

Mecanismos de busca: conteúdo e relevância

Buscar, essa é uma das palavras que principia um dos meios pelos quais o ser humano se mantém até os dias atuais. O homem sempre necessitou buscar. Buscou entender sua origem, o sentido de sua existência, alimentos, moradia, lugares seguros, dentre outros. O ato de buscar está intrinsecamente ligado ao homem e por meio dele a humanidade se desenvolve, tanto os meios materiais como o conhecimento não são, e nunca serão, delegados a alguém sem que o mesmo não os procure. Porém, além da necessidade de buscar, outro fator relacionado e, recorrente dela, se faz presente, o qual recebe o nome de **resultado**. Todos nós quando necessitamos de algo primeiro o buscamos, mas além da busca o resultando é o desejado e, quanto mais rápido e preciso for, melhor nos satisfará.

No mundo virtual não é diferente, entretanto, encontrar determinado conteúdo na web não era uma tarefa simples devido a quantidade de páginas existentes. No máximo, o que poderia ser feito era clicar em links e deixar ser guiado pela sorte. Devido a isto, em 1993 foi criado um dos primeiros buscadores destinado a facilitar as buscas e os resultados desejados, surgia então o **Wandex** desenvolvido por **Matthew Gray** no **MIT** (Instituto de Tecnologia de Massachusetts) o qual se tornou a base de todos os buscadores atuais. Seu conceito básico era a indexação do conteúdo por meio de robôs ou como também são conhecido Spiders (Programas que varrem a internet e retornam com dados que são armazenados). Aparentemente esse tipo de indexação pode parecer algo comum, no entanto foi inovador para sua época, sendo os buscadores até então baseados no cadastro de conteúdo via fonte humana. Desta forma, todo o resultado era posto ao usuário de forma limpa e visivelmente organizada, sendo que, somente eram cadastrados conteúdos relevantes. Mas, com o surgimento dos robôs, um grave problema surge: **o lixo eletrônico**. Quantas vezes você foi ao seu buscador de preferência e ao digitar uma palavra surgem em sua tela conteúdos que você não desejava? Ou empresas querendo vender aquilo que você não quer comprar? Pois bem, isso já foi pior! O Wandex, e outros buscadores baseados no mesmo conceito, como o Altavista, ao revelar os dados para o usuário não se preocupava se o conteúdo seria relevante ou não, simplesmente o lançava na tela como frases completas de palavras chaves que o usuário havia digitado, conteúdo em abundância, relevância quase zero. Pois bem, para sanar esse problema surge 5 anos depois, em 1998, o gigante dos

buscadores, o Google. Mas qual é o seu diferencial? Bem, ao digitar uma palavra, como por exemplo a própria palavra google, o usuário é levado a aproximadamente 7.620.000.000 resultados (7 bilhões e 620 Milhões), contudo, o grande trunfo deste buscador é a relevância. Ao executar a busca, os primeiros resultados que são mostrados baseiam-se na quantidade de pessoas que buscaram a palavra ou frase, clicaram no link, e na quantidade de links externos que a página, a qual será levado o link, é referenciada na web, ou seja, quanto mais links forem redirecionando para a página buscada e quanto melhor for o conteúdo nela postado, mais ela será destacada nas buscas. Outro fato, que é relativamente novo, e a relevância do conteúdo, ou em outras palavras, algo só terá destaque se for algo que realmente tenha “utilidade” e não apenas spam web. Esta acabou sendo a grande “sacada” que diferenciou o Google dos outros buscadores e o destaca até os dias atuais.

Agora, como já conhecemos o funcionamento básico de um buscador, vamos analisar como é considerado o conteúdo na web. A internet pode ser dividida, segundo **Martha Gabriel** em uma de suas palestras sobre o “**Oráculo Digital**”, em web oculta e web visível. A web oculta nada mais é do que o conteúdo que não podemos ver, ou seja, os bancos de dados que não podem ser acessados pelos buscadores ou simplesmente conteúdos que não podem ser indexados pelos mesmos, os quais somam 75% de todas as informações presentes na internet, sobrando 25%, ou seja, a web visível. Dentre estes 25% apenas 10% são indexados pelos buscadores, logo, quando criamos uma pesquisa no google, o qual indexa maior parte destas informações, encontramos apenas 10%, sendo a mesma dividida em várias páginas, as quais, a “maioria dos mortais”, não vai além da terceira, ou em outras palavras, toda a busca que efetivamente fazemos na internet e todo o conteúdo que lemos são menos de 1% de todas as informações visíveis e dispersas no mundo digital. Portanto, aquela celebre frase, “Não está no google não existe” cai por terra. Outro fato, o qual queremos levá-lo a pensar é a relevância do conteúdo. Desde os primórdios da humanidade e logicamente com o surgimento da escrita, todo o conhecimento que não morreu com seus idealizadores foram transcritos em livros. Existe conhecimento e história devido à escrita, e claramente, a internet, mesmo com o grande número de digitalizações, ainda não contém a grande essência do conhecimento, desta forma, nem todo o conteúdo disperso na mesma, sem generalizar, não deve ser encarado como fonte primária e sim como base para pesquisas mais profundas. Nem tudo que você lê na Wikipédia por exemplo, deve ser levado ao seu trabalho de escola ou faculdade como uma das verdades referentes ao

conteúdo desejado.

Mesmo a internet sendo o “**local do conhecimento**”, e disso eu não discordo, devemos saber o que procuramos, referenciar o conteúdo que retiramos, e sobre tudo, manter sempre a ética. Como nas mídias impressas ou fonadas, no mundo da web também não devemos acreditar em tudo que nos é imposto.