



ANTONIO MENEGHETTI FACULDADE - AMF
CURSO DE SISTEMAS DE INFORMAÇÃO

BRUNO VIERA GAI

MECANISMOS DE BUSCA:
UMA ANÁLISE DAS DIFERENÇAS E PRINCIPAIS FUNCIONALIDADES

RESTINGA SECA/RS

2017

ANTONIO MENEGHETTI FACULDADE - AMF
CURSO DE SISTEMAS DE INFORMAÇÃO

BRUNO VIERA GAI

MECANISMOS DE BUSCA:
UMA ANÁLISE DAS DIFERENÇAS E PRINCIPAIS FUNCIONALIDADES

Trabalho de Conclusão de Curso apresentado
como requisito parcial para obtenção do título
de Bacharel em Sistemas de Informação na
Faculdade Antonio Meneghetti-AMF.

Orientador: Profº. Ms. Fábio Sarturi Prass

RESTINGA SECA/RS

2017

FACULDADE ANTONIO MENEGHETTI

Bruno Vieira Gai

MECANISMOS DE BUSCA: UMA ANÁLISE DAS DIFERENÇAS E PRINCIPAIS
FUNCIONALIDADES.

Trabalho de Conclusão de Curso-Monografia, apresentado
como requisito parcial para a obtenção do grau de
Bacharel em Sistemas de Informação, Curso de Graduação
em Sistemas de Informação, Faculdade Antonio
Meneghetti-AMF.

Orientador: Prof. Ms. Fábio Sarturi Prass



Prof. Ms. Fábio Sarturi Prass

Orientador do Trabalho de Conclusão de Curso

Antonio Meneghetti Faculdade



Profª Ms João Otávio Cadó de Matos

Membro da Banca Examinadora

Antonio Meneghetti Faculdade



Prof Dr. Jonas Bulegon Gassen

Membro da Banca Examinadora

Antonio Meneghetti Faculdade

Restinga Sêca, RS, 04 de dezembro de 2017

AGRADECIMENTOS

Agradeço primeiramente a Deus por me dar saúde e ânimo para concluir esta graduação. A meus pais pelo apoio, por acreditarem em mim. A Faculdade Antonio Meneghetti pelos ensinamentos e pelas oportunidades que tive enquanto aluno, e também aos professores por compartilharem seu conhecimento e tempo, quando precisei, sempre foram muito receptivos. Faço um agradecimento especial ao meu orientador, Prof. Fábio Prass que me ajudou muito na conclusão deste trabalho e a Prof^a. Ana Bulegon que me auxiliou de muitas maneiras nesta fase final do curso.

Deixo aqui um agradecimento especial ao meu tio Evandro Gai por toda a paciência que teve para me auxiliar nos momentos de dificuldade e por todo o incentivo que me deu para que eu concluísse esta graduação.

A minha namorada Marciane pela ajuda que me deu no desenvolvimento deste trabalho e em outros durante toda a graduação. Agradeço também a Maria Iraci pela ajuda na revisão do TCC.

Agradeço ao meu avô Nelson que infelizmente não está mais presente fisicamente, mas que está vivo para sempre em meus pensamentos, obrigado pela insistência que teve comigo em querer que eu fizesse uma faculdade e em me mostrar o quanto isso seria importante para o meu futuro.

Também não poderia deixar de agradecer algumas pessoas que me ajudaram no início da graduação, momento que encontrei a maior dificuldade já que estava retornando aos estudos depois de alguns anos: José Cardoso Sobrinho, Claudiane Weber, Fernando Meneghetti.

Obrigado também aos meus colegas pelo companheirismo durante esses anos, destaco alguns que se tornaram mais que colegas, se tornaram amigos: Renan Roos, Pablo Baggiotto, Jonas Bulegon, Vanderlei Kobs, Andreves Dickow.

Obrigado, todos vocês fazem parte desse momento tão esperado por mim.

DEDICATÓRIA

Dedico este trabalho a meus pais, meu avô Nelson e ao meu amigo que lá de cima me acompanha, Tiago Dotto Sachet.

*“As maiores loucuras são as mais sensatas
alegrias, pois tudo que fizemos hoje ficará na
memória daqueles que um dia sonharão em ser
como nós: Loucos, porém, felizes!”*

(Kurt Cobain)

RESUMO

A Internet surgiu em plena Guerra Fria, no ano de 1969 e era chamada de Arpanet, era uma forma dos militares manterem o contato entre si em caso de ataque inimigo. Somente na década de 1990 foi que a Internet começou a alcançar a população em geral. Com o seu surgimento vieram também muitas informações e conteúdos espalhados pela rede, assim, os usuários encontravam dificuldade em procurar determinado assunto. A busca por informações relevantes eram muito difíceis já que não existiam meios de pesquisa precisos. Diante deste problema surgiram os motores de busca, a partir de 1993, com a finalidade de suprir a dificuldade que os usuários encontravam em fazer pesquisas na Internet. Estes motores de busca são programas que procuram por palavras-chaves digitadas por usuários. Dentre os mais conhecidos e populares destacam-se Google, Bing e Yahoo. Essa pesquisa tem o objetivo de mostrar as diferenças e funcionalidades de cada um dos motores de busca e de fazer um estudo detalhado sobre o Google, o qual, segundo pesquisas realizadas, é o mais utilizado. Pretende-se ainda explicar de forma detalhada como acontece a pesquisa nos algoritmos utilizados.

Palavras-chave: Motores de busca; Sites de busca; Buscadores; Google.

ABSTRACT

The Internet emerged in the midst of the Cold War in 1969 and was called the Arpanet. It was a way for the military to maintain contact with each other in the event of an enemy attack. It was only in the 1990s that the Internet began to reach the general population. With its appearance came also a lot of information and content spread over the network, so users found it difficult to look for a certain subject. The search for relevant information was very difficult as there was no precise means of research. Faced with this problem appeared the search engines, from 1993, in order to overcome the difficulty that users encountered in doing searches on the Internet. These search engines are programs that search for keywords typed by users. Among the best known and popular are Google, Bing and Yahoo. This research aims to show the differences and functionalities of each of the search engines and to do a detailed study on Google, which according to research conducted is the most used. It is also intended to explain in detail how research in algorithms used.

Keywords: Search engines; Seekers; Google.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. Funcionamento de um motor de busca | 15 |
| Figura 2. Processo de pesquisa e apresentação do resultado | 16 |
| Figura 3. Fluxograma (Representação gráfica de um algoritmo) | 17 |
| Figura 4. Evolução dos mecanismos de busca | 22 |
| Figura 5. Mecanismos de busca mais utilizados no Brasil | 23 |
| Figura 6. Funcionamento do motor de busca Google | 23 |

LISTA DE TABELAS

Tabela 1. Quadro comparativo entre os motores de busca

28

SUMÁRIO

| | |
|--|-----------|
| LISTA DE FIGURAS | 9 |
| LISTA DE TABELAS | 10 |
| 1. INTRODUÇÃO | 12 |
| 2. OBJETIVOS | 13 |
| 2.1. Objetivo geral | 13 |
| 2.2. Objetivos específicos | 13 |
| 3. JUSTIFICATIVA | 13 |
| 4. REFERENCIAL TEÓRICO | 14 |
| 5. METODOLOGIA | 18 |
| 6. DESCRIÇÃO E ANÁLISE DOS MOTORES DE BUSCA | 18 |
| 6.1. Google | 18 |
| 6.2. Bing | 19 |
| 6.3. Yahoo | 20 |
| 6.4. Como o Google funciona? | 22 |
| 6.5. Googlebot | 24 |
| 6.5.1. Hummingbird | 25 |
| 6.5.2. Google Penguin | 25 |
| 6.5.3. Google Panda | 26 |
| 6.5.4. Page Rank | 27 |
| 7. RESULTADOS | 27 |
| 7.1. Análise dos resultados | 29 |
| 8. CONCLUSÃO | 30 |
| REFERÊNCIAS | 31 |

1. INTRODUÇÃO

Antes das pessoas terem acesso a Internet (final da década de 60 até o final da década de 80) o conhecimento das informações era extremamente limitado. As pessoas que precisavam realizar pesquisas sobre determinados assuntos, naquele tempo, deveriam adquirir livros ou frequentar bibliotecas. Desde o surgimento da internet, em 1969, a quantidade de informações que se pode buscar é imensa, tanto que é difícil apresentar estatísticas numéricas. Na década de 90, quando a Internet tornou-se acessível para pessoas de modo geral, era difícil o usuário conseguir encontrar o conteúdo que procurava, já que as ferramentas (motores) de busca demoraram um pouco para serem desenvolvidas.

Para Cendón (2001), os motores de busca são programas desenvolvidos para buscar algum conteúdo por meio de palavras-chave que são fornecidas pelo usuário. Na atualidade, alguns motores de busca que são acessados na internet são: Google, Yahoo, Bing, Ask, DuckDuckGo, DogPile, Excite, EntireWeb, GigaBlast e Teoma. Esses motores usam diferentes maneiras de procurar por um assunto, o que facilitou e muito a pesquisa dos usuários.

Neste sentido, surge o seguinte questionamento: o que diferencia um motor de busca de outro na coleta de informações na internet? E, por que o Google é o motor de busca mais acessado?

Esta pesquisa apresentará algumas características que permitem diferenciar os motores de busca e estudar a estrutura do Google. Será feita uma apresentação breve sobre os motores de busca: Google, Bing e Yahoo, como cada um deles surgiu e algumas particularidades que eles possuem.

O estudo da estrutura do Google consiste em um detalhamento de cada um dos algoritmos mais significativos, uma explicação sobre como eles funcionam e qual o papel que cada um possui no momento de fazer uma busca, no momento em que o usuário faz a busca por alguma informação. Por fim, a pesquisa também apresentará um quadro comparativo entre os motores de busca estudados, apresentando suas diferenças básicas, não levando em conta o estudo dos algoritmos do Google.

2. OBJETIVOS

2.1.Objetivo geral

O presente trabalho tem como objetivo descrever as funcionalidades e diferenças dos motores de busca: Google, Bing e Yahoo, assim como, um estudo detalhado de como o Google realiza as pesquisas e apresenta os resultados.

2.2.Objetivos específicos

- a) Investigar as características dos motores de busca Google, Bing e Yahoo.
- b) Verificar as diferenças entre os motores de busca Google, Bing e Yahoo.
- c) Investigar os critérios utilizados pelo Google no momento em que o usuário faz a busca por informações.
- d) Elaborar um quadro comparativo com as características e diferenças entre os motores de busca Google, Bing e Yahoo.

3. JUSTIFICATIVA

A escolha desse tema deve-se ao pouco conhecimento dos usuários sobre os motores de busca e suas funcionalidades, fazendo com que sua busca retorne dados que não sejam relevantes. Nesse estudo, serão investigadas as características dos motores de busca Google, Bing e Yahoo bem como suas diferenças.

Também será feito um estudo sobre como o Google realiza as pesquisas e o que possui que faz dele o mais utilizado na atualidade.

Entre diversos sites de busca existentes foi escolhido o nome desses três apresentados anteriormente, pois atualmente são os que estão sendo mais usados pela maior parte dos usuários e também mais conhecidos segundo Goldman (2011) e Taurion (2009).

Tal resultado versa auxiliar o usuário a entender melhor o funcionamento do mesmo, a forma com que cada um deles busca por determinado assunto por trás da tela do computador.

4. REFERENCIAL TEÓRICO

Mecanismos de busca estão se tornando cada vez mais indispensáveis para pesquisas na Internet. Pesquisar através de buscadores virou um hábito. As procuras estão cada vez mais completas e para isso existe um aperfeiçoamento cada vez maior dos motores. Quando se fala em motores ou mecanismos de busca, está se falando de sites que procuram informações em outros sites por meio de uma palavra chave que é introduzida pelo usuário na página web.

Atualmente existem diversos mecanismos de busca. A lista de buscadores existentes não é pequena, mas pode-se citar três que são os mais conhecidos e mais utilizados: Google, Bing e Yahoo. Cada um possui uma particularidade ou uma característica única.

Segundo Silveira (2002) pode-se definir um mecanismo de busca como sendo um banco de dados que auxilia as pessoas na procura de palavras a partir de um termo pesquisado. Se fosse feita uma análise somente na interface desses mecanismos, ou seja, analisar apenas aquilo que o usuário vê na tela do seu computador ou dispositivo móvel, a impressão é que não existe nada de complexo, mas por trás disso existem diversos passos e procedimentos que o site de busca faz até entregar para o usuário o resultado da pesquisa.

De acordo com Peixoto (2008):

“Os motores de busca utilizam software conhecido como ‘aranhas’ ou ‘robots’ que percorrem ‘toda’ a Internet em busca da informação (...) que se pretende (...) sempre que se introduz uma palavra ou conjunto de palavras (...) as bases de dados são percorridas em busca de documentos ou sites que lhe correspondam (...) num processo designado por Web Crawling”.

Resumidamente, esses motores são formados por quatro componentes: o robô que é quem procura e encontra os documentos, o indexador que é responsável por extrair a informação dos documentos, o motor de busca propriamente dito e a interface que o usuário utiliza. Analisando os motores de busca mais utilizados nos dias atuais pode-se concluir que a única diferença visível para o usuário é a interface, o que o usuário vê.

A interface é também um dos critérios usados pelos usuários na hora de escolher qual motor de busca vai utilizar. O usuário que não tem conhecimento de funcionalidades e particularidades mais técnicas fica limitado a analisar somente a interface e a partir daí opta por A ou B. Uma interface simples, sem a presença de publicidade que muitas vezes acaba sendo invasiva, dificilmente deixara de ser escolhida pela maioria dos usuários.

A Figura 1 mostra de forma mais clara o funcionamento de um motor de busca e os componentes básicos: o *spider* ou *robot* que vasculham toda a web em busca de *links* e textos. O indexador que é quem extrai as informações encontradas pelo robô e as armazena em um

banco de dados e o motor de busca propriamente dito e que faz a ligação entre a consulta e o banco de dados contendo as informações. Esse processo é finalizado quando o resultado dessa busca é apresentado para o usuário na interface, na tela do computador ou do dispositivo móvel.

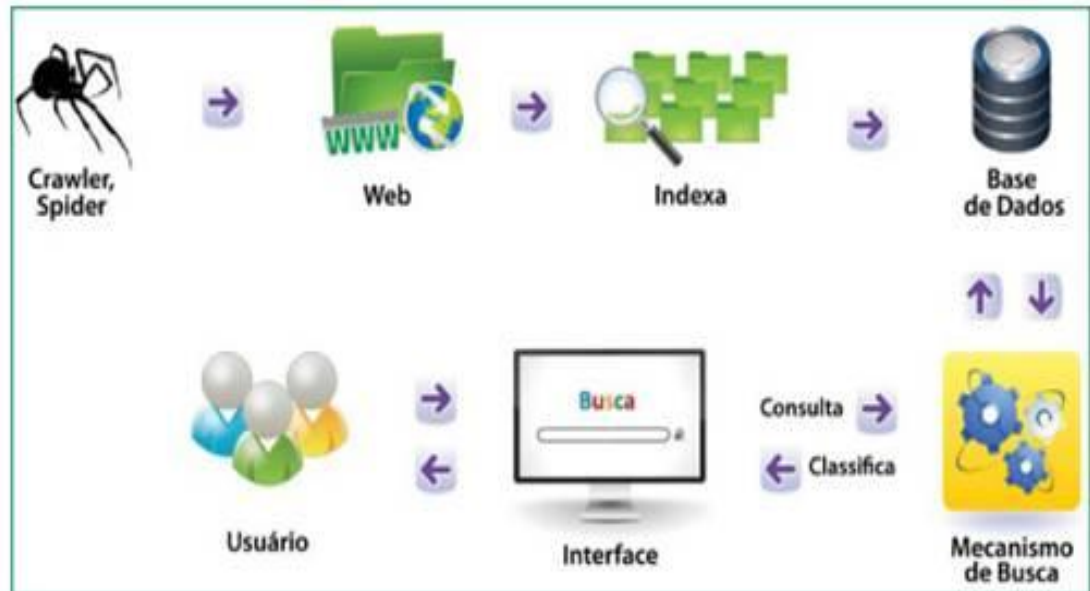


Figura 1. Funcionamento de um motor de busca

Fonte: <https://goo.gl/L86GDY>

O esquema mostrado na imagem é comum para todos os mecanismos de busca, o que os diferencia são os critérios que são usados para realizar as buscas. Esses critérios são determinantes na hora de retornar os resultados das pesquisas realizadas pelos usuários, o sucesso de um motor de busca se dá à relevância dos resultados.

De uma maneira mais clara, a Figura 2 traz um esquema do processo de pesquisa (*query*) e apresentação do resultado:

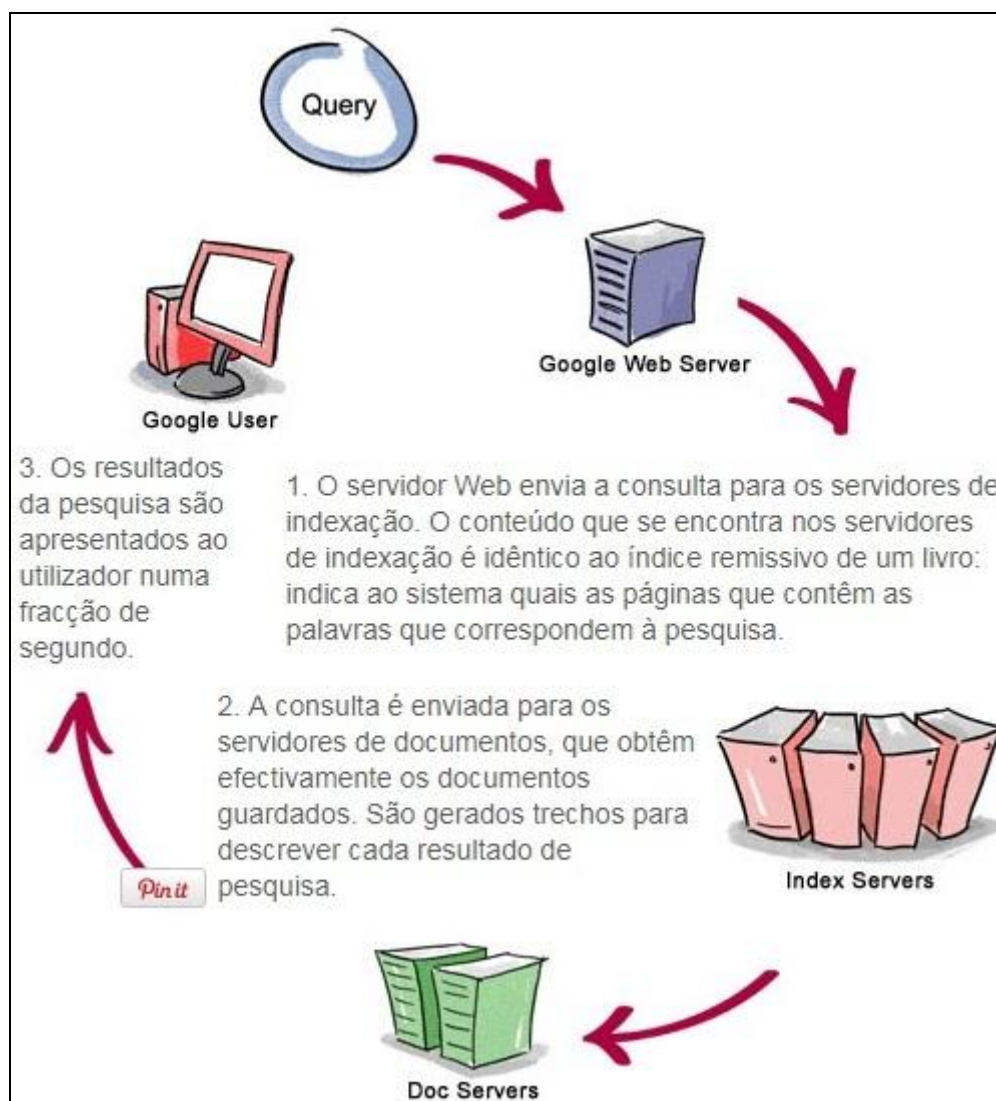


Figura 2. Processo de pesquisa e apresentação do resultado

Fonte: Fonte: <https://goo.gl/My11gu>

Tecnicamente falando, podem-se citar outras diferenças entre eles: o tamanho da base de dados e a frequência com que ela é atualizada, os critérios que são usados para indexação (inclusão da palavra no índice), como os dados são ordenados, como é feita a inclusão de páginas e ainda o recurso de busca que cada um oferece. Todas essas tarefas são executadas através de algoritmos.

Na área da computação, o conceito de algoritmo é definido como uma série de instruções detalhadas, um passo a passo que mostra de forma explícita várias operações, esse é um dos conceitos mais básicos dessa área. Para Bhargava (2017), um algoritmo é um conjunto de instruções que realizam uma tarefa. O computador também trabalha com algoritmos, de uma forma muito rápida, imperceptível.

No dia a dia também se usa algoritmos, muitas vezes de forma inconsciente, para a realização das mais diversas tarefas e também para a resolução de problemas. Pode-se dizer que um algoritmo é uma espécie de “receita” e assim como ela, o algoritmo é finito. Um algoritmo pode seguir caminhos diferentes dependendo da situação, como observado na Figura 3.

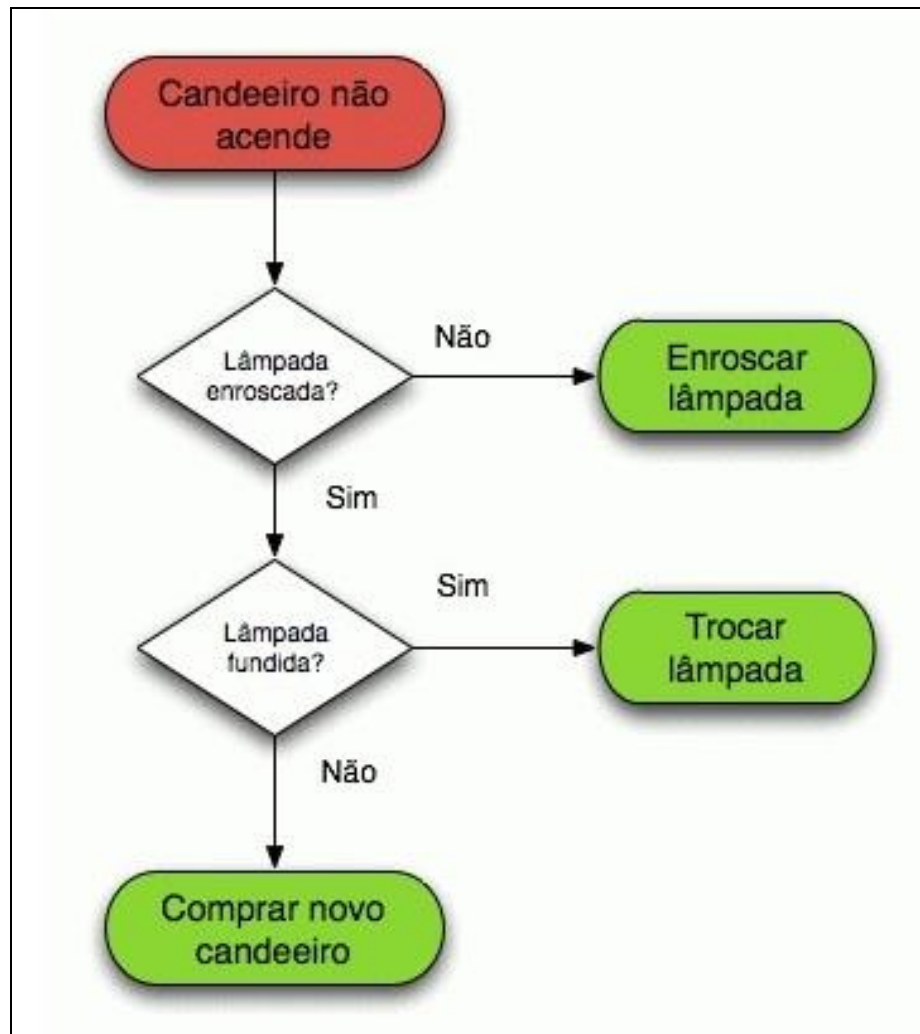


Figura 3. Fluxograma (Representação gráfica de um algoritmo)

Fonte: <https://goo.gl/4fYCaA>

A Figura 3 demonstra a aplicação de um algoritmo no dia a dia, a representação é chamada de fluxograma e os losangos representam as decisões a serem tomadas para executar determinado passo. Esta representação através do fluxograma mostrou como trocar uma lâmpada.

5. METODOLOGIA

A metodologia deste trabalho é de caráter qualitativo. Trata-se de um estudo de caso que visa apresentar uma breve introdução sobre cada motor de busca pesquisado, descrevendo suas características e diferenças. Será feito um estudo mais aprofundado sobre a forma com que o Google faz a busca, utilizando artigos científicos, livros e conteúdos relevantes da Internet. Sobre este tipo de pesquisa, Fonseca (2002) aponta o seguinte:

“Um estudo de caso pode ser caracterizado como um estudo de uma entidade bem definida como um programa, uma instituição, um sistema educativo, uma pessoa, ou uma unidade social. Visa conhecer em profundidade o como e o porquê de uma determinada situação que se supõe ser única em muitos aspectos, procurando descobrir o que há nela de mais essencial e característico. O pesquisador não pretende intervir sobre o objeto a ser estudado, mas revelá-lo tal como ele o percebe. O estudo de caso pode decorrer de acordo com uma perspectiva interpretativa, que procura compreender como é o mundo do ponto de vista dos participantes, ou uma perspectiva pragmática, que visa simplesmente apresentar uma perspectiva global, tanto quanto possível completa e coerente, do objeto de estudo do ponto de vista do investigador”.

Para obter os dados da pesquisa foram realizados estudos usando artigos científicos, livros e materiais encontrados na internet e com fonte confiável. Após uma breve introdução sobre como se originaram os motores de busca citados aqui, foi feito um estudo sobre como o Google faz suas buscas na internet, explicando seu funcionamento por trás da tela que o usuário vê, entendendo assim, por que o Google é o mais utilizado e também o melhor motor de busca da atualidade.

Depois, foi elaborado um quadro comparativo dos três motores de busca com particularidades citadas na primeira parte do estudo, não levando em conta dados mais técnicos presentes na segunda parte.

6. DESCRIÇÃO E ANÁLISE DOS MOTORES DE BUSCA

6.1. Google

O Google foi fundado por Larry Page e Sergei Bin no ano de 1998 e hoje é o nome mais lembrado quando o assunto é busca na Internet. Possui uma interface limpa e sem a presença de anúncios, sem fotos de fundo, o que ajuda a aumentar sua preferência já que quem possui baixa conexão com a Internet popularmente dará a preferência a um buscador que carregue rapidamente, é o caso do Google. Segundo Goldman (2011), a ideia do Google não é criar

produtos ricos em detalhes, os designers apenas incluem os recursos que as pessoas precisam para atingir suas metas. Também é possível navegar pela suíte (conjunto) de aplicativos do Google.

Ainda na página inicial, é possível configurar filtros, idioma dos resultados de pesquisa, número de resultados exibidos por página e uma configuração que somente o Google possui: exibir ou não as sugestões de pesquisa enquanto o usuário digita uma palavra. No Google é possível pesquisar por imagens, notícias, mapas, vídeos, Shopping (vitruine de produtos de lojas cadastradas), livros e voos. O número de resultados de buscas feitas usando o Google (nesse exemplo) é muito maior que os resultados de Bing e Yahoo juntos. Quando é digitada a palavra “Azul” no Google, 786.000.000 de resultados são encontrados.

Uma observação importante de se fazer é a que o Google corrige uma palavra que é digitada de forma errada evitando que o usuário faça uma pesquisa utilizando a grafia da palavra incorreta, Bing e Yahoo não corrigem as palavras, fazendo com que a busca fique mais difícil na ausência de um acento na palavra.

A indexação é um meio muito eficiente de encontrar um documento em uma base de dados de forma organizada. A identificação do arquivo se deve ao número de palavras representativas e particularidades que são associadas a ele, permitindo ganhar mais relevância e muitos acessos. A indexação das páginas é feita pelo rastreador Web (Crawler) que após a indexação é responsável por manter a base de dados atualizada. Todos os motores de busca presentes na pesquisa possibilitam ao usuário indexar páginas gratuitamente. A indexação pode ser feita incluindo grande parte das páginas de cada site ou apenas a página inicial.

Ao abrir a página de indexação do Google, é possível ler uma pequena introdução sobre as políticas de indexação do Google e uma espécie de “manual” que mostra de forma resumida o processo de indexação. A página de indexação do Google é intuitiva, lendo atentamente as instruções o usuário é capaz de indexar uma página sem muitos problemas.

Na página de indexação também é possível remover conteúdo do índice do Google através da URL (Uniform Resource Locator) que significa Localizador Padrão de Recursos e é a descrição do site no navegador, endereço virtual, link. Significa retirar os endereços virtuais que estão indisponíveis ou que não recebem atualizações.

6.2. Bing

O Bing é um motor de busca que foi desenvolvido pela Microsoft em 2009, substituiu os antigos MSN Search, Live Search e Windows Live Search. Sua tela inicial exibe um fundo

com diversas fotos em alta resolução, o que pode ser um impedimento para quem possui uma conexão com a Internet limitada, impedindo uma busca rápida e fazendo usuários partirem para outros motores de busca mais rápidos. É possível acessar os aplicativos Bing a partir da tela inicial.

Também é possível configurar filtros e resultados na página inicial assim como o Google, mas como foi mencionado acima, não é possível configurar as sugestões de pesquisa e nem corrigir a grafia das palavras na hora da pesquisa. No Bing é possível pesquisar por imagens, vídeos, mapas e notícias. O número de resultados de buscas feitas usando o Bing (nesse exemplo) é igual ao resultado usando o Yahoo, mas muito inferior quando comparado ao Google. Quando é digitada a palavra “Azul” no Bing, 15.300.000 resultados são encontrados.

O Bing está buscando lugar no mercado firmando algumas parcerias como a Apple, Amazon e os sites AOL e Yahoo (que usa em seu buscador a tecnologia do Bing). Também vale destacar que seu uso cresceu bastante com o lançamento do sistema operacional da Microsoft, o Windows 10 e já era o buscador oficial dos *gadgets* (miniaplicativos) da Microsoft.

A página de indexação do Bing também é intuitiva e facilita o entendimento do usuário, basta somente que o mesmo saiba o que significa uma URL. A ferramenta de busca do Bing e seus aplicativos estão presentes também na página de indexação, também existe publicidade na página de indexação do Bing, diferente do Google.

6.3. Yahoo

O Yahoo foi desenvolvido por Jerry Yang e David Filo em 1994 quando eles buscavam informações sobre um concurso que envolvia a Liga de Basquete Universitária e perceberam que a busca por informações era difícil e que eles poderiam suprir essa necessidade criando um site de buscas. O endereço yahoo.com leva o usuário a um site com notícias e atualidades e também possibilita o usuário a pesquisar através da barra de pesquisa que se encontra na parte superior da página. Já o endereço br.search.yahoo.com leva o usuário a uma tela de buscas parecida com o Google, sem anúncios e com fundo branco na sua tela inicial. Atualmente o Yahoo usa a tecnologia do Bing em seu buscador, ou seja, a forma como o Yahoo busca as informações é a mesma usada no Bing, a diferença está somente na tela que o usuário vê.

Assim como Google e Bing, o Yahoo possui as mesmas configurações exceto as que somente o Google possui que é a correção da grafia das palavras no momento da pesquisa e a configuração das sugestões de pesquisa. Além dessas configurações, existe uma barra na parte superior da página aonde é possível acessar notícias, atualidades e algumas plataformas. No Yahoo é possível pesquisar por imagens, vídeos, notícias, respostas e Shopping (vitrine de produtos de lojas cadastradas).

O número de resultados de buscas feitas usando o Yahoo (nesse exemplo) é igual ao resultado usando o Bing até por ele usar a mesma tecnologia, mas muito inferior quando comparado ao Google. Quando é digitada a palavra “Azul” no Yahoo, 15.300.000 resultados são encontrados.

O Yahoo passou a perder sua popularidade, pois deixou de ser um motor de busca no momento em que começou a dar prioridade aos anúncios e a deixar as pessoas em segundo lugar, ou seja, preferiam que pessoas anunciassem no Yahoo ao invés de pesquisarem por algo. Por outro lado, o Google priorizou os clientes e ao invés de dar espaço para anúncios, tornou eles relevantes nos resultados das pesquisas, ou seja, tornou os anúncios úteis para os usuários de acordo com a pesquisa.

A tela de indexação do Yahoo é um pouco diferente, a página traz os objetivos do Yahoo Search, o usuário deverá clicar onde diz: Sugira o seu site gratuitamente. Depois disso o usuário é direcionado a página de indexação do Bing, já que o Yahoo usa sua tecnologia no buscador.

A base de dados dos motores de busca é atualizada com frequência para a adição de novas páginas e também para a exclusão ou modificações. É preciso esse cuidado, pois muitas vezes a página que o motor de busca traz como resultado pode não existir mais ou pode estar diferente desde o primeiro acesso, assim não sendo relevante naquela busca.

Cada motor de busca possui uma estratégia para permanecer atualizado, alguns sites podem demorar semanas para serem adicionados na base de dados já que é crescente o número de novos sites na rede. A ordem dos resultados é feita por um algoritmo de ordenação que usa alguns critérios para isso: número de termos da consulta na página, ocorrência de certa palavra em uma página, proximidade dos termos...

Se uma página possui mais termos do que a outra ela se torna mais relevante. Informar como é feito esse critério de ordenação em cada mecanismo é uma tarefa extremamente difícil já que é esse o principal diferencial de cada motor de busca.

6.4. Como o Google funciona?

Como foi descrito nos tópicos anteriores, pesquisas demonstram que o Google é o motor de busca mais usado no mundo. As buscas realizadas através dele retornam realmente o que o usuário necessita. Segundo Goldman (2011), o Google proporciona os resultados mais relevantes, o algoritmo do Google explora com eficácia todo o conhecimento coletivo da internet. A Figura 4 mostra a evolução Google nos Estados Unidos nos anos de 2004 e 2012:

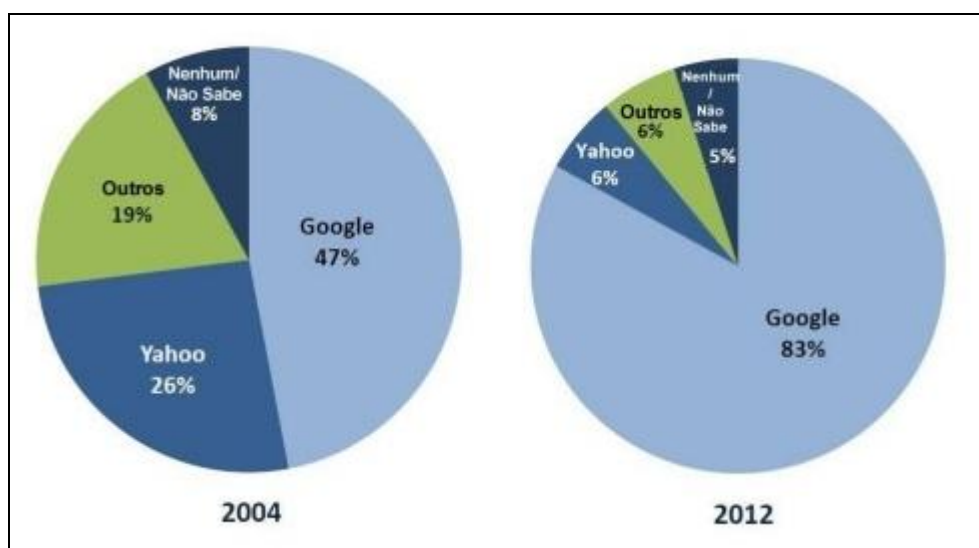


Figura 4. Evolução dos mecanismos de busca
Fonte: <https://goo.gl/LtJev2>

Na Figura 5 os três motores de busca estudados nesta pesquisa aparecem como os mais utilizados no Brasil, foram levantadas estatísticas de novembro de 2016 até novembro de 2017. Em primeiro lugar aparece o Google com 92,23% dos acessos, seguido por Bing com 2,75% de acessos e Yahoo com 2,15% de acessos.

A popularidade do Google é grande, pois o usuário realmente consegue achar o que procura em sua busca ainda na primeira página. A exatidão nos resultados tem uma explicação. Para Pariser (2012, p.36) o Google “monitorea todo e qualquer sinal que consiga obter de nós. [...] Além disso, naturalmente, os termos que pesquisamos trazem grandes revelações sobre nossos interesses”.

O Google instala, ainda, cookies de rastreamento nos computadores capazes de colher informações pessoais. Em realidade, fala-se em cerca de 57 sinais, mas, estrategicamente, o Google revela poucos deles. Dentre os sinais já revelados, encontramos: histórico de pesquisa; região do usuário; sinais sociais (Google +); endereço IP (Internet Protocol ou Protocolo de internet); tipo de computador usado; tipo de navegador usado; idioma.

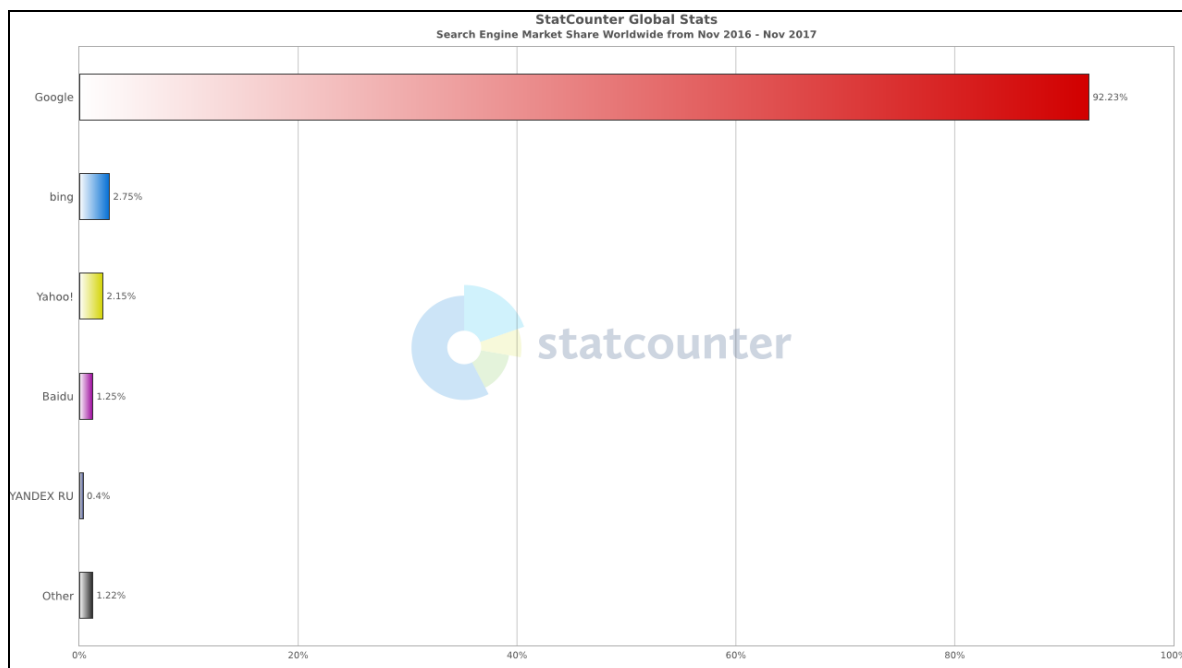


Figura 5. Mecanismos de busca mais utilizados no Brasil

Fonte: <https://goo.gl/qQvdRU>

A estrutura em todos os motores começa com o rastreamento (*crawling*) e a indexação das páginas que é a associação de palavras ou particularidades a uma página, assim permitindo através de uma palavra-chave ser identificada entre milhares de outras.

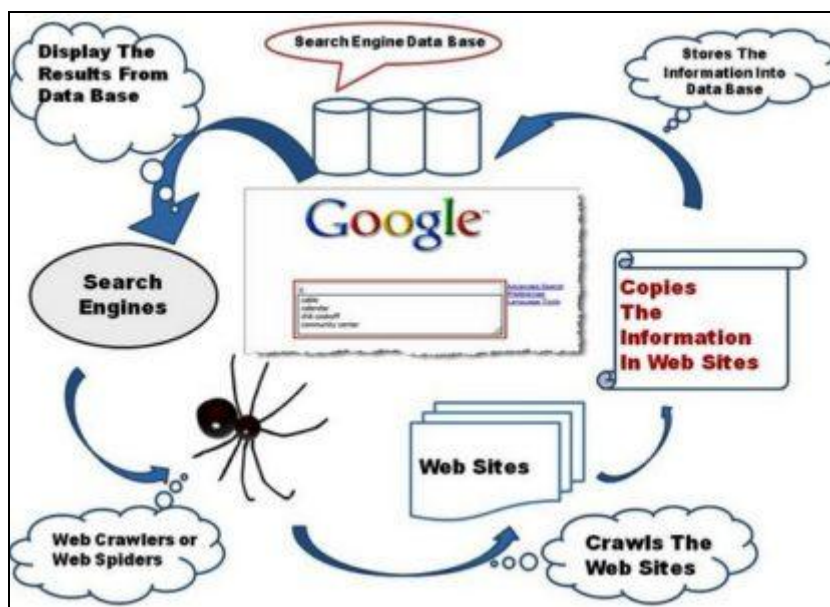


Figura 6. Funcionamento do motor de busca Google

Fonte: <https://goo.gl/4s6Ysr>

A figura 6 reforça como é o funcionamento do Google. O robô (também chamado de aranha) vasculha a internet em busca de conteúdo e encontra, o indexador extrai essas informações colhidas pelo robô e armazena em um banco de dados. Quando o usuário digita a palavra a ser pesquisada na interface, o mecanismo de busca traz essa informação do banco de dados em poucos segundos.

O Google possui diversos algoritmos que trabalham por trás da pesquisa, alguns não são divulgados por razões estratégicas da empresa, por outro lado, existem alguns conhecidos no meio da computação, que são os que tornam a busca no Google mais precisa e mais relevante que em outros mecanismos de busca. Em outras palavras, os critérios usados para classificar a posição das páginas no resultado da pesquisa. A seguir esses algoritmos serão apresentados.

6.5. Googlebot

O rastreador do Google é chamado de Googlebot e tem como finalidade descobrir novas páginas devidamente atualizadas para serem incluídas no índice do Google. O Googlebot usa um processo de algoritmos que vai determinar qual site será rastreado, qual a frequência e também a quantidade de páginas que vão ser buscadas em cada um dos sites. O rastreamento começa com uma lista de URLs de páginas da Web originadas de processos anteriores de rastreamento. O Googlebot analisa cada um desses sites detectando seus links (endereços) e incluindo na lista de páginas a serem rastreadas. Esse processo ajuda a manter o índice do Google atualizado já que é capaz de detectar se existem alterações feitas em sites existentes, se existem novos sites e também sites inativos.

No momento em que o usuário faz uma busca, o algoritmo do Google começa a trabalhar na procura de indícios que possam fazer com que aquilo que o usuário digitou seja compreendido trazendo assim resultados relevantes do índice. O algoritmo analisa o que foi digitado de diversas formas: a maneira que a pesquisa foi feita (se foi por texto, imagem, voz...), apresentando resultados enquanto o usuário está digitando (Google Instant), a ortografia (que como foi falado anteriormente, é corrigida em caso de erro de grafia), previsões de pesquisa que são mostradas enquanto o usuário digita (*Autocomplete*), sinônimos, significado mais profundo de palavras pesquisadas (*Query Understanding*).

Depois dessa análise, os dados são classificados a partir de mais de 200 fatores que não são revelados pela empresa por razões estratégicas. O Googlebot é constantemente atualizado com base nos resultados obtidos.

6.5.1. Hummingbird

O Hummingbird (beija-flor em inglês) é um algoritmo cujo objetivo é melhorar o entendimento do propósito que o usuário tem ao realizar determinada pesquisa, seu nome deve-se ao fato de ele ser preciso e rápido. O Google fez uma melhoria nesse algoritmo já que as buscas móveis (existe diferença já que o Google leva em conta a localização) e por voz vem crescendo, bem como as buscas baseadas em perguntas (como fazer? onde é mais perto?...).

Com essa melhoria no algoritmo, o Google é capaz de fornecer resultados mais precisos baseando-se no contexto, compreendendo a finalidade da busca. A ideia do algoritmo é fazer buscas com palavras-chave mais difusas, igual a alguém falando, ao invés de uma busca com foco definido e curto como alguém digitando. Em outras palavras, o Google passará a interpretar os resultados de forma mais inteligente, uma busca semântica.

Um motor de busca semântico tenta fazer sentido aos resultados da pesquisa com base no contexto, ele identifica automaticamente os conceitos que estruturam os textos. Os motores de busca que se concentram na frequência do aparecimento de palavras no texto (que são convencionais) são menos propensos a compreender os significados.

6.5.2. Google Penguin

Alguns chamam de atualização e outros de mudança algorítmica, o Google Penguin é uma série de atualizações que foram feitas no algoritmo de busca e que condenam links que foram conseguidos de forma não natural (compra) em páginas que não tem um conteúdo relevante. O ranqueamento de um site significa a quantidade de links que apontam para o conteúdo do mesmo mostrando que tem autoridade. O Penguin pune os sites que estão posicionados no ranqueamento de forma ilegal.

Alguns fatores que levam os links a serem punidos são: links provenientes de sites de baixa qualidade, que têm o mesmo texto de âncora semelhante, que foram construídos usando um *boot* ou ferramenta, que foram pagos ou incentivados, que vem de países estranhos, que foram construídos em alta qualidade em um curto período de tempo, de países questionáveis.

O Google Penguin é atualizado em tempo real, quando encontra um spam ele não o usa como fator de classificação, ele o ignora.

6.5.3. Google Panda

O Panda é um algoritmo desenvolvido para detectar os sites com intenções suspeitas, conteúdo duplicado e ruim, superficial, através de métricas que indicam qual a qualidade do conteúdo, recebeu o nome do funcionário do Google, Navneet Panda. É uma espécie de filtro de qualidade que filtra os sites que são “spams” e os que podem realmente ser úteis para o usuário, diminuindo a presença de sites de baixa qualidade nos resultados das pesquisas.

Alguns dos problemas que contribuíram para o desenvolvimento do Google Panda foram:

- Páginas com pouco texto ou um conjunto de páginas com algumas frases em cada uma;
- Conteúdos copiados que na internet aparecem em mais de um lugar (duplicação);
- Páginas que possuem um conteúdo pouco detalhado e com pouco valor aos leitores;
- Conteúdo produzido por fontes que não são consideradas definitivas ou que não são verificadas, como exemplo pode-se citar um site de compra desconhecido e que dificilmente algum usuário forneceria informações como número do cartão de crédito;
- Páginas de baixa qualidade agregadas a partir de outros sites;
- Conteúdos gerados por usuários desqualificados, um exemplo disso são os blogs que são duplicados ou que possuem erros de grafia e informações sem a devida autorização;
- Páginas contendo mais anúncios que conteúdo;
- Conteúdo de baixa qualidade em torno de links de afiliados pagos;
- Sites bloqueados por pessoas diretamente nos resultados das pesquisas por meio de uma extensão instalada no navegador;
- Páginas que dizem responder determinada pergunta, mas ao serem clicadas nos resultados de pesquisa não conseguem cumprir com o prometido.

6.5.4. Page Rank

Sem dúvidas é um dos algoritmos mais importantes já desenvolvidos. O Page Rank é um dos métodos que o Google usa para determinar a relevância da página. Esse algoritmo faz a avaliação das páginas levando em conta a quantidade de links que existem para uma página a partir de outras páginas e também a qualidade do site, individualmente e não como um todo. O Page Rank não classifica sites que são proibidos.

Esse algoritmo foi desenvolvido pelos fundadores do Google, Larry Page e Sergey Brin enquanto cursavam a Universidade de Stanford no ano de 1998, com o propósito de ajudar a determinar a relevância ou importância de uma página. Mesmo assim é possível manipular esse algoritmo atribuindo links descontextualizados com o objetivo da página, modificando a ordenação dos resultados no Google e induzindo a resultados que não são relevantes, processo este que ficou conhecido como “Googlebombing”.

O Google remove alguns resultados decorrentes desse processo, mas mesmo assim esse fato ainda acontece. O algoritmo pode ser aplicado a qualquer coleção de entidades com citações e referências recíprocas. Em resumo, o algoritmo do Page Rank computa a probabilidade de um usuário chegar em determinado site clicando em *back links* (links recebidos por site ou página web) aleatórios. Quanto maior o resultado, maior a chance de aparecer na página principal do Google.

7. RESULTADOS

Antes de considerar os dados obtidos a partir do levantamento apontado, dar-se-á evidência no quadro comparativo entre os motores de busca analisados, destacando as propriedades de cada um conforme a Tabela 1. Esses dados contidos na tabela são fruto do resultado das pesquisas feitas e encontram-se detalhados na análise dos motores de busca.

Tabela 1. Quadro comparativo entre os motores de busca

| <i>CARACTERÍSTICAS E FUNCIONALIDADES</i> | <i>GOOGLE</i> | <i>BING</i> | <i>YAHOO</i> |
|---|----------------------|--------------------|---------------------|
| Ano de fundação | 1998 | 2009 | 1994 |
| Interface carrega rapidamente | X | | X |
| Suíte de aplicativos na página | X | X | X |
| Notícias e atualidades na página | | X | X |
| Configurar filtros dos resultados de pesquisa | X | X | X |
| Configurar idioma dos resultados de pesquisa | X | X | X |
| Configurar número de resultados exibidos por página | X | X | X |
| Exibir ou não sugestões de pesquisa | X | | |
| Pesquisa de imagens | X | X | X |
| Pesquisa por imagens (importando) | X | | |
| Pesquisa de notícias | X | X | X |
| Pesquisa de mapas | X | X | |
| Pesquisa de vídeos | X | X | X |
| Pesquisa de livros | X | | |
| Pesquisa de vãos | X | | |
| Shopping (produtos de lojas cadastradas) | X | X | |
| Pesquisa de respostas | | | X |
| Pesquisa por voz | X | | |
| Corrige uma palavra digitada de forma errada | X | | |
| Ferramenta de inserção de texto na tela inicial | X | | |
| Indexar páginas gratuitamente | X | X | X |
| Maior base de dados | X | | |
| Exemplo de resultado em números | 786.000.000 | 15.300.000 | 15.300.000 |

A tabela mostra de forma resumida toda a introdução feita acerca dos três motores de busca estudados: Google, Bing e Yahoo. O “X” presente em cada característica ou funcionalidade significa que aquele motor de busca possui tal recurso, além de informar o ano de cada motor e o número de resultados que cada um traz se a palavra “Azul” fosse digitada nos três motores de busca estudados, esse exemplo é explicado no desenvolvimento do trabalho.

Os resultados apresentados são originados de pesquisas científicas em livros, revistas eletrônicas, artigos e conteúdos relevantes da internet. A maior parte dessas particularidades e características pode ser vista na página inicial de cada motor de busca.

7.1. Análise dos resultados

Com base na coleta de dados e nas estatísticas estabelecidas pelo site StatCounter e pesquisas bibliográficas dos autores Goldman (2011) e Taurion (2009), bem como todas fontes relevantes da internet, conclui-se que o Google é o motor de busca mais lembrado pelos usuários na atualidade. Possui mais recursos e sua busca é mais precisa que outros motores de busca existentes na rede. Existem diversos motores de busca nesse mercado competitivo, mas certamente será difícil que algum chegue perto do sucesso do Google. Como foi apresentado anteriormente, fazem parte da lista dos motores de busca mais conhecidos além do Google, Bing e Yahoo.

O Bing é o que mais se aproxima do Google, uma vez que o Yahoo utiliza a mesma tecnologia para retornar suas buscas. O Google possui um algoritmo que faz uma busca semântica, o Hummingbird, por outro lado a Microsoft dona do Bing, em 2008 adquiriu a empresa Powerset que estava trabalhando no desenvolvimento de um motor de busca que trabalhasse com busca semântica.

A relevância dos resultados no Bing é diferente, o Bing considera relevante que as palavras chave apareçam no domínio, já o Google dá prioridade a quantidade de links recebidos pela página e também a idade dos links. Ainda sobre relevância, o Bing dá mais importância para trechos de conteúdo da página do que para a *meta description tag* (espécie de “mini-resumo”) que é uma espécie de marcação das páginas e que é direcionada para os motores de busca, o problema disso é um retorno de busca pouco interessante para o usuário.

O Google acha relevante a *meta description tag* e evita fugir dela, esse “mini-resumo” é retornado ao usuário como *snippet* (fragmentos do texto) e em algumas vezes completa esse resumo com trechos do conteúdo, o Bing não faz isso. Quando eles não existem, os motores de busca são os responsáveis por criar o *snippet* e o Google se mostra mais eficiente que o Bing, mostra resultados mais relevantes. O Bing mostra o primeiro texto que encontra na página, então, seria mais prático que as palavras chaves aparecessem no começo do texto como uma *meta description tag* do que na página já que o Bing acertaria na construção do *snippet*.

O Google confia mais nos links de entrada das páginas do que o Bing que em muitos casos posiciona melhor uma subpágina do que a própria página. O Google além de corrigir a grafia das palavras, tem um dicionário de sinônimos e abreviaturas muito mais amplo que o do Bing. Ao final desta análise conclui-se de fato que o Google é o motor de busca mais

completo e que da os resultados mais relevantes para o usuário, graças a todos esses pontos apresentados anteriormente fazendo dele o mais usado.

8. CONCLUSÃO

Como visto no início do trabalho, antes de os motores de busca surgirem, o acesso à informação era difícil. Quando a internet passou a ser acessível para todas as pessoas, tais motores de busca levaram alguns anos até serem desenvolvidos. Atualmente existem diversos motores de busca que trazem para o usuário toda e qualquer informação disponível na internet da forma mais objetiva possível. Através de pesquisas que foram apresentadas no decorrer do trabalho, chegou-se a conclusão de que os motores de busca mais conhecidos na atualidade são: Google, Bing e Yahoo.

O presente trabalho teve como objetivo apresentar os três motores de busca mais utilizados na atualidade e por consequência os mais conhecidos, bem como um estudo dos algoritmos do Google, que são responsáveis por entregar ao usuário o resultado de suas pesquisas. Foram apresentadas as características de cada buscador, assim como suas particularidades.

Depois de colher informações relevantes dos três motores de busca abordados, foi feita uma tabela que mostra as particularidades de cada um deles. Ao final do estudo, foi feita uma análise acerca dos algoritmos dos motores de busca Google e Bing, o Yahoo não apareceu no estudo final porque utiliza a tecnologia do Bing nas suas buscas.

Os resultados reforçam as imagens apresentadas no trabalho e os resultados, o Google é o motor de busca mais utilizado pelos usuários. A justificativa dessa escolha se da graças à eficiência dos seus algoritmos e a relevância dos resultados que ele apresenta, prova disso é a análise feita dos resultados. Comparado aos dois motores de busca que foram estudados, o Google leva vantagem em todos os pontos abordados.

REFERÊNCIAS

AGÊNCIA MESTRE. Bing SEO VS. Google SEO. Disponível em: <https://www.agenciamestre.com/artigos-premium/bing-seo-vs-google-seo/>. Acesso em: 17 de Novembro de 2017.

BHARGAVA, Aditya Y. Entendendo Algoritmos. Um guia para programadores e outros curiosos. São Paulo: Novatec, 2017.

CALDEIRA, Fátima Hassan. O mecanismo de busca do Google e a relevância na relação sistema-usuário. Porto Alegre: PUCRS, 2015. Revista Digital.

CANALTECH. Afinal, o que é um algoritmo e o que isso tem a ver com computação?. Disponível em: <https://canaltech.com.br/produtos/Afinal-o-que-e-um-algoritmo-e-o-que-isso-tem-a-ver-com-computacao/>. Acesso em: 04 de Outubro de 2017.

CANALTECH. Conheça a história dos buscadores e veja como o Google alcançou o topo. Disponível em: <https://canaltech.com.br/materia/internet/conheca-a-historia-dos-buscadores-e-veja-como-o-google-alcancou-o-topo-47289/>. Acesso em: 04 de Maio de 2017.

CENDÓN, Beatriz V. (2001). Ferramentas de busca na Web. Disponível em: <http://www.scielo.br/pdf/ci/v30n1/a06v30n1>. Acesso em 29 de Março de 2017.

CMSWiRE. A diferença real entre Google e Yahoo. Disponível em: <http://www.cmswire.com/cms/web-content/the-real-difference-between-google-and-yahoo-005300.php>. Acesso em: 10 de Maio de 2017.

CONVERSION. Indexação. Disponível em: <http://www.conversion.com.br/otimizacao-de-sites-seo/on-page/indexacao/>. Acesso em: 11 de Maio de 2017.

DICAS DE PROGRAMAÇÃO. O que é Algoritmo?. Disponível em: <https://goo.gl/4fYCaA>. Acesso em: 04 de Outubro de 2017.

ELITE STRATEGIES. Google Penguin. Disponível em: <https://elite-strategies.com/learn-seo/google-penguin/>. Acesso em: 10 de Outubro de 2017.

EXAME. Faz piadas com o Bing? Então é melhor você ler isto. Disponível em: <http://exame.abril.com.br/tecnologia/acha-que-o-bing-e-uma-piada-entao-e-melhor-voce-ler-isso/>. Acesso em: 10 de Maio de 2017.

FB FERRAMENTAS BLOG. As diferenças entre os motores de busca. Disponível em: <http://www.ferramentasblog.com/2012/08/as-diferencas-entre-os-motores-de-busca.html>. Acesso em: 22 de Maio de 2017

FERNANDES, Rogério Paulo Muller et al. Panorama atual do uso dos mecanismos de busca na Web. XIII Encontro Nacional de Pesquisa em Ciência da Informação – XIII ENANCIB 2012. GT 8 Informação e Tecnologia. Comunicação oral. Disponível em: <https://goo.gl/LtJev2>. Acesso em: 26 de Outubro de 2017.

FONSECA, J. J. S. Metodologia da pesquisa científica. Fortaleza: UEC, 2002. Apostila.

GOLDMAN, Aaron. Nos bastidores do Google: tudo o que sei sobre marketing aprendi com o Google. São Paulo: Saraiva, 2011.

I-TÉCNICO. Motor de busca: O que é?. Disponível em: <http://www.i-tecnico.pt/motor-de-busca-o-que-e/>. Acesso em: 04 de Outubro de 2017.

LIRAVIEIRA CORRÊA, Dominique; TRZESNIAK, Piotr; NONATO MACEDO SANTOS, Raimundo. A interpretação semântica de textos científicos em português: Uma perspicácia em Ciência da Informação. Universidade Federal de Pernambuco: Bahia, 2016. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/viewFile/3993/2562>. Acesso em 05 de Outubro de 2017.

MAGALHÃES, Terezinha Moreira de. O Motor de Busca Google e o Algoritmo PageRank. Minas Gerais: FESJ, 2015. Revista eletrônica.

MALVASTYLE. Search Engine Web Crawler. Disponível em: <https://goo.gl/4s6Ysr>. Acesso em: 05 de Dezembro de 2017.

MARKETINGDECONTEÚDO. Nova atualização do Google Penguin: o que mudou no SEO. Disponível em: <https://marketingdeconteudo.com/atualizacao-do-google-penguin/>. Acesso em: 24 de Setembro de 2017.

MOZ. Google Panda. Disponível em: <https://moz.com/learn/seo/google-panda>. Acesso em: 11 de Outubro de 2017.

NEIPATEL. O Guia Definitivo das Atualizações no Algoritmo Google. Disponível em: <https://neilpatel.com/br/blog/o-guia-definitivo-das-atualizacoes-no-algoritmo-do-google/>. Acesso em: 26 de Setembro de 2017.

PARISER, Eli. O filtro invisível: o que a internet está escondendo de você. Tradução Diego Alfaro. Rio de Janeiro: Jorge Zahar, 2012. Disponível em: <https://books.google.com.br/books?id=R9zb5nAfsW8C&printsec=frontcover&dq=O+filtro+invis%C3%ADvel&hl=ptBR&sa=X&ei=A36kU6xc57KxBLekgfgK&ved=0CCYQ6AEwAA#v=onepage&q=O%20filtro%20invis%C3%ADvel&f=false>. Acesso em: 26 de Outubro de 2017.

PEIXOTO, P. (2008). Motores de busca. Disponível em: http://www4.fe.uc.pt/fontes/pesquisa_na_internet/motores_busca/motores_de_busca.htm. Acesso em: 15 de Abril de 2017.

PET NEWS. História da Computação. Disponível em: http://www.dsc.ufcg.edu.br/~pet/jornal/abril2014/materias/historia_da_computacao.html. Acesso em: 14 de Agosto de 2017.

PORTAL WEBMARKETING. Como funciona um motor de busca?. Disponível em: <https://goo.gl/My11gu>. Acesso em: 04 de Outubro de 2017.

SEOMASTER. Google Panda Update: O que mudou?. Disponível em: <http://www.seomaster.com.br/blog/google-panda-update-o-que-mudou>. Acesso em: 24 de Setembro de 2017.

SEOMASTER. O que é o Hummingbird Update do Google e Como Revolucionou o SEO. Disponível em: <http://www.seomaster.com.br/blog/o-que-e-o-hummingbird-update-do-google-e-como-revoluciona-o-seo>. Acesso em: 24 de Setembro de 2017.

SILVEIRA, Marcelo. Web marketing: usando ferramentas de busca. São Paulo: Novatec, 2002.

STATCONTEUR. **Global Stats**: Search Engine Market Share Worldwide - September 2017. Disponível em: <https://goo.gl/qQvdRU>. Acesso em: 03 de Dezembro de 2017.

SUA PESQUISA.COM. História da Internet. Disponível em: <http://www.suapesquisa.com/internet/>. Acesso em: 14 de Agosto de 2017.

TAURION, Cezar. Cloud computing: computação em nuvem: transformando o mundo da tecnologia da informação. Rio de Janeiro: Brasport, 2009.

UOL. Folha de S. Paulo: Internet foi criada em 1969 com o nome de “Arpanet” nos EUA. Disponível em: <http://www1.folha.uol.com.br/folha/cotidiano/ult95u34809.shtml>. Acesso em: 14 de Agosto de 2017.