

PESQUISAS NA WEB: ESTRATÉGIAS DE BUSCA

Elias Estevão Goulart
Annibal Hetem Júnior

Resumo

A *World Wide Web* tem sido utilizada amplamente para a busca e seleção de informações, resultando em um de seus principais empregos como suporte para atividades acadêmicas e profissionais. Este trabalho apresenta um estudo sobre as estratégias de busca de informações na *World Wide Web*, visando analisar e comparar os resultados de uma pesquisa exploratória com estudo similar realizado na Universidade de Telaviv. Apresenta-se nove formas possíveis de buscas e como elas foram utilizadas nos estudos comparados. Como resultado, são apresentadas as mais efetivas e sugere-se melhor treinamento dos usuários para o conhecimento das técnicas apresentadas.

Palavras-chave

Estratégias de busca; Internet; *World wide web*

SEARCHING ON THE WEB: SEARCH STRATEGIES

Abstract

The World Wide Web has been largely used for searching and selecting information, and is one of the most important tools to support academic and professional activities. This work presents a study about information search strategies on the world wide web, seeking to analyze and compare the results of a similar exploratory research implemented at Telaviv University. It presents nine possible ways of information search and how they were compared in both studies. As a result, the most effective of the strategies are presented and users training are suggested as the best way to make them aware of the discussed techniques.

Keywords

Search strategies; Internet; *World wide web*

1 INTRODUÇÃO

A *Web* (mais precisamente a *World Wide Web*) tem crescido de forma assustadora desde seu início em meados dos anos 90, contando atualmente com cerca de 10 bilhões de páginas (NACHMIAS, 2002).

Ela auxilia nas atividades profissionais, comerciais, educacionais e culturais de mais de 300 milhões de “internautas” no mundo todo que realizam mais de 250.000.000 acessos diários aos conteúdos *online* (NETRATINGS, 2005).

Essa imensa massa de informações traz consigo uma dificuldade inerente que é a busca por informações específicas. Por maior e mais completo que este conjunto de informações possa ser, ele será inútil caso não se consiga encontrar o que se precisa.

Inúmeras ferramentas de busca estão disponíveis, mas sua utilização não é trivial, pois estudos têm demonstrado a frustração dos usuários na pesquisa por informações, mencionando índices de até 70% (NACHMIAS, 2002).

Outros estudos mostram que apenas 60% dos usuários encontram “na maioria das vezes” o que buscam, enquanto 2,6% nunca encontram, contra 21% que declaram sempre encontrar com sucesso suas buscas (DENNIS, 2002). Para estes pesquisadores “um modo mais eficiente de pesquisa deve ser encontrado”.

A *Web* tem participado efetivamente das atividades rotineiras de milhões de usuários dos sistemas computacionais e sua aplicação aos locais de trabalho tem sido indispensável em muitos casos, principalmente nos meios educacionais. Estas atividades podem ser prejudicadas quando se desperdiça horas em buscas ineficientes, portanto análises devem ser empreendidas para se compreender melhor os processos de pesquisa implementados pelos usuários.

Esta temática não é nova, tendo sido foco de estudos na área da biblioteconomia (NACHMIAS, 2002), contudo com o advento da informatização e da *Web* novas formas de pesquisa e estratégias de busca devem ser implementadas.

Os conteúdos de informação nas páginas da Internet variam, de acordo com a diversidade temática das respectivas bases de dados oferecidas em *sites*. Algumas são orientadas para um determinado assunto, enquanto outras são orientadas para a missão da instituição que as desenvolvem, como, por exemplo, a base da agência americana sobre o meio ambiente *Environmental Protection Agency* (EPA), que não se limita à indexação de documentos apenas em química, toxicidade ou saúde pública, porém amplia suas áreas temáticas para outros aspectos relacionados com sua missão de controle e regulamentação do meio ambiente. Esses e outros fatores devem ser levados em consideração no planejamento das estratégias de busca, assegurando a qualidade e eficácia na recuperação da informação.

O entendimento dos processos de busca são primordiais para a melhoria da efetividade das pesquisas, pois o tempo consumido com estas atividades chega a 70% do total de acesso à Internet, com cerca de 5 ou mais horas semanais (SULLIVAN, 2000).

Uma forma de melhorar o entendimento sobre o processo de busca é estudar o comportamento do pesquisador, analisando as habilidades e condições necessárias para uma busca de sucesso.

Vários estudos teóricos foram criados para a recuperação da informação como uma complexidade cognitiva. Allen (1991) propôs quatro tipos de conhecimento que afetam os processos de pesquisa; (a) conhecimento geral; (b) conhecimento do sistema; (c) conhecimento da tarefa de pesquisa e (d) conhecimento do domínio.

Hannafin e Hill (1997) empregando uma abordagem mais ampla encontraram cinco tipos de conhecimentos necessários:

- 1) conhecimento metacognitivo: consciência do processo cognitivo necessário

- 2) orientação percebida dentro do sistema
- 3) julgamento de capacidade para a tomada de ações
- 4) conhecimento do sistema
- 5) conhecimento acerca do conteúdo sendo pesquisado

As formas de busca indicam estratégias empregadas pelos usuários, sendo definidas aqui como um plano contemplando uma série de ações visando encontrar uma informação. Como exemplo, uma simples estratégia de busca seria a utilização de um site de buscas (*Google, Yahoo!*, ...) onde se digita o termo “Mona Lisa” e se recebe uma listagem das páginas cadastradas e que contém algum relacionamento com aquele termo. Continua se seguindo para alguma página recebida e assim por diante, até encontrar o que procura ou desistir.

Deve-se considerar ainda os *sites* de busca dirigidos que agregam *links* para páginas de uma temática especial (por exemplo <http://www.dianakacker.com/resdoc> voltada para as áreas de humanidades e ciências sociais. Estes *sites* também são conhecidos como portais específicos.

Este trabalho procura elaborar uma releitura do estudo realizado por Nachmias (2002) por meio da realização de um experimento similar com vistas a uma análise comparativa.

No referido estudo foram apresentadas nove formas de se conduzir uma busca, a saber:

- 1) Busca por Palavra-Chave:

Implementado pela digitação direta do assunto da pesquisa. Por exemplo, digitar as palavras “Mona Lisa”.

- 2) Busca por Definições Amplas:

Realizada por meio de uma pesquisa ampla, mais genérica. Por exemplo, procurar por arte ou pintura para encontrar a Mona Lisa.

3) Busca Complexa:

Obtida pelo emprego de um cruzamento de palavras e indica uma abordagem mais sofisticada de busca. Por exemplo, buscar por “Louvre”, “Mona Lisa” e “Pintura”, simultaneamente.

4) Uso de Conhecimento Geral:

Implementada pela utilização de conhecimentos associados ao assunto de interesse, porém não mencionados no problema em foco. Por exemplo, procurar por Leonardo da Vinci para encontrar uma imagem da pintura da Mona Lisa.

5) Convenção de Computador:

Emprego de convenções particulares do vocabulário computacional. Por exemplo, a busca pelos sufixos dos arquivos digitais (.gif, .jpg, ...).

6) Busca Booleana:

Esta forma emprega elementos da lógica booleana, os conectores E (AND), OU (OR), dentre outros, para formar termos de busca mais complexos. Por exemplo, “Louvre” and “Mona Lisa”.

7) Emprego de Diretório:

Utiliza a busca em diretórios ou catálogos de tópicos como o do *Yahoo!*, por exemplo.

8) Acesso a Portal Específico:

Faz uso de portais específicos, previamente conhecidos, para encontrar o assunto desejado. Por exemplo, para encontrar uma página com a imagem de Mona Lisa, procurar em www.artnews.com.

9) Digitação Direta:

A busca é realizada simplesmente pela digitação de um link específico, por exemplo, www.monalisa.com.

Este trabalho visa, como objetivo, analisar as estratégias de busca de informações na *Web*, implementando uma pesquisa exploratória comparativa.

2 METODOLOGIA

O estudo foi realizado com 49 estudantes integrantes de uma turma noturna do 3º ano de graduação do curso de Licenciatura em Computação do Centro Universitário Fundação Santo André, município de Santo André no estado de São Paulo.

A metodologia empregada foi similar à realizada no trabalho publicado por Nachmias (2002), constando da proposição aos alunos de 3 tarefas de busca na *World Wide Web* – WWW, cada uma com um tema específico para serem completadas em 45 minutos máximos. Os temas propostos foram:

- a) Página com o poema Descobrimento (de Mario de Andrade);
- b) Página com foto e histórico do Cristo na cidade do Rio de Janeiro;
- c) Página do jogador Manuel Francisco do Santos com Pelé e biografia resumida.

Os temas foram escolhidos contendo a língua portuguesa como referência para não inserir dificuldade adicional aos estudantes e além do título, nenhuma informação adicional foi fornecida aos alunos (p.ex. Garrincha, ou Cristo Redentor).

A pesquisa foi realizada nas dependências dos laboratórios de informática do Centro Universitário, no segundo semestre do ano de 2005. Os microcomputadores utilizados eram compatíveis com a linha PC, conectados em rede e ligados a servidor de acesso com *proxy*.

O software empregado para a coleta das navegações dos alunos foi o SurfSpy (2002), em sua versão de demonstração, plenamente apta para o experimento. O programa registrou as informações sobre a data/hora, usuário, número da máquina, número IP do computador e *links* visitados em cada estação. A versão demo do software pode ser obtida gratuitamente na página Internet <http://www.tucows.com>.

Antes de iniciar as buscas, os alunos preencheram um formulário para a designação de seu perfil sócio-econômico contendo nome, idade, sexo, informações sobre trabalho e experiência anterior com a Internet, e receberam as orientações sobre o procedimento para a realização das tarefas, que constaram do registro em papel, junto ao mesmo formulário, da identificação da máquina que operavam, bem como dos sites encontrados para as tarefas solicitadas.

3 RESULTADOS:

3.1 Perfil dos participantes

O estudo foi realizado com 49 participantes, com faixa etária entre 19 e 35 anos (média de 23,6 anos), sendo que 45 deles trabalham durante o dia (91,8%) e 66% utilizam a Internet no trabalho e 95,9% do total possuem acesso à Internet em casa. A pesquisa indicou que a média do grupo foi de 6,6 anos de utilização da Internet, com 1,8 ano de desvio padrão que indica a dispersão das medidas (em torno de 67% dos participantes utilizavam entre 4,8 e 8,4 anos), com máximo de 10 e mínimo de 3 anos.

A figura 1 indica a origem do primeiro contato com a Internet pelos participantes, onde se pode observar que a maioria (51%) teve contato em casa.

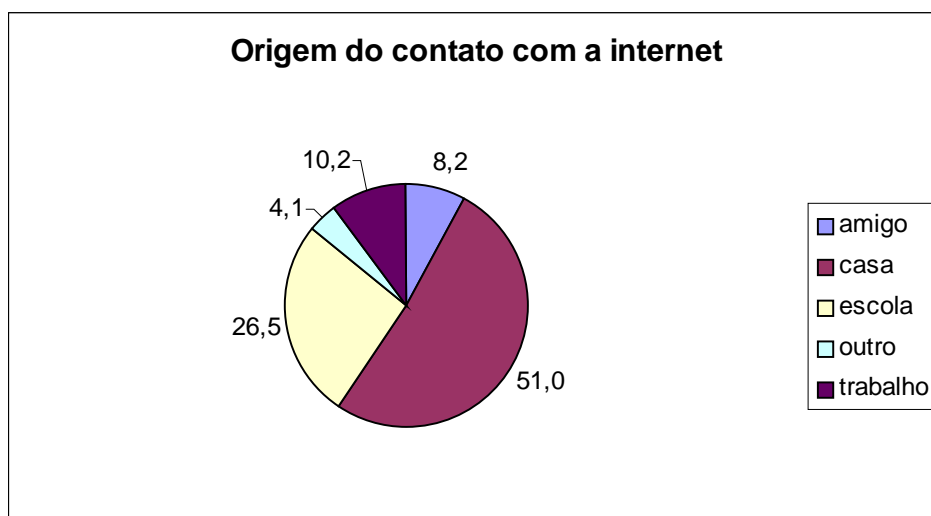


Figura 1: Distribuição percentual do primeiro contato com a Internet.

3.2 Caracterização da buscas

A tabela 1 indica as características gerais das buscas realizadas pelos participantes, onde se observa que a média do tempo das atividades de busca foi de 16,6 minutos para pesquisar 39,6 páginas em média, resultando em 2,4 minutos por página. A pesquisa completa mais longa tomou 37 minutos com a navegação em 101 páginas para a realização das 3 tarefas.

Tabela 01 – Características das buscas realizadas pelos estudantes. Fundação Santo André, 2005.

	Minutos	Páginas	Págs/Min.
Média	16,6	39,6	2,40
Desvio Padrão	7,3	20,4	0,58
Máximo	37	101	3,4
Mínimo	3	10	0,89

A pesquisa conteve 147 tarefas (49 participantes com 3 tarefas cada) com uma média de 11,29 páginas pesquisadas (desvio padrão de 10), com máximo de 47 páginas e mínimo de 2 por tarefa.

A figura 2 mostra as tarefas completadas no total de 120 tarefas (81,6%). Das 27 tarefas incompletas (18,4%), 16 participantes não realizaram 1 tarefa (59,3%), 4 participantes deixaram de completar 2 tarefas (29,6%) e 1 participante não completou quaisquer das 3 tarefas solicitadas (11,1%).

3.3 Estratégias empregadas

Em relação às estratégias empregadas que resultaram em completar as tarefas, a figura 3 mostra a distribuição das buscas empregadas. Constatou-se que a Busca Complexa foi responsável por 66,7% das tarefas concluídas com êxito, estando em segundo lugar no sucesso a Busca por Palavra-Chave (17,5%). As demais estratégias de sucesso foram o Conhecimento Geral (14,2%) e o Uso de Diretório e o Portal Específico com 0,8% do sucesso nas buscas.

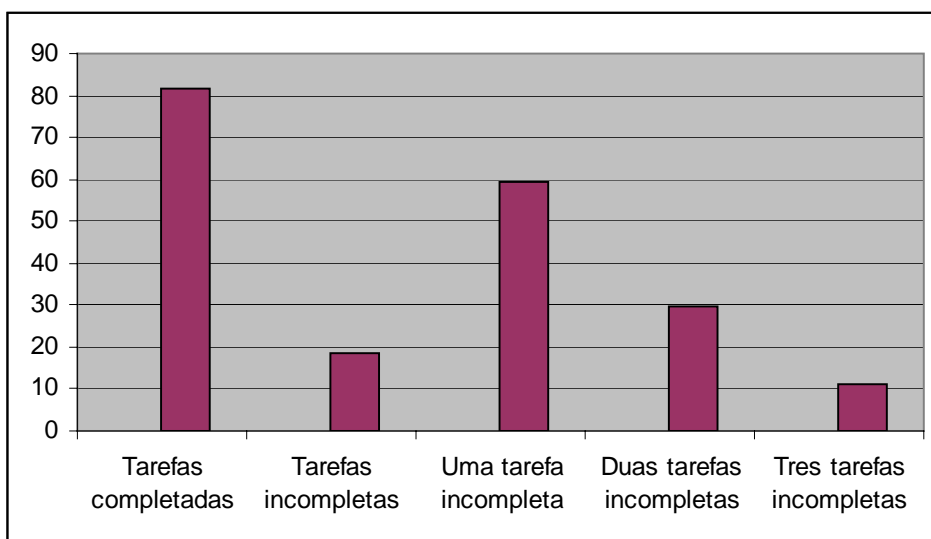


Figura 2 – Distribuição percentual das tarefas pelos estudantes. Fundação Santo André, 2005.

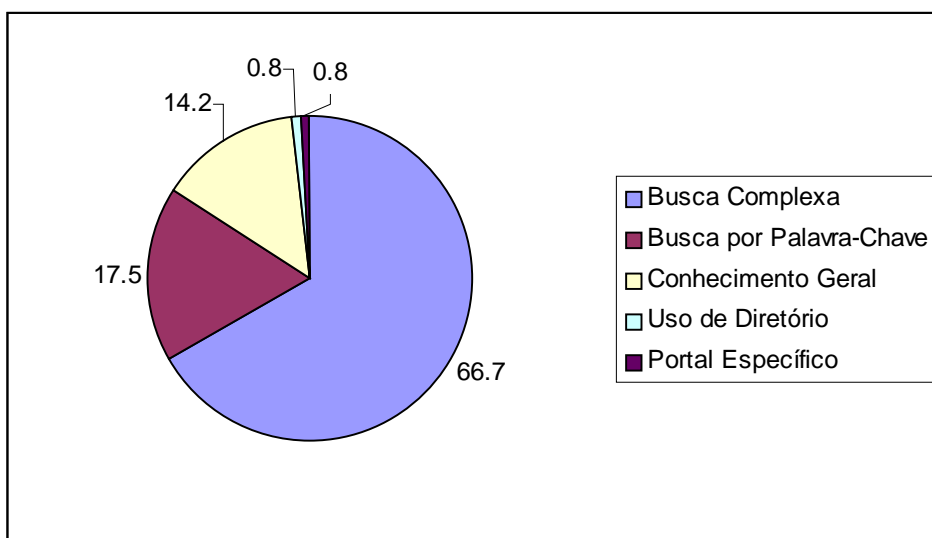


Figura 3 – Distribuição percentual das estratégias de busca empregadas. Fundação Santo André, 2005.

4 DISCUSSÃO

Os resultados mostram que a busca de informações na rede Internet é uma tarefa não-trivial, cuja complexidade pode exigir o emprego de várias estratégias.

A complexidade da *Web*, sua vasta coleção de assuntos, a imensa quantidade de relacionamentos entre as informações, obtidas pelos mecanismos da hipermídia, fornecem aos resultados das buscas alta variabilidade, ou seja, é apresentada uma grande coletânea de *links* a páginas com possíveis associações à busca realizada. Por exemplo, a busca pela expressão ‘Cristo Redentor’ produziu mais de 1.000.000 de *links*.

No estudo anterior (NACHMIAS 2002) obteve-se um tempo médio de 12 minutos por tarefa, sendo que neste estudo este tempo foi de 16,6 minutos. Dentre os motivos para esta diferença podem estar o melhor preparo dos estudantes (eles eram de curso de pós-graduação), maior velocidade nos serviços de conexão à rede Internet, ou pelo fato de as pesquisas deste estudo terem sido realizadas com *sites* em português, ou seja, em menor quantidade de páginas sobre os assuntos pesquisados.

De forma similar, ambos os estudos demonstraram alta variabilidade no tempo de pesquisa, sendo que se obteve aqui de 3 a 37 minutos, enquanto no estudo anterior foi de 1 a 60 minutos. Uma justificativa possível pode ser a grande quantidade de informações, pois algumas buscas resultam em milhares de *links* possíveis, causando desmotivação e senso de desorientação nos usuários. Isto permite o questionamento de se ter efetividade no emprego da Internet como principal fonte de estudos e pesquisas.

Por outro lado, diferentemente do estudo anterior, obteve-se a conclusão de 81,6% das tarefas propostas, enquanto que naquele estudo foi de 60,3%. Ainda, a estratégia mais utilizada no estudo anterior foi o Uso de Conhecimento Geral, enquanto que neste estudo encontrou-se a Busca Complexa como a mais empregada.

O que motiva a escolha de um método de busca ou outro não faz parte deste estudo, mas está ligado às raízes culturais de cada indivíduo e a aspectos psicológicos. Enquanto as opções iniciais de busca podem ser instintivas, ou quase automáticas para um indivíduo experiente na ferramenta, o ponto crítico a ser lembrado é que há a possibilidade de satisfazer a busca de várias maneiras. A capacidade de resolver o problema por uma pessoa não treinada é via de regra limitada a um padrão de comportamento estereotipado individual. As opções iniciais servem apenas para ar o primeiro passo em direção ao que se deseja. Mas esta, por sua vez, não é predeterminada pelos conhecimentos coletivos e nem se caracteriza sempre numa determinada ação particular.

De forma similar, ambos os estudos mostraram que os usuários tendem a empregar duas ou três estratégias diferentes em suas pesquisas, pois parece que o ato de busca pode levar os pesquisadores em diferentes direções, dependendo dos *links* acionados e das páginas visitadas em cada passo da pesquisa. Ainda, como resultado similar, as buscas de sucesso tomaram, aproximadamente, metade do tempo e de passos quando comparada com aquelas onde não se obteve sucesso.

5 CONCLUSÕES

A partir deste estudo pode-se concluir que, pelo comportamento dos participantes, o emprego adequado das estratégias de busca parece ser a principal fonte dos sucessos nas tarefas de busca, bem como em sua eficiência.

A *Web* é interpretada como um sistema de recuperação de informações, uma ferramenta para organizar as informações, uma coletânea de livros e um canal de comunicação. A Internet já faz parte da vida das pessoas e não é apenas um recurso isolado. Rieh (2004) sugere que a tendência mais significativa é a “domesticação” da Internet, cujo emprego passou do serviço para o lar com uso mais prazeroso do que econômico.

Estas considerações permitem avaliar a importância das estratégias de busca para a realização de tarefas pessoais, sendo que os usuários necessitam de auxílio na busca para seu uso eficiente.

Melhor avaliação do comportamento e das habilidades dos usuários é requerida para se construir mecanismos de buscas mais eficientes, e não apenas focalizar os esforços nos algoritmos de pesquisa, como menciona Nachmias (2002).

Ainda, de igual forma, conclui-se que é necessário uma nova maneira de se mapear os conhecimentos disponíveis na *Web* para que as buscas possam ter maior sucesso. O emprego dos diretórios ou catálogos pode ter sua eficiência significativamente aumentada.

Finalmente, a educação dos usuários é fator decisivo nos processos de busca, pois quanto melhor é entendido o processo de busca, o funcionamento dos mecanismos de busca, as estratégias mais eficazes, melhores podem ser os resultados obtidos.

6 REFERÊNCIAS BIBLIOGRÁFICAS:

ALLEN, B. (1991). **Cognitive research in science: implications for design**. Annual review of information science and technology.

NACHMIAS, R.; GILAD, A. (2002). **Needle in a hyperstack: searching for information on the World Wide Web.** Journal of Research on Technology in Education, vol. 34, n° 4, p.475-486.

DENNIS, S.; et al (2002). **Web searching: a process-oriented experimental study of three interactive search paradigms.** Journal of the American Society for Information Science and Technology, vol. 53, n° 2; p.120-133.

HANNAFIN, M.; HILL, J. (1997). **Cognitive strategies and learning from the World Wide Web,** vol. 47, n° 4.

KIM, Kyung-Sun; ALLEN, B. (2002). **Cognitive and task influences on Web searching behavior.** Journal of the American Society for Information Science and Technology, vol. 53, n° 2; p.109-119.

RIEH, S. Y. (2004). **On the Web at home: information seeking and Web searching in the home environment.** Journal of the American Society for Information Science and Technology, vol. 53, n° 8; p.743-753.

NETRATINGS (2005). Disponível em http://www.nielsen-netratings.com/news.jsp?section=dat_gi. Acesso em 01/11/2005.

SULLIVAN, D. (2005). **Survey reveals search habits.** Disponível: <http://searchenginewatch.com/sereport/00/06-realnames.html>. Acessado em 10/11/2005.

Elias Estevão Goulart

Professor doutor e pesquisador da Universidade Municipal de São Caetano do Sul - SP.

eegoulart@yahoo.com.br

Annibal Hetem Júnior

Professor doutor e pesquisador da Fundação Santo André.

annibal@fsa.br