

# UNIDADE 4

## RECUPERANDO INFORMAÇÃO NA *WEB* USANDO OS MECANISMOS DE BUSCA GERAIS

---

### 4.1 OBJETIVO GERAL

Apresentar as ferramentas de busca avançada dos principais mecanismos de busca gerais e a forma de utilizá-las.

### 4.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

- a) utilizar os formulários de busca avançada de mecanismos de busca gerais para recuperar informações;
  - b) aplicar análise conceitual, lógica booleana e as ferramentas de busca avançada, em conjunto, para atender às demandas por informação.
- 





## 4.3 LAPIDANDO NOSSAS BUSCAS

Apesar das limitações que vimos na Unidade 3, os mecanismos de busca gerais são SRIs dos mais utilizados, tanto para resolver problemas do cotidiano das pessoas quanto para atividades profissionais e comerciais. Nesta Unidade, que é fundamentalmente prática, vamos exercitar, com a bagagem que adquirimos nas unidades anteriores – em especial, na Unidade 2 –, a formulação de buscas utilizando os recursos avançados dos mecanismos de busca gerais.

**Figura 64 – Que tal uma busca mais avançada para refinar nossa demanda por informação?**



Fonte: domínio público.<sup>17</sup>

## 4.4 RELEMBRANDO CONCEITOS

Antes de começarmos a tratar do tema principal desta Unidade, vamos sumarizar as principais limitações dos mecanismos de busca gerais que vimos em outras unidades.

Na unidade anterior, vimos que os mecanismos de busca realizam uma indexação automática, ou seja, eles indexam as páginas da *web* através da extração de palavras nelas encontradas, que são armazenadas em uma base de dados associada ao endereço da página. No entanto, na Unidade 2, nós vimos que a unidade de significado de uma expressão feita em linguagem natural não é a palavra, mas o conceito, e que este pode ser formado por uma ou mais palavras, por exemplo: atestado de bons antecedentes. Nossa questão com a busca utilizando os mecanismos de

<sup>17</sup> Autor: Hillebrand Steve, U.S. Fish and Wildlife Service. Disponível em: <<http://www.public-domain-image.com/free-images/people/number-of-tourists-with-binoculars-tour-of-the-environment-725x483.jpg>>. Acesso em: 3 ago. 2021.

busca gerais é como passar das palavras para esses *conceitos*. Uma das formas simples de fazer isso, já estudada na Unidade 2, é submeter um conceito formado por mais de uma palavra entre aspas, assim: “atestado de bons antecedentes”. A maioria dos mecanismos de busca aceita uma expressão entre aspas e esse será nosso recurso básico para fazer buscas por conceitos e não por palavras.

No entanto, muitos mecanismos de busca gerais possuem outros recursos para a realização de pesquisas mais precisas, e é isso o que veremos a seguir.

## 4.5 FORMULÁRIOS DE BUSCA AVANÇADA

Os mecanismos de busca gerais dispõem do que chamamos de formulários de busca avançada, que, em vez de uma simples caixa de entrada, onde o usuário entra com as palavras-chave, possuem outras opções, que facilitam a especificação de detalhes da informação desejada.

Agora, vamos explicar essas opções, usando como exemplo o formulário de busca avançada do *Google*, cujos campos são mostrados na Figura 65; nela, foi incluído um número para cada campo.

Figura 65 – Formulário de busca avançada do *Google*

Fonte: *Google*.<sup>18</sup>

Vejamos, agora, a função de cada campo do formulário:

- a) campo 1: serão recuperadas páginas que contenham TODAS as palavras entradas neste campo; é como se fosse feito um E booleano entre todas as palavras;

<sup>18</sup> Disponível em: <[http://www.google.com.br/advanced\\_search](http://www.google.com.br/advanced_search)>. Acesso em: 3 ago. 2021.

- b) campo 2: será recuperada uma expressão ou frase exata que tenha sido entrada neste campo; é uma das opções de busca por um conceito como “atestado de bons antecedentes” ou “síndrome de Down”, sem necessidade de se utilizar aspas;
- c) campo 3: serão recuperadas páginas que contenham *qualquer uma* das palavras entradas neste campo; é como se fosse feito um OU booleano entre todas as palavras;
- d) campo 4: serão recuperadas páginas que *não* contenham nenhuma das palavras entradas neste campo; é como se fosse feito um NÃO booleano de todas as palavras;
- e) campo 5: este campo será usado quando você desejar recuperar páginas dentro de um intervalo numérico. O uso mais comum seria recuperar páginas em um intervalo de datas. Repare, no entanto, que isso só será efetivo se as páginas contiverem suas datas de criação em seus respectivos campos de metadados;
- f) campo 6: este campo será usado quando você desejar recuperar páginas em determinado idioma por meio de uma lista de opções (clcando na seta do lado direito, uma janela com a lista se abrirá);
- g) campo 7: este campo permite especificar, por meio de uma lista de opções, o país ou região onde uma página foi publicada. O país no qual uma página é publicada é identificado pelo seu terminador do nome de domínio. Por exemplo: <http://www.uff.br> é uma página publicada no Brasil (.br). Se o campo não for preenchido, serão recuperadas páginas de qualquer país;
- h) campo 8: este campo permite selecionar páginas atualizadas a partir de determinada data e escolhidas a partir de uma lista de opções. As opções são “em qualquer data”, “nas últimas 24 horas”, “na última semana”, “no último mês” e “no último ano”;
- i) campo 9: este campo permite selecionar páginas que sejam de determinado tipo de domínio, entre os vários tipos existentes na *web*. Por exemplo, se desejo recuperar somente páginas comerciais, posso especificar, neste campo, o domínio .com; se desejo recuperar páginas de órgãos do governo brasileiro, posso especificar aí o domínio .gov.br.
- j) campo 10: este campo permite especificar em que lugar da página os termos de busca devem ser encontrados. As opções a serem especificadas são termos em “qualquer lugar da página”, termos encontrados somente “no título da página”, termos encontrados “no texto da página”, termos encontrados “no URL da página” ou termos encontrados “em *links* para a página”;
- k) campo 11: através deste campo, pode-se ativar a filtragem de conteúdo sexualmente explícito ou sobre violência, por meio do filtro *SafeSearch*. Ao ativarmos tal filtro, conteúdos sexualmente explícitos são eliminados das páginas de resultado do *Google*;
- l) campo 12: neste campo, podemos especificar o tipo de arquivo que queremos recuperar. Por exemplo, somente arquivos .pdf, somente planilhas .xls etc. As opções são:
- qualquer formato;
  - *Adobe Acrobat PDF* (.pdf);
  - *Adobe Postscript* (.ps);

## Domínio

Nome que serve, em Informática, para localizar e identificar conjuntos de computadores na internet. O conceito de domínio foi concebido com o objetivo de facilitar a memorização dos endereços de computadores, já que, sem eles, teríamos que memorizar uma sequência grande de números. Domínios que não são registrados não podem ser encontrados na internet, sendo que qualquer pessoa física ou jurídica pode ter um domínio. Os detalhes sobre a formalização dos registros dependem da legislação de cada país.

Fonte: FUNDO..., [20--?].



- *Autodesk DWF* (.dwf);
- *Google Earth KML* (.kml);
- *Google Earth KMZ* (.kmz);
- *Microsoft Excel* (.xls);
- *Microsoft PowerPoint* (.ppt);
- *Microsoft Word* (.doc);
- *Formato Rich Text* (.rtf);
- *Shockwave Flash* (.swf).

m) campo 13: neste campo, podemos especificar se queremos recuperar somente páginas que possuam uma licença específica de uso, por exemplo, páginas cujo conteúdo seja liberado segundo a licença de uso *Creative Commons*. As opções do campo são: “não filtrados por licença”, “sem restrições de uso ou compartilhamento”, “sem restrições de uso ou compartilhamento, mesmo comercialmente”, “sem restrições de uso, compartilhamento ou modificação” e “sem restrições de uso, compartilhamento ou modificação, mesmo comercialmente”.

Em vez de usar esses campos, as opções listadas podem ser acionadas por comandos ou símbolos que o *Google* reconhece. Essas outras opções são descritas ao final do formulário de busca avançada do *Google*, nos links que aparecem na sequência da expressão “Você também pode...”.

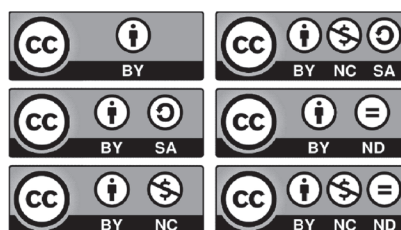


## Multimídia

Antes de começar a estudar esta disciplina, você já tinha ouvido falar nas licenças de uso *Creative Commons* (Figura 66)? Elas não são as únicas que oferecem uma forma padronizada de atribuir autorização de direito de autor e direitos conexos a obras da criação humana, mas podemos afirmar que é uma das mais conhecidas.

Se você quiser se aprofundar na função das licenças fornecidas por essa organização, seu desenho, fundamentação e também os diferentes tipos de licença, não deixe de visitar o endereço: <<https://br.creativecommons.org/licencas/>>.

Figura 66 – Licenças *Creative Commons*



Fonte: Pixabay.<sup>19</sup>

<sup>19</sup> Disponível em: <<https://pixabay.com/pt/creative-commons-licen%C3%A7as-%C3%ADcones-783531/>>. Acesso em: 3 ago. 2021.

Como pode ser visto, são oferecidos, no formulário de busca avançada do *Google*, uma série de recursos que permitem especificar com maior precisão as características das páginas que desejamos recuperar. Outros mecanismos de busca, como o *Yahoo*, também têm formulários de busca avançada com opções semelhantes (você pode encontrar o mecanismo de busca exemplificado em: <<http://br.yhs4.search.yahoo.com/web/advanced?ei=UTF-8&fr=altavista&p=\>>).)

#### 4.5.1 Agora é mão na massa!

A partir de agora, apresentaremos, junto a quadros e figuras, um conjunto de demandas por informação, expressas em linguagem natural, que usaremos para descrever como fazer as buscas usando os recursos de busca avançada do *Google* ou do *Yahoo*. Caso tenha disponível com você um dispositivo conectado à internet, procure acompanhar o passo a passo na prática.

Para que se tenha maior precisão na busca, será necessário utilizar o que foi aprendido, na Unidade 2, sobre análise conceitual e conectivos booleanos.

- a) **Exemplo 1** – queremos fazer uma busca por: longevidade média de portadores de síndrome de *Down*. Análise conceitual (Quadro 11):

Quadro 11 – Conceitos e termos (Exemplo 1)

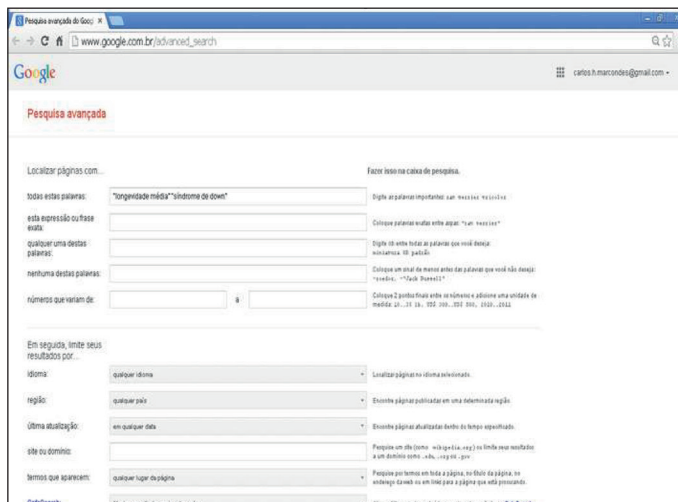
CONCEITOS	TERMOS
longevidade média	longevidade média
síndrome de <i>Down</i>	síndrome de <i>Down</i>

Fonte: produção do próprio autor.

Veja que é desnecessário inserir o conceito/termo “portadores”, porque a longevidade média relacionada à síndrome de *Down* só pode ser a dos indivíduos que são seus portadores.

Expressão booleana: “longevidade média” E “síndrome de *Down*”.

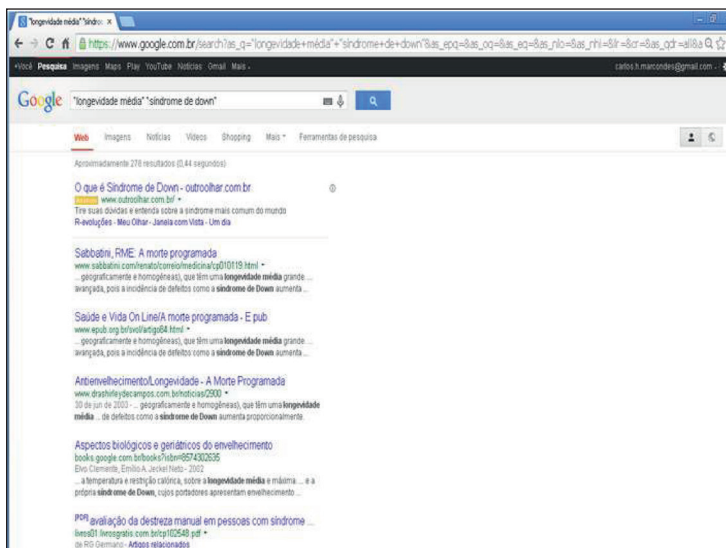
Figura 67 – Estratégia de busca do Exemplo 1, utilizando o formulário de busca avançada do *Google*



Fonte: captura de tela da página do *Google*.

Veja os resultados obtidos na Figura 68:

**Figura 68 – Resultados de busca do Exemplo 1, utilizando o formulário de busca avançada do Google**



Fonte: captura de tela da página no Google.

Se fizéssemos a busca utilizando a caixa *esta expressão ou frase exata*, obteríamos 203.000 resultados, contra 278 altamente pertinentes da tentativa anterior. A diferença pode ser atribuída ao fato de não ser revelado como o Google processa suas consultas. Isso deixa claro que a pesquisa por informações usando um mecanismo de busca geral requer, sobretudo, muita experiência.

- b) **Exemplo 2** – faremos a busca por: cultura negra no Brasil.  
Análise conceitual (Quadro 12):

**Quadro 12 – Conceitos e termos (Exemplo 2)**

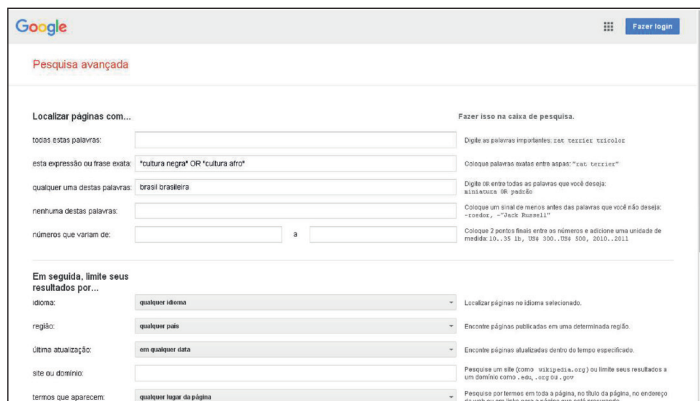
CONCEITOS	TERMOS
Cultura negra	Cultura negra, cultura afro
Brasil	Brasil, brasileira

Fonte: produção do próprio autor.

Expressão booleana: (cultura negra OU cultura afro) E (Brasil OU brasileira).



Figura 69 – Estratégia de busca do Exemplo 2, utilizando o formulário de busca avançada do Google

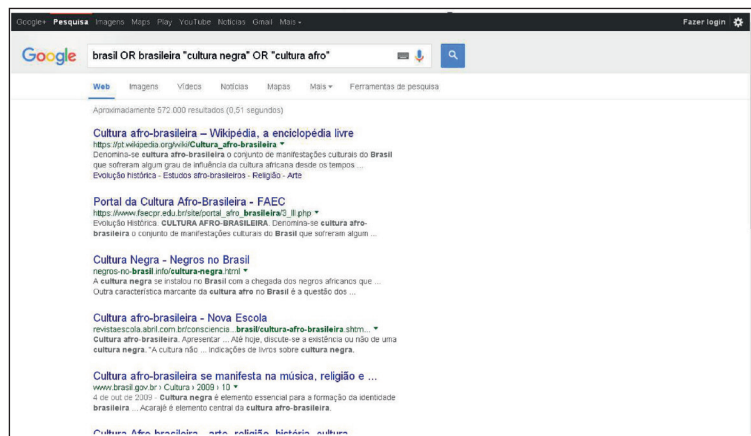


Fonte: captura de tela da página no Google.

Repare como na caixa *esta expressão ou frase exata* conseguimos entrar com dois termos entre aspas (recurso já conhecido por nós), “cultura negra” e “cultura afro”, unidos pelo operador booleano OR. Usamos também a caixa *qualquer uma das palavras* para entrar com os termos “Brasil” e “brasileira”, formados, cada um, por apenas *uma* palavra.

Veja, na Figura 70, os resultados obtidos:

Figura 70 – Resultados da busca do Exemplo 2, utilizando o formulário de busca avançada do Google



Fonte: captura de tela da página no Google.

Pode-se observar, já na primeira página de resultados desta busca, como eles foram recuperados com os quatro termos utilizados, que são os que aparecem no texto (embaixo do endereço), destacados em negrito.

- c) **Exemplo 3** – a proposta, aqui, será encontrar páginas que tenham *link* para a página do periódico *Perspectivas em Ciência da Informação*, que pode ser encontrado em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci>>.

Nesse caso, usaremos o operador de pesquisa *link*, que é utilizado para encontrar páginas vinculadas a um endereço específico. A lista e a descrição dos operadores de pesquisa podem ser encontradas,

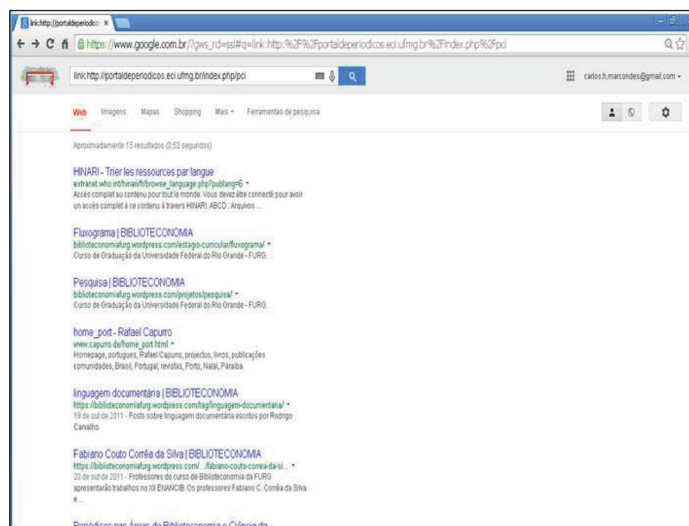


no caso do *Google*, no fim do formulário de busca avançada, após a expressão: “Você também pode...”. Basta clicar no *link* “Encontrar páginas semelhantes ou que vinculam a um URL” e você encontrará a explicação sobre o uso de cada um desses operadores.

Em nosso exemplo, portanto, digitaremos na caixa de busca: *link*: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci>.

Veja os resultados na Figura 71:

**Figura 71 – Resultados da busca do Exemplo 3, utilizando ferramentas de busca avançada do *Google***



Fonte: captura de tela da página no *Google*.



## 4.5.2 Atividade

Avalie as demandas por informação a seguir e explique como você as atenderia, se estivesse usando o formulário de busca avançada do *Google*:

- a) notícias sobre atentado terrorista em Londres;

---

---

- b) informações sobre a ocorrência de Malária fora do continente africano;

---

---

- c) artigos sobre *crack* na língua inglesa, em formato .pdf.;

---

---

- d) notícias sobre o combate à dengue no Brasil, em *sites* de órgãos do *Governo Federal*.
- 
- 

### Resposta comentada

Respostas:

- a) "atentado terrorista" e Londres, na caixa de pesquisa "Todas estas palavras";
- b) malária, na caixa de pesquisa "Todas estas palavras" e "continente africano", África, na caixa de pesquisa "Nenhuma destas palavras";
- c) *crack*, na caixa de pesquisa "Todas estas palavras", escolher inglês na caixa de pesquisa "idioma" e escolher *Adobe Acrobat PDF* (.pdf) na caixa de pesquisa "tipo de arquivo";
- d) "combate à dengue" e Brasil, na caixa de pesquisa "Todas estas palavras" e .gov.br na caixa de pesquisa "site ou domínio".



## 4.5.3 Atividade

Vamos ver se você consegue fazer uma busca sozinho? Mas, para realizar esse desafio, você precisará ter à disposição um dispositivo com acesso à internet.

Sua missão é encontrar voos entre as cidades do Rio de Janeiro e São Paulo, com preços que variem entre R\$100,00 e R\$300,00. Uma dica: utilize operadores de pesquisa.

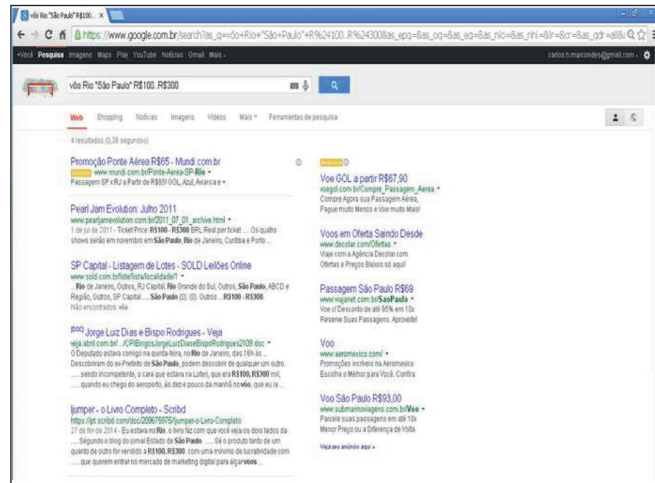
### Resposta comentada

Para encontrar resultado que demande um intervalo numérico, a melhor estratégia é utilizar um operador de pesquisa. Neste caso, usaremos o símbolo ".." para definir um intervalo de preços. Ou seja, você pode digitar, na caixa de busca, algo como: voo Rio "São Paulo" R\$100..R\$300. Veja os resultados que encontramos em nossa busca (Figura 72):

Semestre

6

Figura 72 – Resultados da busca



Fonte: captura de tela da página de resultados do Google.

## 4.6 CONCLUSÃO

O que aprendemos sobre análise conceitual, na Unidade 2, juntamente com as facilidades fornecidas pelos formulários de busca avançada dos mecanismos de busca gerais, vistas nesta quarta Unidade, permitem uma melhor utilização desses SRIs. Com esse conhecimento, é possível que tenhamos maior precisão e refinamento na especificação das demandas de informação dos usuários quando utilizamos os mecanismos de busca no dia a dia do nosso trabalho.

## RESUMO

Os mecanismos de busca gerais oferecem recursos muito mais sofisticados para especificar, com precisão, as demandas por informação dos usuários, por meio dos seus formulários de busca avançada. Nesta Unidade, exercitamos diversos recursos oferecidos pelo formulário de busca avançada do *Google*. Outros mecanismos de busca têm recursos semelhantes.

# REFERÊNCIA

---

FUNDO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO.  
**FAPESP**, São Paulo, [20--?]. Disponível em: <<http://fapesp.org/faq/>>.



## Sugestão de Leitura

---

LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 60-71, maio/ago. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12909.pdf>>. Acesso em: 14 jul. 2015.

---

