

A evolução da Web – Sistemas de busca

Izaquiel Queiroz ¹, Henrique Silva²

¹ Graduando SI na UFRPE-UAST.

E-mail: izaquielqueiroz@live.com

² Graduando SI na UFRPE-UAST.

E-mail: henrique.silva.eversom@gmail.com

Resumo: Com o avanço dos sistemas de buscas da internet foi possível a recuperação de informações na gigantesca rede mundial de internet. No início eram feitas manualmente a catalogação dessas informações para a melhor organização e utilização pelos cientistas e também por usuários. Com o passar do tempo essa catalogação não acompanhava mais o crescimento contínuo da rede, surgindo assim os robôs e sistemas buscadores de informações.

Abstract: With the advancement of the internet search systems was possible to recover information in the gigantic world wide internet. At first they were made manually cataloging such information for better organization and use by scientists and also by users. Over time this cataloging did not follow more the continued network growth, thus resulting robots and seekers of information systems.

Os sistemas de buscas – evolução em 20 anos.

Rodrigo e Sandra (2008) citam um texto de Aaron Wall, onde em 1945 Vannaver Bush publicou na revista The Atlantic Monthly um artigo que levava questões sobre o desenvolvimento de um mecanismo que reunisse o conhecimento humano, de forma organizada e em larga escala, para que este pudesse ser disponibilizado de forma rápida e eficiente. Mas é claro que os recursos da época não tornariam isso possível, mas a evolução da computação viabilizou em parte a realização dessas ideias.

Segundo Rodrigo e Sandra(2008) o processo de busca de informações na internet é quase tão antigo quanto a própria Internet[...] e a popularização dos motores de busca está em consonância com a necessidade de se organizar o volume de informações disponíveis na web que cresce exponencialmente.

De acordo com Cláudio (2007), cabe às ferramentas de busca a tarefa de analisar e indexar os documentos existentes na Web e, posteriormente,

armazená-los em uma base de dados de modo que a cada pesquisa solicitada pelo usuário, suas necessidades sejam atendidas.

Segundo Regina (2008), os catálogos ou diretórios por assunto precederam os programas de busca e constituíram a primeira tentativa de estruturar e recuperar recursos na Web. Estes foram criados quando era possível fazer manualmente a coleta das informações e catalogá-las de acordo com seu tipo de informação. Já os sistemas de buscas surgiram quando não era mais possível a coleta das informações manualmente na Web devido as proporções que esta atingia. Estas ferramentas criam seus bancos de dados automaticamente utilizando softwares conhecidos como spiders ou robôs.

De acordo com Rodrigo e Sandra (2008), os sistemas de busca derivaram da ideia dos pesquisadores que utilizavam a tecnologia criada por Berners Lee, a World Wide Web, reuniam suas páginas dicas de websites e/ou páginas preferidas. Rodrigo e Sandra (2008) também cita que à medida que a variedade de assuntos ia crescendo, alguns sites passaram a se dedicar exclusivamente à tarefa de catalogar outros sites, os diretórios. Como exemplo, o surgimento do Yahoo! Directory, que reunia uma quantidade de links que eram descritos manualmente. Deu-se início então à cobrança pela inclusão de sites comerciais. Mas como as pesquisas já não estavam suprimindo as necessidades dos usuários e a manutenção estarem ficando cada vez mais cara, por serem feitas por pessoas, surgem então os robôs, os sites puramente de pesquisa, a especialização de sites pesquisa e a web semântica.

Segundo Rodrigo e Sandra (2008), os robôs, também chamados de bots, crawlers ou spiders são programas que automatizam o processo de navegação e surgiram da dificuldade em avaliar manualmente uma grande quantidade de sites e taxonomias. Os robôs fazem o mesmo papel que uma pessoa faria manualmente na construção de um diretório em larga escala, só que de forma automatizada. Partem de um site inicial verificando e acessando todos os links disponíveis em cada página acessada. Com isso cada buscador dita suas regras para a indexação dos conteúdos encontrados na web. Um exemplo de Sandra, é quando o robô visita uma página, ele associa o endereço da mesma às palavras nela contidas, como um índice remissivo que contém todas as palavras do livro.

De acordo com Regina (2008), os robôs percorrem a rede iniciando em um site determinado pelo administrador coletando, analisando e indexando em um bando de dados informações, dentre elas, conteúdo, armazenando cada uma das palavras encontradas e seguindo seus links para outras páginas. Seguindo links para outros links e assim sucessivamente adicionando endereços e conteúdos ao banco de dados. E diferente dos catálogos ou diretórios, os sites são incluídos no bando de dados sem nenhuma classificação ou descrição de

seu conteúdo. Os robôs voltam aos sites em intervalos regulares para manter o sistema sempre atualizado.

Sobre a classificação da evolução da Web Cláudio (2007) aponta uma apresentação feita por Hoeschl(2006), onde há uma classificação quanto aos mecanismos de busca da Web. A primeira geração, relaciona-se com diretórios e catálogos (Yahoo e similares) que eram feitos manualmente por profissionais da informação. A segunda geração, surgem a automatização com os robôs digitais(Altavista). A terceira geração veio com os metabuscadores, juntando em um único resultado informações de vários motores de busca (MetaMiner). Na quarta geração os resultados ficam ainda mais refinados (All the Web). E a quinta geração corresponde a atualidade, tendo como grande exemplo o Google, utilizando diversas tecnologias, dentre as principais, a PageRank. A sexta geração compreende a junção de vários tipos d arquivos diferentes em um mesmo processo de busca (A9). Por fim, a sétima geração de ferramentas de busca traz consigo uma tecnologia baseada em ontobuscadores que mescla tecnologias inteligentes com conhecimentos milenares e filosóficos.

Segundo Rodrigo e Sandra (2008), PageRank são utilizadas em sites de busca para determinar o grau de autoridade de uma página. Sendo ranqueada as páginas mais consultadas, determinando assim sua importância.

Assim como Rodrigo e Sandra(2008), Cláudio(2007), cita que o PageRank ranqueia as páginas em importância levando em consideração a estruturas de links. Contabilizando quantas vezes determinado link é acessado ou quantas vezes ele é encontrado com outras páginas da Web.

Sobre os ontobuscadores, Cláudio (2007) cita uma tecnologia chamada de OntoWeb, que é um buscador inteligente baseado em ontologias e técnicas de inteligência artificial, capaz de "pensar" enquanto seleciona as informações. Seu grande diferencial é o uso de semânticas e estruturas valorativas para contextualizar as buscas e refinar resultados.

Referências

Goulart, Rodrigo, and Sandra Portella Montardo. "Os mecanismos de busca e suas implicações em comunicação e marketing." *Líbero* 11.21 (2008): p-119. Acessado dia: 06 de junho de 2015. Disponível em: <http://www.revistas.univerciencia.org/index.php/libero/article/viewArticle/5404>

Schons, Claudio Henrique. "O volume de informações na internet e sua desorganização: reflexões e perspectivas." *Informação & Informação* 12.1 (2007). Acessado dia 07 de junho de 2015. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewArticle/1748>



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO – UFRPE
UNIDADE ACADÊMICA DE SERRA TALHADA - UAST

Branski, Regina Meyer. "Recuperação de informações na Web." Perspectivas em ciência da informação 9.1 (2008). Acessado dia 08 de junho de 2015.

Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/351>