



**Luiz Fernando Fernandes de Albuquerque**

**Avaliação de algoritmos online para seleção de  
links patrocinados**

**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. Eduardo Sany Laber

Rio de Janeiro  
Dezembro de 2009



**Luiz Fernando Fernandes de Albuquerque**

**Avaliação de algoritmos online para seleção de links patrocinados**

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. Eduardo Sany Laber**

Orientador

Departamento de Informática — PUC-Rio

**Prof. Marcus V. S. Poggi de Aragão**

Departamento de Informática - PUC-Rio

**Prof. Claudson Ferreira Bornstein**

Departamento de Ciência da Computação - UFRJ

**Prof. José Eugênio Leal**

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 11 de Dezembro de 2009

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

### **Luiz Fernando Fernandes de Albuquerque**

Graduou-se em Tecnologia de Computação no ITA em 1984.

#### Ficha Catalográfica

Albuquerque, Luiz Fernando Fernandes de

Avaliação de algoritmos online para seleção de links patrocinados / Luiz Fernando Fernandes de Albuquerque; orientador: Eduardo Sany Laber. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2009.

89 f: il. ; 30 cm

1. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2009.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Links patrocinados. 3. Algoritmos online. 4. Máquinas de busca. 5. Internet. I. Laber, Eduardo Sany. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Ao meu avô, Oswaldo Pinto Fernandes.

## Agradecimentos

À minha mãe, Marlene Fernandes de Albuquerque.

Ao meu irmão, Tom, e meus sobrinhos, Pedro e Luisa.

Aos meus avós, Arlette, Oswaldo, Júlio, Edith e Francisco.

Ao meu orientador, Eduardo Laber.

Aos professores da Comissão Examinadora, Poggi e Claudson.

Aos professores e funcionários da PUC-Rio.

Aos meus familiares, amigos, amigas e colegas da PUC-Rio.

A Deus, pelos dons necessários para realizar este trabalho.

## Resumo

Albuquerque, Luiz Fernando Fernandes de; Laber, Eduardo Sany. **Avaliação de algoritmos online para seleção de links patrocinados**. Rio de Janeiro, 2009. 89p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Links patrocinados são aqueles que aparecem em destaque nos resultados de pesquisas em máquinas de busca na Internet e são grande fonte de receita para seus provedores. Para os anunciantes, que fazem ofertas por palavras-chave para aparecerem em destaque nas consultas dos usuários, são uma oportunidade de divulgação da marca, conquista e manutenção de clientes. Um dos desafios das máquinas de busca neste modelo de negócio é selecionar os anunciantes que serão exibidos a cada consulta de modo a maximizar sua receita em determinado período. Este é um problema tipicamente online, onde a cada consulta é tomada uma decisão sem o conhecimento prévio das próximas consultas. Após uma decisão ser tomada, esta não pode mais ser alterada. Nesta dissertação avaliamos experimentalmente algoritmos propostos na literatura para solução deste problema, comparando-os à solução ótima offline, em simulações com dados sintéticos. Supondo que o conjunto das consultas diárias obedeça a uma determinada distribuição, propomos dois algoritmos baseados em informações estocásticas que são avaliados nos mesmos cenários que os outros algoritmos.

## Palavras-chave

Links patrocinados. Algoritmos online. Máquinas de busca. Internet.

## Abstract

Albuquerque, Luiz Fernando Fernandes de; Laber, Eduardo Sany.  
**Online algorithms analysis for sponsored links selection.**  
Rio de Janeiro, 2009. 89p. MSc Dissertation — Departamento de  
Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Sponsored links are those that appear highlighted at Internet search engine results. They are responsible for a large amount of their providers' revenue. To advertisers, that place bids for keywords in large auctions at Internet, these links are the opportunity of brand exposing and achieving more clients. To search engine companies, one of the main challenges in this business model is selecting which advertisers should be allocated to each new query to maximize their total revenue in the end of the day. This is a typical online problem, where for each query is taken a decision without previous knowledge of future queries. Once the decision is taken, it can not be modified anymore. In this work, using synthetically generated data, we do experimental evaluation of three algorithms proposed in the literature for this problem and compare their results with the optimal offline solution. Considering that daily query set obeys some well known distribution, we propose two algorithms based on stochastic information, those are evaluated in the same scenarios of the others.

## Keywords

Sponsored links. Sponsored search. Online algorithms. AdWords.  
Search engines. Internet.

## Sumário

|     |  |           |
|-----|--|-----------|
| 1   | Links Patrocinados   | <b>12</b> |
| 1.1 | Introdução   | 12        |
| 1.2 | Descrição do Problema  | 13        |
| 1.3 | Trabalhos Relacionados   | 14        |
| 1.4 | Contribuições  | 16        |
| 1.5 | Organização do Texto   | 17        |
| 2   | Algoritmos Online  | <b>19</b> |
| 2.1 | Caracterização   | 19        |
| 2.2 | Análise Competitiva  | 19        |
| 2.3 | Estratégias de Solução   | 20        |
| 2.4 | Avaliação de Algoritmos Online                                     | 21        |
| 3   | Seleção de Links Patrocinados com Restrição de Limite de Exibições | <b>22</b> |
| 3.1 | Especificação do Problema  | 22        |
| 3.2 | Solução Ótima  | 22        |
| 3.3 | Competitividade dos Algoritmos Online                              | 25        |
| 4   | Algoritmos Implementados   | <b>27</b> |
| 4.1 | Algoritmo Guloso   | 27        |
| 4.2 | Algoritmo AdWords  | 29        |
| 4.3 | Algoritmo Primal-Dual  | 31        |
| 4.4 | Algoritmos Preditivos  | 36        |
| 5   | Avaliação dos Algoritmos   | <b>39</b> |
| 5.1 | Simulação do Ambiente Online                                       | 39        |
| 5.2 | Geração dos Dados de Entrada                                       | 42        |
| 5.3 | Medição do Desempenho  | 46        |
| 5.4 | Avaliação de Cenários com Anunciantes Similares                    | 47        |
| 5.5 | Análise Gráfica dos Algoritmos no Tempo                            | 65        |
| 5.6 | Avaliação de Cenários para Anunciantes Diferenciados               | 77        |
| 6   | Considerações Finais   | <b>84</b> |
| 6.1 | Sobre o Desempenho dos Algoritmos                                  | 84        |
| 6.2 | Sobre o Uso de Predição nos Algoritmos                             | 86        |
| 6.3 | Trabalhos Futuros  | 87        |
|     | Referências Bibliográficas   | <b>88</b> |



## Lista de figuras

|      |  |    |
|------|--|----|
| 1.1  | Exemplo de links patrocinados no Google                | 12 |
| 3.1  | Solução ótima por fluxo em redes                       | 24 |
| 3.2  | Solução ótima por programação inteira                  | 25 |
| 4.1  | Algoritmo Guloso para seleção de links                 | 28 |
| 4.2  | Algoritmo AdWords para seleção de links                | 30 |
| 4.3  | Solução por programação linear pelo método Primal-Dual | 32 |
| 4.4  | Algoritmo Primal-Dual para seleção de links            | 34 |
| 4.5  | Algoritmo Preditivo para seleção de links              | 38 |
| 4.6  | Algoritmo Aleatório para seleção de links              | 38 |
| 5.1  | Cenário 1 - Gráfico 1                                  | 67 |
| 5.2  | Cenário 2 - Gráfico 1                                  | 68 |
| 5.3  | Cenário 2 - Gráfico 2                                  | 69 |
| 5.4  | Cenário 3 - Gráfico 1                                  | 70 |
| 5.5  | Cenário 3 - Gráfico 1                                  | 71 |
| 5.6  | Cenário 4 - Gráfico 1                                  | 72 |
| 5.7  | Cenário 4 - Gráfico 2                                  | 74 |
| 5.8  | Cenário 4 - Gráfico 3                                  | 75 |
| 5.9  | Cenário 5 - Gráfico 1                                  | 76 |
| 5.10 | Cenário 5 - Gráfico 2                                  | 77 |

## Lista de tabelas

|      |   |    |
|------|---|----|
| 4.1  | Ofertas e limites do exemplo de algoritmo guloso        | 28 |
| 4.2  | Exemplo de melhor caso do algoritmo Guloso              | 28 |
| 4.3  | Exemplo de pior caso do algoritmo Guloso                | 28 |
| 4.4  | Exemplo de funcionamento do algoritmo AdWords           | 31 |
| 4.5  | Exemplo de competitividade do algoritmo AdWords         | 31 |
| 4.6  | Exemplo de funcionamento do algoritmo Primal-Dual       | 35 |
| 4.7  | Solução ótima para utilização nos algoritmos preditivos | 36 |
| 5.1  | Arquivos de entrada e saída                             | 40 |
| 5.2  | Tempo de execução de um dia de processamento            | 42 |
| 5.3  | Consultas por distribuição de Zipf                      | 44 |
| 5.4  | Exemplo de experimentos e medição de um cenário         | 47 |
| 5.5  | Cenários  | 47 |
| 5.6  | Parâmetros fixos e variáveis para anunciantes similares | 48 |
| 5.7  | Cenário 1: $b_{max} = 20; Q_a \leq 5$                   | 50 |
| 5.8  | Cenário 1: $b_{max} = 70; Q_a \leq 5$                   | 50 |
| 5.9  | Cenário 1: $b_{max} = 20; Q_a \geq 15$                  | 51 |
| 5.10 | Cenário 1: $b_{max} = 70; Q_a \geq 15$                  | 52 |
| 5.11 | Cenário 2: $b_{max} = 20; Q_a \leq 5$                   | 53 |
| 5.12 | Cenário 2: $b_{max} = 70; Q_a \leq 5$                   | 54 |
| 5.13 | Cenário 2: $b_{max} = 20; Q_a \geq 15$                  | 55 |
| 5.14 | Cenário 2: $b_{max} = 70; Q_a \geq 15$                  | 55 |
| 5.15 | Cenário 3: $b_{max} = 20; Q_a \leq 5$                   | 57 |
| 5.16 | Cenário 3: $b_{max} = 70; Q_a \leq 5$                   | 57 |
| 5.17 | Cenário 3: $b_{max} = 20; Q_a \geq 15$                  | 58 |
| 5.18 | Cenário 3: $b_{max} = 70; Q_a \geq 15$                  | 59 |
| 5.19 | Cenário 4: $b_{max} = 20; Q_a \leq 5$                   | 60 |
| 5.20 | Cenário 4: $b_{max} = 70; Q_a \leq 5$                   | 61 |
| 5.21 | Cenário 4: $b_{max} = 20; Q_a \geq 15$                  | 61 |
| 5.22 | Cenário 4: $b_{max} = 70; Q_a \geq 15$                  | 62 |
| 5.23 | Cenário 5: $b_{max} = 20; Q_a \leq 5$                   | 63 |
| 5.24 | Cenário 5: $b_{max} = 70; Q_a \leq 5$                   | 63 |
| 5.25 | Cenário 5: $b_{max} = 20; Q_a \geq 15$                  | 64 |
| 5.26 | Cenário 5: $b_{max} = 70; Q_a \geq 15$                  | 64 |
| 5.27 | Parâmetros dos gráficos                                 | 65 |
| 5.28 | Parâmetros para anunciantes diferenciados               | 78 |
| 5.29 | Cenário 1 - Anunciantes por Zipf de potência 0.5        | 79 |
| 5.30 | Cenário 1 - Anunciantes por Zipf de potência 1          | 79 |
| 5.31 | Cenário 2 - Anunciantes por Zipf de potência 0.5        | 80 |
| 5.32 | Cenário 2 - Anunciantes por Zipf de potência 1          | 80 |
| 5.33 | Cenário 3 - Anunciantes por Zipf de potência 0.5        | 81 |
| 5.34 | Cenário 3 - Anunciantes por Zipf de potência 1          | 81 |
| 5.35 | Cenário 4 - Anunciantes por Zipf de potência 0.5        | 82 |
| 5.36 | Cenário 4 - Anunciantes por Zipf de potência 1          | 82 |
| 5.37 | Cenário 5 - Anunciantes por Zipf de potência 0.5        | 83 |

|      |  |    |
|------|--|----|
| 5.38 | Cenário 5 - Anunciantes por Zipf de potência 1           | 83 |
| 6.1  | Estatística sobre desempenho dos algoritmos              | 85 |
| 6.2  | Melhores algoritmos                                      | 85 |
| 6.3  | Estatística sobre desempenho dos algoritmos sem previsão | 85 |
| 6.4  | Melhores algoritmos sem previsão                         | 86 |