VPF) + (2,5 \times CPR) + (2,5 \times CSR) + (3 \times AVG)}{10} =$$

$$\frac{(2 \times 7,3) + (2,5 \times 6,7) + (2,5 \times 3,3) + (3 \times 6,0)}{10} = 5,8$$

Fazendo os mesmos cálculos para os demais usuários, temos:

$$\text{Joana: } \frac{(2 \times 0) + (2,5 \times 0) + (2,5 \times 0) + (3 \times 4,0)}{10} = 1,2$$

$$\text{Natália: } \frac{(2 \times 2,7) + (2,5 \times 6,7) + (2,5 \times 3,3) + (3 \times 4,0)}{10} = 4,2$$

$$\text{Taíssa: } \frac{(2 \times 5,5) + (2,5 \times 10) + (2,5 \times 10) + (3 \times 10)}{10} = 9,1$$

$$\text{Rosa: } \frac{(2 \times 10) + (2,5 \times 3,3) + (2,5 \times 0) + (3 \times 0)}{10} = 2,8$$

Desta forma, calculamos o índice  $RR_{\text{htfcopas}}$  para todos os usuários heterossexuais de copas do sexo feminino. Para ordenar o resultado basta verificar os valores indicados pelo índice:

1. Taíssa –  $RR_{\text{htfcopas}}$  9,1
2. Maria –  $RR_{\text{htfcopas}}$  5,8
3. Natália –  $RR_{\text{htfcopas}}$  4,2
4. Rosa –  $RR_{\text{htfcopas}}$  2,8
5. Joana –  $RR_{\text{htfcopas}}$  1,2

As tabelas com as pontuações VPF, CPR, CSR, AVP, AVN e BLQ são alimentadas *on-line*, conforme os usuários interagem com o sistema. Uma vez ao dia a tabela que armazena os  $RR_{\text{área}}$  recebe atualização dos índices, através da execução do programa atualizarRanking.php<sup>63</sup>.

De posse dos valores de  $RR_{\text{área}}$ , as abordagens de recomendações adotadas no sistema 2coracoes podem usufruir da opinião dos usuários para realizar a ordenação dos resultados de busca.

### **5.2.6. Execução da Pesquisa Experimental**

A fase de execução da pesquisa experimental ocorre após as etapas de desenvolvimento e implementação. A Figura 30 retrata esta fase.

---

<sup>63</sup> Disponível para consulta no Anexo IV.

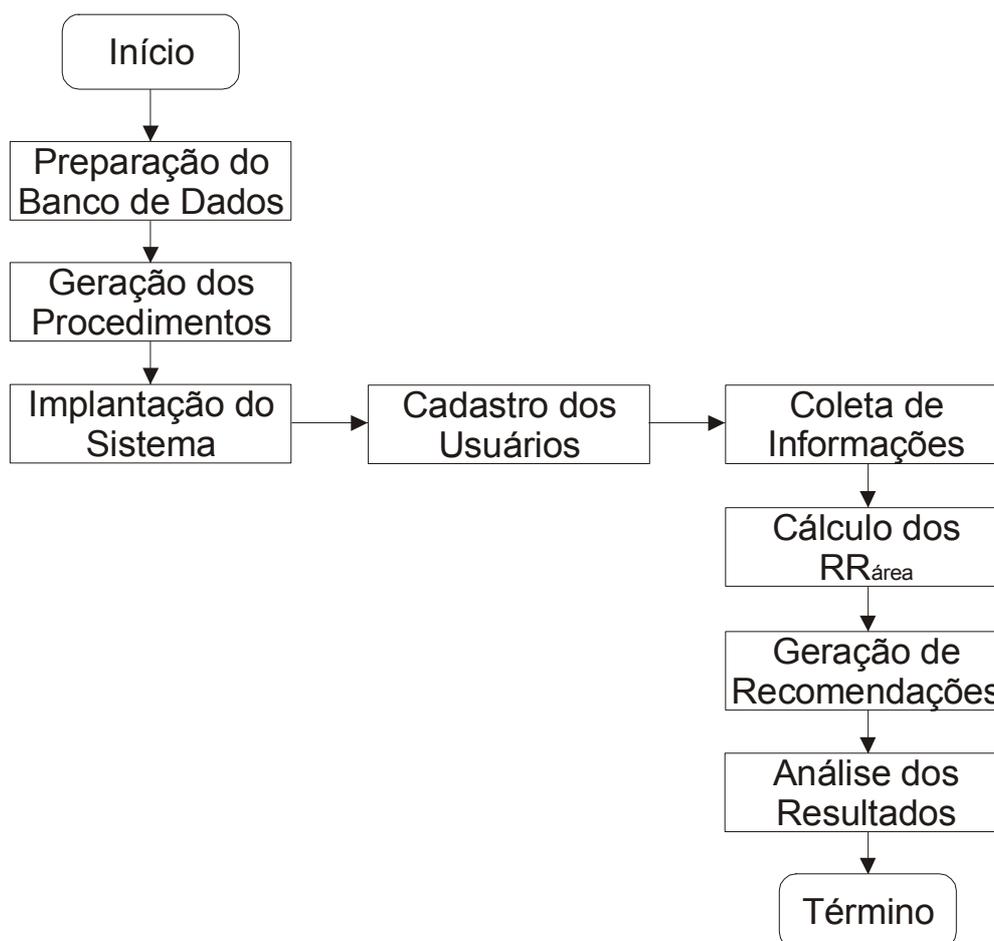


Figura 30 - Fluxo de execução da pesquisa experimental

Fonte: elaboração própria

A preparação do ambiente de banco de dados envolve a alocação de espaço para os dados que serão armazenados (conhecido como *tablespace*), a criação dos usuários administradores do banco e a criação da estrutura de dados apresentada no modelo (Anexo V).

Na geração de procedimentos de banco de dados, é realizada a concessão das autorizações de acesso à estrutura de dados e a implantação das regras de sistema, em código PHP. Montada toda a estrutura em produção, é realizado um teste de sistema, o banco de dados é limpo após o teste e o sistema é declarado implantado.

São enviados convites aos usuários, que alimentam o banco com seus cadastros. As informações são coletadas dinamicamente, de forma *on-line*. O cálculo do  $RR_{\text{área}}$ , que é realizado diariamente, é fundamental, pois é

responsável pela geração dos índices que fornecem ao sistema os *rankings* de recomendação por área de interesse, os quais serão utilizados pelas abordagens implementadas no módulo de recomendação.

Os resultados apresentados pela geração de recomendações na interface de consulta “busca pré-definida”, através do trabalho das abordagens de recomendação implementadas, formam a base para a análise de resultados. Os resultados apresentados nas interfaces de consulta “busca rápida” e “busca por apelido” não serão avaliados, pois se tratam de busca booleana simples, sem a utilização do módulo de recomendação.

No próximo capítulo, a análise de resultados será amplamente discutida.

## **6. Análise dos Dados**

Este capítulo tem como objetivo avaliar os módulos fundamentais referentes à pesquisa experimental e os resultados apresentados pelo sistema. Os módulos fundamentais são:

- Módulo de cadastro
- Módulo de consulta
- Módulo de recomendação

No módulo de cadastro, listaremos todos os campos disponíveis para elaboração do cadastro do usuário — cadastro que é realizado pelo próprio usuário do sistema. Desses campos, serão identificados quais foram selecionados para a elaboração de recomendações e o porquê da escolha desses elementos. Essa seleção tem interface com o módulo de recomendação, que, por sua vez, interage com o módulo de consulta na busca e apresentação de resultados.

O resultado, isto é, o que é esperado deste projeto, será mostrado na análise dos resultados. Antes de se realizar as considerações finais, uma transposição das interfaces será realizada para o sistema Currículo Lattes, mostrando que ele poderia da mesma forma operar segundo o modelo genérico de sistema de recomendação de perfis de usuário.

### **6.1. Análise dos Módulos**

O primeiro passo para utilização do sistema deste projeto é o cadastramento. Sem um perfil de usuário não é possível se autenticar no sistema, não sendo possível, portanto, realizar uma busca no mesmo. E, sem o perfil de usuário, não seria possível para o sistema realizar recomendações, pois estas baseiam-se no perfil do usuário.

A etapa de cadastro foi dividida em três, é uma tarefa rápida, para evitar que possíveis usuários do sistema não a realizem com pouca atenção ou má vontade, e possui poucos campos obrigatórios.

## Geral

Página 1 de 3

Nome:  !

Apelido:  ?

E-mail:

Senha:

Idioma do Perfil: Portugues ▾

Data de Nascimento:  (Ano com 4 dígitos)

Sexo:  Homem  Mulher

Orientação Sexual: heterossexual ▾ ?

Papel: Prefiro não dizer ▾

Relacionamento: Prefiro não dizer ▾

Sexo seguro: Prefiro não dizer ▾

Meu Naípe: Perfil Copas - Romance ▾

**Contato**

MSN:  !

Telefone:  !

**Hábitos**

Bebe: Não ▾

Fuma: Não ▾

Malha: Não ▾

**Localização**

País:  Brasil  Exterior

Figura 31 - Cadastro (página 1 de 3)

Fonte: elaboração própria

A primeira página do cadastro, mostrada na Figura 31, apresenta os seguintes campos obrigatórios: nome, apelido, endereço de *email*, senha, data de nascimento, sexo, orientação sexual e naípe. Os demais campos de cadastro foram idealizados a partir de páginas de cadastro de outros *sites* de relacionamento, como o MSN Paquera<sup>64</sup> ou Par Perfeito<sup>65</sup>.

<sup>64</sup> Disponível em <<http://match.br.msn.com/index.aspx>>

## Geral

Nome:  !

Apelido:  ?

E-mail:

Senha:

O seu nome não será divulgado no site. Apenas o apelido aparecerá para outros usuários

Figura 32 - Observação do campo "nome"

Fonte: elaboração própria

Alguns campos possuem uma exclamação ao lado, com informações sobre o campo que podem ser de interesse do usuário. No caso das exclamações da página 1 do cadastro, elas estão presentes em 3 campos: nome, endereço de contato instantâneo (*instant messenger*, ou *msn*) e número de telefone. Ao passar o mouse sobre as exclamações, as mensagens mostradas são as seguintes:

- Nome: “O seu nome não será divulgado no *site*. Apenas o apelido aparecerá para outros usuários”;
- Endereço de contato instantâneo: “O seu *msn* não será divulgado no *site*. Você só poderá enviá-lo para os usuários que piscarem para você”;
- Número de telefone: “O seu telefone não será divulgado no *site*. Você só poderá enviá-lo para os usuários que piscarem para você”.

Essas preocupações concernem à privacidade do usuário, que é uma das características fundamentais do sistema, constando, inclusive, nos termos de uso (veja Anexo II).

Outros campos possuem uma interrogação ao lado. Na página 1 do cadastro são: apelido e orientação sexual. As informações apresentadas são as seguintes:

<sup>65</sup> Disponível em <<http://www.parperfeito.com.br>>

- Apelido: “Preencha este campo corretamente, pois não será possível alterá-lo!”
- Orientação sexual: “Preencha este campo corretamente, pois não será possível alterá-lo!”

A garantia de que o apelido não pode ser alterado é o que facilita a busca por apelido por outros usuários, que tenham memorizado um perfil segundo seu apelido. O identificador de um perfil é o endereço de *email* — embora não possa ser alterado posteriormente, isso não é realçado na página 1 do cadastro, para evitar poluição visual.

É de suma importância avisar o usuário de que não pode alterar a sua orientação sexual porque nos termos é garantido ao usuário que somente outros usuários de sua mesma orientação sexual possuem acesso a seu perfil.

O campo “idioma do perfil” não é obrigatório e vem selecionado em português. Foram disponibilizados os idiomas português e inglês somente. A idéia é para uma posterior evolução do sistema, comportando mais de uma língua, atraindo usuários não só do Brasil. Embora nesta fase todo o sistema esteja funcionando apenas para a língua portuguesa, todos os rótulos dos campos e telas estão armazenados em uma tabela de língua, e, futuramente, bastará ao usuário selecionar outro idioma que o sistema automaticamente traduzirá todas as telas.

“Data de nascimento” é um campo crítico. O sistema não permite o cadastro de menores de idade, sendo que o usuário é responsável pela veracidade dessa informação.

### Localização

País:  Brasil  Exterior

Estado: GO

Cidade: ANAPOLIS

Bairro: Nova Vila

Figura 33 - Funcionalidade do campo "localização"

Fonte: elaboração própria

Futuramente os campos de hábito poderão ser utilizados para melhor filtragem de perfis. O campo localização, mostrado na Figura 33, está implementado com tecnologia AJAX, que permite reformulação da interface sem a necessidade de atualização da página HTML. Ao selecionar "Brasil", a interface automaticamente disponibiliza os Estados da Federação, as cidades do Estado selecionado e o campo "bairro" para ser preenchido manualmente. A recomendação do sistema leva em conta apenas o Estado (UF). Se fosse selecionado "Exterior", a apresentação seria a seguinte:

### Localização

País:  Brasil  Exterior

Figura 34 - Seleção de "Exterior" na localização

Fonte: elaboração própria

Da mesma forma que o idioma do sistema atualmente trabalha só com o português, o campo de localização para usuários estrangeiros também precisa evoluir em uma próxima versão. Atualmente, só é disponibilizado um campo para preenchimento do país do usuário que está realizando seu cadastro.

Para a realização de recomendações, são utilizados da página 1 do cadastro os seguintes campos: naípe (área de interesse do usuário) e

localização. Identificados os campos de interesse, vamos passar à página seguinte.

## Social

Página 2 de 3

**Meus dados**

Altura: 1,70m a 1,72m ▼

Biotipo: Médio ▼

Pelo corporal: Prefiro não dizer ▼

Cor dos olhos: Prefiro não dizer ▼

Etnia: Prefiro não dizer ▼

Cabelo: Prefiro não dizer ▼

Estilo: Prefiro não dizer ▼

Decoração mais marcante: Prefiro não dizer ▼

**Minha formação**

Ocupação: Ciências Sociais e Humanas ▼

Escolaridade: Ensino Superior ▼

Próximo

Figura 35 - Cadastro (página 2 de 3)

Fonte: elaboração própria

Na página 2 do cadastro, nenhuma das informações é de caráter obrigatório. Porém facilitam muito na filtragem de recomendação: é a seleção desses dados na busca pré-definida — acrescentados de “localização” e “idade entre” — que realiza a filtragem nos resultados recuperados.

Por fim, a página 3 do cadastro apresenta campos onde o usuário tem livre contexto para falar sobre si. No topo, ele dá um título para o campo texto onde pode narrar o que queira (até 1024 caracteres). Há um campo onde ele pode descrever o que procura em uma pessoa, vários campos sobre suas predileções — filmes, livros, locais que frequenta. Também é disponibilizada uma seleção de interesses-padrão, que poderão ser utilizados futuramente na recomendação de perfis.

Nenhum dos campos desta página é utilizado para recomendar ou filtrar resultados. A página 3 do cadastro é mostrada nas figuras abaixo (parte I e II). Campos não obrigatórios nesta etapa são um ponto-chave para garantir que os usuários não fiquem entediados no preenchimento do cadastro e que possam, posteriormente, preencher essas informações ao editar seu perfil.

## Pessoal

Página 3 de 3

Título introdutório

Texto introdutório

Procuo uma pessoa

Prato Preferido

Filme Preferido

Paixões

Idiomas que falo

Programa de TV

Locais que frequenta

Interesses

<input type="checkbox"/> Academia / Ginástica	<input type="checkbox"/> Golf
<input type="checkbox"/> Acampamento / Excursões	<input type="checkbox"/> Iatismo / Navegação
<input type="checkbox"/> Alpinismo	<input type="checkbox"/> Jardinagem
<input type="checkbox"/> Artes / Exposições	<input type="checkbox"/> Literatura / História
<input type="checkbox"/> Artes Maciais	<input type="checkbox"/> Museus / Galerias
<input type="checkbox"/> Astrologia / Exoterismo	<input type="checkbox"/> Música - Alternativa
<input type="checkbox"/> Atletismo	<input type="checkbox"/> Música - Blues Jazz
<input type="checkbox"/> Baseball / Softball	<input type="checkbox"/> Música - Clássica / Ópera
<input type="checkbox"/> Billar / Sinuca / Dardo	<input type="checkbox"/> Música - Cristã / Gospel
<input type="checkbox"/> Box / Lutas	<input type="checkbox"/> Música - Dance / Eletrônica / House
<input type="checkbox"/> Canto / Instrumentos Musicais	<input type="checkbox"/> Música - Latina

Figura 36 - Cadastro (página 3 de 3) - parte I

Fonte: elaboração própria

As interrogações apresentam dicas de como preencher os campos imediatamente próximos.

Locais que frequenta

Interesses

<input type="checkbox"/> Academia / Ginástica	<input type="checkbox"/> Golf
<input type="checkbox"/> Acampamento / Excursões	<input type="checkbox"/> Iatismo / Navegação
<input type="checkbox"/> Alpinismo	<input type="checkbox"/> Jardinagem
<input type="checkbox"/> Artes / Exposições	<input type="checkbox"/> Literatura / História
<input type="checkbox"/> Artes Maciais	<input type="checkbox"/> Museus / Galerias
<input type="checkbox"/> Astrologia / Exoterismo	<input type="checkbox"/> Música - Alternativa
<input type="checkbox"/> Atletismo	<input type="checkbox"/> Música - Blues Jazz
<input type="checkbox"/> Baseball / Softball	<input type="checkbox"/> Música - Clássica / Ópera
<input type="checkbox"/> Billar / Sinuca / Dardo	<input type="checkbox"/> Música - Cristã / Gospel
<input type="checkbox"/> Box / Lutas	<input type="checkbox"/> Música - Dance / Eletrônica / House
<input type="checkbox"/> Canto / Instrumentos Musicais	<input type="checkbox"/> Música - Latina
<input type="checkbox"/> Carros / Motocicletas	<input type="checkbox"/> Música - New Age
<input type="checkbox"/> Ciclismo	<input type="checkbox"/> Música - Pop / Ritm & Blues
<input type="checkbox"/> Cinema / Televisão	<input type="checkbox"/> Música - Rock
<input type="checkbox"/> Comidas e Vinhos	<input type="checkbox"/> Música - Sertanejo
<input type="checkbox"/> Computadores / Internet	<input type="checkbox"/> Música - World Music
<input type="checkbox"/> Corrida de Carros	<input type="checkbox"/> Natureza
<input type="checkbox"/> Cricket	<input type="checkbox"/> Obras de caridade / Voluntariado
<input type="checkbox"/> Culinária	<input type="checkbox"/> Política
<input type="checkbox"/> Dança	<input type="checkbox"/> Religião / Teologia
<input type="checkbox"/> Danceterias / Bares / Nightclubs	<input type="checkbox"/> Teatro / Ballet
<input type="checkbox"/> Esportes Aquáticos	<input type="checkbox"/> Tennis / Squash / badminton
<input type="checkbox"/> Esportes de Inverno	<input type="checkbox"/> Volleyball / Basketball
<input type="checkbox"/> Moda / Beleza / Vida Social	<input type="checkbox"/> Yoga / Meditação
<input type="checkbox"/> Futebol / Rugby	

Figura 37 - Cadastro (página 3 de 3) - parte II

Fonte: elaboração própria

Um cadastro bem realizado é etapa fundamental para qualquer sistema de recuperação de informações eficiente. Finalizada esta etapa, a próxima tela disponibiliza ao usuário o módulo de consulta, que em um primeiro momento torna obrigatória a seleção de pelo menos um interesse para a realização de uma filtragem nos resultados recuperados.

## Bem-vindo, Tiago

⚠ O Dois Corações é um sistema de recomendações inteligente, e para que funcione, é necessário que você defina uma busca de acordo com seus interesses. Portanto, selecione os campos abaixo.

Busca pré-definida

<b>Altura:</b>	<input type="checkbox"/> Anão a 1,40m	<b>Localização:</b>	<input type="checkbox"/> AL	<input type="checkbox"/> PB
	<input type="checkbox"/> 1,41m a 1,65m		<input type="checkbox"/> AM	<input type="checkbox"/> PE
	<input type="checkbox"/> 1,66m a 1,70m		<input type="checkbox"/> AP	<input type="checkbox"/> PI
	<input type="checkbox"/> 1,70m a 1,72m		<input type="checkbox"/> BA	<input type="checkbox"/> PR
	<input type="checkbox"/> 1,73m a 1,75m		<input type="checkbox"/> CE	<input type="checkbox"/> RJ
	<input type="checkbox"/> 1,76m a 1,78m		<input type="checkbox"/> DF	<input type="checkbox"/> RN
	<input type="checkbox"/> 1,79m a 1,81m		<input type="checkbox"/> ES	<input type="checkbox"/> RO
	<input type="checkbox"/> 1,82m a 1,84m		<input type="checkbox"/> GO	<input type="checkbox"/> RR
	<input type="checkbox"/> 1,82m a 1,83m		<input type="checkbox"/> MA	<input type="checkbox"/> RS
	<input type="checkbox"/> 1,84m a 1,85m		<input type="checkbox"/> MG	<input type="checkbox"/> SC
	<input type="checkbox"/> 1,86m a 1,87m		<input type="checkbox"/> MS	<input type="checkbox"/> SE
	<input type="checkbox"/> 1,88m a Gigante		<input type="checkbox"/> MT	<input type="checkbox"/> SP
			<input type="checkbox"/> PA	<input type="checkbox"/> TO
<b>Biotipo:</b>	<input type="checkbox"/> Esquelético		<input type="checkbox"/> Exterior	
	<input type="checkbox"/> Magro/Pequeno	<b>Pelo corporal:</b>	<input type="checkbox"/> Nenhum	
	<input type="checkbox"/> Magro a Médio		<input type="checkbox"/> Pouco	
	<input type="checkbox"/> Médio		<input type="checkbox"/> Médio	
	<input type="checkbox"/> Médio a Definido		<input type="checkbox"/> Muito	
	<input type="checkbox"/> Médio a Gordinho		<input type="checkbox"/> Raspado/Depilado	
	<input type="checkbox"/> Definido		<input type="checkbox"/> Aparado	
	<input type="checkbox"/> Definido a Musculoso	<b>Cor dos olhos:</b>	<input type="checkbox"/> Azul	
	<input type="checkbox"/> Musculoso		<input type="checkbox"/> Verde	
	<input type="checkbox"/> Bombado		<input type="checkbox"/> Castanho Claro/Mel	
	<input type="checkbox"/> Gordinho		<input type="checkbox"/> Castanho	
	<input type="checkbox"/> Gordo/Grande		<input type="checkbox"/> Cinza	
	<input type="checkbox"/> Obeso		<input type="checkbox"/> Preto	
<b>Etnia:</b>	<input type="checkbox"/> Indígena	<b>Cabelo:</b>	<input type="checkbox"/> Nenhum	
	<input type="checkbox"/> Asiática		<input type="checkbox"/> Pouco	
	<input type="checkbox"/> Negra		<input type="checkbox"/> Normal	
	<input type="checkbox"/> Caucasiana		<input type="checkbox"/> Longo	
	<input type="checkbox"/> Hispânica/Latina		<input type="checkbox"/> Raspado	
	<input type="checkbox"/> Oriente médio		<input type="checkbox"/> Rente/Estilo Soldado	
	<input type="checkbox"/> Várias		<input type="checkbox"/> Anelado	
<b>Estilo:</b>	<input type="checkbox"/> Alternativo		<input type="checkbox"/> Caracóis/Estilo anjinhos	
	<input type="checkbox"/> Casual		<input type="checkbox"/> Enrolado	
	<input type="checkbox"/> Esportivo		<input type="checkbox"/> Cachos	
	<input type="checkbox"/> Estudante		<input type="checkbox"/> Bombril	
	<input type="checkbox"/> Executivo/Formal	<b>Decoração mais marcante:</b>	<input type="checkbox"/> Brincos	
	<input type="checkbox"/> Fashion/Moderno		<input type="checkbox"/> Brinco no nariz	
	<input type="checkbox"/> Gótico		<input type="checkbox"/> Piercings no corpo	
	<input type="checkbox"/> Leather		<input type="checkbox"/> Tattoos	
	<input type="checkbox"/> Mauricinho/Patty		<input type="checkbox"/> Outros	
	<input type="checkbox"/> Militar		<input type="checkbox"/> Nenhuma	
	<input type="checkbox"/> Na moda			

Idade entre:  e

Em ordem de:

Figura 38 - Tela de seleção de interesses

Fonte: elaboração própria

Vale notar que a área de interesse do usuário, isto é, o naipe, foi uma informação coletada da página 1 do cadastro, o que já permitiria a geração de recomendações desde a finalização da etapa de cadastro — neste ponto já se encontra disponível o perfil completo do usuário e sua área de interesse. Por que, então, não gerar recomendações sem a seleção de interesses?

As recomendações não são geradas imediatamente após o fim da etapa de cadastro porque precisamos coletar mais informações para prover o correto funcionamento dos algoritmos de recomendação. Se o usuário começa a navegar no sistema sem definir pelo menos um interesse na interface de busca pré-definida, fica mais difícil, por exemplo, encontrar um usuário semelhante a ele na abordagem de recomendação *Cross-Sell*. A seleção de interesses ajuda a entender melhor os grupos de usuário, favorecendo o trabalho do módulo de recomendação.

Vencida a seleção de interesses, o sistema apresenta a tela principal, que se trata da interface de busca pré-definida. Conforme pode ser observado na Figura 39, o sistema apresenta as recomendações imediatamente acima dos campos disponíveis de seleção de interesse.



Figura 39 - Tela principal do sistema 2coracoes

Fonte: elaboração própria

O sistema apresenta até 3 recomendações por vez. Nessa apresentação a abordagem baseada em conhecimento está sendo utilizada em conjunto com as abordagens baseada em conteúdo e *Top-N*, quer dizer, regras do sistema são utilizadas em conjunto com técnicas de filtragem e *rankings* de perfis de usuários para apresentar recomendações que venham atender à necessidade do usuário.

A abordagem *Top-N* não é utilizada explicitamente, não estão disponíveis para o usuário, por exemplo, *rankings* de páginas mais visitadas ou perfis que receberam melhores avaliações positivas. A fim de evitar o lugar comum — a apresentação dessas listas em um menu de navegação —, o sistema se aproveita dos cálculos de RR e utiliza a abordagem *Top-N* encapsulada no módulo de consulta para a apresentação dos resultados.

Os resultados recuperados pela outra abordagem, *Cross-Sell*, que foi adaptada para recomendar perfis de usuário em *sites* de relacionamento, pode ser conferida na Figura 40.

The image shows a user profile interface for 'FemininoHetero'. The profile details include:

- Nome:** Perfil Feminino Hetero
- Idade:** 19 anos
- Data de Nascimento:** 24/05/1987
- Sexo:** Mulher
- Orientação Sexual:** heterossexual
- Papel:** Prefiro não dizer
- Relacionamento:** Prefiro não dizer
- Sexo seguro:** Prefiro não dizer
- Meu Naibe:** Perfil Coringa - Mais de um ou todos acima
- Procurando alguém com a idade entre:** 18 e 19 anos
- Contato:** MSN: Não disponível, Telefone: Não disponível

On the left side, there is a navigation menu with options: Principal, Perfil, and Ver álbum. Below this is a section titled 'Perfis semelhantes' which contains a profile card for 'AnaPascal, 19 anos, Brasil, GO, Copas - Romance'. This profile card is circled in orange, and an orange arrow points from it towards the main profile details.

Figura 40 - Resultado apresentado pela abordagem Cross-Sell

Fonte: elaboração própria

Com apoio das técnicas das demais abordagens de recomendação — regras do sistema, filtragem, estatísticas e outras informações — a abordagem *Cross-Sell* realiza sua tarefa de identificar perfis semelhantes e apresentá-los no perfil que está sendo consultado. Foi definido que apenas um perfil é mostrado no menu do lado esquerdo, a fim de evitar poluição visual.

Muitos dos critérios adotados para o projeto não são ideais para um sistema comercial, mas, para a realização de um projeto acadêmico, são suficientes. Não adotar a abordagem *Top-N* de forma explícita, por exemplo, contraria toda uma legião de *sites* de relacionamento do mercado. Adotar a *Top-N* explícita, no entanto, não traria nenhuma diferença para os resultados do projeto. Para adequar ao prazo, foi tomada a decisão de projeto de não adotar a abordagem *Top-N* explícita neste momento.

Como o sistema se apóia na busca booleana como estratégia fundamental para a recuperação de informação, os resultados que podem

ser filtrados por operadores lógicos são alcançados invariavelmente. Outros registros que não são alcançados pela busca booleana podem aparecer como resultado através do uso da abordagem *Cross-Sell*, na tentativa de satisfazer a necessidade de informação do usuário.

As demais abordagens de recomendação trabalham em conjunto para organizar a apresentação dos resultados e o *Ranking* de Recomendação auxilia o sistema na ordenação dos registros para apresentação, de acordo com a área de interesse — grupo de usuários com interesses semelhantes.

## **6.2. Análise dos Resultados**

Segundo Torres Jr. (2004), as técnicas de filtragem colaborativa e filtragem baseada no conteúdo — e abordagens híbridas utilizando ambas as técnicas — se mostraram estratégias eficientes para a recuperação de informação em sistemas de recomendação nos domínios em que foram estudados. Torres Jr. (2004) demonstrou em seus estudos que a utilização das abordagens de recomendação é útil para a recuperação de artigos científicos e ainda provou que algoritmos diferentes são mais adequados para a recomendação de cada determinado tipo de documento (relatórios são diferentes de livros, que são diferentes de pareceres, que são diferentes de resenhas. Cada tipo de documento possui características próprias que influenciam a forma de o sistema gerar recomendações).

Este trabalho propôs estudar a recomendação de perfis de usuários, fazendo uso das técnicas conhecidas da Ciência da Computação e trabalhando a apresentação dos resultados nos moldes da Ciência da Informação. Os registros que foram recuperados não possuem características diferentes, todos são representações dos usuários — são os perfis dos usuários — e os indicadores são comuns entre eles. Não foi preciso gerar algoritmos diferentes para a recuperação de perfis com área de interesse diferente, ou com diferente orientação sexual. Filtros apenas foram usados para distinguir os grupos e os algoritmos trabalharam de forma igual para todos os registros.

A recuperação de informação foi possível utilizando-se abordagens de recomendação no domínio da recuperação de perfis de usuário.

O modelo genérico foi idealizado antes de se mostrar que era possível a utilização de recomendação para a recuperação de perfis de usuário. O estudo foi realizado sobre um subdomínio deste, a recuperação de currículos. Identificando os indicadores, criou-se um modelo genérico que foi utilizado para implementar o sistema em outro subdomínio da recuperação de perfis de usuário: a recuperação de perfis de usuário em *site* de relacionamentos. A transposição do sistema implementado será mostrada a seguir, logo após a análise dos resultados, confirmando que o modelo genérico é adaptável ao domínio/subdomínio aplicado.

Não foi possível verificar se usuários com diferentes níveis de experiência percebem as recomendações de maneira diferente. Não houve tempo hábil para a execução das enquetes, o que seria extremamente útil para verificar a satisfação do usuário quanto às respostas oferecidas pelo sistema. Esta atividade ficou para trabalho futuro.

### **6.3. Transposição do Sistema Implementado para o Currículo Lattes**

Esta etapa foi dividida em duas partes. Na primeira, serão levantados os indicadores que podem fazer parte da recomendação de currículos no sistema Currículo Lattes. Na segunda parte, as telas do sistema 2coracoes serão adaptadas para o contexto da recuperação de currículos, de forma a visualizar como a apresentação dos resultados pode ser adaptada a um sistema de recuperação de currículos. O resultado esperado dessa transposição é descrito na terceira seção, criada para este fim.

#### **6.3.1. Identificação dos Indicadores**

Os identificadores para um sistema de recuperação de currículos foram levantados durante o estudo do modelo genérico para a recomendação de perfis de usuário.

Este projeto, a princípio, seria aplicado e estudado sobre a plataforma Lattes, e os estudos iniciais foram voltados para este objetivo. No entanto, não houve retorno do CNPq quanto aos contatos realizados. Em um segundo momento — a menos de seis meses da finalização do projeto — um encontro com o CNPq foi possível, porém o estudo já estava direcionado para a recuperação de perfis de usuários em *sites* de relacionamento, com o propósito da criação do modelo genérico de recomendação de perfis de usuário.

Este encontro não foi em vão e ajudou a isolar os indicadores necessários para a recuperação de currículos e, também, a visualizar os benefícios que um sistema de recomendação traria para os usuários internos da plataforma Lattes. O sistema Currículo Lattes permite a consulta de currículos para usuários externos, mas, internamente, ele é utilizado na identificação de especialistas que fornecem pareceres e avaliam projetos acadêmicos de todo o país. Haveria muitos ganhos na utilização de um sistema de recomendação para os usuários internos, e isso ficou evidente durante a reunião. Voltaremos a esta linha de raciocínio ao tratar os resultados esperados na aplicação do modelo de recomendação no sistema Currículo Lattes, na Seção 6.3.3 (p.146).

Segundo o modelo genérico de recuperação de perfis de usuário, os indicadores se dividem em dois grupos: (i) variáveis qualitativas e (ii) variáveis quantitativas.

As variáveis qualitativas para a recomendação de currículos são: (i<sub>1</sub>) nível do usuário — que pode ser nível acadêmico ou nível profissional. Cada nível possui um peso associado, que vai refletir no cálculo do índice  $RR_{\text{área}}$ , conforme visto em capítulos anteriores. As Tabela 3 (p.79) e Tabela 4 (p.80) apresentam os diferentes níveis acadêmicos e profissionais —, (i<sub>2</sub>) área de interesse — que pode ser uma ou mais de uma. Para cada área de interesse é calculado um  $RR_{\text{área}}$ . O  $RR_{\text{total}}$ , que é o índice geral da recomendação de um currículo, pode ser representado pela média aritmética dos  $RR_{\text{área}}$ s, ou pelo primeiro índice RR obtido sem fazer a distinção por área

de interesse. Essa escolha cabe ao administrador do sistema — e ( $i_3$ ) tempo de atuação na área de interesse<sup>66</sup>. É importante, neste ponto, explicitar que, quanto mais específica é uma área de interesse, melhor é para o sistema identificar os usuários de opinião mais relevante nesta área.

As variáveis quantitativas para a recomendação de currículos são extraídas das informações do perfil do usuário — neste caso, de seu currículo cadastrado no sistema. O administrador pode selecionar o que acha mais relevante. Baseando-se no estudo de Cazella e Álvares (2005), foram sugeridas as seguintes variáveis como indicadores quantitativos para a recomendação de currículos: ( $ii_1$ ) publicações em revistas ou jornais, ( $ii_2$ ) publicações em eventos, ( $ii_3$ ) livros publicados, ( $ii_4$ ) supervisão de estudantes e ( $ii_5$ ) participação em bancas. A estas sugestões podem ser acrescentadas outras variáveis de indicadores de produção, tudo a critério do administrador do sistema de recomendação<sup>67</sup>.

Estas são as variáveis a serem trabalhadas na recomendação de currículos, e foi demonstrado em capítulos anteriores como são calculados os índices de *Ranking* de Recomendação, a ferramenta chave para selecionar os perfis (currículos) de usuários que prometem opinião mais relevante.

### 6.3.2. Telas

Adequando a proposta do modelo genérico ao atual sistema Currículo Lattes, modificações seriam necessárias tanto na apresentação das telas do sistema quanto nos algoritmos que tratam as consultas às informações.

Primeiramente, se o foco está voltado para o Currículo Lattes de consulta ao domínio público, seria necessário exigir uma autenticação no sistema antes da realização de consultas, permitindo assim capturar

---

<sup>66</sup> Espera-se que profissionais que tenham atuado por mais tempo em determinada área possuam melhor conhecimento acerca deste campo.

<sup>67</sup> Até mesmo variáveis qualitativas podem ser acrescentadas, como o “nível do curso de pós-graduação onde é docente”. O administrador do sistema de recomendação deve fazer um estudo minucioso do que é ou não importante da necessidade de informação de seus usuários.

informações implícitas e explícitas da interação do usuário com o sistema. Sem isso, dados das consultas anônimas não seriam confiáveis para gerar recomendações para os grupos.

Tomando como premissa que o usuário já se identificou para o sistema, os ganhos que as recomendações trariam para o Currículo Lattes seriam: (a) Ordenação do resultado das consultas por *Ranking* de Recomendação, (b) identificação e apresentação de perfis semelhantes, (c) *feedback* dos usuários quanto aos perfis de pesquisadores semelhantes a ele.

O item (a) facilitaria identificar os pesquisadores com maior relevância de opinião. Ajudaria, por exemplo, na seleção da composição de bancas para avaliação de projetos do CNPq, uma das carências do uso interno do sistema. Essa vantagem — a ordenação por perfis segundo a relevância de opinião — se mostra extremamente interessante do ponto de vista técnico.

No dia-a-dia o CNPq utiliza outros critérios, tais como áreas de atuação, projetos realizados, dentre outros, e critérios excludentes, como não ter sido orientador ou professor, não participar da equipe do projeto, não pertencer à mesma instituição, não ter participado de projeto conjunto ou de redação de artigo em conjunto. Como o número de variáveis é maior, mais difícil é estabelecer pesos para elas. O melhor é permitir que o usuário selecione as variáveis e seus pesos para o cálculo do *Ranking* de Recomendação.

**Resultado da Busca**

**Resultados encontrados: 3127**  
Parâmetros de busca: Buscar por "inteligência artificial"

[Nova consulta](#) [Notas](#) [Opções de visualização](#)

- 1. Fernando Mendes de Azevedo**   
[Currículo](#) [Grupos de Pesquisa](#)  
Bolsista de Produtividade Desen. Tec. e Extensão Inovadora 1D || Orientador de Mestrado || Orientador de Doutorado  
Maior nível de formação: Doutorado  
Primeira área de atuação: Engenharias - Engenharia Biomédica  
Instituição: Universidade Federal de Santa Catarina  
Email: azevedo@ieb.ufsc.br
- 2. Pedro Paulo Balbi de Oliveira**   
[Currículo](#) [Grupos de Pesquisa](#)  
Orientador de Mestrado |  
Maior nível de formação: Doutorado  
Primeira área de atuação: Ciências Exatas e da Terra - Ciência da Computação  
Instituição: Universidade Presbiteriana Mackenzie  
Email: pedrob@mackenzie.br
- 3. Ricardo Miranda Barcia**  
[Currículo](#)



Figura 41 - Busca pelo assunto "inteligência artificial"

Fonte: elaboração própria, a partir de consulta ao Currículo Lattes

A Figura 41 apresenta os resultados de uma consulta ao sistema pelo assunto "inteligência artificial". O critério para ordenação dos resultados não é claro, portanto não é possível que atenda a necessidade do usuário.

Realizando nova consulta, cerca de um minuto depois, para verificar se a apresentação dos resultados era ou não randômica, foi interessante constatar que diminuíram dois currículos do número de perfis<sup>68</sup>:

<sup>68</sup> Uma terceira consulta foi realizada e o número caiu para 3124. Talvez o sistema estivesse passando por uma manutenção, ou então há muitos usuários com interesse em inteligência artificial deletando seus perfis ou desistindo dessa área de interesse.

**Resultado da Busca**

**Resultados encontrados: 3125**

Parâmetros de busca: Buscar por 'inteligência artificial'

[Nova consulta](#) [Notas](#) [Opções de visualização](#)

- 1. Fernando Mendes de Azevedo**  
[Currículo](#) [Grupos de Pesquisa](#)  
Bolsista de Produtividade Desen. Tec. e Extensão Inovadora 1D || Orientador de Mestrado || Orientador de Doutorado  
Maior nível de formação: Doutorado  
Primeira área de atuação: Engenharias - Engenharia Biomédica  
Instituição: Universidade Federal de Santa Catarina  
Email: azevedo@ieb.ufsc.br
- 2. Pedro Paulo Balbi de Oliveira**  
[Currículo](#) [Grupos de Pesquisa](#)  
Orientador de Mestrado |  
Maior nível de formação: Doutorado  
Primeira área de atuação: Ciências Exatas e da Terra - Ciência da Computação  
Instituição: Universidade Presbiteriana Mackenzie  
Email: pedrob@mackenzie.br
- 3. Ricardo Miranda Barcia**  
[Currículo](#)  
Orientador de Doutorado  
Maior nível de formação: Doutorado  
Primeira área de atuação: Engenharias - Multidisciplinar  
Instituição: Instituto Virtual de Estudos Avançados  
Email: rbarcia@vias.org.br



**Carlos Chagas**

Figura 42 - Redução do resultado na mesma consulta por assunto

Fonte: elaboração própria, a partir de consulta ao Currículo Lattes

A única interação que o sistema permite após a apresentação dos resultados está no *link* “Opções de visualização”, que é um filtro para selecionar os campos a serem mostrados, e/ou apresentação de período mais recente de produção científica, como mostrado na Figura 43.

**Opções de Visualização de Currículos**

**Opções de Visualização de Currículos:**

<p><b>Informações Pessoais</b></p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Endereço</li> <li><input checked="" type="checkbox"/> Formação Acadêmica/Titulação</li> <li><input checked="" type="checkbox"/> Atuação profissional</li> <li><input checked="" type="checkbox"/> Áreas de atuação</li> <li><input checked="" type="checkbox"/> Idiomas</li> <li><input checked="" type="checkbox"/> Prêmios e títulos</li> </ul> <p><b>Informações sobre produções bibliográficas</b></p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Artigos publicados</li> <li><input checked="" type="checkbox"/> Livros e capítulos</li> <li><input checked="" type="checkbox"/> Trabalhos em eventos</li> <li><input checked="" type="checkbox"/> Texto em jornal ou revista</li> <li><input checked="" type="checkbox"/> Outras produções bibliográficas</li> </ul> <p><b>Período da produção</b></p> <p><input checked="" type="radio"/> Todo o período</p> <p><input type="radio"/> A partir do ano <input style="width: 50px;" type="text"/></p> <p><input type="checkbox"/> Marcar todos</p>	<p><b>Informações sobre produção técnicas</b></p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Softwares</li> <li><input checked="" type="checkbox"/> Produtos</li> <li><input checked="" type="checkbox"/> Processos</li> <li><input checked="" type="checkbox"/> Trabalhos técnicos</li> <li><input checked="" type="checkbox"/> Outras produções técnicas</li> </ul> <p><b>Informações sobre demais produções/trabalhos</b></p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Produção artística/cultural</li> <li><input checked="" type="checkbox"/> Orientações concluídas</li> <li><input checked="" type="checkbox"/> Orientações em andamento</li> <li><input checked="" type="checkbox"/> Demais Trabalhos</li> </ul> <p><b>Outras Informações</b></p> <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Dados complementares</li> <li><input checked="" type="checkbox"/> Outras informações relevantes</li> </ul> <p><b>Período da atuação Profissional</b></p> <p><input checked="" type="radio"/> Todo o período</p> <p><input type="radio"/> A partir do ano <input style="width: 50px;" type="text"/></p>
---	---

Não mostrar essa janela da próxima vez.

*Figura 43 - Opções de visualização*

*Fonte: elaboração própria, a partir de consulta ao Currículo Lattes*

Para confirmar que o sistema não permitia filtrar os resultados por período da produção, foi alterado o campo “A partir do ano” para 2003, e depois clicado em nova consulta. Confirmou-se que “Opções de visualização” é apenas um filtro sobre o que será mostrado após clicar em algum dos currículos apresentados. O mais curioso é que o resultado caiu para 3122, sendo que um dos primeiros perfis da lista de resultados não mais aparecia.

**Resultado da Busca**

**Resultados encontrados: 3122**  
Parâmetros de busca: Buscar por 'inteligência artificial'

[Nova consulta](#) [Notas](#) [Opções de visualização](#)

- Fernando Mendes de Azevedo**   
[Currículo](#) | [Grupos de Pesquisa](#)  
Bolsista de Produtividade Desen. Tec. e Extensão Inovadora 1D || Orientador de Mestrado || Orientador de Doutorado  
Maior nível de formação: Doutorado  
Primeira área de atuação: Engenharias - Engenharia Biomédica  
Instituição: Universidade Federal de Santa Catarina  
Email: azevedo@ieb.ufsc.br
- Ricardo Miranda Barcia**  
[Currículo](#)  
Orientador de Doutorado  
Maior nível de formação: Doutorado  
Primeira área de atuação: Engenharias - Multidisciplinar  
Instituição: Instituto Virtual de Estudos Avançados  
Email: rbarcia@vias.org.br
- Anita Maria da Rocha Fernandes**   
[Currículo](#) | [Grupos de Pesquisa](#)



**Carlos Chagas**

Figura 44 - Mudança nos resultados da busca por assunto

Fonte: elaboração própria, a partir de consulta ao Currículo Lattes

Na consulta pelo perfil desaparecido, verificou-se que ainda consta “inteligência artificial” como palavra-chave, porém o perfil sofrera atualização no mesmo dia da consulta. O sistema Currículo Lattes altera a ordenação baseando-se na data de atualização? Essa dúvida não poderá ser sanada neste momento, ficando registrada para futura investigação.

Declarado como as recomendações auxiliariam o Currículo Lattes na ordenação dos resultados, o item (b) mostra como a prática da abordagem *Cross-Sell* facilitaria encontrar especialistas semelhantes. Foi feita uma busca por usuários interessados em recuperação da informação. Os resultados estão apresentados na Figura 45.

**Resultado da Busca**

**Resultados encontrados: 2629**  
Parâmetros de busca: Buscar por 'recuperação da informação'

[Nova consulta](#) [Notas](#) [Opções de visualização](#)

- Marcos Luiz Cavalcanti de Miranda**   
[Currículo](#) [Grupos de Pesquisa](#)  
Maior nível de formação: Doutorado  
Primeira área de atuação: Ciências Sociais Aplicadas - Multidisciplinar  
Instituição: Universidade Federal do Estado do Rio de Janeiro  
Email: mlmiranda@unirio.br
- Edilson Damasio**   
[Currículo](#)  
Maior nível de formação: Mestrado  
Primeira área de atuação: Ciências Sociais Aplicadas - Multidisciplinar  
Instituição: Universidade Estadual de Maringá  
Email: edamasio@uem.br
- Fernanda Mirelle de Almeida Silva**  
[Currículo](#) [Grupos de Pesquisa](#)  
Maior nível de formação: Não informado  
Primeira área de atuação: Ciências Sociais Aplicadas - Ciência da Informação  
Instituição: Universidade Federal da Paraíba  
Email: fmirelle@gmail.com



**Milton Santos**

*Figura 45 - Resultados da consulta por "recuperação da informação"*

*Fonte: elaboração própria, a partir de consulta ao Currículo Lattes*

Selecionou-se o primeiro currículo da lista para trabalhar a tela de forma a apresentar a utilização da abordagem *Cross-Sell*. Abaixo segue a tela original (Figura 46):

**Marcos Luiz Cavalcanti de Miranda**

possui graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (1989), mestrado (1997) e doutorado (2005) em Ciência da Informação pela Universidade Federal do Rio de Janeiro. Atualmente é Diretor da Escola de Biblioteconomia, Coordenador do Curso de Especialização em Organização do Conhecimento para Recuperação da Informação e professor adjunto da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) atuando nos Cursos de Bacharelado em Biblioteconomia e em Arquivologia, Especialização em Organização do Conhecimento para Recuperação da Informação e Mestrado em Museologia e Patrimônio da UNIRIO. Tem experiência na área de Biblioteconomia e Ciência da Informação, com ênfase em Organização do Conhecimento, trabalhando principalmente nos seguintes temas: organização do conhecimento, ambientes Web, Museologia, ensino e formação de bibliotecários. É Membro do Conselho Editorial das revistas Arquivística.net (1808-4826) e Revist@CRB-7.  
(Texto informado pelo autor)

Última atualização do currículo em 12/02/2007  
Endereço para acessar este CV:  
<http://lattes.cnpq.br/8713013619609185>

**Certificado pelo autor em 12/02/07**

**Outros links:**  
[Diretório de grupos de pesquisa](#)  
[Univ Federal do Rio de Janeiro](#)

[Dados pessoais](#)    [Formação acadêmica/Titulação](#)    [Atuação profissional](#)    [Linhas de pesquisa](#)  
[Áreas de atuação](#)    [Idiomas](#)    [Projetos de pesquisa](#)    [Membro de Corpo Editorial](#)  
[Eventos](#)    [Orientações](#)    [Produção em C.T & A](#)    [Bancas](#)

Figura 46 - Primeiro currículo da consulta por "recuperação da informação"

Fonte: elaboração própria, a partir de consulta ao Currículo Lattes

Interessante observar que o Currículo Lattes está em constante evolução. No início deste projeto as consultas não mostravam fotos dos usuários, tampouco recomendava os *links* ao lado, onde mostra o pesquisador no diretório de grupos de pesquisa e o *link* para a universidade a qual o pesquisador está vinculado. A estratégia utilizada na criação desses *links* foi a abordagem baseada no conteúdo.

Com o uso da abordagem *Cross-Sell*, o sistema poderia apresentar o currículo da forma mostrada na Figura 47.

**Marcos Luiz Cavalcanti de Miranda**

possui graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (1989), mestrado (1997) e doutorado (2005) em Ciência da Informação pela Universidade Federal do Rio de Janeiro. Atualmente é Diretor da Escola de Biblioteconomia, Coordenador do Curso de Especialização em Organização do Conhecimento para Recuperação da Informação e professor adjunto da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) atuando nos Cursos de Bacharelado em Biblioteconomia e em Arquivologia, Especialização em Organização do Conhecimento para Recuperação da Informação e Mestrado em Museologia e Patrimônio da UNIRIO. Tem experiência na área de Biblioteconomia e Ciência da Informação, com ênfase em Organização do Conhecimento, trabalhando principalmente nos seguintes temas: organização do conhecimento, ambientes Web, Museologia, ensino e formação de bibliotecários. É Membro do Conselho Editorial das revistas Arquivística.net (1808-4826) e Revist@CRB-7. (Texto informado pelo autor)

Última atualização do currículo em 12/02/2007  
Endereço para acessar este CV:  
<http://lattes.cnpq.br/8713013619609185>

**Certificado pelo autor em 12/02/07**

**Outros links:**  
[Diretório de grupos de pesquisa](#)  
[Univ Federal do Rio de Janeiro](#)

**Currículos semelhantes**

- Maria del Rosario Girardi**  
Universidade Federal do Maranhão
- Edilson Damasio**  
Universidade Estadual de Maringá
- Nair Yumiko Kobashi**  
Universidade de São Paulo

**Dados pessoais**   **Formação acadêmica/Titulação**   **Atuação profissional**   **Linhas de pesquisa**  
**Áreas de atuação**   **Idiomas**   **Projetos de pesquisa**   **Membro de Corpo Editorial**  
**Eventos**   **Orientações**   **Produção em C,T & A**   **Bancas**

Figura 47 - Uso da abordagem Cross-Sell no Currículo Lattes

Fonte: elaboração própria, adaptado de consulta ao Currículo Lattes

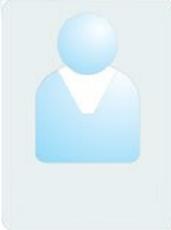
Para identificar perfis semelhantes poderiam ser trabalhadas variáveis, como o argumento da busca — neste caso, “recuperação da informação” —, índices semelhantes de  $RR_{\text{área}}$  e a filtragem de campos que justifiquem a indicação de um usuário semelhante — campos estes que poderiam ser elaborados pelo administrador do sistema, mas que permitissem modificações pelo usuário<sup>69</sup>, garantindo a satisfação de sua necessidade de informação.

O terceiro item — item (c), *feedback* dos usuários quanto aos perfis de pesquisadores semelhantes a ele — se trata de uma coleta explícita da interação do usuário com o sistema. No sistema implementado por este

<sup>69</sup> através de um menu de configuração, por exemplo.

projeto, o 2coracoes, utilizamos uma interface com os símbolos de positivo e negativo para que o usuário informasse se o perfil retornado em um resultado de busca era ou não de seu interesse, nos moldes de “gostei” ou “não gostei”. De forma semelhante, aplicando esta funcionalidade na avaliação de currículos no sistema Currículo Lattes, poderíamos utilizar os mesmos símbolos nos moldes de “tem relação com minha área de atuação” ou “não tem relação com minha área de atuação”.

Explicando melhor a situação descrita no parágrafo anterior, o sistema disponibilizaria uma funcionalidade que permitiria a interação do usuário na avaliação de currículos que podem estar relacionados à sua área de atuação. Esta funcionalidade não seria mostrada em todos os currículos visualizados, somente naqueles que o sistema identificasse como currículos semelhantes ao do usuário. Na prática essa variável não poderia ser tratada como quantitativa, seria uma variável qualitativa que ajudaria o sistema a identificar grupos específicos dentro das áreas de interesse identificadas automaticamente, melhorando cada vez mais o sistema na geração de recomendações mais precisas.



### Marcos Luiz Cavalcanti de Miranda

possui graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (1989), mestrado (1997) e doutorado (2005) em Ciência da Informação pela Universidade Federal do Rio de Janeiro. Atualmente é Diretor da Escola de Biblioteconomia, Coordenador do Curso de Especialização em Organização do Conhecimento para Recuperação da Informação e professor adjunto da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) atuando nos Cursos de Bacharelado em Biblioteconomia e em Arquivologia, Especialização em Organização do Conhecimento para Recuperação da Informação e Mestrado em Museologia e Patrimônio da UNIRIO. Tem experiência na área de Biblioteconomia e Ciência da Informação, com ênfase em Organização do Conhecimento, trabalhando principalmente nos seguintes temas: organização do conhecimento, ambientes Web, Museologia, ensino e formação de bibliotecários. É Membro do Conselho Editorial das revistas Arquivística.net (1808-4826) e Revist@CRB-7.  
(Texto informado pelo autor)

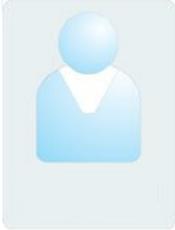
Última atualização do currículo em 12/02/2007  
Endereço para acessar este CV:  
<http://lattes.cnpq.br/8713013619609185>

**Outros links:**  
[Diretório de grupos de pesquisa](#)  
[Univ Federal do Rio de Janeiro](#)

Este currículo está relacionado à sua área de atuação? 👍 sim 👎 não



**Currículos semelhantes**



**Maria del Rosario Girardi**  
Universidade Federal do Maranhão



**Edilson Damasio**  
Universidade Estadual de Maringá



**Nair Yumiko Kobashi**  
Universidade de São Paulo

<a href="#">Dados pessoais</a>	<a href="#">Formação acadêmica/Titulação</a>	<a href="#">Atuação profissional</a>	<a href="#">Linhas de pesquisa</a>
<a href="#">Áreas de atuação</a>	<a href="#">Idiomas</a>	<a href="#">Projetos de pesquisa</a>	<a href="#">Membro de Corpo Editorial</a>
<a href="#">Eventos</a>	<a href="#">Orientações</a>	<a href="#">Produção em C,T &amp; A</a>	<a href="#">Bancas</a>

Figura 48 - Funcionalidade para avaliação de currículos semelhantes

Fonte: elaboração própria, adaptado de consulta ao Currículo Lattes

Além da forma ilustrada na Figura 48, poderia ser acrescentada na tela inicial do sistema uma aba “avaliar perfis semelhantes”, um convite para os usuários auxiliarem o sistema indicando se os currículos apresentados estão ou não relacionados à sua área de atuação. A cada modificação em um dos currículos avaliados, seria necessária nova avaliação. Na Figura 49 segue a tela inicial do sistema, onde poderia ser acrescentada a aba sugerida.



Figura 49 - Tela inicial após autenticação no Currículo Lattes (parte superior)

Fonte: elaboração própria, a partir de consulta ao Currículo Lattes

### 6.3.3. Resultados Esperados

A implantação da coleta explícita, aliada ao índice  $RR_{\text{área}}$  e demais modificações no Currículo Lattes para torná-lo efetivamente um sistema de recomendação, permitiria gerar resultados mais satisfatórios, haja vista que os próprios usuários podem validar se determinado currículo está ou não associado a sua área de atuação, além de ficar clara a forma de ordenação dos resultados apontados pelo sistema.

Fora isso, o sistema solucionaria, de pronto, a carência interna identificada junto aos colaboradores do CNPq, na seleção de especialistas. Para o público em geral seria um grande facilitador na localização de talentos e contatos. Não que todas as necessidades dos usuários possam ser atendidas, pois, como já foi dito, os usuários são fábricas de sonhos.

Independentemente dos resultados esperados pela transposição do modelo para o sistema Currículo Lattes, fica registrado que foi possível transpor o modelo genérico de recomendação de perfis de usuário — verificando-se os resultados alcançados no sistema resultante do modelo aplicado em um *site* de relacionamentos — em um subdomínio diferente, o da recuperação de currículos, de acordo com proposta estabelecida para este projeto. A meta de generalização foi alcançada nos termos para a qual foi idealizada.

## 6.4. Considerações Finais

Confirmou-se que as abordagens de recomendação podem ser utilizadas para a recuperação de perfis de usuário e que é possível criar um modelo genérico para a recuperação de perfis de usuário em sistemas de

recomendação — validação realizada após a transposição do sistema 2coracoes para o Currículo Lattes. Não foi possível verificar a percepção do sistema por usuários reais por dois fatores: primeiro, não houve tempo hábil para a construção de enquetes e validação junto aos usuários. Segundo, ainda que houvesse tempo, não seria possível distinguir níveis de usuário em um sistema de recomendação de perfis em *sites* de relacionamento. Esta avaliação poderia ser feita em um sistema como o da recomendação de currículos, onde fica clara a diferença entre os níveis de usuário, seja ela acadêmica ou profissional.

O máximo que poderia ser alcançado na criação de enquetes no atual estágio do projeto seria a validação de satisfação no uso deste específico sistema de recomendação — o sistema 2coracoes, subproduto do trabalho —, que se diferencia dos tradicionais sistemas de recuperação de informações disponíveis no mercado. Em um estudo posterior poderia ser realizado um levantamento para verificar se a necessidade de informação dos usuários no uso de um sistema de recomendação em *site* de relacionamentos está sendo satisfeita, tarefa que ficará como sugestão para trabalhos futuros.

## 7. Conclusão

O objetivo geral, propor um modelo de recuperação de perfis de usuário baseado em abordagens de recomendação, foi satisfatoriamente alcançado. Abordagens de recomendação puderam ser utilizadas para a recuperação de perfis de usuário.

Confirmou-se que é válido o uso de recomendações na recuperação de perfis de usuário, através da demonstração de que um modelo genérico de recomendação de perfis de usuário pode encontrar resultados que não seriam alcançados por um modelo tradicional de sistema de recuperação de informação.

Os sistemas tradicionais de recuperação de perfis de usuário, baseados em busca booleana, limitam a abrangência dos resultados de busca — obviamente, pois é o que fornece o uso da lógica booleana como estratégia de consulta. A utilização adequada de recomendações com objetivo de satisfazer a necessidade de informação do usuário necessita apresentar formas de colher o *feedback* do usuário, de forma que após apresentar os primeiros resultados de busca, informações possam ser coletadas para a realização de uma nova iteração de busca, com resultados mais adequados.

No sistema apresentado por este projeto, esta coleta é feita implicitamente, de maneira que resultados não antes alcançados pelo uso da lógica booleana pudessem, agora, ter a abrangência apresentada pela Figura 7 (pág. 35). No primeiro resultado da consulta, são utilizados, em conjunto, a lógica booleana, técnicas de filtragem e o *ranking* de recomendação na apresentação dos resultados. Assim que o usuário clica em um primeiro perfil para visualizá-lo, o sistema coleta implicitamente essa informação e cruzando-a com informações armazenadas da busca pré-definida, regras do sistema identificam perfis potencialmente semelhantes ao que está sendo consultado e que possam vir a satisfazer a necessidade do usuário. Nesta atividade, perfis que não foram apresentados dentre os resultados da busca

inicial podem ser alcançados e apresentados nos moldes da abordagem *Cross-Sell*, explicada em seções anteriores<sup>70</sup>.

A abrangência do resultado final é sempre maior que a apresentada por sistemas tradicionais. E a coleta implícita é possível porque o usuário tende a navegar dentre os resultados da busca inicial para visualizar informações mais detalhadas sobre os perfis apresentados. Assumindo que os perfis que são clicados possuem alguma natureza subjetiva de interesse, estes cliques aumentam os valores da variável quantitativa “visitantes do perfil” — uma das variáveis utilizadas para o cálculo do *ranking* de recomendação — que influencia a apresentação dos resultados da busca original. O ciclo foi estabelecido: as informações de consulta retroalimentam o sistema, que as utiliza para aprimorar a geração de recomendações, que apresentam resultados, que, quando consultados novamente, realimentam o sistema e assim por diante.

O estudo dos procedimentos de recuperação da informação baseados no uso de recomendações ampliaram o leque de recursos disponíveis para trabalhar os sistemas em prol da necessidade de informação dos usuários, consolidando, assim, a recomendação como estratégia importante para a recuperação de informação.

O modelo apresentado como produto deste trabalho foi criado a partir das necessidades identificadas para o sistema Currículo Lattes e utilizado no subdomínio de recuperação de perfis de usuário diferente da recuperação de currículos, mostrando que a generalização do modelo depende tão somente da identificação de indicadores flexíveis, podendo, assim, considerar o modelo genérico. A transposição do modelo genérico, a partir do sistema 2coracoes, subproduto deste trabalho, para o Currículo Lattes consolidou a proposta do modelo. O sistema foi implementado em um subdomínio diferente do estudado, o modelo genérico foi aplicado neste subdomínio e, em seguida, o sistema em produção forneceu os argumentos necessários para a realização da transposição.

---

<sup>70</sup> O perfil semelhante é apresentado na coluna da imagem do perfil que está sendo consultado. Veja Figura 40.

A contribuição para a Ciência da Informação fica evidente em função da carência de abordagens equivalentes. Ainda que não tenha sido possível avaliar o sistema junto aos usuários, espera-se que a eficiência e robustez identificadas na utilização de recomendações influenciem a aplicação das abordagens de recomendação em domínios ainda não explorados, carentes dos benefícios intrínsecos dessas estratégias.

A avaliação da satisfação do usuário é, sem dúvida, um esforço que merece atenção, pois permitirá a validação de tudo que aqui foi estudado. Considera-se esta uma atividade fundamental para um trabalho futuro. Outra atividade a ser pesquisada seria verificar como as abordagens de recomendação podem ser aplicadas junto a sistemas de processamento de linguagem natural.

## Referências

- BAEZA-YATES, R.; Ribeiro-Neto, B. **Modern Information Retrieval**. Addison-Wesley and ACM Press, 1999, cap.10.
- BARBOSA, Alexandre. Buscapé e BondFaro anunciam fusão. **Portal Estadão**. São Paulo, 29 mai.2006. Disponível em <<http://www.estadao.com.br/tecnologia/internet/noticias/2006/mai/29/114.htm>>. Acesso em 12 abr.2006.
- CAPRA, Fritjot. **A teia da vida**. São Paulo: Cultrix, 1997. 256 p.
- CAZELLA, S. C.; ALVARES, L. O. C. **Modeling User's Opinion Relevance to Recommending Research Papers**. Technical Report. 2005, 9 p.
- CNPQ. **O que é o Currículo Lattes?** Disponível em <[http://lattes.cnpq.br/conheca/con\\_apres.htm](http://lattes.cnpq.br/conheca/con_apres.htm)>. Acesso em 12 jul.2005.
- CROFT, W. Bruce. **What do people want from Information Retrieval**. Disponível em <<http://www.dlib.org/dlib/november95/11croft.html>>. Acesso em 31 nov.2004.
- EADS, Stefani. The Web's Still-Unfulfilled Personalization Promise. **BusinessWeek Online**. New York, 4 ago. 2000. Disponível em <[http://www.businessweek.com/bwdaily/dnflash/aug2000/nf2000084\\_506.htm](http://www.businessweek.com/bwdaily/dnflash/aug2000/nf2000084_506.htm)>. Acesso em 30 ago.2006.
- FAYYAD U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Data Mining to Knowledge Discovery: An Overview**. In: Advances in Knowledge Discovery and Data Mining, AAAI Press – The MIT Press, Cambridge. 1996.
- FISCHER, C. G.; WIVES, L. K. **Recomendação de Conteúdo em Fóruns Eletrônicos** In: Anais do 3º Fórum de Inteligência Artificial. Canoas: ULBRA, 2005. p. 8-11. Disponível em <<http://www.ulbra.tche.br/~forumia/anais.pdf>>. Acesso em 20 abr.2006.
- FÓRUM de discussão. **Google Toolbar and Desktop Search Applications**. Disponível em <<http://www.webmasterworld.com/forum80/106.htm>>. Acesso em 1 dez.2004.
- GOLDBERG, David; NICHOLS, David; OKI, Brian M.; TERRY, Douglas. **Using Collaborative Filtering to weave an information Tapestry**. Communications of the ACM. New York, v.34, n.10, p. 61-70, dez., 1992.
- GROSSER, K. **Human networks in organizational information processing**. Annual Review of Information Science and Technology, 1991, cap.26, p.349-402.
- LANCASTER, F.W. **Indexação e Resumos: Teoria e Prática**. Briquet de Lemos, 2ª ed, 2004.
- LE COADIC, Y.F. **A Ciência da Informação**. Brasília: Briquet de Lemos, 2ª ed, 2004.

LOH, Stanley. **Descoberta de Conhecimento em Textos**. Porto Alegre: Monografia — Universidade Federal do Rio Grande do Sul, Curso de Pós-Graduação em Ciência da Computação. 1999. 154 p.

LUHN, H. P. The automation derivation of information retrieval encodements from machine-readable texts. In: A. Kent (Ed.) **Information Retrieval and Machine Translation**. 1961, Vol.3, Pt. 2, p. 1021-1028. New York Interscience Publication.

MARQUES, T. M.; LADEIRA, M. **Relevância: A necessidade de recomendações na recuperação de informação** In: I COMPSULMT/2005 - I Congresso de Computação do Sul do Mato Grosso, Rondonópolis. Anais do 1º Congresso de Computação do Sul de Mato Grosso. Rondonópolis: EdUFMT, 2005. v.I. p.170-179.

MIZARO, S. A. **Cognitive Analysis of Information Retrieval**. In: Information Science: Integration in Perspective – CoLIS2, 1996, Copenhagen, Denmark. Proceedings. The Royal School of Librarianship. p.233-250.

MOONEY, R. J.; ROY, L. **Content-based book recommending using learning for text categorization**. Proceedings of the Fifth ACM Conference on Digital Libraries, 2000, p.195–204.

MUKAYAMA, Hiro. **Technical Aspect of Next Generation Digital Library Project**. Disponível em <<http://www.dl.ulis.ac.jp/ISDL97/proceedings/hiro/hiro.html>>. Acesso em 31 nov.2004.

NICHOLS, D. M.; TWIDALE, M. B.; PAICE, C. D.; **Recommendation and Usage in the Digital Library**. Disponível em <<http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/recommend.html>>. Acesso em 1 dez.2004.

NIST. **TIPSTER Text Program**. Disponível em <[http://www.itl.nist.gov/iaui/894.02/related\\_projects/tipster](http://www.itl.nist.gov/iaui/894.02/related_projects/tipster)>. Acesso em 14 ago. 2006.

ONODA, Mauricio. **Metodologia de Mineração de Dados para Análise do Comportamento de Navegar num Web Site**. Rio de Janeiro: Tese (Doutorado). Universidade federal do Rio de Janeiro, COPPE/UFRJ. 2006. 215 p.

PEREIRA, Vânia Lúcia da Cunha. **Sistemas de Redução da Informação: uma (ir)recuperação metodologicamente configurada**. Rio de Janeiro: Dissertação (Mestrado). ECO/UFRJ-IBICT/CNPQ. 1994. 95 p.

POPESCU, A.; UNGAR, L. H.; PENNOCK, D. M.; LAWRENCE, S. **Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments**. San Francisco: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI-2001), 2001, p.437-444.

RAMOS, Mônica Gomes. **O Uso da Teoria de Função de Crença em Sistemas de Recuperação de Informação**. Brasília: Dissertação (Mestrado) – Universidade de Brasília, Curso de Pós-Graduação em Ciência da Informação. 1999. 234 p.

RESNICK, P. et al. **GroupLens**: An open architecture for collaborative filtering of netnews. Chapel Hill: Proceedings of the Computer Supported Collaborative Work Conference, 1994.

RESNICK, Paul; VARIAN, Hal R.. **Recommender Systems**. Communications of the ACM. Nova Iorque, v. 40, n. 3, p. 56-58, mar., 1997.

RIJSBERGEN, C. J. Van. **Information Retrieval**. Disponível em <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>. Acesso em 1 out.2004.

ROBREDO, Jayme. **Documentação de Hoje e de Amanhã**: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas. 4ª ed rev. e ampl. Brasília, Edição de autor, 2005.

SCHAFER, J. B.; KONSTAN, J.; RIEDL, J., **Recommender systems in e-commerce**, Proceedings of the 1st ACM conference on Electronic commerce, 1999.

SCHAFER, J.B. et al. **E-Commerce Recommendation applications**. Data Mining and Knowledge Discovery, Kluwer Academic Publishers, Hingham, MA,USA, Vol 5. Issue 5, 2001.

SILVA, E. L. da; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 3. ed. rev. atual. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2001. 121p.

SMALL, H.; GRIFFITH, B. **The Structure of the scientific literature I**. *Science Studies*, 1974.

STEVEN, Chau. **How does the Google Toolbar work?** Disponível em <[http://www.httprevealer.com/usage\\_google.htm](http://www.httprevealer.com/usage_google.htm)>. Acesso em 1 dez.2004.

SUN MICROSYSTEMS, Inc. **Java Micro Edition Technical Articles**. Disponível em <<http://developers.sun.com/techttopics/mobility/allarticles/>>. Acesso em 14 ago.2006.

TORRES JR., Roberto Dias. **Combining Collaborative and Content-based Filtering to Recommend Research Papers**. Porto Alegre: Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Computação. 2004. 66 p.

VIVISIMO. **About us**. Disponível em <<http://vivisimo.com/html/about>>. Acesso em 30 ago.2006.

VOORHEERS, Ellen M. **Overview of TREC 2002**. Disponível em <<http://trec.nist.gov/pubs/trec11/papers/OVERVIEW.11.pdf>>. Acesso em 11 abr.2006.

YUWONO, Budi.; LAM, Savio L.; YING, Jerry H.; LEE, Dik L. **A World Wide Web Resource Discovery System**. Proceedings of the 4<sup>th</sup> International World Wide Web Conference. 1995. 23 p.

## **Anexo I – Estrutura do Perfil**

Este anexo representa o documento inicial de estrutura do perfil para construção do módulo de cadastro no protótipo deste projeto. O documento foi repassado à equipe de desenvolvimento para especificação e implementação.

### **Estrutura do Perfil**

- 1 Básico (preenchimento obrigatório)
  - 1.1 Apelido (mais de 5 caracteres)
  - 1.2 Idioma do Perfil
    - Português
    - Inglês
    - Espanhol
  - 1.3 Data de Nascimento
    - Dia (DD)
    - Mês (MM)
    - Ano (AAAA)
  - 1.4 Sexo
    - Homem
    - Mulher
  - 1.5 Orientação Sexual
    - Hetero
    - Homo
  - 1.6 Papel (opcional)
    - Prefiro não dizer
    - Ativo
    - Passivo

- Versátil
- 1.7 Relacionamento (opcional)
- Prefiro não dizer
  - Sozinho
  - Comprometido
  - Casado
  - Enrolado
- 1.8 Sexo Seguro (opcional)
- Prefiro não dizer
  - Sempre
  - Muitas vezes
  - Às vezes
  - Nunca
- 1.9 Procurando alguém com idade entre
- Disponibilizar duas caixas de seleção, sempre começando em 18. A segunda caixa de seleção deve marcar um número maior que a primeira (obrigatório).
- 1.10 Meu Naípe (marcar o que se aplica, obrigatório)
- Perfil Copas - Romance
  - Perfil Ouros – Amizade
  - Perfil Espadas – Atividades a dois
  - Perfil Paus – Atividades em grupo/Outras atividades
  - Perfil Coringa – Mais de um ou todos acima
- 2 Localização Geográfica (preenchimento obrigatório)
- 2.1 País
- Brasil
  - Exterior

o Se exterior, preencher o nome do país usando caracteres maiúsculos e não acentuados. Substituir os hífen por espaços, se presentes, para armazenar no banco de dados. Ex: Açores ficaria ACORES. Ula-Ula Island ficaria ULA ULA ISLAND.

## 2.2 Estado (Só se aplica se o país for Brasil)

- Acre
- Alagoas
- Amapá
- Amazonas
- Bahia
- Ceará
- Distrito Federal
- Espírito Santo
- Goiás
- Maranhão
- Mato Grosso
- Mato Grosso do Sul
- Minas Gerais
- Pará
- Paraíba
- Paraná
- Pernambuco
- Piauí
- Rio de Janeiro
- Rio Grande do Norte
- Rio Grande do Sul
- Rondônia
- Roraima

- Santa Catarina
- São Paulo
- Sergipe
- Tocantins

### 2.3 Cidade

- Preencher usando caracteres maiúsculos e não acentuados. Substituir os hífen por espaços, se presentes. Ex: Goiânia ficaria GOIANIA, São José do Capu-açu ficaria SÃO JOSE DO CAPU ACU.

### 2.4 Bairro

- Preencher usando caracteres maiúsculos e não acentuados. Substituir os hífen por espaços, se presentes. Ex: Vila-do-Jacaré ficaria VILA DO JACARE.

## 3 Contato (só o campo Email é de preenchimento obrigatório)

### 3.1 Email (preenchimento obrigatório, nunca será disponibilizado!)

### 3.2 Messenger (só será disponibilizado para os usuários que você permitir, após o usuário te mandar piscadela ou sinal de positivo)

- Meu MSN Messenger é

### 3.3 Telefone (só será disponibilizado para os usuários que você permitir, após o usuário te mandar piscadela ou sinal de positivo)

- Celular (código do país + código de área + número) Ex: +55 (61) 9233 9999

## 4 Social (preenchimento obrigatório)

### 4.1 Bebe

- Não
- Socialmente
- Regularmente

### 4.2 Fuma

- Não
- Socialmente
- Regularmente

#### 4.3 Malha (academia)

- Não
- Há menos de um ano
- Entre um e três anos
- Entre três e cinco anos
- Há mais de cinco anos

#### 5 Aparência (alguns campos são obrigatórios)

##### 5.1 Altura (obrigatório)

- Anão a 1,50m
- 1,51m a 1,60m
- 1,61m a 1,65m
- 1,66m a 1,70m
- 1,70m a 1,72m
- 1,73m a 1,75m
- 1,76m a 1,78m
- 1,79m a 1,81m
- 1,82m a 1,84m
- 1,82m a 1,83m
- 1,84m a 1,85m
- 1,86m a 1,87m
- 1,88m a Gigante

##### 5.2 Biotipo (obrigatório)

- Esquelético
- Magro/Pequeno
- Magro a Médio
- Médio (normal)
- Médio a Definido

- Médio a Gordinho
- Definido
- Definido a Musculoso
- Musculoso
- Bombado
- Gordinho
- Gordo/Grande
- Obeso

### 5.3 Pelo Corporal (opcional)

- Prefiro não dizer
- Nenhum
- Pouco
- Médio
- Muito
- Raspado/Depilado
- Aparado

### 5.4 Cor dos Olhos (opcional)

- Prefiro não dizer
- Azul
- Verde
- Castanho Claro / Mel
- Castanho
- Cinza
- Preto

### 5.5 Etnia (opcional)

- Prefiro não dizer
- Indígena

- Asiática
- Negra
- Caucasiana
- Indiana
- Hispânica/Latina
- Oriente Médio
- Várias

#### 5.6 Cabelo (opcional)

- Prefiro não dizer
- Nenhum
- Pouco
- Curto
- Normal
- Longo
- Raspado
- Rente/Estilo soldado
- Anelado
- Caracóis/Estilo anjinho
- Enrolado
- Cachos
- Bombril

#### 5.7 Estilo (opcional)

- Prefiro não dizer
- Alternativo
- Casual
- Esporte
- Estudante

- Executivo/Formal
- Fashion/Moderno
- Gótico
- Leather
- Mauricinho/Patricinha
- Militar
- Na moda
- Punk
- Rap & hip-hop
- Travestismo/Drag

#### 5.8 Decoração Corporal (marcar o que se aplica, opcional)

- Prefiro não dizer
- Brincos
- Brinco no nariz
- Piercings no corpo
- Tatoos
- Outros
- Nenhuma

#### 6 Formação (preenchimento obrigatório)

##### 6.1 Ocupação (escolher a que melhor se aplica)

- Administração
- Alimentação
- Artes
- Ciências Biológicas, Bioquímicas e da Saúde
- Ciências Exatas, Físicas e Químicas
- Ciências Jurídicas
- Ciências Sociais e Humanas

- Comércio
- Comunicação
- Ensino
- Engenharia
- Eventos
- Finanças
- Indústria
- Militar
- Moda, cabelo e maquiagem
- Política
- Serviços Gerais
- Serviço Público
- Setor Primário (agropecuária, caça e pesca)
- Técnico de Nível Médio
- Informática/Tecnologia da Informação
- Turismo

## 6.2 Escolaridade

- Ensino fundamental
- Estudante do ensino médio
- Ensino médio
- Estudante do ensino superior
- Ensino superior
- Estudante de pós-graduação
- Pós-graduação
- Estudante de mestrado
- Mestrado
- Estudante de doutorado

- PhD/MD/Pós-doutorado
- 7 Pessoal (opcional)
- 7.1 Resumo (todo os campos abaixo são do tipo texto)
  - Título introdutório
  - Texto introdutório - descreva quem você é e o que gosta ou não (*link*: dicas de como preencher)
  - Descreva o que está procurando em uma pessoa (*link*: dicas de como preencher)
  - Prato Preferido
  - Filme Preferido
  - Paixões
  - Idiomas que falo
  - Programa de TV
  - Clube/Discoteca
  - Bar
- 7.2 Interesses (marcar o que se aplica, opcional)
  - Academia / Ginástica
  - Acampamento / Excursões
  - Alpinismo
  - Artes / Exposições
  - Artes Marciais
  - Astrologia / Exoterismo
  - Atletismo
  - Baseball / Softball
  - Bilhar / Sinuca / Dardo
  - Box / Lutas
  - Canto / Instrumentos Musicais
  - Carros / Motocicletas

- Ciclismo
- Cinema / Televisão
- Comidas e Vinhos
- Computadores / Internet
- Corrida de Carros
- Cricket
- Culinária
- Dança
- Danceterias / Bares / Nightclubs
- Esportes Aquáticos
- Esportes de Inverno
- Fazer Compras / Novidades
- Futebol / Rugby
- Golf
- Iatismo / Navegação
- Jardinagem
- Literatura / História
- Museus / Galerias
- Música – Alternativa
- Música – Blues / Jazz
- Música – Clássica / Ópera
- Música – Cristã / Gospel
- Música – Dance / Eletrônica / House
- Música – Latina
- Música – New Age
- Música – Pop / Ritm & Blues
- Música – Rock

- Música – Sertaneja
- Música – World Music
- Natureza
- Obras de Caridade / Voluntariado
- Política
- Religião
- Teatro / Ballet
- Tennis / Squash / Badminton
- Viagens & Turismo
- Volleyball / Basketball
- Yoga / Meditação

### 7.3 Perfil de quem eu quero

- Disposição dos campos conforme os seguintes rótulos:
  - Altura
  - Biotipo
  - Localização
  - Pelo Corporal
  - Etnia
  - Estilo
  - Cor dos Olhos
  - Cabelo
  - Decoração mais marcante

# Anexo II – Termos de Uso do Sistema

Termos e condições de uso<sup>71</sup>

Os presentes termos e condições de uso se aplicam para registro no sistema de recomendações 2coracoes. Se você não concorda com os presentes termos e condições de uso, não deverá efetuar ou manter seu cadastro. Leia com atenção.

## 1. Aceitação dos termos e condições de uso

1.1 Ao marcar a opção "Declaro que li, entendi e aceitei as condições do contrato do usuário" para registrar ou manter cadastro no *site*, considera-se que estes termos e condições de uso do sistema de recomendações 2coracoes foram aceitos, que você é maior de 18 anos e que está de acordo com a nossa Política de Privacidade.

1.2 Reservamo-nos o direito de alterar e atualizar os presentes termos e condições a qualquer tempo. As revisões dos presentes termos e condições serão comunicadas na página inicial do sistema.

## 2. Descrição do estado de usuário e de assinante

2.1 O usuário pode usufruir dos serviços gratuitamente. No futuro, algumas áreas do *site* poderão exigir o pagamento de uma assinatura de usuário. Nesta fase não há serviços exclusivos para assinantes.

## 3. Idade para o consentimento

3.1 O usuário precisa ter pelo menos 18 anos para se registrar. Se descobirmos ou tivermos motivos para suspeitar que você não tenha mais de 18 anos de idade, reservamo-nos o direito de suspender ou terminar imediatamente e sem aviso prévio o seu cadastro no sistema.

## 4. Senha, Apelido e Segurança

4.1 O usuário é responsável por manter sua senha e apelido, e é responsável por todas as atividades que ocorrerão através deles. Não dispomos de meios para verificar a identidade das pessoas que utilizam a área dos usuários e não podemos ser responsabilizados se a sua senha ou apelido forem utilizados por outras pessoas.

4.2 O usuário concorda em alterar sua senha imediatamente após qualquer utilização não autorizada da senha ou apelido ou sobre qualquer outra falha de segurança que possa verificar.

## 5. As suas garantias

5.1 Você assegura que:

5.1.1 Tem mais de 18 anos de idade;

5.1.2 todas as informações e detalhes disponibilizados — incluindo o cadastro como usuário — são verdadeiros, precisos e atualizados em todos os sentidos e em todos os níveis;

5.1.3 Está de acordo com os termos e condições de uso, incluindo, sem limitações, as restrições relacionadas com a utilização aceitável referidas no parágrafo 6.

5.2 Concorda em nos indenizar e de não nos responsabilizar por quaisquer reclamações ou danos — incluindo quaisquer custos judiciais relacionados — causadas por terceiros em relação a qualquer assunto relacionado ou derivado da sua utilização do sistema ou da sua assinatura por qualquer infração ou suspeita de infração dos presentes termos e condições de uso cometidos por você, ou sua violação de qualquer lei ou direitos de terceiros.

---

<sup>71</sup> Disponível em <<http://www.2coracoes.com/termos.php>>

5.3 Alertamos que não existe recurso que impeça cópia de imagem na internet. Por tal razão, o membro deve estar ciente de que, uma vez inserida na rede virtual qualquer imagem, seja fotografia, figura ou reprodução, é possível sua cópia, reprodução ou impressão por terceiros que acessem o sistema.

5.4 As fotos principais dos perfis dos usuários passam por uma avaliação prévia a fim de evitar conteúdo que possa ser considerado inapropriado pelos administradores do sistema 2coracoes. Essa avaliação não nos responsabiliza sobre o que será disponibilizado. A avaliação é subjetiva e nenhum administrador do sistema 2coracoes é responsável pelo que o usuário decidiu exibir em seu perfil. As fotos colocadas nos álbuns de fotos dos usuários não passam por avaliação dos administradores do sistema, cabendo aos usuários administrar o conteúdo de seus álbuns segundo suas vontades.

5.5 Não nos responsabilizamos por encontros ou relacionamentos travados entre os usuários originados através do *site*.

## 6. Utilização aceitável

6.1 Apoiamos o fluxo livre das informações e idéias através da Internet, e, sob condições normais, não monitoramos de forma ativa a utilização do *site*, incluindo a área de usuários. Tentamos equilibrar este objetivo com as nossas obrigações perante os outros usuários e da própria lei. Por isso, poderão ser realizadas revisões ocasionais nos perfis cadastrados pelos usuários para assegurar a conformidade com os presentes termos, e para responder a ou investigar solicitações de uma autoridade, força policial, queixas ou denúncias. Também requisitamos que você:

6.1.1 Não diga ou faça seja o que for que possa causar indignação, inconveniência, assédio ou ansiedade desnecessária a outras pessoas.

6.1.2 Não divulgue ou promova produtos e serviços próprios ou de terceiros incluindo a distribuição de mensagens não solicitadas (spam);

6.1.3 Não utilize uma linguagem desagradável, ameaçadora ou ofensiva, incluindo, sem restrições, uma linguagem racista, sexista, homofóbica ou explicitamente sexual.

6.1.4 Não insulte outros usuários, direta ou indiretamente;

6.1.5 Não distribua material ilegal, que viole os direitos de autor ou ofensivo, ou quaisquer mensagens ou conteúdos que possam provocar tumultos ou encorajar atividades ilegais ou causar danos a menores.

6.1.6 Não personalize uma outra pessoa ou usuário;

6.1.7 Não transfira arquivos contendo vírus ou qualquer programa ofensivo.

6.1.8 Não utilize o *site* para efetuar qualquer atividade fraudulenta, incluindo "esquemas de pirâmide".

6.1.9 Não acesse ou tente acessar contas de outros usuários, ou penetre ou tente penetrar nas medidas de segurança do sistema.

6.1.10 Não inclua dados de contato pessoal em campos inapropriados. Não utilize em seu perfil linguagem ofensiva ou que faça apologia a praticas ilegais, qualquer conteúdo que traga informação em que se identifica o titular ou que atentem contra a lei.

6.1.11 Não troque de mensagens contendo *links* de *websites*, a fim de que o sistema não seja usado como veículo de SPAM.

6.1.12 Não utilize o *site* para prostituição.

6.2 Podemos tomar qualquer uma ou todas as medidas seguintes à nossa discricão:

6.2.1 Remover qualquer perfil de usuário — incluindo fotografias — ou qualquer material que, do nosso ponto de vista, possa ser inapropriado ou que pensamos ser ilegal, que nos pode colocar em uma posição de responsabilidade ou que possa violar os presentes termos e condições de uso ou onde a lei ou autoridade exigir.

6.2.2 Emitir aos usuários avisos escritos e/ou quaisquer outras ações suplementares que julgamos serem apropriadas em concordância com este parágrafo 6.2, se tais avisos não forem atendidos.

6.2.3 Suspender ou terminar em qualquer altura e sem qualquer notificação o acesso de usuário à área de usuários do *site* ou à conta do usuário.

6.2.4 Informar as autoridades competentes e proporcionar-lhes a informação referente a qualquer atividade suspeita de ser ilegal, ou.

6.2.5 Tomar medidas legais contra um membro ou outro usuário do sistema em relação a qualquer não cumprimento dos termos e condições de uso, violação da política de privacidade ou qualquer atividade ilegal ou suspeita de ser ilegal.

6.2.6 Determinar, caso a caso, qual a ação apropriada para ser tomada contra um usuário.

6.3 Podemos estar obrigados por lei ou regulamentação a acessar, monitorizar ou copiar material enviado por ou para usuários sem qualquer notificação a você.

## **7. Direitos de autor e outros direitos de propriedade intelectual**

7.1 O uso do sistema não lhe confere direitos em relação a direitos de autor, marcas comerciais ou outros dos nossos direitos de propriedade intelectual ou dos direitos de propriedade intelectual de terceiros.

7.2 Em caso algum pode, sem limitação, copiar, reproduzir, publicar, transferir, colocar, difundir, gravar, transmitir, explorar comercialmente, comunicar ao público ou utilizar de outra forma o conteúdo incluído ou disponibilizado no *site*, exceto para o seu uso pessoal e não comercial. Condicionado pelo supra dito, pode transferir partes não substanciais desse conteúdo para o disco rígido do seu computador com a finalidade da sua visualização, desde que não seja feita mais do que uma cópia dessa informação.

## **8. Limitações da responsabilidade**

8.1 Embora tomemos todas as medidas para corrigir quaisquer erros ou omissões o mais breve possível, logo que somos chamados à atenção, não podemos garantir que o sistema esteja ininterruptamente disponível e em condições de plena operacionalidade, nem que a informação no ou disponibilizada pelo *site* esteja livre de erros e omissões.

8.2 O acesso ao sistema *2coracoes* e ao seu conteúdo pode ser suspenso temporariamente e sem qualquer aviso em caso de uma falha do sistema, da necessidade de manutenção ou reparação ou por motivos fora do nosso controle.

8.3 Salvo que nenhuma informação do presente parágrafo 9 restringe os seus Direitos previstos na lei, incluindo o seu direito de receber um serviço de qualidade aceitável, todos os conteúdos e serviços neste *site* são disponibilizados "como estão" e "conforme disponível". Não apresentamos nem damos quaisquer garantias em relação ao *site* ou o seu conteúdo, incluindo, sem limitação, informações disponibilizadas por ou a respeito de outros usuários. Não podemos examinar todos os perfis e entradas dos usuários para assegurar que são adequados e corretos. Qualquer decisão ou ação tomada por você com base nas informações disponibilizadas no ou através do sistema é da sua própria conta e do seu próprio risco.

8.4 Devido ao fato de muitos aspectos tecnológicos do sistema e do conteúdo nele representado ser disponibilizado por ou, de outra forma, dependente de terceiros, não podemos garantir a exatidão, adequação, plenitude, funcionalidade, qualidade satisfatória, adequação para determinada finalidade, ou ausência de vírus ou de outros programas ofensivos seja no conteúdo, ou acessível através deste sistema.

## **9. Geral**

9.1 Não podemos ser responsabilizados por quaisquer falhas por suspensão ou conclusão do acesso ao sistema ou qualquer conteúdo devido a situações de força maior. Uma situação de força maior inclui, sem limitação, falhas da infra-estrutura, intervenção do governo, guerras, tumultos civis, seqüestro, fogo, inundação, acidente, tempestade, relâmpagos, black-out, ataques terrorista, ou ação industrial que nos afete ou aos nossos fornecedores.

9.2 No caso de qualquer um dos presentes termos e condições para usuários forem considerados ilegal, inválido ou de outra forma não aplicável por força da lei, este, no seio e no âmbito da respectiva jurisdição em que o termo for ilegal, inválido ou não aplicável, será considerado nulo e eliminado dos presentes termos e condições, e os restantes termos serão mantidos em vigor e válidos e continuarão obrigatórios e aplicáveis.

9.3 Os presentes termos e condições de uso são regidos e interpretados ao abrigo das leis do Brasil. Disputas que surjam em função dos presentes termos serão sujeitas a jurisdição exclusiva dos tribunais do Brasil.

## **10. Reclamações e comunicações**

10.1 Se tiver qualquer reclamação sobre outro usuário ou qualquer outro aspecto do sistema, ou se tiver qualquer questão ou pretende efetuar qualquer outro tipo de comunicação, favor entrar em contato o suporte do sistema *2coracoes*.

ACEITO TODOS OS TERMOS E CONDIÇÕES DE USO, CONCORDO COM A POLÍTICA DE PRIVACIDADE E CERTIFICO QUE SOU MAIOR DE 18 ANOS.

## **Política de Privacidade**

A presente Política de Privacidade se aplica para registro no sistema de recomendações 2coracoes. Se você não concorda com esta Política de Privacidade, não deverá efetuar ou manter seu cadastro. Leia com atenção.

## **1. Coleta de Informações**

1.1 O sistema 2coracoes coleta informações pessoais quando você se registra no sistema ou utiliza a ferramenta. Podemos combinar as informações fornecidas por você com as regras do sistema ou informações fornecidas por outros usuários para prestar melhores serviços para os usuários, incluindo a personalização de conteúdo para você.

1.2 Todas as informações coletadas servem para aprimorar as funcionalidades do sistema 2coracoes e a qualidade dos nossos serviços.

## **2. Uso da informação**

2.1 Podemos utilizar suas informações pessoais para disponibilizar as requisições que você realizar, incluindo a personalização de conteúdo e propaganda.

2.2 Podemos utilizar suas informações pessoais para auditoria do sistema, pesquisas e análises necessárias para operar e melhorar os serviços e a tecnologia do sistema 2coracoes.

2.3 Podemos compartilhar o agrupamento de informações impessoais a terceiros (outras partes fora do sistema 2coracoes). Informações impessoais se referem a informações gravadas sobre os usuários e coletadas em grupo, de forma que já não reflita individualmente ou faça referência isolada à identificação de um usuário.

2.4 Quando terceiros prestam assistência no processamento das informações pessoais, requisitamos que eles obedeçam nossa Política de Privacidade e qualquer outra medida de segurança e confidencialidade.

2.5 Podemos compartilhar as informações do sistema a terceiros em circunstâncias especiais, como na abertura de processos legais, prevenções de fraudes ou danos, e para garantir a segurança de nossos serviços.

2.6 Podemos coletar informações sobre as interações entre os usuários do sistema, e manter informações associadas aos perfis desses usuários, incluindo endereços de email e conteúdo dos perfis.

## **3. Atualização da Política de Privacidade**

3.1 Reservamo-nos o direito de alterar e atualizar a presente Política de Privacidade a qualquer tempo. As revisões serão comunicadas na página inicial do sistema.

## **4. Suas escolhas**

4.1 Onde for possível, é disponibilizada a você a alteração das informações pessoais registradas. As informações coletadas na utilização da ferramenta não podem ser alteradas, pois já foram processadas juntamente com o que foi coletado dos demais usuários.

4.2 Muitas das informações que você cadastra no sistema serão mostradas no seu perfil de usuário e poderão ser visualizadas por outros usuários do sistema. Fique atento sobre as informações que você permite que sejam vistas pelos outros membros.

4.3 O sistema 2coracoes diferencia a visualização de perfis segundo a orientação sexual do usuário. Usuários que se declaram heterossexuais só podem visualizar perfis de membros heterossexuais, e usuários que se declaram homossexuais só podem visualizar perfis de membros homossexuais. Não há outra opção para orientação sexual, senão heterossexual ou homossexual.

4.4 Avisos do sistema ou interações dos usuários podem provocar o envio de emails para você.

4.5 Você pode excluir sua conta a qualquer momento. Após a exclusão, seu perfil será removido do sistema 2coracoes. Se houver algum problema em nossos servidores, não nos responsabilizamos pela exclusão de sua conta. O usuário fica responsável por verificar se há resíduo da solicitação de exclusão e deve tentar excluir o perfil novamente e, caso não seja possível, entrar em contato com o suporte do sistema.

## Anexo III – Sobre

Este anexo apresenta o documento “Sobre” do *website* 2coracoes.com.

### Sobre o 2coracoes.com

**2coracoes** é um sistema de relacionamentos *on-line* criado para tornar a busca por outro alguém uma tarefa mais ativa e estimulante. As recomendações do **2coracoes** podem ajudá-lo a localizar quem mais combina com você por meio das suas informações e de informações coletadas por usuários que possuem as mesmas afinidades que as suas.

Com o **2coracoes** é fácil conhecer pessoas que tenham os mesmos hobbies e interesses que você, que estejam procurando o mesmo tipo de relacionamento afetivo, o que aqui chamamos de “naípe”. Você pode piscar para quem acha interessante, e se for correspondido, você e o outro enviar seus endereços de messenger e telefone celular. Você decide com quem interagir. Antes de conhecer uma pessoa no 2coracoes, você pode ver seu perfil, conhecer seus favoritos e ler sobre seus gostos e atividades. E o sistema estará trabalhando para sempre indicar a você as pessoas que possam ser de seu interesse. Quanto mais você utiliza o sistema, mais você gera informações que nos permite aprimorar o conhecimento sobre suas preferências e melhorar as recomendações que geramos a você.

Para ingressar no **2coracoes**, cadastre-se no *link* na página principal, e comece a criar seu perfil imediatamente. Você estará pronto para utilizar esse sistema inovador em poucos minutos!

Nossa missão é ajudá-lo a criar relacionamentos com pessoas de mais afinidade. Esperamos que em breve você esteja curtindo mais a sua vida social.

## Anexo IV – Código-fonte do Sistema 2coracoes

Este anexo apresenta o código-fonte dos principais arquivos que formam o sistema 2coracoes. Os arquivos PHP que não foram selecionados realizam apenas controles lógicos e operacionais desnecessários para o entendimento do projeto.

### Arquivo: cadastrar.php

As variáveis são enviadas por POST. Elas são trabalhadas em um array, usando o \$HTTP\_POST\_VARS. As verificações de consistência realizadas na tela de cadastro estão implementadas neste trecho de código.

---

```
<?
include_once('class/tbs_class.php'); // Conecta no MySQL e chama o
TBS

// Define que o valor de passo deve ser enviado apenas por post
$passo = $_POST['passo'];

switch ($passo) {
    case 0:
        // Passo 1
        // Carrega o arquivo HTML do primeiro passo do cadastro
        $template = 'templates/perfil_geral.html';
        break;
    case 1:

        // Verifica os campos do passo 1
        while(list($campo,$valor) = each($HTTP_POST_VARS)) {

            // Nao verifica os campos citados
            if ($campo != "idade_final" and $campo != "ext" and
$campo != "estado" and $campo != "telefone" and $campo != "bebe" and
$campo != "malha" and $campo != "drogas" and $campo != "fuma" and
$campo != "cod_cidade" and $campo != "msn" and $campo != "bairro")
            {
                //Pega todos os campos vazios
                if(!$valor){
                    $erro[$campo] = "*";
                    $valor[$campo] = $valor;
                    // Define a variavel para nao avancar
                    $erros = 1;
                }
            }
        }

        // Verificacao aparte de localizacao
        if(!$POST['pais']){
            $erros = 1;
            $informallocalizacao = 1;
        } else {
            if($POST['pais'] == "Brasil"){
                if(!$POST['estado']){
                    $erros = 1;
                }
            }
        }
    }
}
```

```

                $informallocalizacao = 1;
            }
        } else {
            if (!$_POST['ext']) {
                $erros = 1;
                $informallocalizacao = 1;
            }
        }
    }
    $data = explode("/", $dt_nascimento);
    $d = $data[0];
    $m = $data[1];
    $y = $data[2];

    $res=checkdate($m,$d,$y);
    if ($res != 1){
        $erros = 1;
        $datainvalida = 1;
    }

    //substr($dt_nascimento, 6,4);
    $valor_dt = date("Y")-$y;
    if($valor_dt <= 17){
        $erros = 1;
        $menoridade = 1;
    }
    // se a variavel erro estiver cheia chama o primeiro passo
    if($erros){ $errosmsg = $lang['errosmsg'];
        $template = 'templates/perfil_geral.html';
    }
    // caso nao tenha erros verifica se já existe o email
    else {
        $sql = mysql_query("SELECT email FROM perfil WHERE email
= '$email'");
        $num_rows = mysql_num_rows($sql);
        if($num_rows == 1){
            $errosmsg = $lang['errosmsg2'];
            $erro['email'] = "*";
            $erros = 1;
            $template = 'templates/perfil_geral.html';
        }
    }
    if(!$erros){

        // Passo 2
        $template = 'templates/perfil_social.html';
    }
    break;
case 2:
    //Passo 3
    $template = 'templates/perfil_pessoal.html';
break;
case 3:
    //Passo Final
    /* Separar os |
echo $interesse;
echo "<pre>";
$ar = explode( "|" , $interesse);
$cont = count($ar);
for ($i=0; $i<$cont; $i++) {

```

```

$valor = $ar[$i];
echo "<br>".$lang['interesse'][$valor]['tipo'];
}
*/

while(list($campo,$valor) = each($HTTP_POST_VARS)) {
    // Junta os interesses em x|y|z
    $campo_int = substr($campo, 0,9);
    if($campo_int == "interesse"){
        $interesse = $interesse.'|'.$valor;
    }
    // Monta a SQL
    if ($campo != "passo" and $campo_int != "interesse"){
        // corrige o formato da data para o banco de dados
        if($campo == "dt_nascimento"){
            $dia = substr($valor, 0,2);
            $mes = substr($valor, 3,2);
            $ano = substr($valor, 6,4);
            $valor = $ano.'-'.$mes.'-'.$dia;
        }
        $sqlcampo = $sqlcampo.$valor."', '";
    }
}

// Monta a SQL de dados do usuario para inserir no banco de
dados
$sql =
"', '".$sqlcampo.$interesse."', '',' ',' ',' ',' ',' ',' ',' ',' ',' ',' '";
mysql_query("INSERT INTO perfil values (". $sql.")") or die
("SQL ERROR: ".mysql_error());

// Finaliza o cadastro logando no sistema
$sql = mysql_query("SELECT
id,email,senha,apelido,foto,ori_sexo,naipe,sexo FROM perfil WHERE
email = '$email' and senha = '$senha'");
$num_rows = mysql_num_rows($sql);
while ($row = mysql_fetch_array($sql)) {
    $_SESSION['email'] = $row['email'];
    $_SESSION['id_usr'] = $row['id'];
    $_SESSION['apelido_usr'] = $row['apelido'];
    $_SESSION['ori_sexo_usr'] = $row['ori_sexo'];
    $_SESSION['naipe_usr'] = $row['naipe'];
    $_SESSION['sexo_usr'] = $row['sexo'];
    if(!$row['foto']){ // Define a foto para quem nao tem
foto de acordo com o sexo e o naipe
        $foto =
$_SESSION['sexo_usr'].'_'.$_SESSION['naipe_usr'].'.jpg';
    }
    $_SESSION['foto_usr'] = $foto;
    $_SESSION['logon_usr'] = 1;
    header("Location: perfil.php");
    exit;
}
$template = 'templates/index.html';
break;
}
$tb->LoadTemplate($template);
//Array das traducoes
include_once('class/MergeBlock.php');
// exhibe a pagina

```

```
if ($template) {$tbs->Show();}
```

```
?>
```

**Arquivo: ver.php**

O arquivo ver.php exibe os perfis para visualização. Neste trecho de código são trabalhadas as ferramentas de piscada, bloqueio, avaliações positivas e negativas. As ferramentas trabalham da forma seguinte: é realizado um *refresh* verificando o valor armazenado de piscada ou de outra funcionalidade. Após a verificação, a variável tem valor incrementado e a tabela recebe o valor atualizado (registro no banco de dados).

É importante ressaltar que apenas a piscada, que é o contato primário, está disponível a princípio. Após um usuário piscar para o outro que são liberadas as ferramentas para envio de endereço de contato instantâneo e telefone celular.

---

```
if($_POST['piscar']){
    // se $piscar estiver cheio ele executa os seguintes comandos
    abaixo
    // primeiro verifica se ja existe ou nao uma piscada de x usuario
    para y usuario
    $sql = "select * FROM piscada WHERE usr_1 = '$perfil' and usr_2 =
'$id_usr' and ativo = '1'";
    $ver = mysql_query($sql);
    $rows = mysql_num_rows($ver);
    // se o valor for igual a 1 e porque ja existe
    if($rows == 1){
        // resultados na tela para o tbs
        $status = 1;
        $font_color = "#B40C0C";
        $statusmsg = $lang['interacao'][10];
        $img_status = "piscada.gif";
        //caso o valor for 0 ele registra a nova piscada
    } else {

        // resultados na tela para o tbs
        $status = 1;
        $font_color = "#428837";
        $statusmsg = $lang['interacao'][0];
        $img_status = "piscada.gif";
        // registra na tabela de piscadas que um x usuario piscou para
y usuario
        mysql_query("INSERT INTO piscada values
('','$perfil','$id_usr','1')");

        $sql = " select p_piscada FROM perfil WHERE id = '$perfil'";
        $ver = mysql_query($sql);
        $rows = mysql_num_rows($ver);
        // pucha o valor das piscadas do perfil
        while($result = mysql_fetch_array($ver)) {
            $p_piscada = $result['p_piscada'];
        }
        // Soma a qualificacao
        $p_piscada = $p_piscada + 1;
        // Registra o valor da soma no Banco de dados
        mysql_query("UPDATE perfil SET p_piscada = '$p_piscada' WHERE
id = '$perfil' LIMIT 1") or die ("Erro: ".mysql_error());
    }
}

if($_POST['nmi']){
    // NMI => nao me interessa
```

```

// definicoes do tbs para resultado html
$font_color = "#428837";
$status = 1;
$statusmsg = $lang['interacao'][5];
$img_status = "bloquear.gif";
// registra na tabela bloqueados quem esta sendo bloqueado
$query = "INSERT INTO bloqueados values
('','$perfil','$id_usr)";
mysql_query($query);
// deleta todas as acoes de um usuario para o outro
$query = "DELETE FROM piscada WHERE usr_1 = '$perfil' and usr_2 =
'$id_usr'";
mysql_query($query);
$query = "DELETE FROM borboleta WHERE usr_1 = '$perfil' and usr_2
= '$id_usr'";
mysql_query($query);
$query = "DELETE FROM telefone WHERE usr_1 = '$perfil' and usr_2
= '$id_usr'";
mysql_query($query);
// Registra pontos
$sql = " select p_bloqueos FROM perfil WHERE id = '$perfil'";
$ver = mysql_query($sql);
$rows = mysql_num_rows($ver);
while($result = mysql_fetch_array($ver)) {
    $p_bloqueos = $result['p_bloqueos'];
}
// Soma a qualificacao
    $p_bloqueos = $p_bloqueos + 1;
// Registra o valor da soma no Banco de dados
mysql_query("UPDATE perfil SET p_bloqueos = '$p_bloqueos' WHERE
id = $perfil LIMIT 1") or die ("Erro: ".mysql_error());
}

```

**Arquivo: Perfilsemelhantes.php**

Trata a recomendação de perfis semelhantes, realizando comparações entre informações coletadas do cadastro dos usuários.

---

```
// Verifica se a busca pre definida está definida
if($_SESSION['busca_pre'] == 1){
echo "<link rel=\"stylesheet\" type=\"text/css\"
href=\"templates/ajaxtabs/ajaxtabs_small.css\"/>\n";
include('language/br.inc');
$id_usr = $_SESSION['id_usr'];

// Pega dados para fazer comparacoes do perfil exibido
$sql_perfil = "SELECT sexo,naipe,ori_sexo,relacionamento FROM perfil
WHERE id = '$id_usr'";
$ver_perfil = mysql_query($sql_perfil);
$rows_perfil = mysql_num_rows($ver_perfil);
while($result = mysql_fetch_array($ver_perfil)) {
    $sexo = $result['sexo'];
    $ori_sexo = $result['ori_sexo'];
    if($sexo == 'M'){
        $sexo_oposto = 'F';
    } else {
        $sexo_oposto = 'M';
    }

    $naipe = $result['naipe'];
    $relacionamento = $result['relacionamento'];
}
if($_GET['perfil']){
// Quando abre um perfil faz comparacoes para achar semelhantes
$sql_busca = "SELECT * FROM busca WHERE usr = '$id_usr'";
$ver_busca = mysql_query($sql_busca);
$rows_busca = mysql_num_rows($ver_busca);
if($rows_busca){
    while($result = mysql_fetch_array($ver_busca)) {
        $idade_inicial = $result['idade_inicial'];
        $idade_final = $result['idade_final'];
        $data_atual = date("Y-m-d");
        $idade_inicial = $idade_inicial - 5;
        $dt_nasc_inicial = $data_atual - $idade_inicial."-01-01";
        $idade_final = $idade_final + 5;
        $dt_nasc_final = $data_atual - $idade_final."-12-31";
    }
}

$usr_perfil = $_GET['perfil'];
$sql_perfil = "SELECT * FROM perfil WHERE id = '$usr_perfil'";
$ver_perfil = mysql_query($sql_perfil);
$rows_perfil = mysql_num_rows($ver_perfil);
if($rows_perfil){
    while($result = mysql_fetch_array($ver_perfil)) {
        $altura = $result['altura'];
        $peso = $result['peso'];
        $biotipo = $result['biotipo'];
        $pelo_corporal = $result['pelo_corporal'];
        $olhos = $result['olhos'];
        $etnia = $result['etnia'];
        $cabelo = $result['cabelo'];
        $estilo = $result['estilo'];
    }
}
```

```

        $decoracao_corporal = $result['decoracao_corporal'];
    }
}

// Monta a SQL
$sql = "SELECT * FROM perfil WHERE dt_nascimento >= '$dt_nasc_final'
AND dt_nascimento <= '$dt_nasc_inicial' AND id != '$usr_perfil'";

if($altura){
    $sql = $sql." AND altura = '$altura'";
}

if($peso){
    $sql = $sql." AND peso = '$peso'";
}

if($biotipo){
    $sql = $sql." AND biotipo = '$biotipo'";
}

if($pelocorporal){
    $sql = $sql." AND pelo_corporal = '$pelocorporal'";
}

if($cordosolhos){
    $sql = $sql." AND olhos = '$cordosolhos'";
}

if($etnia){
    $sql = $sql." AND etnia = '$etnia'";
}

if($cabelo){
    $sql = $sql." AND cabelo = '$cabelo'";
}

if($estilo){
    $sql = $sql." AND estilo = '$estilo'";
}

if($cabelo){
    $sql = $sql." AND cabelo = '$cabelo'";
}

if($estilo){
    $sql = $sql." AND estilo = '$estilo'";
}

if($decoracao){
    $sql = $sql." AND decoracao_corporal = '$decoracao'";
}

$sql = $sql."
    AND sexo = '$sexo_oposto'
    AND ori_sexo = '$ori_sexo'
    AND sexo = '$sexo_oposto'
    ORDER BY RAND() limit 3
";

```

```
}  
// Pesquisa no banco de dados  
$sql = mysql_query($sql) or die ("Erro: ".mysql_error());  
  
## O código restante trata apenas a impressão na tela, sendo  
## desconsiderado para apresentar aqui.
```

**Arquivo: Recomendacoes.php**

Trabalha como o código anterior, porém com informações da tabela de busca.

---

```
// montar uma select e retornar os valores
$sql_busca = "SELECT * FROM busca WHERE usr = '$id_usr'";
$ver_busca = mysql_query($sql_busca);
$rows_busca = mysql_num_rows($ver_busca);
if($rows_busca){
while($result = mysql_fetch_array($ver_busca)) {
    $idade_inicial = $result['idade_inicial'];
    $idade_final = $result['idade_final'];
    //
    $data_atual = date("Y-m-d");
    $idade_inicial = $idade_inicial - 2;
    $dt_nasc_inicial = $data_atual - $idade_inicial."-01-01";
    $idade_final = $idade_final + 2;
    $dt_nasc_final = $data_atual - $idade_final."-12-31";
    //
    $estado = $result['estado'];
    $altura = $result['altura'];
    $peso = $result['peso'];
    $biotipo = $result['biotipo'];
    $pelocorporal = $result['pelocorporal'];
    $cordosolhos = $result['cordosolhos'];
    $etnia = $result['etnia'];
    $cabelo = $result['cabelo'];
    $estilo = $result['estilo'];
    $decoracao = $result['decoracao'];

    $sql = "SELECT * FROM perfil WHERE id != '$id_usr' AND
dt_nascimento >= '$dt_nasc_final' AND dt_nascimento <=
'$dt_nasc_inicial'";
    // como agente trabalha com multipla opcoes, aqui construimos o
    restante da sql
    // depois estarei mudando para uma funcao, para evitar esse monte
    de if.

    $altura = $_GET['altura'];
    if($altura){
        while(list($campo,$valor) = each($altura)) {
            if(!$resultaltura){
                $resultaltura = $resultaltura."altura =
'".$valor."' ";
            } else {
                $resultaltura = $resultaltura."OR altura =
'".$valor."' ";
            }
        }
        $resultaltura = "AND (".$resultaltura.") ";
    } else {
        $resultdecoracao = "AND altura like '%" ";
    }

    $estado = $_GET['estado'];
    if($estado){
        while(list($campo,$valor) = each($estado)) {
            if(!$resultestado){
```

```

        $resultestado = $resultestado."estado =
'".$valor.'" ";
    } else {
        $resultestado = $resultestado."OR estado =
'".$valor.'" ";
    }
}
$resultestado = "AND (".$resultestado.") ";
} else {
    $resultdecoracao = "AND estado like '%" ";
}

$pelocorporal = $_GET['pelocorporal'];
if($pelocorporal){
    while(list($campo,$valor) = each($pelocorporal)) {
        if(!$resultpelocorporal){
            $resultpelocorporal =
$resultpelocorporal."pelo_corporal = '".$valor.'" ";
        } else {
            $resultpelocorporal = $resultpelocorporal."OR
pelo_corporal = '".$valor.'" ";
        }
    }
    $resultpelocorporal = "AND (".$resultpelocorporal.") ";
} else {
    $resultdecoracao = "AND pelo_corporal like '%" ";
}

$biotipo = $_GET['biotipo'];
if($biotipo){
    while(list($campo,$valor) = each($biotipo)) {
        if(!$resultbiotipo){
            $resultbiotipo = $resultbiotipo."biotipo =
'".$valor.'" ";
        } else {
            $resultbiotipo = $resultbiotipo."OR biotipo =
'".$valor.'" ";
        }
    }
    $resultbiotipo = "AND (".$resultbiotipo.") ";
} else {
    $resultdecoracao = "AND biotipo like '%" ";
}

$cordosolhos = $_GET['cordosolhos'];
if($cordosolhos){
    while(list($campo,$valor) = each($cordosolhos)) {
        if(!$resultcordosolhos){
            $resultcordosolhos = $resultcordosolhos."olhos =
'".$valor.'" ";
        } else {
            $resultcordosolhos = $resultcordosolhos."OR olhos =
'".$valor.'" ";
        }
    }
    $resultcordosolhos = "AND (".$resultcordosolhos.") ";
} else {
    $resultdecoracao = "AND cordosolhos like '%" ";
}
}

```

```

$etnia = $_GET['etnia'];
if($etnia){
    while(list($campo,$valor) = each($etnia)) {
        if(!$resultetnia){
            $resultetnia = $resultetnia."etnia = '". $valor.'"
";
        } else {
            $resultetnia = $resultetnia."OR etnia =
'". $valor.'" ";
        }
    }
    $resultetnia = "AND (". $resultetnia.) ";
} else {
    $resultdecoracao = "AND etnia like '%" ";
}

$cabelo = $_GET['cabelo'];
if($cabelo){
    while(list($campo,$valor) = each($cabelo)) {
        if(!$resultcabelo){
            $resultcabelo = $resultcabelo."cabelo =
'". $valor.'" ";
        } else {
            $resultcabelo = $resultcabelo."OR cabelo =
'". $valor.'" ";
        }
    }
    $resultcabelo = "AND (". $resultcabelo.) ";
} else {
    $resultdecoracao = "AND cabelo like '%" ";
}

$estilo = $_GET['estilo'];
if($estilo){
    while(list($campo,$valor) = each($estilo)) {
        if(!$resultestilo){
            $resultestilo = $resultestilo."estilo =
'". $valor.'" ";
        } else {
            $resultestilo = $resultestilo."OR estilo =
'". $valor.'" ";
        }
    }
    $resultestilo = "AND (". $resultestilo.) ";
} else {
    $resultdecoracao = "AND estilo like '%" ";
}

$decoracao = $_GET['decoracao'];
if($decoracao){
    while(list($campo,$valor) = each($decoracao)) {
        if(!$resultdecoracao){
            $resultdecoracao =
$resultdecoracao."decoracao_corporal = '". $valor.'" ";
        } else {
            $resultdecoracao = $resultdecoracao."OR
decoracao_corporal = '". $valor.'" ";
        }
    }
    $resultdecoracao = "AND (". $resultdecoracao.) ";
}

```

```

} else {
    $resultdecoracao = "AND decoracao_corporal like '%" ";
}

}

// junta os valores da sql

$sqlarray = $sqlnaipe." ".$resultaltura." ".$resultbiotipo."
".$resulttetnia." ".$resultestilo." ".$resultestado."
".$resultpelocorporal." ".$resultolhos." ".$resultcabelo."
".$resultdecoracao;
$sql = $sql."
    AND sexo = '$sexo_oposto'
    AND ori_sexo = '$ori_sexo'
    ".$sqlarray."
    ORDER BY RAND() limit 3
    ";
}

```

**Arquivo: atualizarRanking.php**

Trabalha as informações coletadas do perfil do usuário para gerar o Ranking de Recomendação. Além disso, atualiza o Ranking na tabela normalizada. As informações coletadas são: naipe (área de interesse), contatos primários e secundários, quantidade de bloqueios, avaliações positivas e negativas.

---

```
<?

// Conexão Banco de dados
include "scripts/cnx_mysql.php";
$con = $link;
// Selecionar Banco de Dados
$db = mysql_select_db( "coracoes_db", $con );
if ( !$db ) {
    echo "Falha ao selecionar banco de dados.";
    exit;
}

// Pesos
$pesosCopas = array(
    "VPF" => 2, // peso para quantidade e visitas
    "CPR" => 2.5, // peso para contatos primarios
    "CSR" => 2.5, // peso para contatos secundarios
    "AVG" => 3, // peso para avaliacoes
    positivas/negativas
    "BLQ" => 0, // peso para bloqueios
    "TOTAL" => 10, // Total de pesos (nao coloquei
    calculado pra ficar mais rápido o processo)
);

// Para cada naipe
// Para cada categoria, calcular o ranking
$arNaipes = array( 1, 2, 3, 4, 5 );
foreach ( $arNaipes as $naipe ) {
    echo "<hr>Calculando naipe $naipe
<br>";

    // Consulta quantide de perfis de uma determinado
nipe
    $sql = "
        select
            id,
            p_positivo - p_negativos as AVG,
            p_bloqueos as BLQ,
            p_piscada as CPR,
            p_msn + p_telefone as CSR,
            p_clicks as VPF
        from perfil
        where naipe = $naipe
    ";
    $queryPerfil = mysql_query( $sql, $con );

    $nPerfis = mysql_num_rows( $queryPerfil );
    echo "Quantidade de naipe $naipe:
    $nPerfis <br>";
    if ( $nPerfis > 0 ) {
```

```

// Consultando valores maximos e mínimos na
tabela
$sql = "
    select
        min(p_clicks) as menorVPF,
        max(p_clicks) as maiorVPF,

        min(p_piscada) as menorCPR,
        max(p_piscada) as maiorCPR,

        min(p_msn) + min(p_telefone)
as menorCSR,
        max(p_msn) + min(p_telefone)
as maiorCSR,

        min(p_positivo-p_negativos)
menorAVG,
        max(p_positivo-p_negativos)
maiorAVG,

        min(p_bloqueos) as menorBLQ,
        max(p_bloqueos) as maiorBLQ
    from perfil
    where naipe = $naipe
";
$query = mysql_query( $sql, $con );

// Atribuindo Valores menores e maiores ao
vetor para os calculos
$valoresMM = array(
    "menorVPF" => mysql_result( $query,
0, "menorVPF" ), // Menor valor de visitas do perfil
    "maiorVPF" => mysql_result( $query,
0, "maiorVPF" ), // Maior valor de visitas do perfil

    "menorCPR" => mysql_result( $query,
0, "menorCPR" ), // Menor valor de contatos primarios
    "maiorCPR" => mysql_result( $query,
0, "maiorCPR" ), // Maior valor de contatos primarios

    "menorCSR" => mysql_result( $query,
0, "menorCSR" ), // Menor valor de contatos secundarios
    "maiorCSR" => mysql_result( $query,
0, "maiorCSR" ), // Maior valor de contatos secundarios

    "menorAVG" => mysql_result( $query,
0, "menorAVG" ), // Menor valor de avaliacoes positivas/negativas
    "maiorAVG" => mysql_result( $query,
0, "maiorAVG" ), // Maior valor de avalicoes positivas/negativas

    "menorBLQ" => mysql_result( $query,
0, "menorBLQ" ), // Menor valor de bloqueios
    "maiorBLQ" => mysql_result( $query,
0, "maiorBLQ" ), // Maior valor de bloqueios
);
if ( $valoresMM["menorAVG"] < 0 ) {
    $valoresMM["menorAVG"] = 0;
}
if ( $valoresMM["maiorAVG"] < 0 ) {

```

```

        $valoresMM["maiorAVG"] = 0;
    }
    echo "Valores
Máximos e Mínimos naipe($naipe) <br>";
    iv( $valoresMM );

    for( $x=0; $x<$nPerfis; $x++ ){
        $valoresPerfil = array(
            "VPF" => mysql_result(
$queryPerfil, $x, "VPF" ), // total de visitas do perfil
            "CPR" => mysql_result(
$queryPerfil, $x, "CPR" ), // total de contatos primarios do perfil
            "CSR" => mysql_result(
$queryPerfil, $x, "CSR" ), // total de contatos secundarios do
perfil
            "AVG" => mysql_result(
$queryPerfil, $x, "AVG" ), // total de avaliacoes
positivas/negativas do perfil
            "BLQ" => mysql_result(
$queryPerfil, $x, "BLQ" ), // total de bloqueios do perfil
        );
        if ( $valoresPerfil["AVG"] < 0 ) {
            $valoresPerfil["AVG"] = 0;
        }
    }
    //
    iv( $valoresPerfil ); echo "<br>";

    // Atualizando campo com os dados
    calculados
    $sql = "
        update perfil set pontos =
". calcularRanking( $valoresPerfil, $valoresMM, $pesosCopas ) ."
        where id = ". mysql_result(
$queryPerfil, $x, "id" ) ."
    ";
    mysql_query( $sql, $con );
}

}

echo "Ranking atualizado.";

function calcularRanking( $valoresPerfil, $valoresMM, $pesos
){
    // Normalizar dados
    foreach( $valoresPerfil as $chave => $valor ){
        if ( $valoresMM[ "maior$chave" ] -
$valoresMM[ "menor$chave" ] < 1 ) {
            $normalizado[ $chave ] =
                (
                    ( $valoresPerfil[
$chave ] - $valoresMM[ "menor$chave" ] ) /
                    1
                ) * 10
        }
    }
}

```

```

        } else {
            $normalizado[ $chave ] =
                (
                    ( $valoresPerfil[
$chave ] - $valoresMM[ "menor$chave" ] ) /
                    ( $valoresMM[
"maior$chave" ] - $valoresMM[ "menor$chave" ] )
                ) * 10
        }
        ;
        if ( $normalizado[ $chave ] < 0 ) {
            $normalizado[ $chave ] = 0;
        }
    }
    //
        iv( $normalizado ); echo "<br>";

    // Média ponderada
    $soma = 0;
    foreach( $valoresPerfil as $chave => $valor ){
        $soma += $normalizado[ $chave ] * $pesos[
$chave ];
    }

    return $soma / $pesos["TOTAL"];
}

function iv( $ar ){
    echo "<pre>";
    print_r( $ar );
    echo "</pre>";
}

?>
<!-- <script> window.location = "resultadoRanking.php" </script> -->

```

## Anexo V – Estrutura do Banco de Dados

As tabelas do banco de dados estão representadas na figura abaixo:



*Figura 50 - Tabelas do banco de dados*

Os campos são os seguintes:

Table Name	Fields
cidades	<ul style="list-style-type: none"> <li>cod_cidade: INTEGER(4)</li> <li>uf: CHAR(2)</li> <li>nome_cidade: VARCHAR(50)</li> </ul>
borboleta	<ul style="list-style-type: none"> <li>usr_1: TINYINT(4)</li> <li>usr_2: TINYINT(4)</li> <li>ativo: TINYINT(4)</li> </ul>
busca	<ul style="list-style-type: none"> <li>usr: INTEGER(12)</li> <li>estado: VARCHAR(30)</li> <li>altura: VARCHAR(30)</li> <li>peso: VARCHAR(30)</li> <li>biotipo: VARCHAR(30)</li> <li>pelocorporal: VARCHAR(30)</li> <li>cordosolhos: VARCHAR(30)</li> <li>etnia: VARCHAR(30)</li> <li>cabelo: VARCHAR(30)</li> <li>estilo: VARCHAR(30)</li> <li>decoracao: VARCHAR(30)</li> <li>emordem: VARCHAR(30)</li> <li>online: INTEGER(1)</li> <li>idade_inicial: INTEGER(2)</li> <li>idade_final: INTEGER(2)</li> </ul>
favoritos	<ul style="list-style-type: none"> <li>id: INTEGER(11)</li> <li>usr_1: INTEGER(11)</li> <li>usr_2: INTEGER(11)</li> </ul>
online	<ul style="list-style-type: none"> <li>timestamp: INTEGER(15)</li> <li>ip: VARCHAR(40)</li> <li>arquivo: VARCHAR(100)</li> </ul>
bloqueados	<ul style="list-style-type: none"> <li>id: INTEGER(11)</li> <li>usr_1: INTEGER(11)</li> <li>usr_2: INTEGER(11)</li> </ul>
fotos	<ul style="list-style-type: none"> <li>id_foto: INTEGER(11)</li> <li>id_perfil: INTEGER(11)</li> <li>foto: VARCHAR(50)</li> <li>comentario: TEXT</li> </ul>
telefone	<ul style="list-style-type: none"> <li>usr_1: TINYINT(4)</li> <li>usr_2: TINYINT(4)</li> <li>ativo: TINYINT(4)</li> </ul>
piscada	<ul style="list-style-type: none"> <li>id: INTEGER(11)</li> <li>usr_1: INTEGER(11)</li> <li>usr_2: INTEGER(11)</li> <li>ativo: INTEGER(11)</li> </ul>
qualificacao	<ul style="list-style-type: none"> <li>qualificacao: INTEGER(11)</li> <li>usr_1: INTEGER(11)</li> <li>usr_2: INTEGER(11)</li> </ul>

Figura 51 - Campos das tabelas - parte 1

perfil	
	id: INTEGER(11)
	nome: VARCHAR(50)
	apelido: VARCHAR(30)
	email: VARCHAR(60)
	senha: VARCHAR(32)
	idioma: VARCHAR(5)
	dt_nascimento: DATE
	sexo: CHAR(1)
	ori_sexo: TINYINT(1)
	papel: TINYINT(1)
	relacionamento: TINYINT(1)
	sexoseguro: TINYINT(1)
	naipe: TINYINT(1)
	idade_final: INTEGER(2)
	idade_inicial: INTEGER(2)
	msn: VARCHAR(150)
	telefone: VARCHAR(150)
	bebe: TINYINT(1)
	fuma: TINYINT(1)
	malha: TINYINT(1)
	drogas: TINYINT(1)
	pais: VARCHAR(30)
	exterior: TINYINT(1)
	estado: CHAR(2)
	cod_cidade: INTEGER(5)
	bairro: VARCHAR(150)
	altura: TINYINT(4)
	peso: TINYINT(4)
	biotipo: TINYINT(1)
	pelo_corporal: TINYINT(1)
	olhos: TINYINT(1)
	etnia: TINYINT(1)
	cabelo: TINYINT(1)
	estilo: TINYINT(1)
	decoracao_corporal: TINYINT(1)
	ocupacao: TINYINT(1)
	escolaridade: TINYINT(1)
	titulo: VARCHAR(200)
	texto: TEXT
	procuro: TEXT
	prato: TEXT
	filme: TEXT
	paixoes: TEXT
	idiomas: TEXT
	tv: TEXT
	frequenta: TEXT
	interesses: VARCHAR(255)
	foto: VARCHAR(30)
	pontos: INTEGER(11)
	p_positivo: INTEGER(11)
	p_negativos: INTEGER(11)
	p_bloqueos: INTEGER(11)
	p_piscada: INTEGER(11)
	p_msn: INTEGER(11)
	p_telefone: INTEGER(11)
	foto_a: INTEGER(1)
	email
	email

Figura 52 - Campos das tabelas - parte II

Os relacionamentos entre as tabelas não foram representados porque um modelo entidade-relacionamento não foi criado para este projeto.

## Anexo VI – Recomendação do Site

O site, no menu superior, possui os *links* abaixo:



Figura 53 - Links do menu superior

O terceiro *link*, “indique o site”, apresenta a janela abaixo depois de clicado:

http://www.2coracoes.com - Indi...

 2corações.com

**Indique para um amigo(a)**

Seu nome

Seu e-mail

Nome do amigo

E-mail do amigo

Coi Internet

Figura 54 – Indicador de recomendação do site

É uma forma de recomendação do próprio *site*. Uma análise estatística das recomendações realizadas por esta janela poderia fornecer uma visão de satisfação do usuário.

## Glossário

**Balanced Scorecard** – Metodologia desenvolvida pelos professores da Escola de Administração de Harvard, Robert Kaplan e David Norton, em 1992. Os métodos usados na gestão do negócio, dos serviços e da infraestrutura baseiam-se normalmente em metodologias consagradas que utilizam a tecnologia da informação como solução de apoio, relacionando-a a gerência de serviços e garantia de resultados do negócio. Os passos dessas metodologias incluem: definição da estratégia empresarial, gerência do negócio, gerência de serviços e gestão da qualidade — implementados através de indicadores de desempenho. Os requisitos para definição de tais indicadores tratam dos processos de um modelo da administração de serviços e busca da maximização dos resultados baseados em quatro perspectivas que refletem a visão e estratégia empresarial: financeira, clientes, aprendizado-crescimento e processos internos.

**Base de Dados** – Sistema de manutenção de registros por computador. Os termos Banco de Dados e Base de Dados são sinônimos neste trabalho.

**CAPES** – A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, mais conhecida pela sigla CAPES, é uma agência de fomento à pesquisa brasileira que atua na expansão e consolidação da pós-graduação stricto sensu (mestrado e doutorado) em todos os estados do Brasil.

**Data Mining** – *Data Mining* ou Mineração de Dados é um conjunto de técnicas que buscam a descoberta de novos conhecimentos através da análise de grandes bases de dados. Talvez a definição mais importante seja a de Usama Fayyad (FAYYAD et al. 1996): “o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis”. Esse processo vale-se de diversos algoritmos que processam os dados e encontram os padrões “válidos, novos e úteis”. Embora os algoritmos sejam capazes de descobrir

padrões “válidos e novos”, ainda não há uma solução eficaz para determinar padrões úteis<sup>72</sup>.

**Documento** – Dentro do presente trabalho, documento é a informação passível de armazenamento em um banco de dados.

**Feedback de Relevância** – Interação do usuário com o sistema de recuperação de informação para alterar os critérios de relevância automática definidos pelo sistema. O *feedback* pode ser implementado através da realização de uma pesquisa inicial com pesos simples, definidos pelo sistema, onde o usuário é convidado a julgar a relevância de alguns itens. Em seguida, O sistema extrai termos de consulta dos itens julgados relevantes. Há uma ordenação dos termos extraídos e alguns são selecionados (automaticamente ou com *feedback* do usuário). Em seguida promove-se um rebalanceamento da importância dos termos originais, os novos termos são categorizados segundo a relevância da informação, e uma nova iteração de pesquisa é realizada.

**Histograma** – Representação gráfica da distribuição de freqüências de uma massa de medições, normalmente um gráfico de barras.

**Informação** – Conjunto de dados que satisfazem a uma determinada necessidade de informação. Possui valor agregado e precisa de intervenção do usuário.

**Metadados** – Metadados, ou Metainformação, são dados capazes de descrever outros dados, ou seja, dizer do que se tratam, dar um significado real e plausível a um arquivo de dados; são a representação de um objeto digital. De acordo com a definição do W3C<sup>73</sup>, metadados são informações localizadas na *web*, inteligíveis por um computador. Mais sinteticamente, podemos dizer que um metadado é um dado utilizado para descrever um dado primário.

**Mineração de Dados** – Consulte o termo *Data Mining*.

---

<sup>72</sup> A definição de "útil" remete à necessidade do usuário.

<sup>73</sup> World Wide Web Consortium, disponível em <<http://www.w3.org>>

**Modelo** – Segundo Pereira (1994), um modelo é uma representação da percepção da realidade.

**Necessidade de informação** – A necessidade de informação, segundo Mizarro (1996), é a falta de conhecimento do usuário para realizar determinada atividade. Para solucionar um determinado problema, o usuário necessita de um determinado conhecimento. Se ele não possui tal conhecimento, ele passa a ter uma necessidade de informação. Essa necessidade deve ser satisfeita, geralmente através de uma consulta a um sistema de recuperação de informação.

**Ontologias** - Para a Ciência da Informação, ontologias são teorias que especificam um vocabulário relativo a certo domínio. Este vocabulário define entidades, classes, propriedades, predicados e funções, e as relações entre estes componentes.

**Perfil de usuário** – Perfil de usuário é como cada usuário é conhecido pelo sistema. Além das informações de registro elaboradas pelo usuário ao se cadastrar, podem ser armazenadas interações percebidas pelo sistema. Isso não quer dizer que seja registrado apenas o comportamento em termos de espaço e tempo, também podem ser as interações com o conteúdo do ambiente. Esses parâmetros são armazenados em um banco de dados.

**Relevância** – Relevância é um critério subjetivo. De acordo com Le Coadic (2004, p.80-82), a relevância é uma relação entre um documento e uma questão. A relevância objetiva é aquela do ponto de vista do sistema e a subjetiva é a do ponto de vista do usuário. A relevância é objeto de estudo contínuo, pois sua multidimensionalidade faz com que nunca haja consenso a respeito de sua conceituação.

**Sistema de informação** – Sistema automatizado — que pode ser denominado como Sistema de Informação Computadorizado —, ou mesmo manual, que abrange pessoas, máquinas, e/ou métodos organizados para coletar, processar, transmitir e disseminar dados que representam informação para o usuário.

**Sistema de recuperação de informação** – Sistema automatizado que visa a organizar, a tratar e a recuperar documentos que possam satisfazer a uma necessidade expressa do usuário. Seu principal problema consiste na representação do documento e da expressão de busca, e seu foco é o usuário.

**Sistema de recomendação** – É um sistema de informação que auxilia o processo de indicar ou receber uma indicação. De acordo com Torres Jr. (2004), um sistema de recomendação geralmente possui: (i) dados prévios: dados armazenados pelo sistema antes do início do processo de recomendação; (ii) dados de entrada: dados que precisam ser inseridos pelo usuário a fim de receber recomendações; (iii) um algoritmo que combine os dados de entrada com os dados prévios para produzir recomendações. Em um sistema real, os dados prévios são obtidos do perfil ou histórico dos usuários, e os dados de entrada são obtidos através das ações que o usuário realiza para obter recomendações.

**Site** – Consulte *Website*.

**Spreading activation** – Nos modelos de busca em rede, a comparação entre consulta e documento não é exata, sendo esse processo de busca um raciocínio de inferência. Nas buscas em redes semânticas, os pesos das associações são dados de acordo com a semântica; a comparação é feita pela técnica de *spreading activation* (propagação de ativação), onde a consulta de um lado da rede estimula documentos de outro, usando os *links*; as restrições impostas limitam o número de nodos a serem ativados; o valor final da ativação representa o valor do status de recuperação. Em linhas gerais, isto quer dizer que uma consulta do usuário — utilizando a técnica de *spreading activation* — ativa nodos que, por sua vez, ativam nodos vizinhos até alcançar documentos, e o *ranking* é dado pelo número de estímulos que cada documento recebe (LOH, 1999, p.66).

**SQL** – Linguagem de Consulta Estruturada (*Structured Query Language*), linguagem de pesquisa declarativa para banco de dados relacional.

**Utilitário** – Ferramenta ou trecho de programa que desempenha funções de auxílio ao usuário.

**Usuário** – Utilizador de um sistema de informação; o único capaz de transformar o conteúdo de um documento em informação.

**Website** – Conjunto de páginas *Web*, isto é, de hipertextos acessíveis geralmente pelo protocolo *http* na Internet.

**Wizard** – Parte do programa que assiste ao usuário executar uma função da melhor forma.