

# Avaliação do Sistema de Busca em Mercados Eletrônicos

Diego Duarte Wagner Meira Jr.  
Dpto. de Ciência da Computação  
Universidade Federal de Minas Gerais (UFMG)  
Caixa Postal 486 – 31.270-901  
Belo Horizonte – MG – Brasil  
{diegomd, meira}@dcc.ufmg.br

Adriano C. M. Pereira  
Dpto. de Computação – Centro Federal de  
Educação Tecnológica de MG (Cefet-MG)  
Av. Amazonas, 7675 – 30.510-000  
Belo Horizonte – MG – Brasil  
adriano@decom.cefetmg.br

## RESUMO

Mercados eletrônicos têm se tornado mais populares nos últimos anos. Este trabalho apresenta uma avaliação do sistema de busca no mercado eletrônico brasileiro *TodaOferta*, uma vez que a funcionalidade de busca é uma das mais populares e importantes. A partir da análise de sessões que possuem busca e das que não as contém, observou-se uma superioridade em quantidades do primeiro conjunto, bem como uma maior eficiência na compra de produtos nesse mercado. Os resultados alcançados podem ser úteis para o provedor e usuários do sistema, possibilitando criar serviços personalizados e aperfeiçoar o uso do sistema de busca.

## Keywords

Electronic Markets, E-commerce, Search, Navigation, Web

## 1. INTRODUÇÃO

Nos últimos anos, percebe-se um aumento da popularidade do Comércio eletrônico. Este fenômeno pode ser visto através de mercados eletrônicos (*e-markets*) conhecidos mundialmente, como o *eBay* e *Amazon.com*. Além disso, com o surgimento de um novo conceito, chamado *Web 2.0*, o usuário muda a sua forma de interação com a *Web*, e passa a ser não apenas um espectador, mas também um provedor de informações. Dessa forma, uma grande quantidade de conteúdo está sendo criada e trocada a cada dia, abrindo inúmeras oportunidades de pesquisas nesse novo cenário.

Um dos desafios que surgem com essa mudança é a compreensão da interação entre usuário e o sistema, como, por exemplo, no contexto de mercados eletrônicos. Dentre as informações disponíveis, temos o processo de negociação, que resulta ou não em uma transação, o processo de busca por um produto, etc. Entender o comportamento de usuários diante de uma ferramenta de compra e venda online pode contribuir para uma melhora de eficiência desse sistema, tanto para o seu provedor, quanto para quem o utiliza.

A partir da navegação de usuários em mercados eletrônicos, é possível concentrarmos em algumas funcionalidades de

grande importância, como o processo de busca de produtos. O sistema de busca é um dos mecanismos mais importantes e utilizados nesse ambiente. Os resultados apresentados por esse sistema de busca influenciam diretamente o processo de negociação de produtos. Por isso, é de vital importância compreender o funcionamento da busca em um mercado eletrônico, assim como o seu impacto nas transações.

Neste trabalho realizamos uma análise do sistema de busca em mercados eletrônicos, mais especificamente no *TodaOferta*<sup>1</sup>. Pretende-se realizar uma avaliação detalhada acerca da influência da utilização do sistema de busca sobre as negociações de produtos online. Os resultados dessa pesquisa podem ser úteis tanto para o provedor, quanto para os usuários. Com o entendimento do processo de busca, será possível propor modificações com o intuito de melhorar a eficiência dessa ferramenta, além da criação de serviços de busca personalizados e geração de cargas de trabalho sintética, que podem ser utilizados para a análise de desempenho.

## 2. TRABALHOS RELACIONADOS

Mercados eletrônicos têm apresentado um grande crescimento a cada dia, tanto em volume de transações quanto financeiro. A opção de *Compre-já* (*Buy-it-now*), que permite dar um preço fixo para encerrar o leilão, tem alcançado grande popularidade nos mercados eletrônicos. Durham [7] e Anderson [1] mostraram que vendedores experientes utilizam frequentemente o *Compre-já*, e que esta opção é mais usada quando utilizada por vendedores com alta reputação. No Brasil, esse tipo de negociação direta é muito popular.

A caracterização de carga de trabalho é fundamental para o entendimento e aprimoramento de sistemas *Web*. Há vários estudos que apresentam caracterizações de carga de trabalho de diferentes tipos, tais como servidores *Web* [3], de comércio eletrônico [8, 9], de blogs [6], de vídeo sob demanda [5] e de vídeo ao vivo [10]. Dentre as várias contribuições desses trabalhos, destacamos a criação de valiosos modelos capazes de descrever a carga que chega nesses servidores, essenciais para a geração de carga sintética que, por sua vez, possibilita a realização de experimentação e simulação baseadas em distribuições mais realistas.

Em particular, [5] analisa as requisições de vídeos em um contexto educacional. Mostrou-se que o tempo entre requisições segue uma distribuição Pareto e que a popularidade de objetos de multimídia pode ser modelada pela concatenação de duas distribuições do tipo *Zipf*.

O trabalho [4] apresenta uma análise da carga de trabalho de um serviço de vídeos do UOL, utilizando sessões e requisições ao servidor, além de uma avaliação do perfil de nave-

<sup>1</sup><http://www.todaoferta.com.br>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

gação dos usuários. Foram identificados diferentes perfis de usuários que acessam o sistema, o que pode ser utilizado para prover serviços personalizados aos usuários.

Não foram encontrados trabalhos como este sobre mercados eletrônicos brasileiros, uma vez que este tipo de pesquisa não é muito comum em sistemas nacionais, como o *TodaOferta*, o que reforça a relevância de pesquisas como esta.

### 3. SISTEMA DE BUSCA

#### 3.1 Estudo de Caso

Esta seção faz uma descrição geral dos dados do *TodaOferta*, pertencente ao Universo Online (UOL)<sup>2</sup>. Algumas informações e estatísticas são apresentadas na tabela 1. Devido à confidencialidade desses dados, é importante ressaltar que informações quantitativas não poderão ser apresentadas.

Período da amostra	01/04/2009 a 30/06/2009
Número de categorias de produtos	32
Número de sub-categorias de produtos	2.189
Tipos de negociação	Preço fixo e Leilão

Tabela 1: *TodaOferta* - Informações Gerais

#### 3.2 Caracterização Geral

Esta seção apresenta uma caracterização geral da carga de trabalho (requisições) do mercado em estudo. Inicialmente é importante distinguir os dois tipos de requisições que estão discriminados nos resultados apresentados a seguir. São eles:

- **Humano:** acessos de um usuário comum, que representa um potencial comprador.
- **Robô (ou Bot):** programa ou *script* que possui como objetivo simular ações humanas para inúmeras funcionalidades, dentre elas, a coleta de dados.

No gráfico exposto, serão utilizadas porcentagens relativas. Além disso, a porcentagem de requisições de robôs é de aproximadamente 70% da base utilizada.

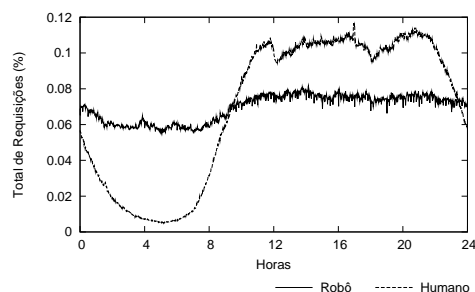


Figura 1: Requisições por minuto

Na figura 1, temos a distribuição de requisições para cada minuto do dia. Como esperado, há uma grande queda nas requisições por parte dos humanos durante o período da madrugada. A partir das 10 horas da manhã, suas requisições já alcançam um patamar maior, estabilizando em determinado valor por boa parte do dia. No final da noite, por volta de 22 horas, as requisições começam a cair novamente. Já os robôs possuem um comportamento diferente e mais estável. Com uma pequena diminuição durante a madrugada, os valores permanecem relativamente constantes durante o dia.

<sup>2</sup><http://www.uol.com.br>

### 3.3 Caracterização Funcional

Nesta seção, são apresentadas as funcionalidades referentes às requisições de usuários no mercado *TodaOferta*. Ao todo, existem 9 diferentes funções, descritas na tabela 2.

Funcionalidade	Definição
Busca	Busca por palavra chave
Busca por Usuário	Listar produtos de um usuário
Página Inicial	Página principal do <i>TodaOferta</i>
Página Categoria	Listar de produtos de uma categoria
Comprar	Processo para compra de um produto
Confirmação da Compra	Finalização de uma compra
Detalhar Anúncio	Informações de um anúncio
Listar Compradores	Listar compradores de um produto
Perfil Usuário	Detalhar usuário

Tabela 2: Funcionalidades do *TodaOferta*

É importante ressaltar que não estão sendo consideradas as funcionalidades referentes a ações de administração e de cadastramento de ofertas no mercado eletrônico.

#### 3.4 Avaliação

Nesta seção, apresentamos uma análise focada no sistema de busca do mercado em questão. Devido a limitação de espaço, os resultados expostos a seguir referem-se ao mês de *abril* de 2009, que conta com milhões de requisições. Inicialmente, é importante definir o conceito de sessão de usuário.

Uma sessão de um usuário é definida como uma série de requisições realizadas a um sítio na *Web* durante um determinado período de tempo [2]. Assim, duas requisições consecutivas pertencem à mesma sessão se o tempo entre elas é menor do que um tempo limite de inatividade, denominado tempo de expiração da sessão. No caso do mercado eletrônico *TodaOferta*, o tempo de sessão é de 17 minutos.

Assim, realizou-se uma avaliação das sessões de usuários humanos focando no sistema de busca. É importante ressaltar que este sistema não se resume a apenas à busca interna do *TodaOferta*, mas também considera inúmeros tipos diferentes, como: busca orgânica, busca patrocinada (*Google AdWords*), e até mesmo busca nas categorias e por vendedores do mercado. Assim, separou-se o conjunto de dados em dois grupos de sessões: aquelas que possuem algum tipo de requisição com característica de busca e o restante.

Inicialmente, já percebe uma superioridade das sessões com busca, uma vez que estas correspondem a 90,02% do total de requisições. Este mesmo comportamento se repete ao considerar a quantidade de sessões, onde as requisições de busca estão presentes em 78,38% do total das sessões.

Outra métrica importante é o tamanho de sessão, que corresponde à quantidade de requisições presente em uma sessão. A média deste valor em sessões com busca é de 3,96 (desvio padrão  $d=28,12$ ), enquanto para o outro conjunto de sessões é de apenas 1,59 ( $d=7,29$ ), menos que a metade do primeiro grupo. Porém, acredita-se que sessões com tamanho 1 agregam pouca informação para a avaliação do mercado. Assim, desconsiderando estes casos, temos valores diferentes a mais reais de médias e desvios, apesar das buscas ainda possuírem sessões com tamanhos maiores. Seu valor chega a 6,48 ( $d=38,17$ ), enquanto sessões sem busca possuem uma média de 3,16 ( $d=13,91$ ). Além da média dos tamanhos, o gráfico a seguir nos permite visualizar melhor a variação dos tamanhos de sessão do mercado.

A partir da figura 2, observa-se uma superioridade de sessões com busca, apresentando ocorrências maiores do que sessões sem busca. Além disso, em ambos os conjuntos,

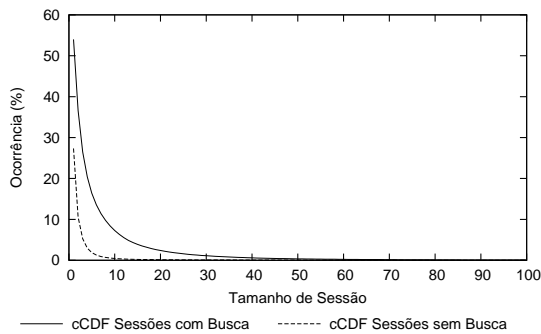


Figura 2: Tamanhos de Sessão

percebe-se um comportamento semelhante com relação à distribuição de ocorrências. Seguindo a Lei de Zipf, temos uma grande concentração de ocorrências de tamanhos de sessão pequenos (menores que 10), enquanto sessões maiores ocorrem pouco no mercado.

Observando mais detalhadamente as sessões, assim como as funcionalidades da tabela 2, acredita-se que três funções merecem mais atenção, uma vez que traduzem, de certa forma, a eficiência do sistema de busca. Estas funções são: *Detalhe Anúncio*, *Comprar* e *Confirmação da Compra*. As análises a seguir são referentes a sessões que contém ou não requisições com essas funcionalidades.

Sobre o detalhamento do anúncio, dentre as sessões com busca, 71,82% das sessões chegam a este estado, contra 82,46% do outro conjunto. Isto pode ser explicado pelo fato das sessões com busca apresentarem um refinamento maior do que o usuário almeja, resultando numa visualização mais objetiva dos anúncios. Sobre o processo de compra, realizando uma análise comparativa entre os dois conjuntos, percebe-se que a quantidade de sessões com busca que possuem esta função é quase o dobro das que não possuem busca. Este comportamento é mais acentuado ao observarmos a confirmação de uma compra, com um valor cinco vezes maior das sessões com busca. Por fim, observa-se que 17,53% das sessões com busca que iniciam o processo de compra chegam a finalizá-lo, enquanto no grupo de sessões sem busca este valor é de apenas 6,41%.

#### 4. CONCLUSÃO E TRABALHOS FUTUROS

Com a popularização do comércio eletrônico através de mercados como *eBay* e *Amazon.com*, e com o usuário passando a ser um provedor de informação (fenômeno chamado de *Web 2.0*), existem uma grande quantidade de informação sendo criada e difundida a cada dia. Assim, no cenário de mercados eletrônicos, temos informações de negociações, de compradores e vendedores, e busca. Neste trabalho, focamos no sistema de busca utilizando como estudo de caso o mercado eletrônico brasileiro *TodaOferta*, pertencente ao UOL.

Distinguindo-se requisições de usuário humanos e robôs, percebe-se uma distinção no comportamento de ambos os tipos através da distribuição destas requisições ao longo do dia. Além disso, foi possível identificar inúmeras funcionalidades atribuídas a cada requisição, permitindo realizar uma análise mais semântica da carga de trabalho.

Com um foco nas requisições de usuários humanos, separou-se as sessões em conjuntos que possuem características de busca (busca no site, busca orgânica, etc.) e que não possuem. Dessa forma, observou-se uma superioridade tanto

em questão de quantidade, quanto em questão de tamanho das sessões com busca. Além disso, com o foco em algumas funcionalidades das requisições, percebeu-se que, apesar do detalhamento de anúncios ser menor, sessões com busca alcançam uma maior quantidade de compras confirmadas do que sessões sem busca, demonstrando assim a sua importância para o mercado eletrônico.

Através deste trabalho, podemos perceber o quão importante é o sistema de busca em um mercado eletrônico, o que justifica a realização de pesquisas como esta. Como trabalhos futuros, espera-se aprofundar mais nos dois conjuntos de sessões citados, com um detalhamento maior para identificarmos o seu impacto no mercado eletrônico em questão. Os resultados desta pesquisa são importantes para o provedor e usuários do sistema, além de se possível propormos uma melhora desta ferramenta. Os resultados também podem ser utilizados para geração de cargas de trabalhos sintéticas.

#### 5. AGRADECIMENTOS

Esta pesquisa é parcialmente patrocinada por Universo OnLine S. A. - UOL ([www.uol.com.br](http://www.uol.com.br)) e parcialmente apoiada pelo Instituto Nacional de Ciência e Tecnologia para a Web (CNPq no. 573871/2008-6), CAPES, CNPq, Finep e Fapemig.

#### 6. REFERÊNCIAS

- [1] S. Anderson, D. Friedman, G. Milam, and N. Singh. Buy it now: A hybrid internet market institution. *Industrial Org.* 0412003, EconWPA, Dec. 2004.
- [2] M. Arlitt. Characterizing web user sessions. *SIGMETRICS Performance Evaluation Review*, 28(2):50–63, 2000.
- [3] M. Arlitt and C. Williamson. Web server workload characterization: the search for invariants. *SIGMETRICS Performance Evaluation Review*, 24(1):126–137, 1996.
- [4] F. Benevenuto, A. Pereira, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Avaliação do perfil de acesso e navegação de usuários em ambientes web de compartilhamento de vídeos. In *WebMedia'09*, Fortaleza, Brasil, 2009. SBC.
- [5] C. Costa, I. Cunha, A. Vieira, C. Ramos, M. Rocha, J. Almeida, and B. Ribeiro-Neto. Analyzing client interactivity in streaming media. In *WWW*, 2004.
- [6] F. Duarte, B. Mattos, A. Bestavros, V. Almeida, and J. Almeida. Traffic characteristics and communication patterns in blogosphere. In *Proc. Int'l Conference on Weblogs and Social Media (ICWSM)*, 2007.
- [7] M. R. Durham, Y. and S. Standifird. ebay's buy-it-now function: Who, when and how. *Topics in Economic Analysis & Policy*, 4(1):28, 2004.
- [8] D. Menasce and V. Almeida. *Scaling for E Business: Technologies, Models, Performance, and Capacity Planning*. Prentice Hall PTR, NJ, USA, 2000.
- [9] A. Pereira, L. Silva, and W. Meira, Jr. Evaluating the impact of reactive workloads on the performance of web applications. In *Proc. of the 25th IEEE Int'l Performance, Computing, and Communications Conf. (IPCCC)*, Phoenix, USA, 2006. IEEE CS.
- [10] E. Veloso, V. Almeida, W. M. Jr., A. Bestavros, and S. Jin. A hierarchical characterization of a live streaming media workload. *ACM Trans. on Network*, 14(1), 2006.