



**Ricardo Niederberger Cabral**

**Um estudo de recomendadores baseados em  
conteúdo e redes sociais**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática da PUC-Rio

Orientador: Prof. Daniel Schwabe

Rio de Janeiro  
Novembro de 2008



**Ricardo Niederberger Cabral**

**Um estudo de recomendadores baseados em  
conteúdo e redes sociais**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão examinadora abaixo assinada.

**Prof. Daniel Schwabe**

Orientador

Departamento de Informática — PUC-Rio

**Prof. Ruy Luiz Milidiú**

Departamento de Informática - PUC-Rio

**Prof. Marcus Vinicius Soledade Poggi de Aragão**

Departamento de Informática - PUC-Rio

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 28 de Novembro de 2008

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

### **Ricardo Niederberger Cabral**

Graduou-se em Engenharia de Computação no Instituto Militar de Engenharia em 2004 e tem atuado desde então como consultor em engenharia de software na Chemtech-Siemens. Possui interesse acadêmico e profissional em sistemas distribuídos, mineração de dados, web semântica e computer-aided-engineering.

#### Ficha Catalográfica

Cabral, Ricardo Niederberger

Um estudo de recomendadores baseados em conteúdo e redes sociais / Ricardo Niederberger Cabral; orientador: Daniel Schwabe. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2008.

v., 90 f: il. ; 29,7 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Recomendadores. 3. Mineração de dados. 4. Redes sociais. 5. Web semântica. I. Schwabe, Daniel. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

# Agradecimentos

Agradeço à minha família e namorada pela compreensão; ao meu orientador pela motivação, discussões e confiança; aos membros da banca pelos comentários e discussões e à Chemtech pelas oportunidades.

## Resumo

Cabral, Ricardo Niederberger; Schwabe, Daniel. **Um estudo de recomendadores baseados em conteúdo e redes sociais**. Rio de Janeiro, 2008. 90p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

São duas as contribuições deste trabalho: (1) analisar a empregabilidade de algoritmos de recomendação para redes sociais. Tais algoritmos de recomendação podem receber como entrada não somente o grafo social destas redes como também características do conteúdo de itens a serem recomendados para usuários finais. Para tal, serão levantadas as principais características das redes sociais e as técnicas de recomendação automática que podem ser empregadas para essas tarefas. Especial atenção será dada à rede social online *Flickr* para compartilhamento de fotos e ao emprego de métricas de semelhança visual entre imagens. A segunda contribuição (2) é a construção de uma *framework* para a modelagem e análise de redes sociais, bem como o estudo do desempenho de algoritmos de recomendação nestes contextos. Estarão nela contidas as melhores práticas adotadas ao longo do estudo, como técnicas para coleta, análise e visualização de dados, classificação de redes sociais e tarefas de recomendação dentro destas, implementação dos algoritmos e arquiteturas de recomendadores. A relevância de tais contribuições advém da enorme quantidade de informação disponível online e a crescente complexidade dos inter-relacionamentos entre esses dados. Sistemas recomendadores nesse contexto podem oferecer grande ajuda para usuários finais.

## Palavras-chave

Recomendadores. Mineração de dados. Redes sociais. Web semântica.

## Abstract

Cabral, Ricardo Niederberger; Schwabe, Daniel. **A study on recommender systems based on content and social networks.** Rio de Janeiro, 2008. 90p. MsC Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation offers two major contributions: (1) to evaluate the suitability of recommender algorithms for social networks. Such recommender algorithms may receive as input not only the social graph of these networks but also content-based data from recommended items. For such, the relevant characteristics of social networks and the most important recommender techniques for these tasks will be surveyed. Special attention is given to the web-based system for social photo-sharing called *Flickr* and to the employment of visual metrics for image similarity. The second contribution (2) is the construction of a framework for the modeling and analysis of social networks, as well as aiding the empirical study of recommender algorithms on these contexts. Also part of this framework are the best practices adopted throughout the work done on this dissertation, such as: techniques for the gathering, analysis and visualization of data; social networks classification; identification and modeling of recommending tasks within these contexts; implementation of algorithms and their architecture. The relevance of such contributions lies on the enormous amount of information available online and on the ever-growing complexity of the relationships between this data. In this context, recommender systems may provide a great aid for end-users.

## Keywords

Recommenders. Data mining. Social networks. Semantic web.

# Sumário

1	Introdução	<b>13</b>
1.1	Problemas a resolver e relevância	13
1.2	Objetivos desse trabalho	14
1.3	Organização da dissertação	15
2	Conceitos	<b>16</b>
2.1	Redes	16
2.2	Redes sociais	16
2.3	Sistemas recomendadores	20
2.4	Frameworks para avaliação de sistemas recomendadores	27
2.5	Filtragem colaborativa	27
2.6	Métricas de desempenho de recomendadores	29
3	Framework para avaliação de sistemas de recomendação	<b>33</b>
3.1	Modelagem da rede social e tarefas de recomendação	33
3.2	Realização de experimentos	43
4	Análise de recomendadores para a rede social Flickr	<b>47</b>
4.1	Conceitos da rede Flickr	48
4.2	Coleta de dados	49
4.3	Caracterização dos dados coletados	51
4.4	Avaliação das possíveis tarefas de recomendação	54
4.5	Recomendação de grupos para fotos	56
4.6	Recomendação de grupos para fotos utilizando conteúdo visual	61
4.7	Recomendação de fotos para usuários	67
4.8	Recomendação de usuários para usuários	67
4.9	Recomendação de grupos para usuários	71
4.10	Resumo dos experimentos realizados	72
5	Conclusão	<b>74</b>
5.1	Resultados e contribuições	74
5.2	Trabalhos futuros	75
A	Diagramas auxiliares	<b>85</b>
B	Módulos de software da framework	<b>88</b>
B.1	Packages	88
B.2	Package recfwk.engine	88
B.3	Package recfwk.filters	89
B.4	Package recfwk.io	89
B.5	Package recfwk.model	89
B.6	Package recfwk.util	90
B.7	Package recfwk.vis	90

## Lista de figuras

3.1	Relação entre recomendadores para redes sociais, agregadores de conteúdo social e redes sociais	35
3.2	Principais classes e relacionamentos da ontologia <i>SIOC-Core</i>	36
3.3	Tarefas básicas de recomendação na ontologia <i>SIOC-Core</i>	37
3.4	Notação gráfica para a representação de redes sociais e tarefas de recomendação	38
3.5	Tarefas de recomendação na rede social <i>Last.fm</i>	40
3.6	Tarefas de recomendação na rede social <i>Orkut</i>	42
3.7	Fluxo de informações entre os componentes da <i>framework</i> proposta	43
4.1	Principais entidades e relacionamentos do serviço <i>Flickr</i>	48
4.2	Frequência de fotos contribuídas por usuário	53
4.3	Relacionamentos entre grupos e fotos	53
4.4	Relacionamentos entre usuários e fotos favoritas	54
4.5	Tarefas de recomendação na rede social <i>Flickr</i>	55
4.6	Experimentos de recomendação realizados na rede <i>Flickr</i> (setas em azul)	56
4.7	Fluxo de dados e entidades envolvidas no processo de avaliação de recomendadores	58
4.8	Desempenho da recomendação de grupos para fotos usando algoritmo <i>Top-N</i> (filtragem colaborativa baseada na co-ocorrência de fotos em grupos)	59
4.9	Desempenho da recomendação de grupos para fotos (comparativo naïve: grupos com maior número de fotos)	60
4.10	Comparativo entre o algoritmo <i>Top-N</i> para recomendação de grupos para fotos e a abordagem naïve (rank de 12 recomendações)	60
4.11	<i>Wavelets</i> da base <i>Haar</i> permitem uma boa caracterização aproximada da imagem com poucos coeficientes	61
4.12	Distribuição da homogeneidade visual dos grupos	63
4.13	Exemplos de imagens do grupo <i>Absolut Red</i>	63
4.14	Exemplos de imagens do grupo <i>BURMA FREE - Birmania Libera</i>	63
4.15	Exemplos de imagens do grupo <i>Roses</i>	64
4.16	Exemplos de imagens do grupo <i>Sunrise, Sunset – Anything Sun!</i>	64
4.17	Desempenho de recomendador de grupos para fotos baseado apenas no conteúdo visual dos grupos	65
4.18	Desempenho da recomendação de fotos para usuários usando duas abordagens naïve: “fotos mais favoritas pelos contatos do usuário” vs. “fotos mais favoritas por todos os usuários”)	68
4.19	Desempenho da recomendação de fotos para usuários usando filtragem colaborativa <i>Top-N</i> , tendo como base co-ocorrência de fotos favoritas em comum entre usuários	68
4.20	Exemplo de tela do protótipo implementado para recomendação de novos contatos para usuários	70



4.21	Desempenho da recomendação de usuários para usuários usando algoritmo de filtragem colaborativa <i>Top-N</i> e abordagem <i>naïve</i> comparativa	71
4.22	Desempenho da recomendação de grupos para usuários usando algoritmo <i>Top-N</i> (filtragem colaborativa baseada na co-ocorrência de usuários em grupos)	72
4.23	Desempenho da recomendação de grupos para usuários (comparativo: grupos com maior número de usuários)	72
A.1	Classes e propriedades da ontologia <i>SIOC-Core</i>	85
A.2	Classes e propriedades da ontologia <i>SIOC-Types</i>	86
A.3	Contexto <i>Flickr</i> completo. Crédito: <a href="http://soldierant.net/">http://soldierant.net/</a>	87

## Lista de tabelas

2.1	Evolução do contexto onde sistemas recomendadores desempenham suas funções	22
4.1	Resumo do volume total de dados coletado do serviço <i>Flickr</i>	51
4.2	Grupos com maiores e menores homogeneidades visuais. Maiores valores indicam maior grau de semelhança entre as fotos do grupo (valores entre 0 e $\sim 6$ ).	64
4.3	Resumo dos experimentos realizados com recomendadores para a rede <i>Flickr</i>	73

## Listagens

3.1	Exemplo de query SPARQL para recuperar itens de um usuário	44
3.2	Exemplo de uso da API Jena para execução de queries SPARQL	44
3.3	Exemplo de <i>construct</i> SPARQL para traduzir triplas RDF de uma ontologia para outra . . . . .	45
4.1	Exemplo de triplas RDF descrevendo dados <i>Flickr</i> . . . . .	49

*In the name of the best within you, do not sacrifice this world to those who are its worst. In the name of the values that keep you alive, do not let your vision of man be distorted by the ugly, the cowardly, the mindless in those who have never achieved his title. Do not lose your knowledge that man's proper estate is an upright posture, an intransigent mind and a step that travels unlimited roads. Do not let your fire go out, spark by irreplaceable spark, in the hopeless swamps of the approximate, the not-quite, the not-yet, the not-at-all. Do not let the hero in your soul perish, in lonely frustration for the life you deserved, but have never been able to reach. Check your road and the nature of your battle. The world you desired can be won, it exists, it is real, it is possible, it's yours.*

**Ayn Rand**, *Atlas Shrugged*, p.979.