

4 Incongruências entre os mecanismos de busca e a aprendizagem¹

Em sua publicação mais recente, o Comitê Gestor da Internet no Brasil (2013) identificou que 96% dos professores de Ensino Médio e Fundamental utilizam a internet na busca de conteúdos que serão trabalhados em sala de aula, assim como 82% dos alunos fazem uso do recurso para suas pesquisas escolares. Está vivenciando-se “[...] o desenvolvimento de [...] uma configuração social pautada num modelo digital de pensar, criar, produzir, comunicar [e] aprender [...]” (COMITÊ GESTOR DA INTERNET NO BRASIL, 2011, p. 28).

Por constituírem o meio preponderante de acesso à conteúdos no ambiente digital, os mecanismos de busca passaram a possuir profunda relação com os processos de aprendizagem e disseminação do saber. Os algoritmos de seleção de documentos utilizados por essas ferramentas tornaram-se responsáveis pela valoração positiva ou negativa dos conteúdos. Desse modo, é essencial que seja realizada uma análise da compatibilidade dos critérios de relevância adotados pelos mecanismos de busca com os aspectos cognitivos envolvidos na aquisição de conhecimento.

4.1. O conceito de relevância do Google

Ao interagir com o mecanismo de busca do Google, se está sempre limitado à camada mais externa da ferramenta. Possui-se acesso somente às interfaces visuais e não se percebe a dimensão dos processos que estão envolvidos no delineamento dos documentos apresentados como resposta aos questionamentos. Para cada pesquisa realizada, sequências de instruções computacionais previamente estabelecidas, conhecidas como algoritmos, dão forma ao conceito de

¹Conteúdos deste capítulo também foram publicados no artigo:

PENNA, Leonardo; QUARESMA, Manuela. Incongruências entre os mecanismos de buscas e a aprendizagem. In: Estudos em Design, v. 23, n. 3, p. 101-111, 2015.

relevância que orienta a geração dos resultados. Uma lógica particular é construída com base em pressupostos específicos sobre como se deve identificar um conhecimento e seus componentes mais importantes (Gillespie, 2014).

Como não existe uma métrica independente para o que realmente são os resultados mais relevantes de uma determinada consulta, os desenvolvedores devem decidir se o resultado parece "certo" e ajustar o algoritmo para atingi-lo [...] (Gillespie, 2014, p. 9, tradução nossa).

Por constituir seu maior diferencial frente ao mercado e, portanto, um patrimônio de valia comercial, a Google não revela detalhes a respeito do conjunto de mais de 200 indicadores (Google, 2014) utilizados pela sua ferramenta de busca durante o processo de seleção e ordenamento dos resultados. Uma compreensão mais extensa da lógica computacional inerente ao seu funcionamento apenas pode ser alcançada de forma indireta, pelo levantamento de dados que sugerem como os algoritmos operam e pelo posterior relacionamento das informações obtidas com aquelas de domínio público. Sob essa ótica, pode-se adotar como ponto inicial de análise da questão o estudo de Couvering (2007), em que foram avaliadas motivações e fundamentações que norteiam o trabalho dos responsáveis e desenvolvedores dos sistemas de pesquisa.

Por meio de uma série de entrevistas conduzidas com produtores dos principais mecanismos de busca e de alguns outros de menor expressividade (Google, Yahoo!, MSN, Ask Jeeves, AOL, Excite, Lycos, Infoseek, WebCrawler, entre outros), Couvering procurou revelar qual era o discurso que embasava as ações e estratégias de alocação de recursos nesse segmento do mercado tecnológico. A análise do material obtido nos contatos com funcionários ligados diretamente às decisões relativas à concepção e implementação dos códigos computacionais de busca permitiu que fosse constatada a existência de duas argumentações majoritárias: uma derivada da linguagem do *marketing* e outra pautada no pensamento científico-tecnológico.

A visão permeada pelo *marketing*, centrada no enquadramento da ferramenta de pesquisa como um negócio que envolve custos, está inserida em um mercado competitivo e deve gerar renda. As decisões relativas ao funcionamento do mecanismo de busca, incluindo as mudanças de algoritmo, estão usualmente associadas à mensuração da qualidade dos resultados, entendidos como positivos

sempre que alcançam a satisfação do consumidor e estimulam, desse modo, o crescimento da empresa.

O posicionamento científico-tecnológico identificado pela pesquisadora aponta uma defesa predominante dos procedimentos decorrentes da experimentação científica objetiva envolvendo constantes mensurações e testes. Dentro dessa lógica, as modificações da ferramenta devem estar sempre atreladas a dados que apontem um aumento da taxa de respostas adequadas às requisições dos usuários.

Embora o estudo de Couvering tenha sido efetuado em outra realidade temporal e tecnológica, pode-se encontrar reverberações dos discursos identificados por ela nas diretrizes que orientam os procedimentos de desenvolvimento das ferramentas de busca na atualidade. Evidenciando seus processos avaliativos bastante diretos e uma correlação entre a satisfação do consumidor e o conceito de relevância do buscador, a Google (2012) disponibiliza aos responsáveis pela avaliação da qualidade das pesquisas uma escala que estabelece uma hierarquia de valor para as páginas baseada no número provável de usuários que verão utilidade em seu conteúdo.

Essa associação entre relevância e popularidade pode ser encontrada nas origens da concepção do mecanismo de busca por Brin e Page. Em 1998, os dois publicaram em conjunto um artigo acadêmico descrevendo o PageRank, um algoritmo que veio a se tornar a base da ferramenta de busca criada posteriormente. Conforme documentação da própria empresa:

[Ele] interpreta um *link* da página A para a página B como um voto da página A para a página B. Mas o Google olha além do volume de votos, ou *links*, que uma página recebe: analisa também a página que dá o voto. Os votos dados por páginas ‘importantes’ pesam mais e ajudam a tornar outras páginas ‘importantes’. (Google, 2011)

Ainda hoje, a persistência desse modelo pode ser atestada. Em um estudo realizado no ano de 2013 pela empresa SEOMoz, constatou-se que, dentre os fatores que afetam o posicionamento de uma página no Google, o número de *links* em *sites* externos possui peso maior que a aparição dos termos pesquisados nas primeiras 100 palavras de conteúdo.

De forma indireta, até o valor dos tipos de conteúdo sofre com a necessidade de satisfação da maioria. Em seu manual de avaliação de pesquisas, a Google (2012) informa que palavras-chave que tenham interpretações ambíguas e possam sugerir interesses pornográficos devem ser tratadas de forma distinta das demais. Enquanto o padrão é avaliar a pertinência de um resultado de acordo com o significado mais comum da palavra-chave buscada, para termos associados a conteúdos adultos essa regra é inválida. Nesses casos, mesmo que a palavra seja usada majoritariamente em contextos associados à pornografia, os resultados devem ser julgados com base no significado menos usual do termo pesquisado. A consequência óbvia desse posicionamento da empresa é que os resultados desse segmento terão um *ranking* inferior por poderem ser ofensivos a uma parcela dos consumidores.

Como o foco da ferramenta é ampliar a base consumidora para que o negócio da empresa tenha êxito, a popularidade torna-se um termômetro fundamental na definição da importância dos resultados apresentados pelo mecanismo de busca. Sua preponderância acaba por desqualificar outros indicativos de qualidade como objetividade, imparcialidade, diversidade e representatividade. Sinais de relevância que possam gerar impacto negativo no objetivo primário intencionado com o serviço são negligenciados ou recebem peso reduzido nas fórmulas dos algoritmos.

Mesmo quando se distancia a análise dos aspectos diretamente associados à popularidade, continua-se a encontrar em outras políticas da Google um comportamento bastante característico do discurso entrelaçado às noções de *marketing*. A personalização dos resultados com base na localização do usuário e de seu histórico na web exemplifica essa correlação com clareza.

A fórmula dos gigantes da internet [...] é simples: quanto mais personalizadas forem suas ofertas de informação, mais anúncios eles conseguirão vender e maior será a chance de que você compre os produtos oferecidos. (PARISER, 2012, p. 7).

A realidade virtual é moldada pelos interesses do usuário e, sequencialmente, esses mesmos interesses moldam sua subjetividade, levando-o a reforçar a demanda pelos mesmos tipos de estímulos. Estabelece-se uma câmara de eco individualizada que favorece a supressão do acesso à pluralidade dos saberes. Novamente, as preocupações com a criação de um espaço informacional

que contemple os diferentes posicionamentos frente a um tópico perdem espaço para a necessidade de alcançar a ampliação do número de consumidores.

4.2. Popularidade e os processos de aprendizagem

Embora o Google utilize uma elevada quantidade de indicadores em sua seleção de resultados, viu-se anteriormente que o aspecto mais significativo de seu critério de relevância ainda é a difusão que um determinado material intelectual alcança. Mesmo que seja possível observar uma aparente validade em priorizar os pensamentos dominantes de um determinado momento histórico, essa bolha informativa possui um revés pouco nítido para os que estão envolvidos por seus limites. Conforme será visto a seguir, a repetição de informações — típica nos resultados selecionados de acordo com sua popularidade — pode ser prejudicial para a aquisição de conhecimento, pela sua incongruência com aspectos cognitivos dos indivíduos.

4.2.1. Estímulos repetidos

O modelo neurológico dos indivíduos evidencia a necessidade de permitir acesso ao desconhecido.

Parte do nosso sucesso como espécie pode ser atribuído ao interesse persistente do cérebro por novidades [...]. Ele está constantemente explorando o ambiente em busca de estímulos. Quando um estímulo inesperado surge [...], uma descarga de adrenalina fecha todas as atividades desnecessárias e foca a atenção do cérebro para que ele possa entrar em ação. Reciprocamente, um ambiente que contém estímulos majoritariamente previsíveis ou repetidos [...] reduz o interesse do cérebro no mundo exterior [...]. (SOUSA, 2011, p. 29, tradução nossa)

Conforme amplia-se o contato com informações novas, possibilita-se o enriquecimento dos processos de reflexão e construção do saber. “Uma qualidade maravilhosa do cérebro humano é sua capacidade de combinar itens individuais de muitas maneiras diferentes. À medida que se acumula mais itens, o número de combinações possíveis aumenta exponencialmente” (SOUSA, 2011, p. 56, tradução nossa).

[As] memórias não são armazenadas intactas. Em vez disso, elas são armazenadas em pedaços e distribuídas em locais ao longo do cérebro. A forma, a cor e o cheiro de uma laranja, por exemplo, são categorizados e armazenados em conjuntos de neurônios. Ao ativar esses locais simultaneamente, tem-se uma lembrança dos pensamentos e das experiências envolvendo uma laranja. (SOUSA, 2011, p. 84, tradução nossa).

Os locais de armazenagem que serão selecionados são determinados pelo número de associações que o cérebro faz entre aprendizagens novas e do passado. Quanto mais conexões são feitas, mais compreensão e significado o estudante pode anexar ao novo aprendizado, e mais provável é que ele seja armazenado em redes diferentes.

Seguindo essa linha argumentativa, pode-se inferir que a prevalência de resultados populares nos mecanismos de busca acaba por criar uma situação de empobrecimento intelectual. As correntes de pensamentos populares são valorizadas e, como consequência direta, tem-se um contato repetido com o sabido e uma privação de acesso a novos conteúdos.

4.2.2. Atrofia da divergência

Na utilização dos mecanismos de busca, as informações que não estão alinhadas com um pensamento dominante sobre um tópico não aparecem em posições de destaque e, dessa forma, não se fazem visíveis para as pessoas. E esse processo tende a se agravar com o tempo, uma vez que as páginas que não são encontradas, acessadas e referenciadas, tendem a ocupar posições cada vez piores no *ranking* do sistema. É um mecanismo de atrofia da divergência que vai de encontro com a essência de descentralização e universalidade da internet. Embora, em teoria, o espaço amostral de uma busca no Google esteja na escala de trilhões de páginas (Google, 2013b), a realidade é que parte dessas informações são mantidas inacessíveis em um limbo virtual.

A correlação entre relevância e popularidade, em última instância, torna os mecanismos de busca um entrave para o pensamento divergente. Definido como um processo de geração de ideias em que um indivíduo é confrontado com problemas ou questões para as quais não existe apenas uma resposta (Guilford, 1950; Runco, Dow & Smith, 2006 apud Roue, 2011, p. 4), esse tipo de

pensamento é considerado uma medida direta da criatividade e uma importante característica nos avanços científicos e de engenharia (Kim, 2011).

Em um estudo realizado pela professora de psicologia educacional Kim (2011), foram obtidos resultados que mostraram que crianças até a quinta série possuem a mente mais curiosa e apta a produzir respostas originais. Depois disso, elas tendem a ampliar o grau de pensamento conformista, possivelmente influenciadas pelo estágio de pensamento convencional de Kohlberg (Runco, 2007 apud Kim, 2011, p. 291). De acordo com essa teoria, nesse estágio adolescentes e uma grande parcela dos adultos estão na maior parte do tempo imersos em uma noção de que as regras da sociedade são as bases para o que é certo e errado (Boeree, 2009).

Como os mecanismos de busca recorrem ao critério da popularidade e são utilizados como uma interface de realidade, é nítida sua capacidade de construção de um pensamento convergente no qual existe espaço para apenas uma única resposta, normalmente influenciada pelas expectativas sociais e pelas noções de certo e errado carregadas pelas pessoas. Indiretamente, portanto, essas ferramentas acabam por estimular um processo de supressão da criatividade e inovação.

4.2.3. Heurística da disponibilidade

A heurística da disponibilidade estabelece que os eventos facilmente lembrados são julgados como sendo mais prováveis que os eventos lembrados com maior dificuldade (Tversky & Kahneman, 1973 apud Goldstein, 2011). Utilizada durante o raciocínio indutivo, em que se está buscando generalizações a partir da análise de casos particulares, essa heurística pode levar a falsas conclusões sempre que eventos pouco frequentes se destacarem em nossa memória (Goldstein, 2011).

Em um experimento relatado por Goldstein (2011), Stuart McKelvie (1997) demonstrou esse fenômeno apresentando uma lista aos participantes com 26 nomes, sendo 12 deles famosos (Ex: Mick Jagger) e 14 comuns. Divididos em dois grupos, metade dos sujeitos da pesquisa receberam uma lista em que os nomes famosos eram de mulheres e os comuns de homens. A outra metade interagiu com uma lista oposta, em que os nomes famosos eram de homens e os

comuns, de mulheres. Quando foi requisitado aos participantes que estimassem se existia mais pessoas do sexo masculino ou feminino na lista que leram, a grande maioria afirmou que o gênero com maior representatividade numérica era aquele que continha os nomes famosos. Consistente com a heurística da disponibilidade, o resultado encontrado mostrou que a facilidade de lembrança influenciou os participantes a responderem de forma enviesada a pesquisa.

Esse aspecto cognitivo evidencia que, durante a aprendizagem de tópicos em que não existe consenso sobre a resposta correta, deve-se evitar um contato quantitativamente maior com fatos que defendam uma determinada posição. Para que a aquisição de conhecimento não seja enviesada, as diferentes correntes a respeito do assunto devem, idealmente, possuir representatividade similar.

Os mecanismos de busca, por associarem seu critério de relevância à popularidade, estimulam a utilização da heurística da disponibilidade e aumentam as chances de os usuários alcançarem conclusões equivocadas. Sempre que houver informações concorrentes com difusões distintas, o maior número de resultados da teoria hegemônica acabará colaborando com a formulação de um raciocínio distorcido. Como caso ilustrativo, pode-se imaginar um cenário em que uma pessoa está fazendo uma pesquisa com intuito de descobrir qual doença é compatível com seus sintomas. Por ter tido um surto recente de dengue no país, a maior parte dos registros que o usuário encontra é a respeito dessa doença, embora ela seja bastante incomum para os sintomas que ele está experimentando. Pela representatividade desigual das enfermidades nos resultados da pesquisa, é provável que a conclusão do usuário seja a de que está com dengue, quando na verdade a maior probabilidade é que a resposta seja outra doença.

4.2.4. Predisposição para confirmação

Devido aos algoritmos utilizados na seleção dos resultados, os mecanismos de busca apresentam sempre as visões hegemônicas para os tópicos divergentes, em que existe mais de uma resposta plausível. Exceto pelos casos em que as informações concorrentes possuem representatividade similar, o número de resultados nas primeiras posições tenderá a apresentar conteúdos relativos à corrente de pensamento preponderante. Para além da supressão da divergência

informacional, essa característica dos buscadores cria dificuldades diretas para a aquisição de conhecimento.

Um dos principais obstáculos para o raciocínio preciso é a predisposição de confirmação, a nossa tendência de olhar seletivamente para a informação que está de acordo com nossa hipótese e de ignorar informações que argumentam contra ela (Goldstein, 2011, p. 374, tradução nossa).

Quando um leigo entrar em contato com os resultados sobre um assunto desconhecido, o primeiro entendimento será o hegemônico e, a partir desse momento, seu olhar passará a estar enviesado para confirmação da teoria inicial.

Para exemplificar esse fenômeno, pode-se citar a pesquisa realizada por Charles Lord em 1949 e descrita posteriormente por Goldstein (2011). O experimento envolveu dois grupos, um a favor da pena de morte e outro contra. A cada participante foram apresentados alguns estudos contendo evidências de que a pena de morte tinha efeito sobre o número de assassinatos e outros que mostravam que não havia nenhuma relação. Ao reagirem sobre os estudos, os participantes mantiveram suas posturas iniciais, rotulando as informações contrárias à sua visão como pouco convincentes.

A predisposição de confirmação age como uma venda - o mundo é visto de acordo com as regras consideradas corretas e nunca se é dissuadido deste ponto de vista, pois busca-se apenas evidências que confirmem essas regras (Goldstein, 2011, p. 374, tradução nossa).

Uma possível solução para essa inconsistência entre o aparato cognitivo dos sujeitos e o funcionamento dos mecanismos de busca seria o agrupamento de resultados com conteúdos altamente similares, para que houvesse espaço para teorias divergentes nas primeiras páginas. Dessa forma, existiria uma maior chance de entrar em contato com pensamentos distintos e seria evitado que a tendência de confirmação ocasionasse um impacto relevante no processo de raciocínio.

4.2.5. Retenção de aprendizado

O processo de aquisição de conhecimento não se restringe à exposição de um sujeito a um conteúdo inexplorado.

A atribuição de sentido e significado a um (...) aprendizado ocorre apenas se o aluno tiver tempo suficiente para processá-lo e reprocessá-lo. Este contínuo reprocessamento (...) é um componente crítico na transferência de informação da memória de trabalho para a armazenagem de longo prazo. (Sousa, 2011, p. 92, tradução nossa)

Quando ele ocorre por meio de processamentos elaborativos mais complexos (analíticos, avaliativos e criativos), em que se cria relações do que está sendo estudado com conhecimentos prévios, existe uma maior probabilidade de que o novo aprendizado seja consolidado e armazenado para uso futuro (Sousa, 2011).

Para exemplificar o impacto que a associação de informações pode gerar na aquisição de conhecimento, pode-se citar o experimento relatado por Goldstein (2011) e realizado por John Bransford e Marcia Johnson (1972). Nele, os participantes foram expostos à seguinte passagem:

Se os balões estourassem, o som não seria capaz de chegar uma vez que tudo estaria muito longe do andar correto. A janela fechada também impediria o som de entrar, uma vez que a maioria dos edifícios tendem a ser bem isolados. Uma vez que toda a operação depende do fluxo constante de eletricidade, uma ruptura no meio do fio também poderia causar problemas. É claro que o colega poderia gritar, mas a voz humana não é alta o suficiente para chegar tão longe. Um problema adicional é que a corda pode se romper no instrumento. Então, não haveria acompanhamento para a mensagem. É claro que a melhor situação envolveria uma distância menor. Dessa forma, haveria menos problemas potenciais. Com contato cara a cara, um menor número de coisas poderia dar errado.

Embora o texto tenha sido apresentado a todos fora de seu contexto original, somente uma parcela pôde visualizar previamente uma imagem (ver Figura 3) que complementava e dava sentido a ele. Em um segundo momento, foi requisitado aos sujeitos que tentassem lembrar o trecho lido. Aqueles que tiveram contato com a ilustração tiveram um índice de lembrança duas vezes maior que os demais. Os dois itens combinados facilitaram a criação de uma história que fizesse sentido, resultando numa armazenagem de longo prazo mais eficaz.

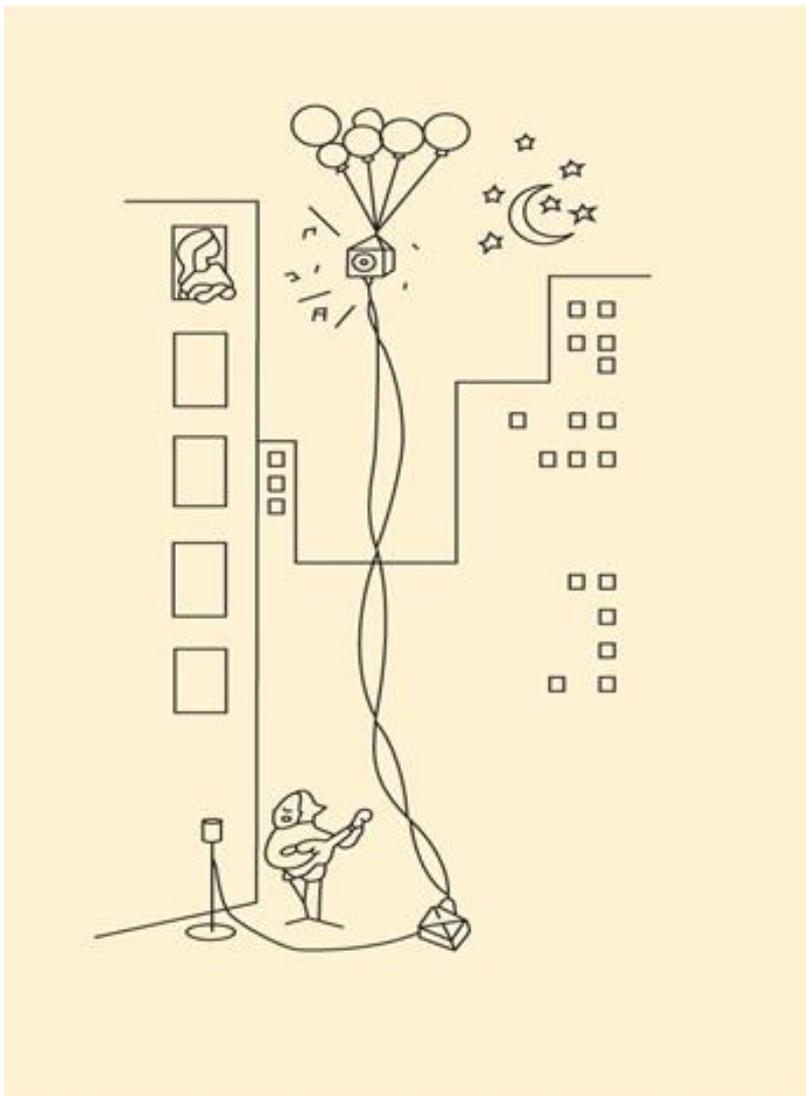


Figura 3 - Imagem utilizada por Bransford e Johnson (1972)

Fonte: J. D. Bransford & M. K. Johnson, "Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall," *Journal of Verbal Learning and Verbal Behavior*, 11, Figure 1, 717– 72., apud, GOLDSTEIN, 2011, p. 180.

Ao sobrepor as características relacionadas à retenção de aprendizado com o funcionamento dos mecanismos de busca, pode-se inferir que o critério de relevância adotado atualmente não apresenta as condições ideais para armazenagem de novos conhecimentos. Como essas ferramentas apresentam os itens mais populares nas posições de maior destaque, é comum que o usuário tenha acesso a conteúdos com grau de similitude elevado. A não ser que existam informações concorrentes com representatividade equivalente, haverá uma carência de informações distintas e divergentes, importantes para um raciocínio

analítico mais profundo e para uma melhor armazenagem na memória de longo prazo.