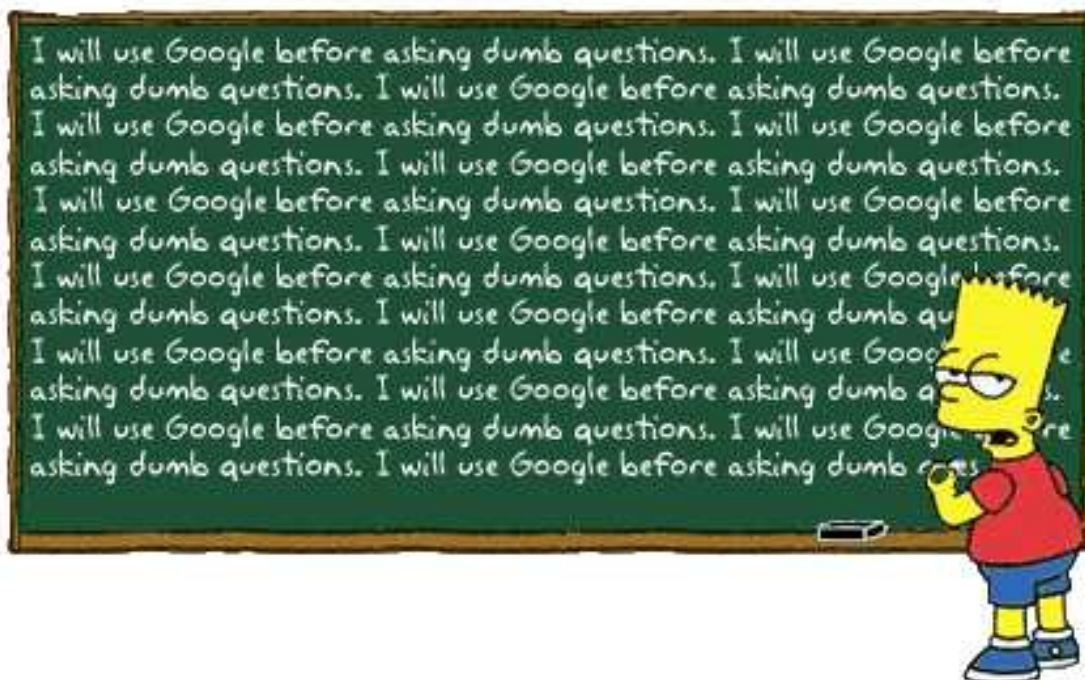


Como pesquisar no google

Carlos Duarte
cgd[]sdf-eu.org
Maio 2003, r2



Introdução

Tradicionalmente, para resolver problemas, as pessoas pensavam neles até eventualmente os solucionarem. Uma vez obtida a solução, tentava-se memorizar o modo como foi obtida, ou seja, a receita geral e não a solução de um caso em concreto. Hoje em dia as coisas mudaram. Perante um problema, a primeira reacção normal é aceder a um motor de pesquisa e tentar encontrar a solução existente mais parecida possível.

“Evoluiu-se” do clássico “resolver e memorizar” para o moderno “procurar e adaptar”. Porquê? Porque é mais fácil, rápido e... pode ser feito.

Antes de haver *internet*, as grandes concentrações de informação encontravam-se quase em **exclusivo** nos livros, revistas e publicações em geral, que, por razões logísticas, são de acesso e consulta mais limitada. Uma limitação óbvia é o facto de não serem pesquisáveis de forma automática.

Com os computadores surgiu um novo espaço para criar, armazenar e manipular informação. Com a *internet* surgiu uma nova forma de a publicar e aceder. Qualquer pessoa ou empresa dispõe de uma enorme facilidade de criar informação — por mais insignificante ou incorrecta que seja — e publicá-la na rede. Paralelamente dispõe também de uma forma simples e eficiente de a consultar (*browsers*, por exemplo) e de a procurar (motores de pesquisa).

Os motores de busca¹ são serviços que nos permitem tentar encontrar a informação que pretendemos,

¹ ou motores de pesquisa, *search engines*.

na quantidade semi-anárquica que está disponível na *internet*. Apesar dos esforços e tecnologias empregues no processo de indexação e pesquisa, uma boa parte do sucesso das nossas procuras continua a depender do tipo de pesquisa que fazemos.

Tendo em conta as considerações de cima, procurar bem significa “resolver” melhor e mais rapidamente problemas. Este texto tenta fornecer algumas pistas e técnicas de como bem procurar, usando o *google*, que é o motor de pesquisa mais popular actualmente.

Técnicas de pesquisa

Uma procura, ou pesquisa, consiste normalmente em introduzir algum (pouco) texto numa caixa e submeter esse pedido a um motor de busca.

O problema mais comum é o texto ser demasiadamente curto, logo muito genérico, encontrando muitas soluções que não nos interessam, a que vulgarmente se chama “lixo”. Adicionar demasiado texto à pesquisa não ajuda, pois passa-se de uma situação demasiada genérica, com muitos resultados, para outra demasiado específica, sem resultados ou com poucos resultados de lixo.

A principal regra é **contextualizar**. Aliás, esta regra é importante em motores de busca, em discurso verbal, escrito, etc... É importante contextualizar.

Nem as pessoas nem os motores de busca sabem o que nós estamos a pensar. Em alguns casos, contudo, esse contexto pode ser omitido. São os casos mais comuns, mais procurados, que os motores normalmente retornam em primeiro. Por exemplo, *Unix* é um sistema operativo inventado no fim dos anos 60 e muito popular na *internet* e na área da computação. Se se procurar apenas por `unix` é natural que as primeiras dezenas, ou mais, de ligações encontradas, sejam relacionadas com esse sistema operativo. Neste caso não é preciso contextualizar. Mas se se quiser encontrar informação sobre a marca de canetas *Unix*, convém adicionar uma ou outra palavra extra (ver abaixo).

O mundo da *internet* é inerentemente técnico. A maior parte dos conteúdos são técnicos, relacionados sobretudo com computação em geral, tecnologias de informação, redes, sistemas operativos, linguagens de programação de computadores... Também existem muitos conteúdos de notícias, que apesar de serem dinâmicos — a notícia de ontem já não o será amanhã — são por vezes arquivados nos próprios *sites* que os produzem, e indexados pelos motores. Em escala talvez um pouco menor, existem materiais científicos e tecnológicos. As outras áreas existem em menor quantidade. É de esperar que as pesquisas retornem resultados mais ou menos abundantes conforme a natureza do tema. Mais irritante é procurar por um dado tema com um dado nome, e encontrar duzentas e três mil referências com o mesmo nome... mas sobre outro tema. Por exemplo, quem quiser obter informações sobre a ilha de Java, na Indonésia, vai aprender da pior forma toda a panóplia e afins sobre essa tecnologia informática :-)

Os resultados referidos ao longo do texto foram obtidos no *www.google.com* usando as pesquisas das caixas, em 28-Maio-2003.

Contexto I

Adicionar uma ou mais (mas não demais!) palavras para fugir aos resultados mais óbvios. Caso do *Unix*: qualquer referência a *Unix* vai parar sempre ao mesmo tipo de páginas, que são as que se relacionam com o sistema operativo com o mesmo nome. Existem outros produtos com o nome *Unix* — embora o **verdadeiro** e puro *Unix* seja o sistema operativo :-) — como canetas!

A pesquisa:

`unix pens office`

retorna logo no primeiro resultado informação para esses *Unixes* alternativos, embora ressalve-se que o URL em causa é da autoria de um dos criadores do sistema operativo *Unix*, que teve a “gentileza” de compilar uma página com informação sobre outros *unixes*.

Contexto II

No caso *Unix* trata-se de tentar encontrar “Unix” em contextos muito raros, pelo menos existentes na *internet*. Esta é uma das situações mais difíceis. Uma variante também complicada consiste em procurar por um tópico específico dentro de um domínio comum. Por exemplo, saber o que significam as letras S/A/P da empresa alemã SAP. O domínio comum e frequente aqui é SAP. Qualquer procura por essa palavra retorna informação genérica sobre a empresa e os seus produtos mais populares. O domínio específico neste caso é o significado das letras. A isso chama-se um acrónimo, por isso a melhor procura a fazer será

```
sap acronym
```

Neste caso, a resposta encontra-se logo no quarto URL retornado:
<http://www.sapdesignguild.org/FAQ.ASP>².

Contexto III

Procurar directamente pelas palavras que queremos pode não ser suficiente para isolar completamente uma busca. Um truque passa por adicionar textos que imaginemos estarem presentes na mesma página, ou textos que acompanhem as nossas palavras. Por exemplo, para saber algo sobre a origem da equação resolvente, procurar por equação resolvente não vai ajudar. Mas como queremos saber coisas sobre o passado, é expectável que as palavras *history* ou *dates back* apareçam nas páginas que nos interessam. De facto,

```
solving equation second degree "dates back"
```

retorna logo como primeiro resultado o *site*
<http://www.ucs.louisiana.edu/~sxw8045/history.htm> que contem alguma informação sobre o assunto.

Procurar dentro dos *sites*

O *google* dispõe de operadores especiais que modificam a forma como a procura é feita (ver opções e operadores). Um deles, *site:*, limite as pesquisas a determinados sites. Desta forma conseguimos limitar o número de resultados retornados e aumentar o seu interesse, desde que conheçamos um fonte de informação sobre o assunto que queremos pesquisar.

- informação sobre *webservices* em *python*, na *O'Reilly Network*:

```
site:oreillynet.com webservices python
```

- notícias sobre *python*, no *Dr. Dobb's Journal*:

```
site:www.ddj.com python news
```

Procura literal

O *google* interpreta o texto que colocamos na caixa de pesquisa como um conjunto de palavras a procurar. Se quisermos que esse texto seja interpretado como uma só frase, devem-se usar aspas a envolvê-la. Útil para encontrar ficheiros ou *sites*.

Remover lixo

Se fizermos uma dada busca que produza resultados com muito lixo (em relação ao que pretendíamos originalmente), um truque para refinar a pesquisa consiste em adicionar vários “-xxx”, que significam: retorna o mesmo, excepto “xxx”. Alguns exemplos:

- remover material Brasileiro

```
xpto -br -brazil
```

² SAP: Systems, Applications, Products (German: Systeme, Anwendungen, Produkte)

- apanhar texto britânicos e não americanos (ingleses têm terminações em “s” e os americanos em “z”): *optimise* vs. *optimize*).

`optimisation -optimization`

De notar que para este tipo de questões linguísticas existem opções específicas incorporadas no *google*, mas a ideia é generalizar o conceito.

Next → Next → Next ...

Não desistir na primeira página de respostas retornada. Nem na segunda...

Quando todos os truques de cima não conseguem localizar algo nos primeiros resultados, é porque a informação que queremos é muito específica e procurada por poucas pessoas. É provável que não exista nenhuma sequência de pesquisa que a localize de imediato. O truque é fazer *next* várias vezes. Já cheguei a encontrar coisas entre as páginas vinte e trinta, por exemplo.

Operadores e opções

Para se procurar por um dado texto, basta inseri-lo na caixa de pesquisa do *google*. Simples. Mas existem alguns operadores e opções que são úteis para refinar as buscas. Os operadores são palavras-chave ou sinais que se colocam na caixa de procura, juntamente com o restante texto, para produzirem efeitos especiais: ou seja, são usados para controlar a pesquisa, e não como elementos a pesquisar. As opções acedem-se nas ligações (*links*) disponíveis.

Operadores de pesquisa

+palavra

Algumas palavras, por serem muito frequentes, são automaticamente ignoradas pelo motor. O + força que essa palavra seja incluída. Por exemplo “how to make bombs” equivale a “make bombs”, porque “how” e “to” são ignorados. Enquanto

`+how +to make bombs`

devolve resultados mais... hum... relacionados com o tema.

-palavra

Rejeita a palavra especificada. Muito útil para filtrar “lixo”. Por exemplo,

`conjuntiva -es`

procura o termo “conjuntiva” tentando eliminar referências espanholas, embora isso pudesse ser feito através das opções linguísticas.

“frase completa”

Procura por frases completas. Uma procura por “how to write good” procura por “how”, “to”, “write” e “good”. Como a procura tenta encontrar páginas que contenham todos os termos, os resultados podem ser satisfatórios, mesmo assim não garantem que a frase aparece como está.

Esta técnica é útil para procurar sequências muito específicas. Um exemplo em concreto são nomes de ficheiros:

`"weblogic.properties"`

encontra todos os ficheiros com este nome que foram indexados pelo motor.

link:url

Procura por páginas que tenham ligações para o URL especificado. Exemplo:

`link:www.compuquali.com`

páginas que apontem para o *site* da CompuQuali.

related:url

Encontra páginas que o *google* pensa serem semelhantes ao URL indicado. Aquilo que o *google*

pensa muitas vezes é interessante. A seguinte busca:

```
related:www.sapo.pt
```

retorna (em 26-maio-2003) os seguintes URLs: *www.sapo.pt*, *www.aeiou.pt*, *forum.sapo.pt*, *www.clix.pt*, *www.busca.net*; o que parece lógico, porque os *sites* estão de facto relacionados.

`site:url`

Restringe as pesquisas a conteúdo no URL (e sub-URLs) indicado. Esta é uma das opções mais úteis, em minha opinião. Permite fazer pesquisa dentro de *sites* normalmente de forma muito mais eficiente do que as funcionalidades de pesquisa do próprio *site*.

```
site:www.sic.pt ponte vasco gama aumento portagem
```

`intitle:palavra`

Pesquisas que contenham a palavra indicada no título da página.

`inurl:palavra`

Páginas cuja palavra indicava esteja presente no URL da página. Útil para encontrar recursos por nome, quando a sua colocação nos *sites* é estruturada. Por exemplo, no *www.freshmeat.net* um *site* que anuncia *software*, os artigos lá expostos têm “/articles/” no URL. Assim:

```
site:freshmeat.net inurl:/articles/
```

dá-nos uma lista de artigos desse *site*.

Opções

Em http://www.google.com/advanced_search existe um página de procura avançada. Estão disponíveis algumas opções que só aqui se conseguem usar, como:

- páginas escritas em determinadas linguagens (inglês, português, ...)
- ficheiros num dado formato (PDF, PS, ...)
- páginas modificadas nos últimos 3, 6 e 12 meses

As outras opções são *front-ends* para os operadores especiais referidos em cima:

- Caixa *domain* → `site:`
- Caixa *similar* → `related:`
- Caixa *links* → `link:`

Outros serviços do google

Além de procuras *web*, o *google* fornece uma variedade de outros serviços, muitos dos quais não são tão amplamente conhecidos como a pesquisa *web*.

- <http://images.google.com/> — procura de imagens
- <http://groups.google.com/> — procura nos grupos de discussão
- <http://directory.google.com/> — serviço de directório: informação classificada por tópicos.
- <http://news.google.com/> — serviço de notícias
- <http://answers.google.com/> — perguntas e respostas
- http://translate.google.com/translate_t — tradução de texto entre umas poucas línguas (inclui o português)

Resumo

+palavra	força inclusão desta palavra na busca
-palavra	remove lixo associado a esta palavra
"uma frase"	procura esta frase por si, não a separa em palavras
link:url	páginas que liguem para este URL
related:url	páginas semelhantes a este URL
site:url	procura apenas neste URL (e sub-URLs)
intitle:palavra	páginas cujo título contenha palavra
inurl:palavra	páginas em que palavra apareça nos seus URLs
contexto	adicionar palavras extra ao que se pretende
next, next	ver as segundas e terceiras páginas de resultados...