

Expressões de busca e o uso de diferentes operadores avançados de pesquisa em um mecanismo de busca

Search queries and the use of different advanced search operators in a search engine

Antonio Carlos Picalho ^{*1}, Luciane Maria Fadel ^{†2} e Alexandre Leopoldo Gonçalves ^{‡3}

¹Universidade Federal de Santa Catarina, Centro Tecnológico, CTC, Florianópolis, SC, Brasil.

²Universidade Federal de Santa Catarina, Departamento de Design e Expressão Gráfica, EGR, Florianópolis, SC, Brasil.

³Universidade Federal de Santa Catarina, Departamento de Computação, CTS/Campus Araranguá, Florianópolis, SC, Brasil.

Resumo

Pesquisar *on-line* é uma atividade presente na vida de muitas pessoas, principalmente no que diz respeito às pesquisas de rotina para resolver problemas do cotidiano. Simultaneamente, o volume de dados disponíveis na *web* cresce diariamente. Filtrar os resultados pode determinar a eficácia desta busca. Os mecanismos de busca também retornam outros elementos além dos resultados, compondo uma interface complexa. Essa complexidade pode dificultar o uso e desviar a atenção do usuário para esses elementos. Diante disso, o presente trabalho buscou examinar quais características nos resultados de pesquisa se destacam a partir do uso de operadores avançados na busca por documentos na *web*. Partindo do pressuposto de que a utilização desses operadores pode sugerir maior agilidade na busca e maior precisão aos resultados, a pesquisa identificou características apresentadas junto aos resultados de várias buscas. Por meio de uma pesquisa experimental, foram coletados e analisados os resultados presentes em 380 capturas de tela de usuários que reproduziram uma busca no Google em cinco cenários diferentes: busca livre; busca normalizada; busca com o operador booleano *AND* e acréscimo de aspas; busca com o operador *filetype:* e busca com o operador *site:*. Os resultados ressaltaram características como anúncios de pesquisa, sugestões de pesquisas já realizadas por outros usuários e quantidade aproximada de resultados. Entende-se que o uso de operadores avançados de pesquisa pode interferir não apenas nos resultados apresentados, mas também na sua forma de apresentação.

Palavras-chave: Buscadores. Ferramentas de busca na web. Recuperação da informação. Operadores booleanos. Google.

Abstract

Online searching is an activity that is present in the lives of many people, especially when it comes to routine searches to solve everyday problems. At the same time, the volume of data available on the web is growing daily, and filtering the results can determine the effectiveness of this search. Search engines also return other elements besides the results, composing a complex interface. This complexity can make it difficult to use and divert the user's attention to these elements. Considering this, the present study sought to examine which characteristics in search results stand out from the use of advanced operators in searching for web documents. Assuming that the use of these operators can suggest greater agility in the search and greater precision to the results, the research identified characteristics presented along with the results of several searches. Through an experimental research approach, the results from 380 user screen captures were collected and analyzed. The users replicated a search on Google in five different scenarios: free search; normalized search; search with the Boolean operator *AND* and the addition of quotation marks; search with the *filetype:* operator; and search with the *site:* operator. The results highlighted features such as search ads, suggestions of searches already performed by other users, and approximate result counts. It is understood that the use of advanced search operators can not only affect the presented results but also their presentation format.

Keywords: Search engines. Web search tools. Information retrieval. Boolean operators. Google.


Linguagem e Tecnologia

DOI: 10.1590/1983-3652.2023.47531

Seção:
Artigos

Autor Correspondente:
Antonio Carlos Picalho

Editor de seção:
Daniervelin Pereira
Editor de layout:
Leonado Araújo

Recebido em:
1 de agosto de 2023
Aceito em:
21 de outubro de 2023
Publicado em:
20 de novembro de 2023

Esta obra tem a licença
"CC BY 4.0".



*Email: tonipicalho@gmail.com

†Email: liefadel@gmail.com

‡Email: a.l.goncalves@ufsc.br

1 Introdução

Em pleno ano de 2023, realizar uma pesquisa de rotina é, para muitos, algo trivial, com práticas tão inseridas no cotidiano que, por mais que se repitam dia após dia, passam despercebidas, levando em consideração o quão automático essas práticas se tornaram. Wu, Zhu e Wu (2022) citam a busca como uma das atividades mais comuns inerentes ao ser humano, que constantemente está tentando satisfazer suas necessidades em diversos assuntos.

Líder mundial, o Google é o mecanismo de busca de maior relevância no mercado nas mais diferentes plataformas com acesso à internet. Dados da empresa irlandesa de análise de tráfego *web* Statcounter (2022) mostram que globalmente, de janeiro de 2009 a dezembro de 2022, a empresa permaneceu no topo, com uma média de 92.58% de uso, sem ameaças ao pódio, já que os demais colocados dividem médias abaixo dos 4%, sendo eles Bing, Yahoo!, Yandex, Baidu e DuckDuckGo, respectivamente. No Brasil, no mesmo período, o percentual sobe para 96.32%.

A própria gigante de buscas define a primeira página de resultados de pesquisa como uma posição valiosa (Google, 2021). Não é à toa que para isso existem inúmeras técnicas orgânicas de otimização como o *Search Engine Optimization* (SEO), além do tráfego pago. Aparecer bem-posicionado nos resultados do mecanismo de busca pode garantir mais acessos e engajamento com o conteúdo disponibilizado.

Tomando, por exemplo, a caixa de pesquisa disponibilizada pelo Google, o usuário digita alguns termos (podendo receber inclusive sugestões do próprio mecanismo de busca sobre como completá-los), clica em “pesquisar” e pronto. O que retorna na tela são inúmeros resultados que, segundo o mecanismo de busca, estão relacionados com aquilo que foi pesquisado. Podem ser muitos, podem ser poucos, podem ser exatamente aquilo que o usuário procura ou também podem nada ter a ver com o que ele precisava. São vários os cenários possíveis.

Em paralelo a esse cenário, o volume de dados disponíveis na *web* cresce dia após dia. Se hoje uma pesquisa qualquer de um documento em formato PDF buscou um dentre 1.000 conteúdos, no próximo ano poderá buscar esse mesmo único documento em um montante de 10.000 conteúdos. Mesmo os mecanismos de busca que captam preferências de buscas dos usuários e as utilizam para otimizar os resultados e a forma como cada um vê os conteúdos pesquisados exigirão de seus robôs uma varredura mais abrangente e exaustiva.

A informação que é recuperada por um usuário mediante o uso de um mecanismo de busca, para além de suprir as suas necessidades informacionais, contribui para a produção de novos conhecimentos (Choo, 2003). Assim sendo, Santos e Viera (2015) refletem que se a informação almejada for de fato encontrada, oportunizando a geração de conhecimentos, há chances desse novo conhecimento ser disponibilizado em rede. Isso corrobora o aumento no volume de dados disponíveis na *web* discutido anteriormente.

Considerando a facilidade em pesquisar qualquer tipo de conteúdo na *web*, sugere-se que a precisão — ou falta dela — nos resultados levanta questões sobre quão criteriosa é a construção e a execução das pesquisas de rotina. Afinal, basta digitar algumas palavras soltas para que muitas possibilidades surjam segundos depois.

Entretanto, mesmo a pesquisa em mecanismos de busca ter se tornado algo relativamente simples de ser realizado, é comum o usuário não encontrar aquilo que precisa, deparar-se com informações incorretas ou demorar mais tempo que o necessário para chegar em algo que seria relativamente simples de encontrar.

Aparentemente, pesquisar na *web*, por mais mecanicamente simples que seja, continua sendo um “exercício de sobrevivência do usuário mais apto” que, ao possuir maiores conhecimentos sobre a busca, se destaca por conseguir extrair o melhor dos mecanismos de busca (Vidmar, 1999, p. 1, tradução nossa). Em 2023, mesmo o padrão de busca *on-line* ter mudado radicalmente com a intensiva introdução de tecnologias de informação e comunicação, a constatação ainda permanece válida. Estudos na área da ciência da informação sobre ‘comportamento de busca de informações’ (*Information-seeking behavior*) seguem tratando dessa aquisição de habilidades para pesquisar na *web* (Mnguni; Kekana, 2022; González Guitián *et al.*, 2022). Nesse quesito, bibliotecários contribuem, ao atuar diretamente com o desenvolvimento de habilidades de busca em todo tipo de fonte, como

buscadores e no desenvolvimento da competência em informação (Lacey, 2022).

Atualmente, alguns mecanismos de busca contam com a possibilidade de combinar palavras-chave com operadores de pesquisa específicos a uma determinada função. Operadores são símbolos ou palavras que são adicionados à pesquisa para tornar os resultados mais precisos (Google, 2013). O uso de tais operadores permite a formulação de buscas complexas se necessário.

Nesse contexto, tem-se um problema de pesquisa a ser explorado: quais características nos resultados de pesquisa se destacam a partir do uso de operadores avançados na busca por documentos na *web*? Dito isso, este trabalho teve como objetivo, testar o uso dos operadores de pesquisa por diferentes usuários.

O pressuposto é que a utilização dos operadores avançados pode apresentar características mais uniformes, o que pode sugerir maior agilidade na busca e maior precisão aos resultados, possibilitando ao usuário maior controle e direcionamento durante uma pesquisa *on-line*.

2 Anúncios nos resultados de pesquisa

A popularidade do Google, em grande parte se deve ao fato da utilização de um algoritmo que estuda os costumes e aproveitamento dos usuários em relação a suas buscas e interação com resultados apresentados, produzindo resultados personalizados de acordo com tais medições (Yus, 2011). Fatores como a geolocalização do usuário detectada pelo tipo de dispositivo utilizado no momento da pesquisa influenciam na forma como o algoritmo entende que cada resultado deve ser classificado.

Os resultados de pesquisa do Google podem apresentar anúncios. Inicialmente, os anúncios pouco ou nada tinham a ver com a pesquisa executada pelo usuário. Com aprimoramento e ao passar dos anos isso mudou e os anúncios passaram a ter relação ou com a pesquisa executada pelo usuário naquele momento ou com o seu perfil e interesses já demonstrados anteriormente em plataformas integradas. Como pode ser observado nas categorias apresentadas na Figura 1.



Figura 1. Exemplo de fatores presentes nas categorias de um perfil de busca.

Fonte: (Google, 2022)

Essas e outras categorias contribuem para a personalização dos resultados de busca apresentados a um determinado usuário. Isso ocorre de acordo com três fatores distintos: informações pessoais que o usuário adicionou a conta do Google; estimativas do mecanismo de busca com base nas atividades em serviços da empresa como um todo (como pesquisas no YouTube) e dados de anunciantes que possuem parceria com o mecanismo de busca, dos quais são baseadas nos rastros digitais deixados pelos usuários, os famosos *cookies* (Rosa, 2022; Google, 2022).

Intituladas por campanhas de pesquisa, palavras-chave específicas são marcadas automaticamente pelo sistema ou manualmente pelo responsável pela campanha (Google, 2018). Os anúncios aparecem

sempre com a tag 'Anúncio' diferenciando-os dos resultados orgânicos da pesquisa.

A Figura 2 demonstra a diferenciação dos resultados, que podem se apresentar em duas posições distintas: em cima ou ao lado dos resultados orgânicos.

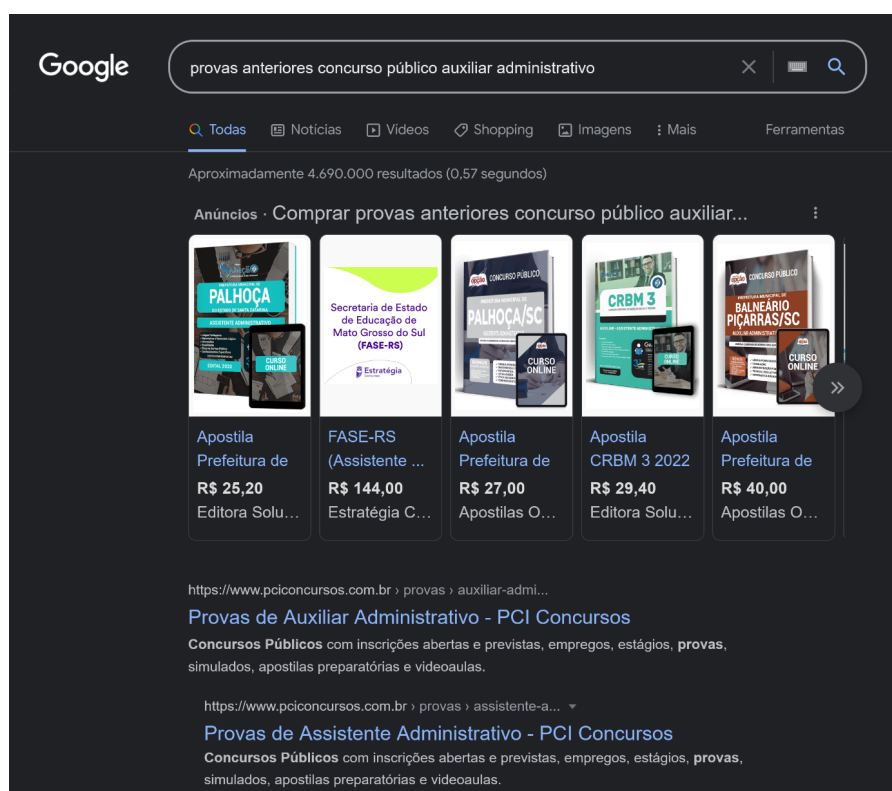


Figura 2. Posição dos anúncios nos primeiros resultados do Google.

Fonte: Dados da pesquisa – coletados dos respondentes

O estudo de Rosa (2022) reforça a ideia de que a maioria dos usuários entende que os anúncios nos resultados de pesquisa são personalizados e apresentados com base em seus interesses, entretanto, possuem conhecimentos mínimos acerca da personalização desses interesses.

Desse modo, apesar de haver certa ciência na forma como esses anúncios aparecem, e considerando que os resultados orgânicos e pagos são diferenciados visualmente pela plataforma, para o usuário que anseia por uma informação, um resultado é um resultado. Portanto, essa diferenciação pode passar despercebida. Questão essa que envolve vários aspectos acerca de como a informação é organizada e apresentada ao usuário na *web*.

Camossi, Amorim e Rodas (2023) discutem questões relacionadas a técnicas de SEO no que diz respeito ao processo de ranqueamento dos resultados apresentados pelos mecanismos de busca, o que, nesse caso, está atrelado também aos anúncios dispostos junto aos resultados orgânicos.

3 Operadores de pesquisa

Operadores de pesquisa são caracteres especiais que, ao serem adicionados a uma expressão de busca, permitem criar uma maior complexidade a fim de obter resultados mais precisos. Em tese, um sistema de recuperação da informação pode transformar qualquer palavra ou símbolo em um operador de pesquisa de acordo com uma finalidade específica de busca dentro do seu banco de dados. Os operadores apresentados são o AND, Aspas, filetype: e site: conforme o Tabela 1.

O AND é um operador booleano que corresponde à intersecção de dois ou mais termos. Numa busca, serve para garantir que o sistema apresentará resultados que contenham obrigatoriamente todos os termos ou expressões ligadas por ele (Picalho; Lucas; Amorim, 2022). No caso da expressão 'google AND Bing AND duckduckgo', os resultados retornarão documentos que citam os três termos descritos.

As aspas " " podem ser utilizadas sempre que existir um termo composto, frase completa ou até

Tabela 1. Descrição e função dos operadores avançados selecionados.

OPERADOR	FUNÇÃO RESUMIDA
AND	Unir dois ou mais termos a fim de recuperar resultados que representem uma intersecção entre eles
Aspas “ ”	Blindar termos compostos ou frases completas a fim de recuperar resultados com todos os termos na ordem exata em que são indicados
filetype:	Indicar ao mecanismo de busca o formato exato de arquivo que deseja
site:	Indicar ao mecanismo de busca o site exato no qual você quer ele busque e apresente resultados

Fonte: Dados da pesquisa (2022).

mesmo um parágrafo todo de texto, desde que seja de interesse do usuário recuperar documentos que constem todos os termos exatamente na ordem nas quais foram informados. Termos compostos como por exemplo: auxiliar administrativo ou ciência da informação podem ser acrescidos de aspas “ ” antes do primeiro e depois do último caractere da expressão para que um mecanismo de busca entenda que a pesquisa requer que ele retorne resultados exatamente como as expressões pedem. O resultado compreenderá apenas aqueles que apresentarem a expressão completa e exatamente na ordem na qual foi apresentada.

O *filetype:* é um operador utilizado para pesquisar documentos que estejam em um formato de arquivo específico. Por exemplo: PDF, DOC, SRT, XLS, MP3, PNG, entre inúmeros outros. É útil para evitar que documentos com uma extensão de arquivo diferente da requerida pelo usuário apareça entre os resultados e pode ser utilizado antes ou depois da expressão de busca, entretanto, pode haver variação na quantidade e apresentação de resultados.

No caso dos exemplos 'modelos de currículo filetype:doc' e 'filetype:doc modelo de currículo'. Tematicamente, os resultados em ambas as expressões são similares. O mecanismo de busca retornará modelos de currículo editáveis em Word já que o formato indicado para a extensão de arquivo requerida foi DOC.

O operador *site:* é utilizado para direcionar a pesquisa realizada dentro da caixa de pesquisa para apenas um site em específico. Todos os resultados apresentados na página do mecanismo de busca virão obrigatoriamente de dentro do site indicado.

4 Metodologia

Adotou-se uma situação de controle representada pela busca normalizada (cenário 2) e outros três cenários (cenário 3, 4 e 5) com diferentes intervenções a fim de observar alterações a partir delas. Neste sentido, o cenário 1 representou uma ação de busca livre efetuada pelo respondente.

As variáveis independentes neste estudo foram os operadores de pesquisa *AND*, aspas, *site:* e *filetype:* que pretendiam influenciar a variável dependente representada pelos resultados visualizados a partir das capturas de tela.

A seleção destes operadores obedeceu a adequação para complementar a expressão de busca criada a partir da narrativa de pesquisa e com base na experiência de um dos autores enquanto bibliotecário. Como a busca visava encontrar documentos, o foco dos operadores foi auxiliar na união das palavras-chave (*AND* e aspas “ ”), na definição do formato de arquivo (*filetype:*) e na fonte de informação (*site:*).

4.1 Respondentes

Os respondentes da pesquisa de campo foram convidados por meio de divulgação do link de acesso ao questionário em grupos de pesquisa e mídias sociais (Instagram e Facebook) do pesquisador.

Não foram impostos critérios relacionados ao perfil pessoal, como idade ou gênero, para qualificar os usuários como aptos a participarem da pesquisa. Os pré-requisitos foram que o respondente tivesse

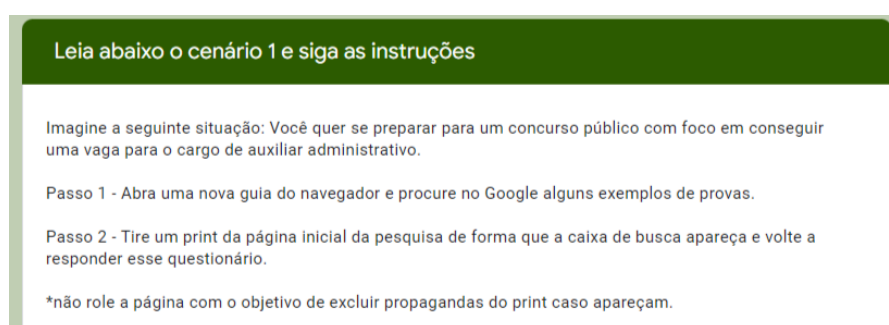
acesso a um dispositivo com conexão à internet (*smartphone*, *tablet*, notebook ou computador de mesa) e possuísse uma conta no Gmail que estivesse ativa no momento das buscas e fosse informada para que o questionário pudesse ser respondido apenas uma vez.

4.2 Procedimento

Foram desenvolvidos cinco testes de cenários de busca distintos, sendo que todos eles tratavam do mesmo tema: provas de concurso público para o cargo de auxiliar administrativo. A escolha ocorreu por conta da necessidade de criar uma busca fictícia que tivesse por objetivo a recuperação de algum tipo de documento na *web*, neste caso, modelos de provas para um concurso público. O cargo de auxiliar administrativo foi definido por ser geralmente um cargo de nível médio, ampliando as possibilidades de resultados se comparado a cargos de nível superior.

Não era possível avançar para o próximo cenário sem antes ter realizado a atividade proposta pelo cenário anterior e cada seção juntamente com suas orientações era visualizada de forma única.

O cenário 1 (ver Figuras 3 e 4) consistia em apresentar a temática ao respondente e instruí-lo a se imaginar na situação de uma pessoa interessada em encontrar provas de concurso públicos já realizados, para o cargo de auxiliar administrativo.

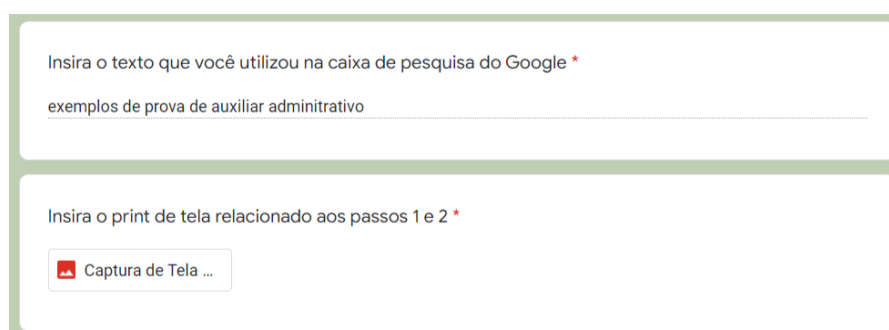


The image shows a green header with the text "Leia abaixo o cenário 1 e siga as instruções". Below the header, the text reads: "Imagine a seguinte situação: Você quer se preparar para um concurso público com foco em conseguir uma vaga para o cargo de auxiliar administrativo." It then lists two steps: "Passo 1 - Abra uma nova guia do navegador e procure no Google alguns exemplos de provas." and "Passo 2 - Tire um print da página inicial da pesquisa de forma que a caixa de busca apareça e volte a responder esse questionário." A note at the bottom states: "*não role a página com o objetivo de excluir propagandas do print caso apareçam."

Figura 3. Cenário 1 do questionário de coleta de dados.

Fonte: Dados da pesquisa – coletados dos respondentes

Nesse momento, o objetivo era coletar quais palavras-chave o usuário considerava relevantes e essenciais para realizar a busca diante do cenário apresentado. Foi solicitado que o usuário inserisse a expressão de busca executada em sua página do Google e realizasse o *upload* da captura de tela com os resultados apresentados ao executá-la.



The image shows two input fields. The first field is labeled "Insira o texto que você utilizou na caixa de pesquisa do Google *" and contains the text "exemplos de prova de auxiliar administrativo". The second field is labeled "Insira o print de tela relacionado aos passos 1 e 2 *" and contains a button labeled "Captura de Tela ...".

Figura 4. Campo de resposta do cenário 1.

Fonte: Dados da pesquisa – coletados dos respondentes

Em todos os cenários foi orientado que o usuário realizasse a captura de tela sem considerar o aparecimento de anúncios, não sendo, portanto, necessário que ele rolasse a página de forma que só aparecessem resultados orgânicos em relação à busca realizada.

No cenário 2 da Figura 5, em vez de deixar a busca livre para que o usuário definisse as palavras-chave e a formulação das expressões de busca de acordo com seus próprios critérios, foi fornecida uma expressão única consolidada, para que ele copiasse, colasse em seu mecanismo de busca e executasse

a pesquisa. A prática uniformizou a busca para todos os respondentes e teve por objetivos comparar os primeiros resultados de cada um deles.

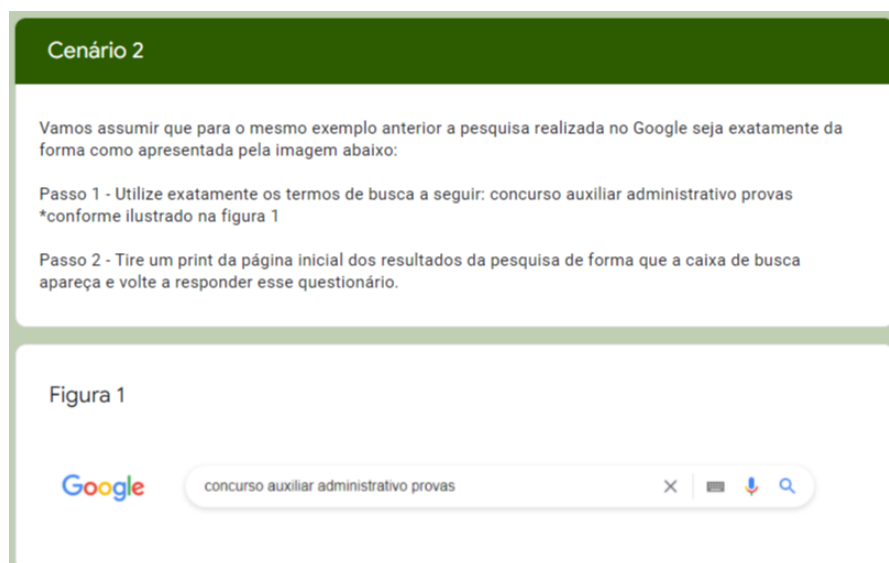


Figura 5. Cenário 2 do questionário de coleta de dados.

Fonte: Dados da pesquisa – coletados dos respondentes

A partir do cenário 2, apenas a captura de tela era requerida ao respondente, e a forma de execução da pesquisa deveria vir acompanhada de um exemplo visual sobre como a caixa de pesquisa do Google deveria estar preenchida quando ele a realizasse.

O cenário 3 (ver Figura 6) introduz o uso de operadores ao determinar que o usuário utilizasse o operador booleano *AND* na expressão de busca apresentada.

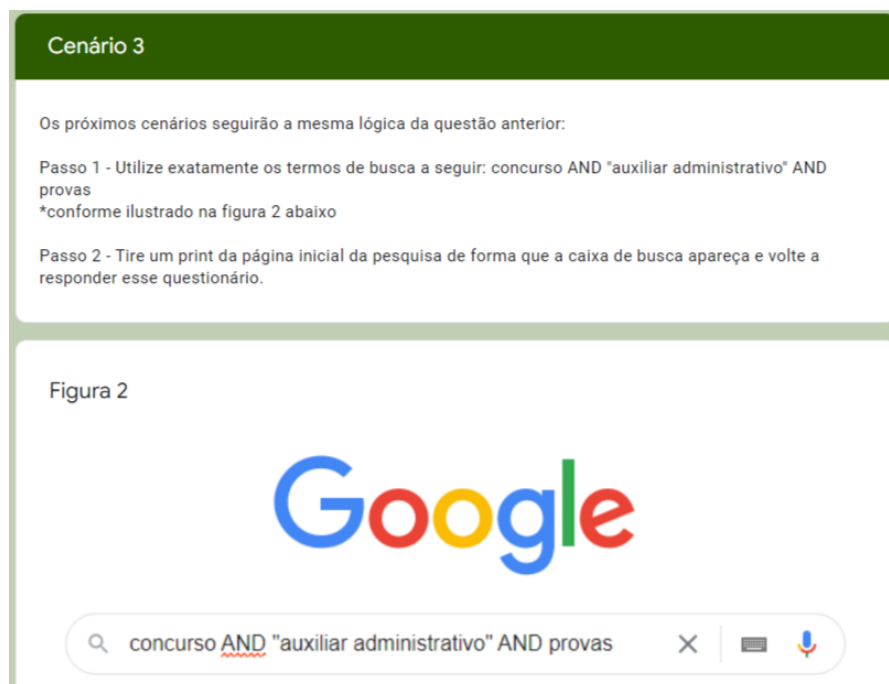


Figura 6. Cenário 3 do questionário de coleta de dados.

Fonte: Dados da pesquisa – coletados dos respondentes

Mesmo que toda busca no Google de certa forma aplique o operador booleano *AND* em sua execução, a ideia era testá-lo de forma mandatária, visando garantir somente resultados que obrigatoriamente apresentem as três expressões (*concurso; auxiliar administrativo; provas*) juntas.

Além do uso do operador booleano *AND*, nesse cenário, outro operador também foi empregado de forma combinada. As aspas “ ”.

No cenário 4 (ver Figura 7) era requerido que os respondentes utilizassem o operador *filetype:pdf* para que, ao aplicar a mesma pesquisa proposta pelos cenários 2 e 3, fossem restringidos os resultados para somente aqueles que apresentassem documentos de extensão no formato PDF (*Portable Document Format*).

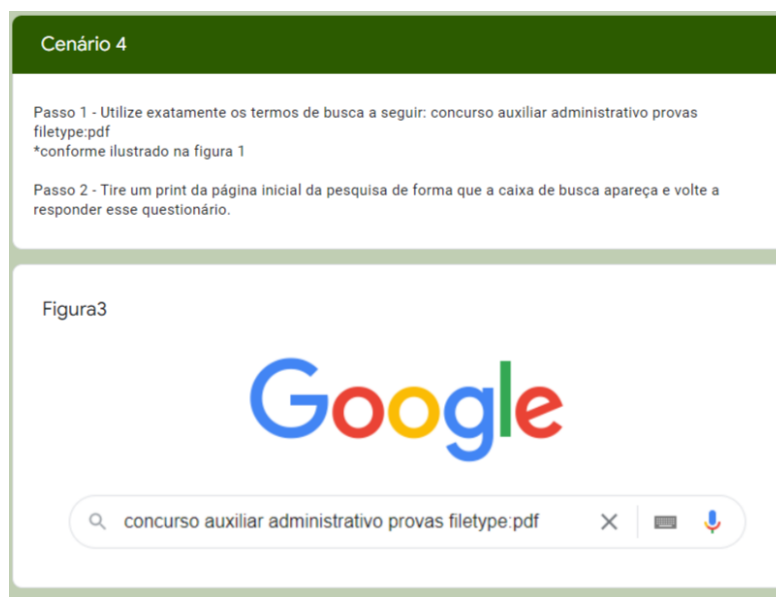


Figura 7. Cenário 4 do questionário de coleta de dados.

Fonte: Dados da pesquisa – coletados dos respondentes

Ao utilizar o operador de formato *filetype:* e restringir resultados apenas em PDF, esperava-se uma maior uniformização dos resultados nos diferentes perfis dos usuários respondentes do questionário.

Por último, o cenário 5 (ver Figura 8), introduziu o operador *site:* para que a mesma pesquisa reproduzida nos três cenários anteriores estivesse direcionada para resultados que viessem de um único *site*.

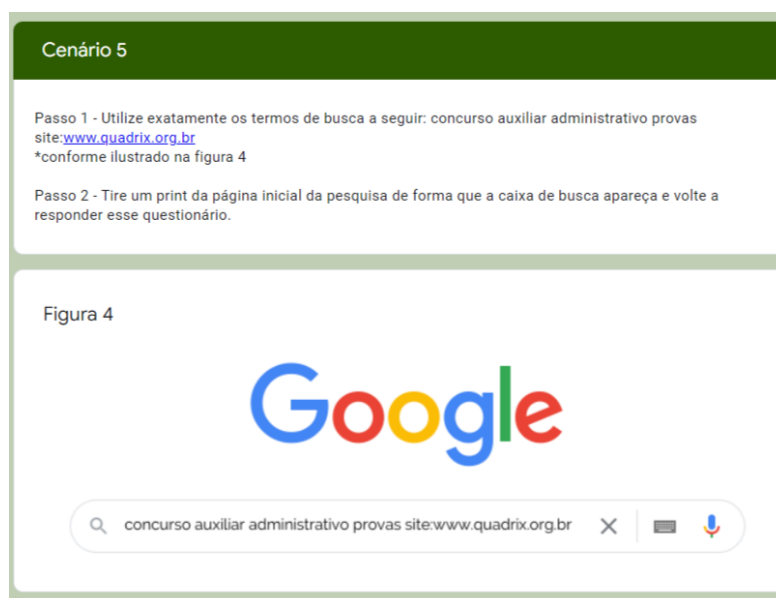


Figura 8. Cenário 5 do questionário de coleta de dados.

Fonte: Dados da pesquisa – coletados dos respondentes.

A escolha do *site* utilizado ocorreu ao levar em consideração a preferência por uma banca de

concursos que atuasse com provas para cargos de nível médio, em que o cargo de auxiliar administrativo estivesse presente. A banca selecionada para o exemplo da expressão de busca com o uso do operador foi o Instituto Quadrix. A organização sem fins lucrativos desenvolve e aplica provas para Conselhos Federais, Conselhos Regionais, Hospitais e Fundações, nos quais o cargo é recorrente, além de estar presente com atuação em todo Brasil, assegurando a abrangência nacional e variedade que o exemplo necessitava.

O cenário 5 apresentou o maior nível de especificidade na aplicação de operadores de pesquisa, restringindo mais ainda os resultados apresentados aos diferentes usuários respondentes do questionário.

4.3 Coleta de dados

O questionário foi construído na plataforma Google Forms e aplicado no segundo semestre de 2022. Ao todo, obteve um total de 76 respostas registradas durante o período.

A Tabela 2 resume a quantidade de capturas de tela que foram excluídas e quantas compuseram o total de cada cenário de busca descrito.

Tabela 2. Conjunto de dados pré e pós-seleção.

CENÁRIO DE BUSCA	COLETADOS	EXCLUÍDOS	ANALISADOS
Busca livre	76	00	76
Busca normalizada	76	04	72
Operador booleano AND	76	05	71
Operador filetype:	76	04	72
Operador site:	76	03	73
TOTAL	380	16	364

Fonte: Dados da pesquisa (2022).

Cada respondente enviou cinco arquivos de imagem resultando num total de 380 capturas de tela coletadas. Destes, 16 não puderam ser analisadas. A exclusão ocorreu por problemas como: captura da tela não correspondente ao solicitado, expressão de busca formulada incorretamente ou pesquisa realizada em outro mecanismo de busca diferente do solicitado. Assim sendo, o *corpus* da pesquisa analisado resultou em 364 capturas de tela.

5 Contribuições para a ciência aberta

Para Lecardelli (2020, p. 17) “compartilhar informações científicas é primordial para potencializar os avanços da ciência, poupar retrabalho, possibilitar reproduzir pesquisas para validação e permitir o reuso de informações por novos interessados para novas descobertas”. Portanto, se tem como compromisso, neste trabalho, contribuir para a ciência aberta, permitindo a disponibilização dos dados científicos gerados no desenvolvimento da pesquisa.

Em conformidade com a Lei Geral de Proteção de Dados (LGPD) em relação ao tratamento de dados em território nacional, todas as 380 capturas de tela foram anonimizadas, fotos de perfil, nomes de usuário e outras guias abertas no navegador foram anuladas. Além disso, os arquivos de imagem tiveram seus metadados relacionados a propriedade e informações pessoais foram removidas.

As planilhas criadas a partir das capturas de tela coletadas foram disponibilizadas no repositório de dados científicos abertos Zenodo, mantido pela *European Organization for Nuclear Research* (CERN). A escolha foi realizada por conta da possibilidade de depósito gratuito dos dados por parte dos autores, acrescida da atribuição do identificador persistente *Digital Object Identifier* (DOI) que concede singularidade e permanência a objetos digitais na internet.

Dito isso, os dados abertos desta pesquisa encontram-se disponíveis em: <https://doi.org/10.5281/zenodo.7434768>.

6 Análise dos resultados

Os dados geográficos demonstraram que 93% dos respondentes estavam em alguma localidade do Brasil no momento de preenchimento do questionário, e 7% estavam em outros países, sendo eles: Portugal, Colômbia, Argentina e Canadá.

Dos que estavam no Brasil, o Sul apresentou maior incidência com 69% do total. Seguido pelo Sudeste com 22%. As demais regiões Nordeste, Centro Oeste e Norte ficaram com 6%, 3% e 0%, respectivamente (ver Figura 9).

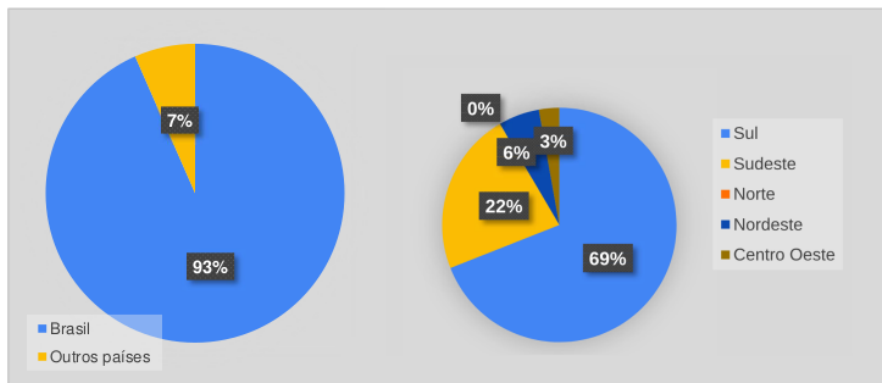


Figura 9. Dados de geolocalização dos respondentes.

Fonte: Dados da pesquisa (2022).

Em relação ao tipo de dispositivo que foi utilizado pelos respondentes para completar o questionário, a divisão ficou em 55% para notebook ou computador de mesa e 45% para smartphone (ver Figura 10). Nenhum respondente utilizou tablet para completar a pesquisa.

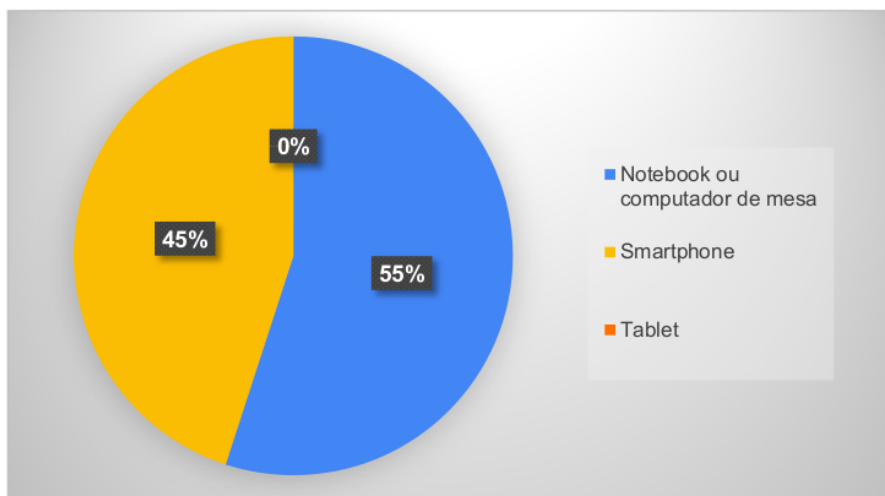


Figura 10. Tipo de dispositivo utilizado pelos respondentes.

Fonte: Dados da pesquisa (2022).

A maioria respondeu ao questionário por meio do navegador Google Chrome, representando 78% do total. Para os demais, 11% utilizaram o Safari, 5% o Mozilla Firefox, 3% o Opera. Microsoft Edge, Brave Browser e Samsung Internet atingiram 1% cada (ver Figura 11).

Outro dado coletado foi acerca da utilização ou não de bloqueadores de anúncios nos navegadores utilizados durante o preenchimento do questionário (ver Figura 12).

De modo geral, 8% dos respondentes afirmaram não utilizar nenhum tipo de bloqueador de anúncio, enquanto 17% que utilizam ou não souberam responder, sendo essa última parcela distribuída em 13% para o uso do *AdBlock*, 1% para o *uBlock* e 3% não souberam responder por dúvidas se essa é uma ferramenta que estava instalada em seu navegador de forma automática ou por terceiros sem que eles soubessem.

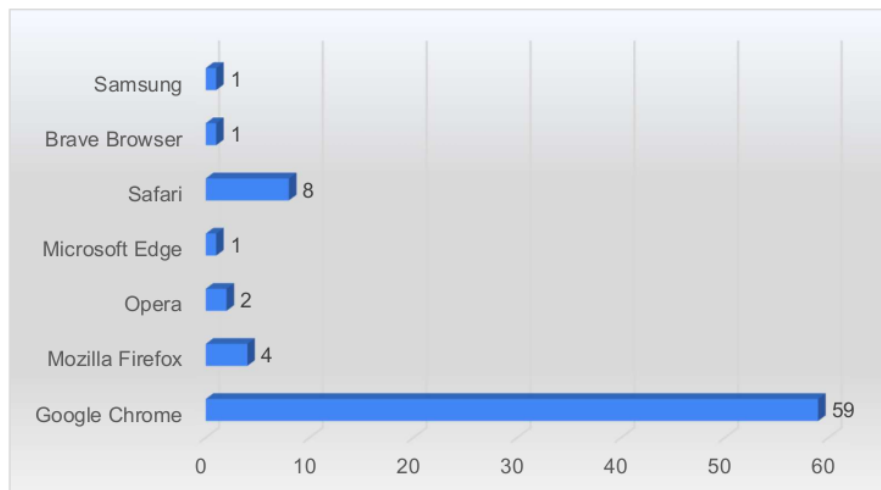


Figura 11. Navegador de internet utilizado pelos respondentes.

Fonte: Dados da pesquisa (2022).

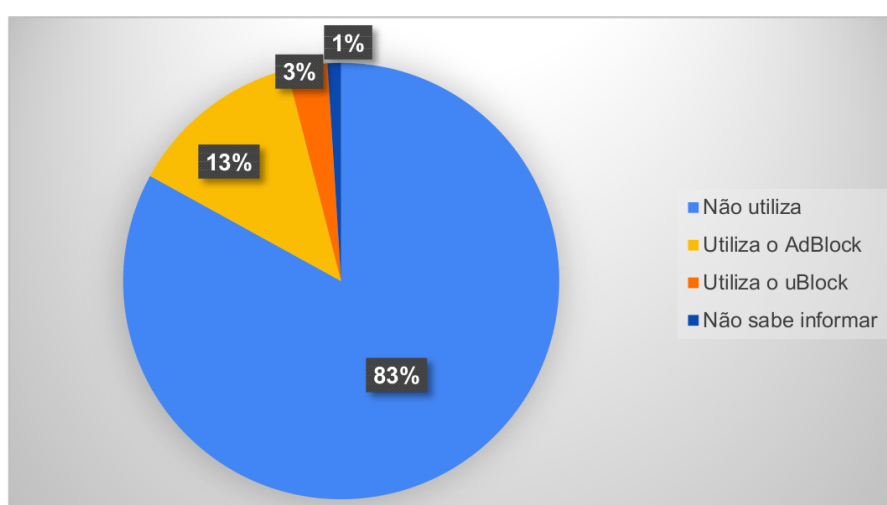


Figura 12. Presença de bloqueador de anúncios no navegador.

Fonte: Dados da pesquisa (2022).

6.1 Busca livre

A busca livre corresponde ao cenário 1, considerando a narrativa apresentada no questionário para que o usuário realizasse a pesquisa livremente. Portanto, o primeiro resultado diz respeito às escolhas dos termos efetuadas por cada respondente (ver exemplo Figura 13).

O cenário 1 solicitava se imaginar na situação de uma pessoa interessada em encontrar provas de concurso públicos já realizados, para o cargo de auxiliar administrativo.

Os três termos com maior ocorrência na expressão de busca livre criada pelos respondentes foram *auxiliar administrativo* com 69 usos (91% dos usuários), seguido de *prova* com 63 (83% dos usuários) e *concurso* com 59 (78% dos usuários). Sem levar em consideração: plurais e variações, ainda assim os três termos citados continuam com maior ocorrência dentro das pesquisas, nessa mesma ordem.

A maioria dos respondentes entendeu a busca livre da mesma forma que a busca normalizada. Isto é, grande parte dos respondentes utilizou os mesmos termos empregados posteriormente na busca com termos pré-indicados.

Além disso, também é possível notar o acréscimo de outros termos complementares à expressão de busca na nuvem de palavras da Figura 14. Ao todo, 55% dos usuários utilizaram algum termo adicional em relação aos três principais citados. Desses, apenas 12% empregaram algum tipo de operador avançado por conhecimento próprio.

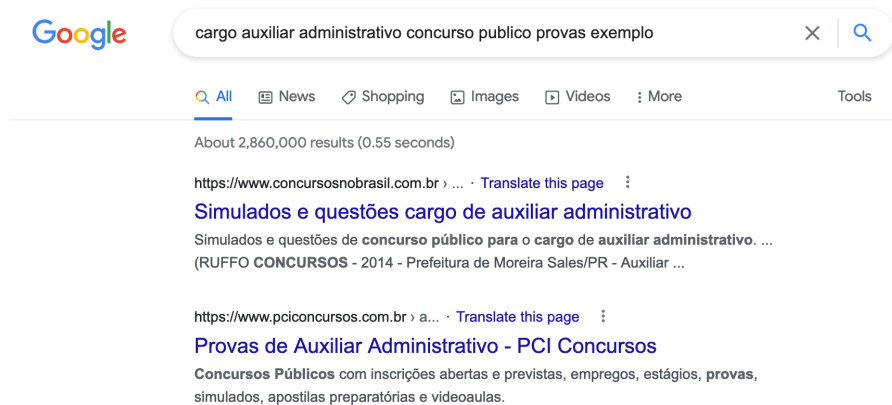


Figura 13. Execução da busca livre.

Fonte: Dados da pesquisa – coletado dos respondentes (2022).

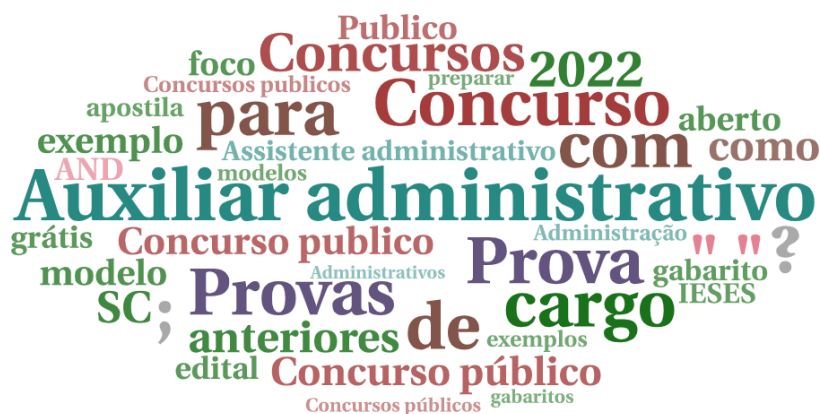


Figura 14. Nuvem de palavras com os termos da busca livre.

Fonte: Dados da pesquisa (2022).

As *stop words*¹ não foram ignoradas, pois fazem sentido diante do contexto. Mesmo sendo palavras consideradas pela computação como vazias e com uma função semântica praticamente irrelevante para a lógica de indexação e apresentação do conjunto de resultados pelos mecanismos de busca, as *stop words* foram mantidas com intuito de representar a necessidade sentida pelos respondentes em inseri-las na sua expressão de busca livre.

Contradizendo a questão da total irrelevância dos resultados e tendo por base a observação das capturas de tela coletadas neste cenário, principalmente no que diz respeito a quantidade e ordem dos resultados, toda preposição, conjunção ou até um caractere a mais na expressão de busca, fez diferença (mesmo que não contextualmente) para o mecanismo de busca. Conseqüentemente, apresentou resultados diferentes daqueles sem tais acréscimos. Uma busca que utilizou a expressão *provas concurso público* apresentou resultados diferentes de uma pesquisa como *provas de concursos público* e, ainda diferentes, de uma pesquisa como *prova de concurso público*, mesmo que a diferença seja apenas um acento agudo.

Em relação ao aspecto visual, os tamanhos representam a frequência com que os termos apareceram. As cores dividem os termos em sete categorias: a cor azul petróleo (■) refere-se aos termos relacionados ao cargo em questão; a cor roxa (■) para termos relacionados ao documento; a cor vermelha (■) para termos relacionados ao evento; a cor marrom (■) para termos de ligação; a cor verde (■) para outros termos relacionados; a cor rosa (■) para operadores de pesquisa avançados e a cor cinza (■) para sinais ortográficos. O grau de opacidade de cada uma delas corresponde à

¹ *Stop words* ou palavras-vazias são termos que possuem pouco significado num contexto tais como preposições, artigos, conjunções e outras (IBICT, 2015).

importância dentro da categoria pertencente. Todas as cores foram escolhidas aleatoriamente.

Consequentemente, a partir de uma busca livre, os resultados são diversificados, e como já dito, cada acento, plural ou qualquer outro caractere, pode conduzir a resultados diferentes.

6.2 Busca normalizada

O cenário 2 apresentou uma busca pronta designada como: concurso auxiliar administrativo provas e solicitou que o respondente a replicasse em sua própria caixa de pesquisa (ver exemplo Figura 15).

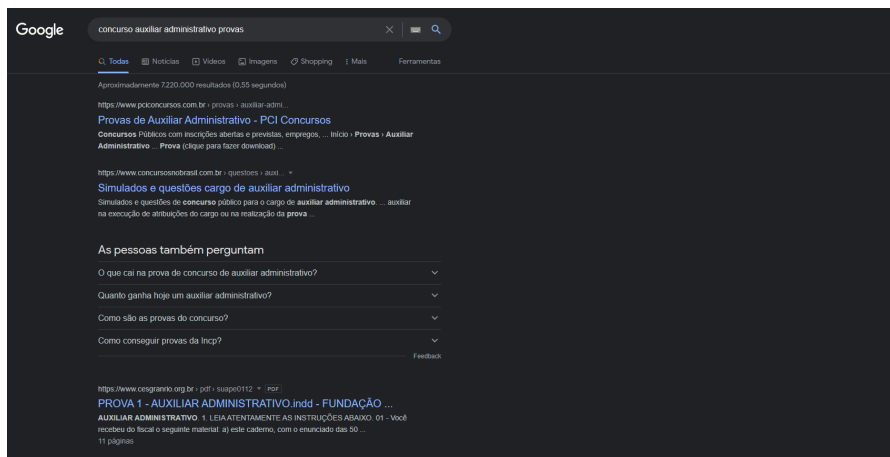


Figura 15. Execução da busca normalizada.

Fonte: Dados da pesquisa – coletado dos respondentes (2022).

A busca normalizada apresentou resultados mais uniformes que a busca livre. Das 72 capturas de telas válidas analisadas, os resultados em notebooks ou computadores de mesa variaram entre aproximadamente 5.500.000 e 8.000.000 de resultados recuperados. Apenas uma busca teve 34.100.000 links, sendo que o usuário estava no Brasil e informou utilizar um bloqueador de anúncios no navegador Google Chrome.

Em apenas seis capturas de tela notou-se a presença de propagandas no topo dos resultados orgânicos. Esses 8% do total da categoria foram pesquisados utilizando o navegador Google Chrome. Nenhum deles apresentou sugestões de outras pesquisas realizadas no tema, que estavam presentes em outros 83% das capturas de tela, todas sem publicidade. Sugere-se uma relação entre a presença de publicidade e o aparecimento de recomendações de outras pesquisas já realizadas. No entanto, os demais cenários não confirmaram tal hipótese se comparada a pesquisas com o uso de operadores avançados, já que os cenários 2, 3 e 4 tiveram pesquisas que apresentaram publicidade juntamente com sugestões de pesquisas feitas por outros usuários e não apresentaram dados suficientes para comprovar uma correlação entre esses dois tópicos.

6.3 Busca com o operador booleano AND

O cenário 3 foi o primeiro a acrescentar um operador de pesquisa e apresentou uma expressão de busca com o uso do *AND*. O operador booleano funcionou com apenas seis respondentes de um total de 71. Isto porque para 65 usuários, a expressão de busca foi alterada pelo sistema. A alteração substituiu o primeiro booleano *AND* pela preposição *DE* como é possível verificar na Figura 16.

O mecanismo de busca considera todas as pesquisas realizadas com mais de um termo como já acrescidas do operador booleano *AND* em sua expressão. Portanto, um usuário que pesquisa por 'provas concurso auxiliar administrativo' já tem todos esses quatro termos multiplicados logicamente em decorrência do operador.

Em vista disso, entende-se que nesse caso o sistema realizou uma correção textual ao não interpretar o *AND* como um operador e sim como mais um termo de busca de valor textual semelhante aos termos 'concurso' ou 'provas', por exemplo. Como o objetivo era avaliar a efetividade do booleano com diferentes usuários numa mesma busca e não o conhecimento sobre o buscador e possíveis ações corretivas, a orientação foi que a captura de tela fosse registrada sem nenhuma interferência do

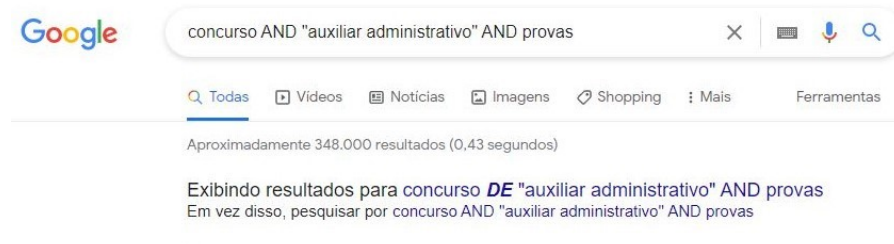


Figura 16. Erro de pesquisa com o operador booleano *AND*.

Fonte: Dados da pesquisa – coletado dos respondentes (2022).

usuário, que poderia informar ao mecanismo de busca para não efetuar a troca e refazer a pesquisa da forma proposta inicialmente.

As pesquisas em que o booleano foi aceito pelo mecanismo de busca não apresentaram os avisos da Figura 16 e tiveram resultados exibidos normalmente conforme a Figura 17.

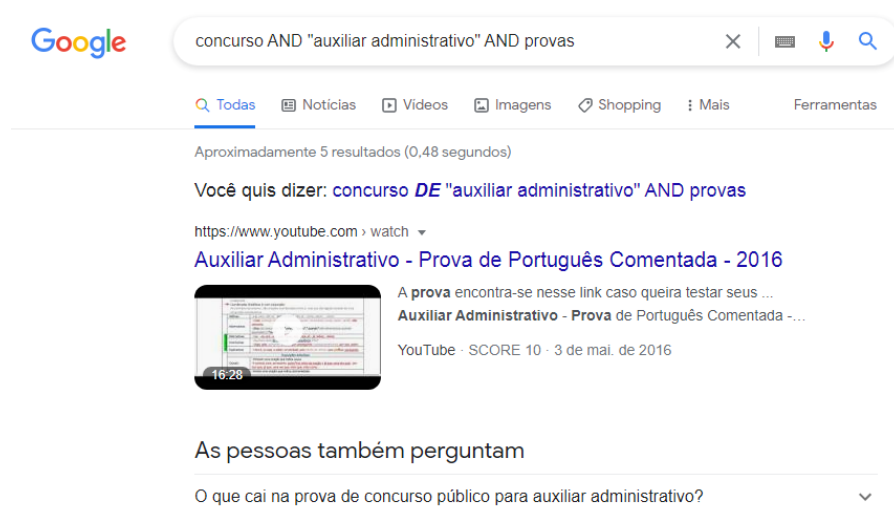


Figura 17. Pesquisa bem-sucedida com o operador booleano *AND*.

Fonte: Dados da pesquisa – coletado dos respondentes (2022).

Após análise, ao que parece, somente cinco usuários tiveram suas pesquisas validadas pelo mecanismo de busca como uma pesquisa que requer ao sistema o uso forçado e explícito do operador booleano *AND* por duas vezes. As cinco pesquisas não apresentaram nenhum anúncio no topo ou ao lado dos primeiros resultados, sendo que nenhum deles afirmou estar utilizando um bloqueador de anúncios.

Todas as cinco, ainda que o mecanismo de busca tenha executado a expressão, foram questionadas pelo sistema se a pesquisa de fato era a que foi informada ou se a intenção era a pesquisa concurso DE “*auxiliar administrativo*” *AND provas*, executada automaticamente pela grande maioria dos respondentes dessa categoria.

Das seis capturas de tela excluídas da categoria, duas delas, a princípio, pareciam ter o operador booleano executado corretamente. No entanto, o número de resultados aproximados ficou aquém do esperado para uma expressão de busca tão específica, ambos acima de 350 mil. Outro ponto percebido foi que duas dessas capturas de tela estavam relacionadas a mecanismos de busca configurados em outros idiomas que não o português brasileiro, uma delas estava em inglês e outra em espanhol.

Levando em consideração que das buscas válidas, apenas 7% obtiveram êxito, sugere-se que, para a maioria dos usuários, o operador *AND* não tende a funcionar de início em buscas com apenas um termo composto acrescido de aspas junto a outros dois termos únicos. Entretanto, funciona implicitamente por ser um parâmetro empregado em todas as buscas com vários termos.

Como se trata de uma função que já ocorre logicamente em toda pesquisa realizada no sistema,

ao acessar a aba de pesquisa avançada do Google, somente outros operadores booleanos com o OR e o NOT são disponibilizados aos usuários, conforme a Figura 18.

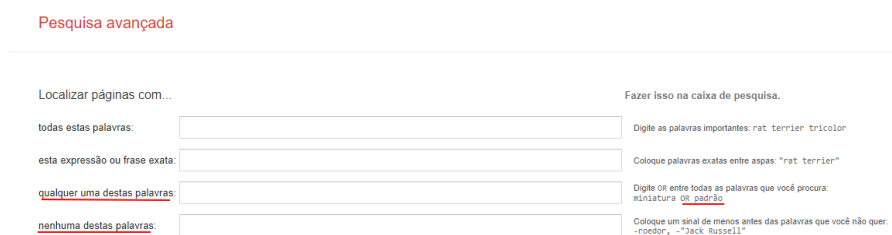


Figura 18. Guia de pesquisa avançada do Google.

Fonte: (Google, 2022)

O mecanismo de busca permite localizar resultados que contenham qualquer uma das palavras, usando o *OR* e com nenhuma das palavras citadas, utilizando o *NOT*, na forma de '-' (sinal de menos).

6.4 Busca com o operador avançado *filetype*:

O cenário 4 apresentou o operador avançado *filetype*: com foco em obter precisão nos formatos dos documentos recuperados. A expressão de busca com o uso de operador *filetype*: funcionou com 70 respondentes do total de 72 válidos. Dos excluídos, um apresentou a mensagem de que houve resultados que coincidiram com a expressão pesquisada, no entanto, o operador de pesquisa precisaria ser incluído naquele resultado em si, o que denota que o sistema o interpretou como um termo simples e não um comando de busca avançado como é possível observar na Figura 19.

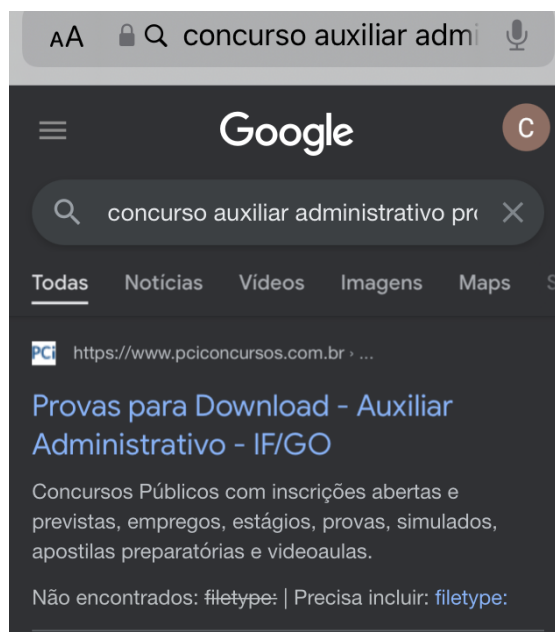


Figura 19. Erro no uso do operador *filetype*:

Fonte: Dados da pesquisa – coletados dos respondentes (2022).

Neste caso, o usuário não especificou o formato de arquivo requerido e indicado no questionário, que no caso era PDF. Qualquer outra extensão de arquivo, caso fosse especificada, não seria listada pelo mecanismo de busca como possível inclusão de termo já que ele entenderia como um comando direto mesmo que o resultado fosse zero arquivos recuperados.

Nas 70 demais buscas é possível observar uma etiqueta após a URL de cada resultado, confirmando que o documento hospedado naquele site é na extensão requerida na expressão de busca, no caso, PDF (ver Figura 20).

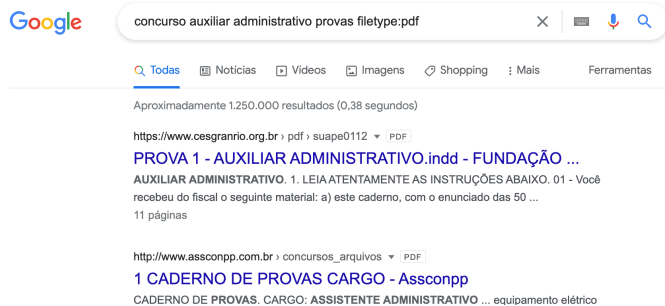


Figura 20. Etiqueta com extensão de arquivo.

Fonte: Dados da pesquisa – coletados dos respondentes (2022).

Em geral, também é possível afirmar que, para a amostra analisada, o operador não influenciou na aparição ou não de anúncios e suas respectivas posições em tela ou sugestões de pesquisas. Em relação a quantidade aproximada de resultados em pesquisas realizadas em *notebooks* ou computadores de mesa foi mantida a variação entre 1.000.000 e 1.600.00 de resultados. Apenas dois destoaram desse número, um para mais e outro para menos, o que é justificado pela geolocalização dos dois usuários que estavam fora do Brasil no momento das buscas.

6.5 Busca com o operador avançado *site*:

O cenário 5 contava com o uso do operador *site*: a fim de especificar ao mecanismo de busca a fonte principal de informação dos resultados a serem apresentados. Os resultados mostram que o operador *site*: funcionou nas 73 capturas de tela válidas analisadas dentro da categoria.

Com o operador avançado *site*: foi possível observar o maior nível de similaridade entre os primeiros resultados e as perguntas sugeridas pelo próprio mecanismo de busca como pesquisas similares que outros usuários realizaram.

Em relação aos anúncios, este foi o operador que mais influenciou na presença de publicidade dentre os resultados. De todas as 73, somente uma apresentou um anúncio no topo, o que gerou controvérsias tendo em vista que o usuário em questão afirmou utilizar um bloqueador de anúncios em seu navegador (ver Figura 21).

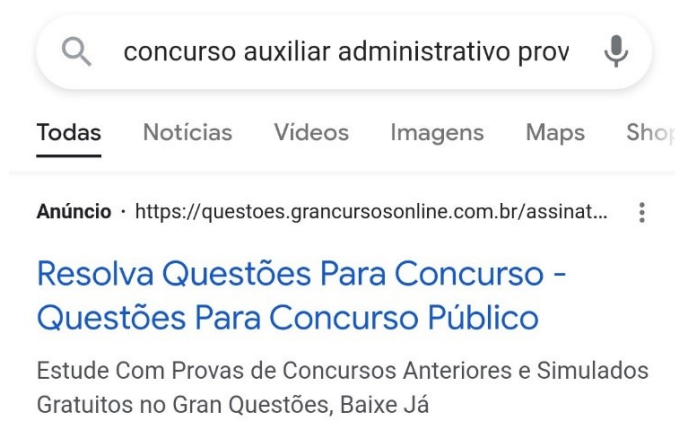


Figura 21. Etiqueta com extensão de arquivo.

Fonte: Dados da pesquisa – coletados dos respondentes (2022).

Esse mesmo respondente que indicou o utilizar bloqueadores de anúncios, também apresentou outra captura de tela com a presença de anúncios na categoria do operador *filetype*:. Nesse caso, ou a informação cedida por ele em relação ao uso de um bloqueador de anúncios estava incorreta ou o funcionamento deste bloqueador foi comprometido em níveis mais avançados de busca com a presença de operadores como *filetype*: e *site*:.

A quantidade aproximada de resultados em pesquisas realizadas em *notebooks* ou computadores

de mesa variou entre 900 e 2300 resultados, sendo, portanto, o operador que mais trouxe exatidão aos resultados, retornando para o usuário um número consideravelmente baixo em comparação aos outros quatro cenários nos quais havia resultados na casa dos milhões. Ocorrência que condiz com a funcionalidade do operador, já que seu uso permite apontar somente para resultados advindos de uma única fonte. Nesse caso, a quantia é relativa, mas a fonte de busca é precisa.

6.6 Observações gerais

Em 10 das 380 capturas de tela, os respondentes, por conta própria, rolaram a tela e realizaram a ação. Algumas imagens trazem apenas resultados abaixo da caixa de pesquisa, o que leva a crer que anúncios foram ignorados e desconsiderados pelos respondentes como resultados válidos frente a expressão de busca pesquisada.

Em oito capturas de tela, mesmo os usuários que informaram utilizar um bloqueador de anúncios, tiveram anúncios exibidos em suas pesquisas. Seis deles ao utilizarem o operador *filetype*:. O uso dos bloqueadores não interferiu na quantidade de resultados aproximados informada pelo mecanismo de busca em comparação aos que não utilizaram.

Os diferentes navegadores, bem como o uso de *smartphones* ou *notebooks* e computadores de mesa, não manifestaram características que visivelmente interferissem no uso de funcionamento dos operadores propostos na pesquisa.

As características que mais se destacaram nas buscas foram a presença de anúncios, sugestões de outras pesquisas já realizadas que tinham relação com a pesquisa atual e resultados aproximados em *notebooks* ou computadores de mesa.

Em relação a presença de anúncios no topo dos resultados, calculou-se a porcentagem de pesquisas que apresentaram anúncios dentre os respondentes que informaram não utilizar nenhum tipo de bloqueador de anúncios (ver Tabela 3).

Tabela 3. Porcentagem de anúncios nos cenários normalizados.

CENÁRIO DE BUSCA	ANÚNCIOS
Busca normalizada	9%
Busca com o AND e as aspas	0%
Busca com o filetype:	39%
Busca com o site:	2%

Fonte: Dados da pesquisa (2022).

Nota-se que uso de alguns operadores avançados pode favorecer a aparição de anúncios de pesquisa à medida que ou obrigatoriamente o anunciante tenha usado todas as palavras-chave da intersecção resultante do uso do operador booleano *AND* acrescido as aspas, ou esteja atrelado ao *site* indicado como fonte única de pesquisa do operador *site*:

Se a busca realizada pelo operador *filetype*: indicar um formato de arquivo atrelado ao produto anunciado, existe a possibilidade deste produto se encaixar nos resultados como anúncio de pesquisa (ver Figura 22). Isso está atrelado a personalização descrita no estudo de Rosa (2022) e as técnicas de SEO empregadas na indexação de conteúdos web, exploradas no estudo de Camossi, Amorim e Rodas (2023).

Nos dados coletados, o formato PDF sempre esteve ligado a anúncios de apostilas, provavelmente comercializada nesse formato de arquivo. Na sequência, calculou-se a porcentagem das capturas de tela em que as sugestões de pesquisa apareciam dentro dos cenários normalizados (ver Tabela 4).

Não houve uma grande variação que possa inferir correlação entre o aparecimento ou não de sugestões (ver Figura 23) a partir do uso dos operadores avançados testados. No entanto, as sugestões quase sempre presentes, corroboram o dito por Santos e Viera (2015) ao trazer a ideia de que encontrar resultados satisfatórios para uma pesquisa, levará a produção de novos conhecimentos, que são disponibilizados na web e conseqüentemente atrelados às pesquisas que o precederam, como as sugestões.

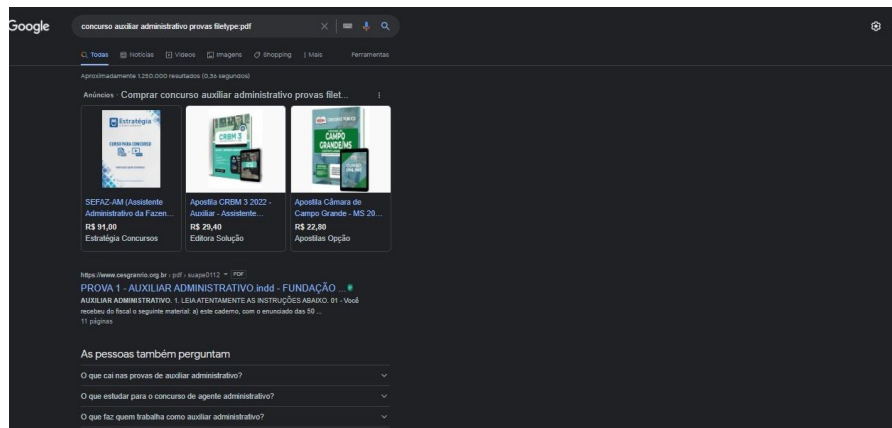


Figura 22. Anúncios de apostilas usando o operador *filetype:*.
Fonte: Dados da pesquisa – coletados dos respondentes (2022).

Tabela 4. Porcentagem das sugestões de pesquisa nos cenários normalizados.

CENÁRIO DE BUSCA	SUGESTÕES
Busca normalizada	83%
Busca com o AND e as aspas	100%
Busca com o filetype:	76%
Busca com o site:	92%

Fonte: Dados da pesquisa (2022).

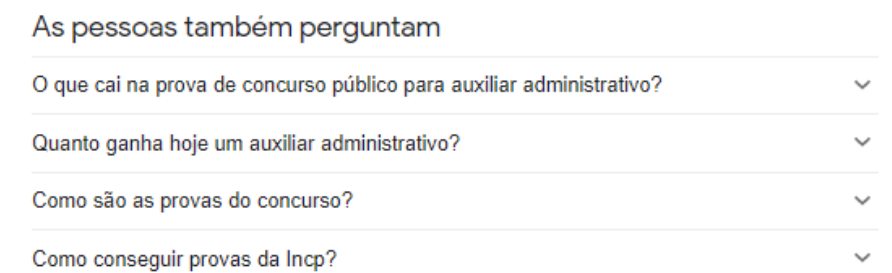


Figura 23. Exemplo de sugestões de pesquisas já realizadas.
Fonte: Dados da pesquisa – coletados dos respondentes (2022).

Por fim, calculou-se a mediana entre as pesquisas realizadas em notebooks e computadores de mesas, no que diz respeito à quantidade de resultados aproximados (ver Tabela 5).

Tabela 5. Mediana dos resultados nos cenários normalizados.

CENÁRIO DE BUSCA	MEDIANA DOS RESULTADOS APROXIMADOS
Busca normalizada	7.400.000
Busca com o AND e as aspas	4
Busca com o filetype:	1.270.000
Busca com o site:	1.930

Fonte: Dados da pesquisa (2022).

Os resultados apontam para uma redução na quantidade total de resultados ao utilizar operadores avançados de pesquisa se comparado a uma busca sem eles. Afinal, quanto mais específica for a expressão, menor é a quantidade de documentos dentro do índice em que o mecanismo de busca realizará a varredura à procura de resultados precisos.

Essa possibilidade de aumentar a precisão dos resultados por meio do uso de operadores endossa a importância de conhecê-los e saber empregá-los quando preciso. Resultados como os apresentados por Mnguni e Kekana (2022) trazem o Google como a ferramenta de busca mais utilizada por estudantes universitários para encontrar informações, sendo um público que necessita desses conhecimentos diante do volume de resultados apresentados em grande parte das pesquisas e como consequência da transformação do ato de pesquisar, proporcionada pelas tecnologias de informação e comunicação como os mecanismos de busca (González Guitián *et al.*, 2022).

7 Considerações finais

Mecanismos de busca estão em constante aprimoramento para melhorar a experiência de pesquisa dos usuários. Muitas vezes a forma simples funciona, pois o usuário insere um ou mais termos, clica em buscar e navega pelos resultados apresentados em tela. Todavia, é importante conhecer recursos para que, caso venha a precisar, possa recorrer a estratégias específicas visando encontrar um conjunto menor ou mais padronizado de resultados para avaliar, fazendo com que tal tarefa possua mais assertividade e precisão. O sucesso de um processo de busca na internet pode ser analisado desde a construção da expressão de busca — com ou sem o uso de operadores —, até o resultado apresentado em tela. Este resultado ainda pode apresentar elementos de *feedback* que apontam sucesso ou fracasso daquilo que foi solicitado pelo usuário. Ou seja, se a expressão construída de fato foi entendida pelo mecanismo de busca e funcionou.

Diante disto, e após as observações realizadas neste trabalho, é possível apresentar duas reflexões. A primeira se refere ao *funcionamento da expressão de busca*. Ao efetuar uma expressão de busca com operadores booleanos, e o retorno não apresentar resultados, pode-se imaginar que a expressão de busca não funcionou.

Porém, uma pesquisa retornar zero resultados não quer dizer que a expressão de busca estava incorreta. Ela pode ter funcionado sem nenhum erro de sintaxe e constatado que não há nada ali correspondendo ao que foi pesquisado.

Já num segundo cenário reflexivo, usa-se uma expressão de busca com operadores booleanos, e nesse caso, suponha que a pesquisa retornou com 23 resultados.

Essa segunda reflexão, é sobre a *precisão dos resultados*. Uma pesquisa retornar resultados não quer dizer que a expressão de busca foi precisa. A depender da situação, alguns sistemas podem ignorar operadores ou retornar resultados com partes da expressão. Usuários mais desavisados podem achar que a expressão foi executada exatamente como ele construiu, quando na verdade o sistema usou outros caminhos para apresentar os resultados.

Esse último caso fica explícito nas pesquisas utilizando o operador booleano *AND* que compuseram esse estudo. No caso, o mecanismo de busca informou que estava exibindo resultados para uma expressão de busca que não a solicitada. Contudo, nem sempre isso acontece e cabe sempre ao usuário examinar se aqueles resultados que retornaram equivalem a ideia inicial de expressão que ele elaborou e construiu.

Isto é importante e se relaciona com todo o processo de observações realizado neste trabalho à medida que tais operadores avançados não estão presentes na interface inicial da tela de busca e nem são apresentados aos usuários como sugestões. Assim, demandam que os usuários tenham um prévio conhecimento de como funcionam e para que servem, de modo a evitar um uso inadequado ou ineficaz dos operadores.

Das características nos resultados de pesquisa que se destacam a partir do uso de operadores avançados na busca por documentos na *web* estão os anúncios de pesquisa, as sugestões de pesquisas que já foram realizadas anteriormente por outros usuários e a quantidade aproximada de resultados.

Em relação aos anúncios na página de resultados, nota-se que o uso de operadores avançados de pesquisa pode favorecer seu aparecimento como também os inibir. Para aparecer a criação dos anúncios possivelmente precisará estar alinhada ao objetivo do operador, como por exemplo, um formato de documento compatível com o produto anunciado na *web*.

No que diz respeito às sugestões de pesquisas já realizadas, os operadores não parecem interferir em seu aparecimento, sendo estas teoricamente associadas ao conteúdo da busca em si a partir

somente dos termos de busca selecionados pelo usuário.

Por último, a quantidade de resultados aproximados demonstrou que o operador booleano *AND* acrescido do uso de aspas “ ”, tem potencial para reduzir os resultados, seguidamente do operador *site*: e por último o *filetype*:. Todos os quatro operadores utilizados apresentaram redução na quantidade aproximada de resultados se comparados a mesma pesquisa sem o uso deles. Com isso, sugere-se que o uso de operadores avançados de pesquisa pode diminuir a quantidade de resultados apresentados aos usuários e, portanto, direcionar com maior precisão a expressão de busca para encontrar os documentos desejados.

Os mecanismos de busca alteram a maneira como executam uma expressão de busca em relação à forma como os usuários a executam. Assim, a substituição do operador *AND* por *DE* ou mesmo, sempre considerar um *AND* entre cada palavra da busca, são evidências desta influência. E aqui é importante salientar que o operador booleano *AND*, descrito ou não na expressão de busca, funciona de forma implícita em mecanismos de busca como o Google, justamente por sua função de intersecção de temas ser algo basilar na lógica de busca dessas ferramentas.

Os usuários por vezes sentem necessidade de incluir outros termos na expressão de busca que vão além de operadores de pesquisa ou das palavras-chave principais relacionadas à temática de uma pesquisa. Termos como preposições, conjunções e demais qualificadores como *datas*, *locais* e *formatos*, reforçando a ideia introduzida pelo *AskJeeves* nos anos 90 de que a pesquisa pode ser expressa da forma de diálogo entre humanos, ou seja, com linguagem natural.

Os algoritmos de pesquisa de fato tornam cada tela de resultado particular. Pesquisas bem específicas aparentemente reduzem as diferenças, mas não a ponto de torná-las idênticas. Aliás, esse não é o objetivo, afinal, a personalização de interesses por meio de algoritmos é um dos grandes trunfos do *marketing* digital, com alto valor comercial ao considerar buscas, cliques e vendas.

O uso de imagens, ou seja, das capturas de telas como unidade de dados, pode gerar adversidades na pesquisa principalmente pela dificuldade em estipular o perímetro da captura, fazendo com que a dimensão dessas seja distinta de dispositivo para dispositivo, o que resulta em mais ou menos informações em cada um deles. Problema esse que poderia ter sido contornado se as páginas tivessem sido salvas de forma completa em formato *HTML*.

Pesquisas futuras podem realizar discussões a respeito dos mecanismos de busca semânticos baseados em modelos pré-treinados como BERT e GPT. Estes modelos já são realidade e, a partir de buscas mais contextuais, proporcionam uma experiência do usuário que se assemelha a ideia explorada pelo *AskJeeves* na década de 90 ao prenciar o uso de um mecanismo de busca com a consulta de um especialista humano. O ChatGPT, o GPT-4 incorporado ao Bing e o Bard do Google estão na corrida das inteligências artificiais e interferindo ativamente na forma de fazer pesquisa *online*.

Além disso, os dados abertos disponibilizados nessa pesquisa possibilitam novas análises mais aprofundadas acerca do uso de bloqueadores de anúncios, tipos de navegadores atrelados e diferentes buscadores e localização geográfica como elementos passíveis de interferência nos resultados de pesquisa apresentados aos usuários.

Referências

CAMOSSI, Gustavo; AMORIM, Cineia Josefina Da Rocha; RODAS, Cecilo Merlotti. Conceito de Relevância e as Técnicas de Search Engine Optimization: suas implicações no ranqueamento nos mecanismos de busca. *Brazilian Journal of Information Science: research trends*, v. 17, e023009, jan. 2023. ISSN 1981-1640. DOI: 10.36311/1981-1640.2023.v17.e023009. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/12982>. Acesso em: 20 nov. 2023.

CHOO, Chun Wei. *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. São Paulo: Ed. SENAC, 2003.

GONZÁLEZ GUITIÁN, María Virginia; DE ZAYAS PÉREZ, María Rosa; NÚÑEZ GRILLO, María Mercedes; RODRÍGUEZ ARIAS, Kadiuska; GONZÁLEZ SÁNCHEZ, Ana Leonor. Comportamiento informacional en Internet en estudiantes de Ciencias de la Información, Universidad de Holguín. *Información, cultura y sociedad*, n. 46, p. 107–128, jun. 2022. ISSN 1851-1740, 1514-8327. DOI: 10.34096/ics.i46.11250. Disponível em: <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/11250>. Acesso em: 20 nov. 2023.

GOOGLE. *Refinar pesquisas do Google - Ajuda da Pesquisa do Google*. [S. l.: s. n.], 2013. Disponível em: <https://support.google.com/websearch/answer/2466433?hl=pt>. Acesso em: 20 nov. 2023.

GOOGLE. *Escolher o tipo certo de campanha - Ajuda do Google Ads*. [S. l.: s. n.], 2018. Disponível em: <https://support.google.com/google-ads/answer/2567043?hl=pt-BR>. Acesso em: 20 nov. 2023.

GOOGLE. *Exibir anúncios na primeira página - Ajuda do Google Ads*. [S. l.: s. n.], 2021. Disponível em: <https://support.google.com/google-ads/answer/2472742?hl=pt-BR>. Acesso em: 20 nov. 2023.

GOOGLE. *Anúncios e dados*. [S. l.: s. n.], 2022. Disponível em: <https://safety.google/privacy/ads-and-data/>. Acesso em: 20 nov. 2023.

IBICT. *Stop words e sinonimos - IBICT*. 2015. Disponível em: http://wiki.ibict.br/index.php/Stop_words_e_sinonimos. Acesso em: 20 nov. 2023.

LACEY, Patricia. Google is goodish: An information literacy course designed to teach users why Google may not always be the best place to search for evidence. *Health Information & Libraries Journal*, v. 39, n. 1, p. 91–95, mar. 2022. ISSN 1471-1834, 1471-1842. DOI: 10.1111/hir.12401. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/hir.12401>. Acesso em: 20 nov. 2023.

LECARDELLI, Jane. *Dados científicos abertos em agências de fomento à pesquisa: cenário dos planos de gestão de dados (PGD) e princípios FAIR*. 2020. Dissertação (Mestrado) – Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Florianópolis. Disponível em: <http://sistemabu.udesc.br/pergamumweb/vinculos/00008b/00008bdf.pdf>. Acesso em: 20 nov. 2023.

MNGUNI, Khulekani; KEKANA, Matsobane. Web Information Seeking Behaviour of Undergraduate Students of Library and Information Studies at the University of Zululand. *Library Philosophy and Practice (e-journal)*, p. 1–18, jun. 2022. Disponível em: <https://digitalcommons.unl.edu/libphilprac/7174>. Acesso em: 20 nov. 2023.

PICALHO, Antonio Carlos; LUCAS, Elaine Rosangela De Oliveira; AMORIM, Igor Soares. Lógica booleana aplicada na construção de expressões de busca. *AtoZ: novas práticas em informação e conhecimento*, v. 11, p. 1–12, 2022. Disponível em: <https://revistas.ufpr.br/atoz/article/view/81838>. Acesso em: 20 nov. 2023.

ROSA, Bruna Neder. *Publicidade e os limites da privacidade: percepção dos usuários a anúncios personalizados*. 2022. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital, Pontifícia Universidade Católica de São Paulo, São Paulo. Disponível em: <https://tede.pucsp.br/bitstream/handle/25981/1/Bruna%20Neder%20Rosa.pdf>. Acesso em: 20 nov. 2023.

SANTOS, Luana Carla de Moura dos; VIERA, Angel Freddy Godoy. Avaliação da recuperação da informação em acervos digitais de jornais. *Em Questão*, p. 49–73, set. 2015. ISSN 1808-5245. DOI: 10.19132/1808-5245212.49-73. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/52362>. Acesso em: 20 nov. 2023.

STATCOUNTER. *Search Engine Market Share Brazil*. 2022. Disponível em: <https://gs.statcounter.com/search-engine-market-share/all/brazil>. Acesso em: 20 nov. 2023.

VIDMAR, Dale James. Darwin on the Web: The Evolution of Search Engines. *Computers in Libraries*, v. 19, n. 5, p. 1–5, 1999. Disponível em: <http://webpages.sou.edu/~vidmar/Articles/cildarwn.htm>. Acesso em: 20 nov. 2023.

WU, Xindong; ZHU, Xingquan; WU, Minghui. The Evolution of Search: Three Computing Paradigms. *ACM Transactions on Management Information Systems*, v. 13, n. 2, p. 1–20, jun. 2022. ISSN 2158-656X, 2158-6578. DOI: 10.1145/3495214. Disponível em: <https://dl.acm.org/doi/10.1145/3495214>. Acesso em: 20 nov. 2023.

YUS, Francisco. En qué difieren los géneros electrónicos de los analógicos. In: CONGRESO Leer y Escribir en Español en la Red. Aprendizaje, Renovación y Propuestas. Cantabria, Espanha: University of Comillas, 2011.

Contribuições dos autores

Antonio Carlos Picalho: Conceituação, Curadoria de dados, Investigação, Metodologia, Visualização, Escrita – rascunho original, Escrita – revisão e edição; **Luciane Maria Fadel:** Metodologia, Administração de projetos, Metodologia, Supervisão, Escrita – rascunho original, Escrita – revisão e edição; **Alexandre Leopoldo Gonçalves:** Supervisão, Escrita – rascunho original, Escrita – revisão e edição.