

A busca (im)perfeita: Humanos e Técnicas nos caminhos por informação.

The (in)perfect search: Humans and Technics in the pathways for information.

Luzo Vinicius Pedroso Reis¹
José Carlos Leite²
Dolores Cristina Gomes Galindo³

Artigo recebido em para publicação em ago./2014 e aceito para publicação em out./2014

RESUMO

Este ensaio tem como proposta confrontar o discurso da busca perfeita, proferido sobre a evolução das ferramentas de busca na web. Tal discurso enuncia que estamos caminhando em direção a buscadores capazes de oferecer respostas precisas para qualquer tipo de pergunta feita por operadores humanos. Verificaremos neste trabalho como esse discurso se sustenta, bem como os processos de funcionamento das ferramentas de pesquisa na web e o modo como estas se relacionam com os usuários humanos produzindo caminhos por informação. Para tanto, faremos uma revisão bibliográfica do assunto aliada à descrição de relações entre humanos e máquina de busca.

Palavras-chave: Busca Perfeita. Buscadores. Web. Relação humano-técnica.

ABSTRACT

This essay proposes to confront the discourse of the perfect search, said about the evolution of web research tools. Such epeech, states that the searchers are evolving towards capacity offering accurate answers to any question posed by human operators. This work will check how this discourse is sustained and what it hides, how is the actual mode of operation of the search tools on the web and how they relate to human users producing pathways for information. To do so, we will make a literature review of the subject together with the description of the operation of man and search machine.

Keywords: Perfect Search. Search Engines. Web. Technical-human relationship.

¹ Mestre em Estudos de Cultura Contemporânea pela Universidade Federal de Mato Grosso, na linha de pesquisa: comunicação e mediações culturais (2012). Produtor Cultural na mesma Universidade. Interesse por pesquisas nas áreas de tecnologia e sociedade, audiovisual e fotografia. E-mail: luzoreis@gmail.com.

² Doutor em Comunicação e Semiótica pela Pontifícia Universidade Católica de São Paulo (2001) e pós-doutor pelo Instituto Tecnológico de Coimbra. Professor Associado III da Universidade Federal de Mato Grosso. Tem experiência na área de Filosofia, com ênfase em Epistemologia, atuando principalmente nos seguintes temas: colonialidade, naturalismo, historicismo, perspectivismo. E-mail: j.leite@uol.com.br.

³ Doutora em Psicologia Social pela Universidade Católica de São Paulo (PUCSP). Docente do Departamento de Psicologia da Universidade Federal de Mato Grosso (2013-2014). Possui experiência em Psicologia como área interdisciplinar, atuando principalmente nos seguintes interesses de pesquisa: Psicologia e epistemologia; Psicologia, governamentalidade e resistências; Psicologia, tecnociências e arte Contemporânea. E-mail: dolorescristinagomesgalindo@gmail.com.

INTRODUÇÃO

Desde o princípio da internet, a busca por informações é uma das dimensões mais importantes do processo de navegar pela rede. Imagine a enorme quantidade de informações disponíveis nas milhares de memórias que compõem a internet. Imagine agora se essas informações não estiverem de algum modo ordenadas para que as pessoas possam localizá-las. Ao longo da história da web, temos acompanhado a evolução do processo de busca no sentido de tornar a localização dos conteúdos cada vez mais rápida e eficiente. Saímos de uma busca onde o usuário tinha que percorrer enormes catálogos divididos por assuntos - Como o site *Yahoo*, do início da web - para outra em que complexos algoritmos relacionam palavras-chave com um conjunto de informações acumuladas sobre os usuários para fornecer resultados em poucas frações de segundo - como é o caso do atual site de pesquisa Google.

A pretensa velocidade e eficiência conseguidas por essas ferramentas deu margem para o surgimento de um discurso que ganha cada vez mais força: o discurso da busca perfeita. Tal discurso, alicerçado em uma concepção moderna de tecnologia⁴, garante que as ferramentas de busca na web estão cada vez mais inteligentes e que chegará um dia em que essas máquinas serão capazes de responder a qualquer tipo de pergunta feita pelos usuários da rede, mesmo as mais subjetivas. No entanto, algo nos chama a atenção nessa afirmação. Não obstante toda a automatização, rapidez e relevância dos resultados conseguidos atualmente, os usuários continuam, em muitos casos, não encontrando as informações que precisam sem levarem algum tempo em um processo de diálogo com as ferramentas de busca: alterando as palavras-chave, utilizando estratégias de busca avançadas, operadores booleanos⁵ etc. Se é assim, o que o discurso da busca perfeita invisibiliza e mobiliza para continuar a operar?

Existe, de fato, um aumento de velocidade na localização de muitas informações pelos buscadores atuais. Ao mesmo tempo, existe uma série de buscas que demoram para serem realizadas. Este segundo lado, a demora em muitas

⁴ Uma visão que entende a tecnologia como instrumento a serviço do homem em seu intento de manipulação e controle da natureza (FEENBERG, 2010).

⁵ Operadores booleanos são termos oriundos da teoria dos conjuntos da matemática inseridas nas pesquisas para comandar uma busca, restringindo ou ampliando a mesma. Na prática, são símbolos que produzem filtros estratégicos nos buscadores. Quando realizamos, por exemplo, a busca “água” + casa, o símbolo + é um operador booleano que orienta ao buscador o retorno de resultados que contenham a palavra “água” e a palavra “casa” na mesma pesquisa, sem estarem necessariamente contíguas. Da mesma forma, o símbolo “-“ serve para excluir uma palavra da pesquisa (“Brasil”- carnaval retornará resultados para a palavra Brasil que não contenham a palavra carnaval).

pesquisas, é ignorado pelo discurso da busca perfeita. Este discurso, de cunho mercadológico, se interessa no que a tecnologia tem a oferecer ao humano e não em como ela, de fato, funciona. Evoca, portanto, sua dimensão simplista e sedutora, onde a máquina não passa de um objeto para facilitar a vida das pessoas:

Um dia o Google conseguirá responder a perguntas como “que faculdade devo cursar?”. Não demorar alguns anos até que consigamos responder a essas perguntas, ao menos parcialmente. Mas o resultado final é que... o Google vai conseguir dar respostas para às perguntas mais hipotéticas” – Fala de Eric Schmidt, presidente do Google (PARISER, 2012: 37)

“O mecanismo de busca ideal entenderia exatamente o que queremos dizer e nos ofereceria exatamente o que buscamos” – Fala de Larry Page, co-fundador do Google (PARISER, 2012: 35)

Os excertos supracitados exemplificam o que estamos chamando de discurso da busca perfeita. Eles querem nos fazer crer que um dia as ferramentas de busca serão capazes de realizar uma busca ideal, ou seja, oferecer uma resposta precisa para qualquer pergunta realizada. Tal crença, propagada com entusiasmo por empresas e mesmo pelas pessoas de modo geral, revela, antes, a persistência do desconhecimento da técnica pelo humano no contemporâneo. Depositamos na técnica uma esperança de resolução imediata dos nossos problemas, mas ignoramos seu real funcionamento e limites. Continuamos a vê-la como instrumento e não como uma dimensão fundante de nossa própria existência. Apesar de estarmos cercados por máquinas e atrelados a elas mais do que nunca, cultivamos um desconhecimento profundo destes seres, de seu funcionamento e, portanto, do discernimento que nós somos, em última instância, seres técnicos, constituídos do ritmo das máquinas que interagem conosco (SIMONDON, 1958). O caminho para esse entendimento, como revelou Gilbert Simondon, passa pela análise do funcionamento das máquinas, pela observação de seus pontos críticos de funcionamento e pela observação de como o humano participa desse processo. No caso da busca na web que analisaremos, isso significa ver o que está por trás da busca perfeita, ou seja, como, de fato, máquinas de busca e operadores humanos interagem nos caminhos por informação.

Desse modo, há que se investigar duas coisas: Em primeiro lugar; o que alicerça o discurso da busca perfeita. O que faz com que ela seja creditada e valorizada. Para tanto, investigaremos, brevemente, os recentes modos de rastreamento e de acúmulo de informações pelas máquinas do ciberespaço que tornaram possível o aumento das velocidades de busca para diversos casos, em especial com o advento

da chamada web semântica. Em segundo lugar; por que, em outros tantos casos, as buscas continuam demorando em serem realizadas, exigindo um diálogo mais intenso dos pesquisadores com as ferramentas de busca. Nessa segunda parte, olharemos de perto o que acontece desde a postagem das informações na rede até a busca pelas mesmas por usuários diversos. Como elas são processadas e o que impede que os pesquisadores encontrem àquela informação são as perguntas que orientam essa parte do trabalho. Estamos aqui interessados por tudo aquilo que atrapalha a busca perfeita. Ao final poderemos ver, de forma um pouco mais clara, o produto da interação de humanos e máquinas nas buscas pelo ciberespaço e verificar que ao contrário do que se imagina, é a imperfeição, ou antes, a construção de um caminho cheio de desencontros, que parece ser a regra nesse processo.

ORGANIZANDO OS ÍNDICES NA WEB 2.0

A fim de esclarecer minimamente a “caixa preta” das atuais ferramentas de busca⁶, discorreremos nesse momento um pouco sobre o trabalho dos algoritmos que o compõe. Afinal, como realizam a tarefa de organizar as informações da rede para disponibilizá-las aos usuários? A princípio, falaremos sobre o funcionamento dos softwares que compõe os índices⁷ de busca. Depois, comentaremos um pouco sobre a entrada e o papel do usuário nesse processo.

Inicialmente, um programa denominado *crawler*, também conhecido como *spider* ou indexador automático, faz varreduras na web e cria cópias das informações que encontra em um banco de dados. Seu funcionamento é assim descrito por Branski (2000):

Estes programas vasculham a internet visitando os sites, lendo seu conteúdo e seguindo seus links para outras páginas. Alguns índices indexam integralmente o conteúdo dos sites, outros somente o título e um resumo algorítmicamente construído, outros o título e as primeiras linhas do site. De toda forma, cada endereço encontrado é registrado e passa a fazer parte do banco de dados da ferramenta de pesquisa (BRANSKI, 2000: 12).

⁶ No início da *World Wide Web* (web) os buscadores eram bem diferentes dos atuais. Naquele tempo, anos 1990, os usuários entravam no site do buscador e ao invés de dialogarem com um sistema colocando palavras-chave para o retorno de resultados, como acontece hoje, eles visitavam as informações dispostas em categorias previamente editadas, com assuntos como artes, esportes, lazer, política etc. Segundo Batelle, o grande exemplo desse tipo de busca foi o site Yahoo, lançado em 1994 (BATELLE, 2006: 49). O que chama a atenção nesse processo é ainda a pouca automação do processo. Esses catálogos eram produzidos e editados por seres humanos, funcionários do Yahoo que indexavam e catalogavam o que encontravam na rede.

⁷ Índice é o nome da modalidade de buscador que estamos analisando. Ele se caracteriza pelo processo de diálogo com um operador humano que irá inserir termos de busca (palavras-chave) para o retorno da pesquisa. Desse modo, diferencia-se da abordagem dos catálogos de busca, que classificavam os sites por assuntos e oferecia os mesmos aos usuários para que estes, ao modo de uma biblioteca tradicional, procurassem sua informação entrando nas seções.

Como todos os documentos na web estão, de algum modo, conectados, a varredura de um indexador automático consegue, através da passagem de uma conexão a outra, se estender a toda a web. É certo que com o aumento contínuo de conteúdo, decorrente de novas postagens, essa tarefa nunca consegue ser definitiva. Os *crawlers* precisam trabalhar regularmente para manter os bancos de dados atualizados.

Uma vez dentro do banco de dados, que é o índice propriamente dito, as informações capturadas pelos *crawlers* precisam ser melhor ordenadas. Para isso existem algoritmos que trabalham analisando e classificando essas informações, os endereços das páginas, as palavras mais relevantes, as conexões etc. Esse índice é conectado ao site do buscador, que é uma interface entre ele e o usuário. Ao inserir palavras-chave para realizar nossas buscas estamos, portanto, dialogando com os índices das empresas de busca e não diretamente com todo o conjunto de sites da web, é esse o motivo para os resultados retornarem tão rapidamente. O resultado das nossas buscas é uma lista de informações conectadas com o conteúdo original da web que os algoritmos das ferramentas de busca relacionaram com as palavras-chave que digitamos.

Desse modo, podemos concluir que um buscador é composto por três etapas: 1) Rastreamento das informações na web pelos *crawlers*; 2) Armazenamento das mesmas em bancos de dados⁸; e 3) Interface que conecta os usuários ao banco de dados. Em todas essas etapas funcionam ferramentas tecnológicas que trabalham para nos apresentar resultados. Elas procuram as informações na web e as armazenam; analisam-nas, classificam-nas e interagem com usuários humanos relacionando os termos de busca inseridos por estes com todo o seu trabalho anterior. São programas e algoritmos que intervêm de alguma forma nos resultados e, conseqüentemente, nos nossos caminhos em busca das informações.

Quando o usuário encontra-se em um site de busca ele está, portanto, diante de um sistema que já realizou todo um trabalho de ordenação. Trabalho este não mais desenvolvido por operadores humanos, como era nos primeiros catálogos da web, mas por máquinas. Daqui para frente, inicia-se uma nova etapa no processo de busca. Os usuários terão que dialogar com o sistema de ordenamento de informações do site. Para terem sucesso e encontrarem o que buscam, terão que

⁸ Os bancos de dados são chamados de servidores. São supercomputadores com capacidade aumentada de memória. Empresas como o Google têm uma enorme estrutura física para comportar seus servidores.

colocar em um campo em branco palavras-chave que atendam a dois requisitos simultaneamente: a capacidade de traduzir a necessidade da sua busca e a adequação à lógica de ordenação previamente programada pelo sistema.

Essa etapa pode tornar-se uma tarefa nada fácil, especialmente aos neófitos. Mesmo usuários mais experientes, por vezes encontram alguma dificuldade para encontrar informações através dos buscadores. Ainda hoje, com toda a evolução dessas ferramentas, ocasionalmente é necessário refazer as buscas, trocar as palavras-chave, pensar em termos que atendam melhor os requisitos do sistema. Apesar de toda a automação que o processo de busca ganhou, a busca ainda depende, em grande medida, dos usuários e de sua capacidade imaginativa para operar os buscadores.

É claro que se compararmos o grau de participação humana na busca atual com a realizada anteriormente com os catálogos, veremos que foi bastante minimizada. No entanto, ela ainda existe e desempenha papel fundamental. É preciso pensar para chegar ao melhor termo de busca, às melhores palavras-chave, para usar filtros de pesquisa e operadores booleanos. Ao contrário do que popularmente se diz, de que o Google, por exemplo, tem a resposta rápida a todas as nossas perguntas, o que acontece, em uma série de casos, é um processo de jogo com o sistema. De tentativa e erro até o encontro das respostas. Esse processo, todavia, ficou ainda mais enevoado com a mais recente fase da web: a web semântica ou 3.0.

NAVEGADOR CONSUMIDOR: A BUSCA NA WEB SEMÂNTICA

O funcionamento dos índices de busca apresentados acima pertencia à lógica da web 2.0. Hoje, com o advento da chamada web semântica, ou web 3.0, foram implementadas novas tecnologias. Nessa nova lógica, as ferramentas do ciberespaço, especialmente os buscadores, estão cada vez mais inteligentes. Em tese, tratam-se de máquinas capazes de entender quem são seus operadores humanos para adequar a experiência da navegação a esse indivíduo. Temos duas perguntas: Em primeiro lugar, o que é, exatamente, essa inteligência das máquinas? Em segundo: com o aumento da automação na busca, como se dá a participação do usuário no processo?

Segundo Pariser (Op. Cit: 8), desde dezembro de 2009, o Google retorna os resultados que seu algoritmo “sugere ser melhor para cada usuário específico”.

Nesta lógica, duas pesquisas com o uso das mesmas palavras-chave, realizadas por usuários diferentes, podem retornar resultados diferentes, dependendo do filtro a que pertencem. Vamos entender como isso acontece. Para retornar resultados diferentes para cada usuário o Google precisa saber quem eles são. Quais são os seus nomes, onde moram, trabalham, quem são seus amigos, seus namorados e cônjuges, de que gostam, o que e quando consomem, sua orientação sexual, hobbies etc. O conceito é simples: quanto mais informações se tem sobre algo, mais o conhecemos. É através desse princípio do acúmulo de informações sobre os usuários que operam os dispositivos tecnológicos da web semântica e de muitos buscadores e sites atuais, inclusive o Google.

Além de acumular informações sobre o conteúdo dos sites da web, o Google e diversos outros sites acumulam, agora, também informações sobre seus usuários. Os dispositivos que trabalhavam varrendo, indexando e relacionando conexões, fazem na web semântica, além disso, a coleta e análise de informações sobre usuários e seus hábitos de navegação. Na verdade, o acúmulo de informações pelas máquinas não é, em si, algo novo. Uma das primeiras estratégias para a coleta de dados foi a instalação de *cookies* nas máquinas dos usuários: programas que marcam informações sobre a navegação. Inicialmente eles serviam para a gravação de senhas, *logins* e outros dados de acesso para facilitar a vida das pessoas na web. Hoje, os *cookies* gravam diversos outros tipos de informações, como preferências de informação - usados para prover futuros anúncios e conteúdo. Além deste dispositivo, existem diversos algoritmos que trabalham anotando a localização e a quantidade dos cliques dos usuários.

Já o Google, o gigante da rede, adotou como principal estratégia para conhecer os usuários, a disponibilização de serviços gratuitos (e de ótima qualidade) que exigem que as pessoas se conectem à contas particulares. O *G-mail*, serviço de e-mail do Google, é um dos mais populares, sendo uma enorme fonte de informações para a empresa. Ao conectar-se, o usuário é reconhecido como dono do perfil construído a partir de seus hábitos anteriores, que foram coletados durante a navegação pelos serviços da empresa: e-mails no *G-mail*; buscas no Google; vídeos no *YouTube*; notícias no Google notícias etc. Todos os dados oferecidos pelos usuários em troca dos serviços gratuitos do Google ajudaram a empresa a criar uma verdadeira teoria do comportamento dos usuários na internet.

Todos esses dados permitiram que o Google acelerasse o processo de construção de uma teoria sobre a identidade de cada usuário – os temas em que estavam interessados, os links em que clicavam. Em novembro de 2008, o Google já detinha várias patentes para algoritmos de personalização⁹ – códigos capazes de classificar as pessoas em grupos e então adaptar os resultados da pesquisa para que se adequem as preferências das mesmas. As categorias que o Google tinha em mente eram bastante específicas: na patente o Google usou o exemplo de “todas as pessoas interessadas em colecionar dentes de tubarões ancestrais” e “todas as pessoas não interessadas em colecionar dentes de tubarões ancestrais”. As pessoas da primeira categoria que pesquisassem, por exemplo, “incisivos de tubarão branco” obteriam resultados diferentes das pessoas da segunda categoria (PARISER, 2012: 36).

Mas afinal, o que difere os anúncios e resultados de hoje dos anúncios e resultados oferecidos na web 2.0? Antes de o Google personalizar os resultados por perfis de usuário, os resultados orgânicos (aqueles obtidos naturalmente, sem o pagamento para melhorar a classificação) eram os mesmos para todos os usuários. Essa lista de resultados orgânicos era fruto de um critério universal de relevância, no qual o trabalho do conjunto de algoritmos era ordenar os resultados de acordo com o julgamento dos milhões de usuários sobre as milhões de páginas da web. A propaganda, nesse caso, obedecia parcialmente ao mesmo critério de ordenação e se referia ao mesmo conjunto de termos de busca responsáveis pelos resultados orgânicos. Na prática, isso significava que uma busca por “serviços de informática”, por exemplo, retornava os mesmos resultados para todos os usuários situados em uma determinada localidade. Com a personalização da navegação, cria-se um filtro de personalização, e essa mesma busca retorna agora resultados segmentados por subgrupo de usuários. O navegador é também, e agora mais do que nunca, consumidor, pois seu caminho no ciberespaço é cada vez mais orientado pelo oferecimento de anúncios publicitários.

⁹ É importante entender que esse esforço para produzir uma navegação personalizada tenta responder a uma demanda mercadológica. A personalização é uma ótima forma de oferecer tráfego aos anunciantes, porque aumenta a possibilidade de que anunciem para o grupo de pessoas certas, ou seja, aquelas com propensão à compra daquele produto. Descobrir a fórmula para oferecer propagandas segmentadas e com maior garantia de retorno do investimento para outras empresas ajudou o Google a consolidar sua posição de líder de mercado. Pois, como Pariser comenta, “os anúncios contextuais que o Google coloca ao lado dos resultados de pesquisas e em sites são sua única fonte significativa de lucro” (PARISER, 2012: 41). Como a personalização funciona através do acúmulo de informações e da categorização dos internautas, consideramos que Pariser acerta quando afirma que, “recebemos um serviço gratuito, e o custo são informações sobre nós mesmos” (Op. Cit: p.12).

O discurso da máquina Inteligente, capaz de realizar uma busca ideal é potencializado agora na web semântica, como se algo revolucionário estivesse acontecendo. Contudo, se observarmos bem, trata-se do implemento de um filtro de personalização que classifica os usuários com base em dados de navegação acumulados. Obviamente que esse processo trouxe ganhos significativos na localização de conteúdo: encontram-se mais rapidamente muitas coisas e é esse ganho que sustenta o discurso da busca perfeita. No entanto, não nos esqueçamos das outras tantas informações que, a despeito de todos os filtros e grupos específicos que pertencemos nessa lógica de ordenação, ainda não conseguimos encontrar sem algum esforço.

A RESISTÊNCIA DAS MÁQUINAS E A CONSTRUÇÃO DE CAMINHOS POR INFORMAÇÃO

A lógica da personalização da web semântica tem, supostamente, o objetivo de proporcionar uma busca mais eficiente (mais rápida e precisa) aos usuários da internet. A ideia é “limpar” do caminho da pesquisa, todas aquelas informações que são pouco ou nada relevantes com base no que se conhece sobre cada usuário através da análise de seus dados acumulados. Nesse ponto, gostaríamos de explorar alguns aspectos da realidade do funcionamento da busca que revelam uma resistência ao intento do Google, e de outras empresas, de avançar na proposta de personalização rumo a uma busca ideal, sem perdas e resultados inesperados. Resistência que torna a busca asséptica e a pretensão de atingi-la através da personalização da web semântica algo bastante improvável, pelo menos de acordo com o funcionamento dos atuais mecanismos tecnológicos.

Começemos com um dos principais elementos dessa resistência: as práticas *spams*. Consideradas como maliciosas por terem o objetivo de utilizar a estrutura da internet para promover interesses próprios, é, no caso da busca, mais do que uma realidade que queremos evitar, um empecilho ao funcionamento pretendido pelas empresas de busca. Veremos como essas práticas funcionam, como se aproveitam dos algoritmos da busca e como, ao fazerem isso, acabam revelando o funcionamento desses objetos técnicos e a difícil conjunção de fatores que poderiam nos levar à busca perfeita alardeada pelos mentores do Google.

Poderíamos pensar que a existência de *spams* encerra uma das principais características do ciberespaço e da busca: estamos sempre, em alguma medida,

perdidos no dilúvio informacional (LEVY, 1999). Por mais que contemos com ferramentas de ordenação de informações, e estas estejam “nos conhecendo” mais atualmente, persiste certo caos no ciberespaço. Acreditamos que esta desordem é provocada por este ser um espaço de potência, onde ao mesmo tempo em que acontecem tentativas de controle e regulamentação de seu uso (em uma tentativa de limpeza e assepsia) por parte de governos e empresas, ocorrem invenções e estratégias de contra fluxo, por parte de usuários, que atrapalham as lógicas algorítmicas vigentes. Vejamos.

Battelle (Op. Cit.) chama de *spam* a prática de pessoas que utilizam o conhecimento que possuem sobre as ferramentas do ciberespaço para tirarem vantagem pessoal de forma maliciosa. Eles subvertem o que seria o uso adequado dos algoritmos e programas para criarem, geralmente, publicidade para seus produtos e ideias. No caso das ferramentas de busca, esses usuários criam estratégias para que os algoritmos dos buscadores entendam que seus sites merecem estar classificados entre os primeiros resultados. Antes do Google, quando os índices funcionavam basicamente procurando a correspondência entre os termos colocados pelos usuários e as palavras indexadas, sem uma classificação mais apurada, um tipo de *spam* que se tornou bastante comum foi o aplicado pelos sites pornográficos. Eles escondiam dentro de seus sites palavras-chave que eram bastante procuradas, como a palavra carro, por exemplo, de modo que os robôs de busca os indexassem como sendo um site de carros. Imagine uma busca por carros que retorne uma imensa quantidade de pornografia. Mesmo que uma pequena parcela de usuários entre nesses sites, já valeu a pena a aplicação do *spam*. Essa prática no final da década de 1990 foi responsável por um verdadeiro caos no mundo da busca na web. Como comenta Battelle:

(...) Depois que os emissores de *spam* (em particular, a indústria de entretenimento para adultos) perceberam que poderiam capturar tráfego com palavras-chave de alto tráfego como “carros” ocultando-as por todos os seus sites (com frequência, por exemplo, em pequenas cartas com fundo branco), em pouco tempo o modelo entrou em colapso. Foi por essa razão que, no final de 1998, a maioria dos resultados de uma busca por “carros” no Lycos era de sites pornográficos (BATTLELLE, 2006:89).

Esse exemplo trata do *spam* aplicado por internautas com maior grau de conhecimento sobre informática; contudo a prática *spam* pode ser muito mais simples. No *YouTube*, por exemplo, os algoritmos trabalham com as informações oferecidas pela pessoa que envia o vídeo. São os usuários que escolhem as

palavras-chave e fazem a descrição da produção. A prática *spam* aqui consiste na escolha deliberada de palavras-chave que não possuem nenhuma relação com a produção em questão, mas que retornam grande tráfego. O controle do Google nesse caso consiste em analisar as denúncias que porventura aconteçam. Após a denúncia, o vídeo é analisado pela equipe do Google, nomeadamente seres humanos responsáveis por assistirem ao vídeo e julgarem a procedência da denúncia.

Os *spams* introduzem desvios nos percursos de busca por informações na internet. Um *spammer* é considerado, com frequência, um contraventor do ciberespaço, uma pessoa que ignora algum sentido de colaboração em prol de uma vantagem pessoal. Deliberadamente usam o sistema e sua lógica para promoverem vídeos, publicidades etc. Na busca, especialmente no *YouTube*, eles são responsáveis por diversos links enganosos, que contam com uma descrição interessante aos pesquisadores potenciais, mas que quando acessados revelam algo completamente diferente do esperado. A busca perfeita encontra nos *spams* um desvio que requer das empresas a constante introdução de novas tecnologias capazes de identificar os agentes humanos e desmontar as brechas descobertas no funcionamento do sistema.

Contudo, para que o Google consiga oferecer uma busca perfeita não bastaria se livrar dos *spams*. Mesmo que isso fosse uma coisa simples e pudesse ser feita sem grandes prejuízos à liberdade de navegação não resolveria o problema. Pois, não são apenas os *spams* os responsáveis pela “sujeira” no processo de busca pelo ciberespaço. Muitos dos links enganosos não são oriundos de práticas *spams*, mas de uma falha de comunicação entre o pesquisador e as ferramentas de busca.

O que por vezes temos dificuldade de entender, e que os *spammers* parecem ser experts, é que a busca é um processo lógico matemático de processamento de dados. O Google - e qualquer outro buscador - é um conjunto técnico composto pelo funcionamento ordenado de algoritmos que desempenham funções específicas. É uma máquina que processa informações oferecidas por humanos. Isso pode parecer óbvio, mas é comum vermos as pessoas brigando com suas máquinas porque não conseguiram fazê-las funcionar. Chegam ao ponto de tratá-las como se fossem um ser humano: “O Google é burro!”. Uma máquina jamais pode ser adjetivada como inteligente ou não inteligente, pois inteligência é uma qualidade ou propriedade humana (SIMONDON, 2007: 77). Para técnicos de programação, o que ocorre

quando as máquinas não funcionam como gostaríamos é uma falha cuja localização passa por investigar o que ocorreu entre um ser que, em tese, “pensa” e outro que, em tese, “processa” informação, pessoa e objeto técnico.

Não estamos dizendo com isso que a máquina é um mero instrumento, que é passivo às ordens humanas. Na verdade, a técnica é tão ativa que impõe um determinado ritmo de vida, além de demandar informações do pensamento humano. No entanto, quando a informação oferecida não corresponde com a lógica da máquina há falha de comunicação. O ciberespaço está cheio dessas falhas, existe uma infinidade de informações que seriam as informações perfeitas para aquela busca, mas que não aparecem entre os primeiros resultados porque não estão adaptadas à lógica algorítmica vigente. Os *spammers* nada mais fazem do que dialogar bem com a máquina, mesmo que seja para uma finalidade que talvez seja reprovável pelos demais usuários do ciberespaço.

Battele (Op. Cit.) traz um exemplo interessante do que estamos dizendo. Ele relata a experiência do senhor Neil Moncrief, pequeno empresário norte americano, dono de uma loja de calçados grandes - números acima de 44 - que teve durante algum tempo seu site listado como primeiro resultado para quem fizesse uma busca procurando lojas de calçados grandes nos Estados Unidos. O senhor Moncrief relata que logo após a entrada do seu site na rede, o “2bigfeet.com”, seus negócios começaram a prosperar como nunca. Os clientes começaram a surgir de todos os cantos do país, encaminhados ao sr. Moncrief pelo Google. Contudo, subitamente os clientes começaram a desaparecer e as vendas caíram. Imaginando que a razão pudesse ter sido a falência do Google, o sr. Moncrief foi ao computador e fez a pesquisa por calçados grandes no buscador que tanto o tinha ajudado, nada! O site “2bigfeet.com” não aparecia nem mesmo entre os 100 primeiros listados. O que aconteceu foi que o Google modificou seus algoritmos para tentar conter os *spams* que estavam se proliferando em seus resultados, de modo que a nova configuração acabou atingindo também o site do sr. Moncrief, que nada tinha a ver com isso.

Agora vejamos, vamos assumir que o site “2bigfeet.com” tinha exatamente as informações ideais para uma pessoa que estivesse atrás de sapatos grandes. Considerando isso, poderíamos pensar que a mudança que o Google operou nessa época em seus algoritmos acabou distanciando tanto o pesquisador do resultado ideal que praticamente o tornou impossível. De fato, o que o Google faz é uma classificação dos sites de acordo com determinados critérios lógicos. Antes, com o

famoso algoritmo *pagerank* esse critério era a boa conexão dos sites. Como Battelle lembra:

(...) Ele simplesmente revelava a frequentemente feia verdade de quão bem conectado um site poderia ser. Não importando a boa aparência de um site, ou quantos prêmios ele tinha recebido, se ele não fosse conectado por outros sites – idealmente sites também bem conectados –, então, pela estimativa do Google, ele na realidade, não existia. (BATTELLE, 2006: 68).

Cansado de tentar reclamar ao Google sobre o desaparecimento de seu site dos primeiros resultados, o sr. Moncrief não teve outra escolha senão buscar uma alternativa para seu negócio. Depois de algum tempo o site “2bigfeet.com” voltou a ocupar os primeiros resultados. Perguntado sobre como conseguiu isso, o sr. Moncrief respondeu: “Nós limpamos um pouco o site e esperamos pacientemente, trabalhamos para subir!” (BATTELLE, 2006: 142). O que o sr. Moncrief fez foi, na verdade, um esforço de adaptação. Para voltar aos primeiros resultados foi preciso um trabalho para que os novos algoritmos do Google “entendessem” que o site do sr. Moncrief era relevante e merecia estar entre os primeiros listados para a busca de sapatos grandes. Provavelmente esse trabalho consistiu em uma melhoria dos textos do site de forma a destacar algumas palavras-chave, especialmente na primeira página, além de um trabalho de marketing para aumentar as parcerias com outros sites, de modo a aumentar a quantidade de links que levassem ao “2bigfeet.com”¹⁰.

Tudo isso revela que o Google não lê o conteúdo dos sites nem os entende, mas que faz uma ordenação dos mesmos através de determinada lógica. O que a personalização da experiência na web fez foi acrescentar mais um item lógico para o intento de uma busca mais relevante: um filtro de personalização. Esse filtro, porém, não tornou a busca mais inteligente, do ponto de vista de um entendimento humano, apenas se apresenta como mais um recurso para tentar cercar o que seria a informação mais adequada para um determinado grupo de pessoas.

Pelo exposto, poderíamos dizer que uma busca perfeita seria uma combinação improvável de alguns elementos, praticamente impossível de acontecer, como sugere a figura que segue:

¹⁰ Esse tipo de trabalho, responsável por melhorar a classificação das empresas nos resultados do Google, se tornou tão importante que hoje existem empresas e pessoas especializadas na prestação desse serviço. Chamado de SEO, sigla para *search engine optimization* - “otimização em motor de busca”, esse serviço é essencial para quem deseja manter-se entre os primeiros resultados orgânicos do Google. Há que se dizer que existe uma tênue linha que separa o trabalho em SEO do que é praticado pelos *spammers*. Pois, se prestarmos atenção, ambos cumprem a mesma tarefa de entender o funcionamento dos algoritmos para conseguirem promover algo, para dialogarem bem com o sistema técnico da busca. Estabelecer os limites entre uma atividade e outra é mais um desafio para os gestores e reguladores.

Figura 1: A busca perfeita.



Fonte: Autoria própria

O funcionamento da busca perfeita prevê uma comunicação entre humanos e máquinas sem modulações que a desviem em todos os níveis, desde a postagem de conteúdos bem adaptados à lógica algorítmica até pesquisas que, realizadas por alguém conhecedor das lógicas de ordenação, consigam encontrar os termos de busca mais adequados para chegar à informação correta. Obviamente que acidentes acontecem e pode-se chegar ao resultado perfeito sem acertar essa fórmula. Mesmo um pesquisador experiente, ainda pode ter o azar de que, caso a informação exista, não ter sido postada de forma a ser bem interpretada e classificada pelos algoritmos dos buscadores ou ainda que *spams* entrem em sua lista de resultados.

A busca perfeita é algo muito difícil de acontecer do ponto de vista de qualquer usuário ou usuária, mesmo com toda a tentativa de personalização da web semântica. Mas, as imperfeições – descaminhos, pesquisas frustradas etc. - vão sendo incorporadas por meio de correções nos algoritmos. Cria-se uma maquinaria de retroalimentação contínua na qual cada erro ou desvio converte-se em acerto: nada há nada que não possa ser rastreado. Dessa maneira, paradoxalmente, mesmo quase impossível de execução, a busca perfeita permanece como um imperativo que impulsiona os caminhos de usuários e desenvolvedores.

CONSIDERAÇÕES FINAIS

A improbabilidade do encontro da pesquisa perfeita com a informação perfeita torna a busca através dos algoritmos um processo de construção onde se coleta uma informação aqui, outra acolá. Tenta-se melhorar a pesquisa num processo incessante. Mesmo o Google tendo aprimorado a pesquisa na web, tornando-a mais relevante e rápida, não contamos com um sistema de busca que nos leve de imediato onde queremos. As velocidades sem dúvida aumentaram; encontra-se as informações hoje mais rápido que outrora e, no entanto, os usuários permanecem em certa medida tateando no escuro.

Percorremos as conexões do ciberespaço trocando informações com um conjunto de objetos técnicos de modo que nossa habilidade em nos comunicar com eles resolve apenas parte do problema. Ainda dependemos de todos os outros itens da equação. A improbabilidade da reunião de todos eles torna a localização das informações no ciberespaço um processo complexo de busca. Processo que produz seres filtradores, uma agência humano-algoritmos que trabalha para construir caminhos de conhecimento através dos labirintos do ciberespaço. Um agenciamento que se modifica e se constrói a cada mudança técnica, mas que nem por isso produziu homens que não estejam, em alguma medida, perdidos no excesso. Provavelmente a única chance para melhor aproveitarmos as potencialidades das máquinas da web seja aumentar nossa capacidade de entendimento sobre elas, tornando-nos, em algum nível, técnicos de programação ou simplesmente homens técnicos, como dizia Simondon (Op. Cit.).

REFERÊNCIAS

BATTELLE, John. **A busca**. Rio de Janeiro: Elsevier, 2006.

BRANSKI, Regina Meyer. **Localizações de informações na internet: características e formas de funcionamento dos mecanismos de buscas**. Transinformação, Campinas, v. 12, n. 1, p. 11 – 19, 2000.

FEENBERG, Andrew. **Racionalização democrática, poder e tecnologia**. In: NERDER, Ricardo T. (Org.). *Construção Crítica da tecnologia e sustentabilidade*. Brasília, Vol. 1, n. 3, 2010.

LEVY, Pierre. **Cibercultura**. São Paulo: Ed. 34, 1999.

PARISER, Eli. **O filtro invisível: O que a internet está escondendo de você**. Trad: Diego Alfaro. Rio de Janeiro: Zahar, 2012.

SIMONDON, Gilbert. (1958). **El modo de existência de los objetos técnicos**. 1º Ed. Buenos Aires: Prometeo libros, 2007.