Tire o máximo partido do seu conteúdo

Um manual para editores da Web

Google™

Conteúdo

Introdução 2

Uma breve introdução à pesquisa na Web 3

Novidades na pesquisa na Web do Google? 4

O Google vai encontrar o seu Web site? 5

O Google vai indexar o seu Web site? 6

Controlar o que o Google indexa 7

Robots.txt versus meta tags 9

Controlar o armazenamento em cache e snippets 10

O seu Web site apresenta conteúdo exclusivo e útil? 11

Aumentar a visibilidade: práticas recomendadas 12

Centro para webmasters 13

Sitemaps 14

Perguntas frequentes 15

Glossário 19

Introdução

Se procura visibilidade, a Internet é o lugar ideal. Pergunte a qualquer anunciante que tenha aumentado as vendas utilizando anúncios on-line, a um blogger cuja popularidade o tenha levado à publicação de um livro ou a um jornal que tenha ampliado o seu público ao nível internacional graças ao aumento de tráfego.

Recebemos frequentemente perguntas sobre como funciona a pesquisa na Web e como os editores da Web podem maximizar a sua visibilidade na Internet.

Preparámos este folheto para o ajudar a compreender como um mecanismo de pesquisa "percepciona" o seu conteúdo, e como pode afinar a sua presença na Web, disponibilizando o que pretende que seja visualizado – e ocultando o que deseja manter oculto.

Desde sugestões de webmasters e ferramentas on-line a um manual passo-a-passo com perguntas frequentes, este folheto foi desenvolvido tanto para pequenos editores da Web como para proprietários de grandes Web sites.

Assim como a própria Internet evoluiu drasticamente na última década, a abordagem do próprio Google em relação à pesquisa na Web e o seu relacionamento com os proprietários de Web sites também evoluíram. Criámos diversas ferramentas para ajudar os webmasters a maximizar a visibilidade do seu conteúdo, bem como a controlar a forma como as suas páginas da Web são indexadas. Mas há sempre mais que podemos fazer e por isso esperamos que este folheto o incentive a partilhar connosco os seus comentários e a informar-nos sobre o que podemos fazer para tornar a Web um lugar ainda melhor para utilizadores e editores.

- A equipa de webmasters do Google

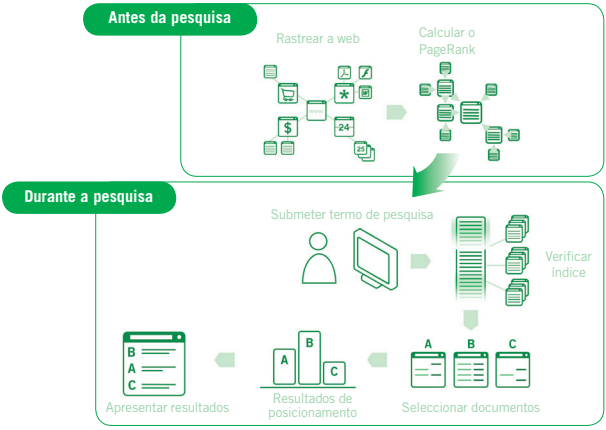
Uma breve introdução à pesquisa na Web: *como funciona*

De forma simples, pode imaginar a Web como um enorme livro, com um índice de grandes dimensões que indica exactamente onde é que tudo está localizado.

O Google tem um grupo de computadores – o "Googlebot" – que estão continuamente a "rastrear" (a navegar) bilhões de páginas na Web. Este processo de rastreio é algorítmico: os programas de computador determinam quais os Web sites que devem ser rastreados, com que frequência e quantas páginas de cada Web site devem ser analisadas. Não aceitamos pagamento para rastrear um Web site com maior frequência; além disso, mantemos a parte de pesquisa perfeitamente independente do nosso serviço de publicidade Google AdWords.

O processo de rastreio do Google inicia com uma lista de URLs de páginas da Web. À medida que o Googlebot navega nestes Web sites, ele detecta os links de cada página e adiciona-os à sua lista de páginas a serem rastreadas. O Googlebot faz uma cópia de cada uma das páginas que rastreia, a fim de compilar um índice de grande proporção de todas as palavras que visualiza. Esta lista também indica o local onde cada palavra surge em cada página.

Quando um utilizador insere um termo de pesquisa, as nossas máquinas pesquisam no índice páginas relacionadas com o mesmo e apresentam ao utilizador os resultados mais relevantes. A relevância é definida por mais de 200 factores, sendo um deles o "PageRank" da página. O PageRank é uma medida de "importância" de uma página com base nos links provenientes de outras páginas. Em resumo, cada página que possua um link para uma página XYZ aumenta o PageRank dessa página XYZ.



O que há de novo na pesquisa na Web do Google?

Embora os princípios básicos da pesquisa na Web tenham permanecido constantes, o Google está continuamente a trabalhar no sentido de melhorar os resultados de pesquisa.

Por exemplo, o que é que há de diferente na pesquisa na Web, desde os últimos cinco anos? Bem, para começar, está muito mais rápida.

Além disso, em comparação a cincos anos atrás, os nossos sistemas de rastreio e indexação estão muito mais inteligentes. Por exemplo, agora navegamos continuamente na Web e programamos visitas a cada página de forma mais inteligente a fim de manter o conteúdo sempre actualizado. Esta abordagem mais eficiente considera o facto do Web site de um jornal on-line, por exemplo, precisar de ser rastreado com maior frequência do que um Web site estático que seja actualizado uma vez por mês. Para mais, permitimos ainda que os webmasters controlem com frequência os seus Web sites são rastreados utilizando as nossas ferramentas para webmasters. De um modo geral, isto resulta num índice mais actualizado e abrangente.

Ainda que actualmente a pesquisa na Web esteja mais rápida e mais eficiente do que nunca, os principais factores para determinar a visibilidade de um Web site nos resultados de pesquisa do Google têm sido uma prioridade desde o dia em que o nosso mecanismo de pesquisa foi inaugurado:

O Google vai encontrar o Web site? (página 5)

O Google vai indexar o Web site? (página 6)

O Web site apresenta conteúdo exclusivo e útil? (página 11)

O Google vai encontrar o seu Web site?

A inclusão nos resultados de pesquisa do Google é gratuita e simples; nem sequer é necessário enviar o seu Web site para o Google. Na verdade, a grande maioria dos Web sites relacionados nos nossos resultados de pesquisa não são enviados manualmente para inclusão, mas localizados e adicionados automaticamente quando o Googlebot rastreia a Web.

Embora o Google rastreie milhares de milhões de páginas,é inevitável que alguns Web sites nao sejam visualizados. Quando isto acontece, geralmente é devido a um dos seguintes motivos:

- o Web site não está associado através de links incluídos noutros Web sites na Web;
- o Web site foi iniciado depois da conclusão do mais recente rastreio do Google;
- o Web site estava temporariamente indisponível ou recebemos uma mensagem de erro quando o tentamos rastrear.

Com as ferramentas para webmasters do Google, como o Sitemaps, poderá verificar se o Web site está incluído no índice do Google ou se recebemos erros quando o tentamos rastrear (consulte a página 14). Também poderá utilizar estas ferramentas para adicionar o seu URL manualmente ao índice do Google, ou fornecer ao Google um Sitemap com mais informações sobre o seu conteúdo. Isso irá ajudar a encontrar novas secções e conteúdos do seu Web site.

O Google vai indexar o seu Web site?

Por vezes, os webmasters descobrem que os seus Web sites não estão a aparecer nos resultados de pesquisa. O problema poderá estar relacionado com a "capacidade de indexação", ou seja, a capacidade de o Google fazer ou não uma cópia de uma página da Web para inclusão nos nossos resultados de pesquisa.

Estrutura e conteúdo

Uma razão para a não inclusão nos resultados de pesquisa está relacionada com a estrutura e o conteúdo da página. Por exemplo, uma página que requer que o utilizador preencha um formulário pode não ser indexada pelo Google. Da mesma forma, uma página utilizando "conteúdo dinâmico" (Flash, JavaScript, frames ou URLs criados dinamicamente) pode não ser indexada facilmente pelos mecanismos de pesquisa. Se quiser confirmar que é este o problema do seu Web site, tente visualizá-lo num navegador de texto, como o Lynx, ou num navegador com imagens, Javascript e Flash desactivados, o que indicará se todo o seu conteúdo está acessível ou não.

Se o seu Web site utilizar muitas imagens, certifique-se de que inclui uma descrição do conteúdo de cada imagem no texto. Isto permite que os mecanismos de pesquisa indexem a imagem correctamente e disponibiliza a imagem para acesso por parte dos utilizadores com problemas na visualização. Também poderá utilizar o texto alternativo da imagem e nomes de arquivos descritivos, como mostrado neste exemplo (que é uma imagem de um logotipo de uma empresa intitulada "Buffy's House of Pies" - algo como "Casa de tortas do Buffy"):

```
<img src="buffyshouseofpies.jpg" alt="Bem-vindo à
Buffy's House of Pies!">
```

URLs

Um outro obstáculo poderia ser o próprio URL. Se houver IDs de sessão ou vários parâmetros no URL ou se o URL efectuar vários redireccionamentos, é possível que o Google não consiga indexar a página.

Servidor e rede

Problemas com o servidor ou com a rede podem impedir o acesso a determinadas páginas do seu Web site. Ao utilizar as ferramentas disponíveis no Centro para webmasters do Google, os editores podem ver a lista das páginas que não puderam ser acedidas pelo Googlebot. Para saber mais sobre o Centro para webmasters, consulte a página 13.

Protocolo de exclusão de robôs

Por vezes, as páginas são bloqueadas pelo Protocolo de exclusão de robôs, um padrão técnico que permite que os editores da Web dêem indicação aos mecanismos de pesquisa para não indexar o conteúdo do Web site (consulte a página 7). Se o seu Web site não estiver a ser exibido nos resultados de pesquisa do Google, verifique se o robots.txt ou uma meta tag não estão a bloquear o acesso aos nossos rastreadores.

Controlar o que o Google indexa

Cada editor da Web tem um objectivo diferente em relação ao que está a tentar alcançar na Internet. Editores de alguns jornais, por exemplo, optaram por disponibilizar o acesso gratuito aos seus artigos mais recentes, oferecendo um serviço premium pago para acesso aos arquivos anteriores. Alguns querem visibilidade em todas as propriedades de um mecanismo de pesquisa (como o Google Mobile, o Google Imagens, etc.), enquanto outros preferem aparecer apenas nos resultados de pesquisa na Web.

Os mecanismos de pesquisa respeitam os desejos dos editores – afinal de contas, o conteúdo é deles. No entanto, não podemos adivinhar estes desejos, de forma que é essencial que os webmasters nos informem como desejam que o seu conteúdo seja indexado. Isto pode ser feito através do Protocolo de exclusão de robôs, uma especificação técnica bem definida que informa aos mecanismos de pesquisa qual o Web site ou partes do mesmo que não devem ser pesquisáveis, bem como quais partes que devem permanecer visíveis nos resultados de pesquisa.

Robots.txt: controlo alargado dentro de cada Web site

A base do Protocolo de exclusão de robôs é um arquivo de texto simples intitulado robots.txt, o qual tem sido o padrão do sector há muitos anos. Com o robots.txt, poderá controlar o acesso a vários níveis, desde um Web site inteiro a directórios individuais, páginas de um tipo específico ou até mesmo páginas individuais.

> Há algumas páginas no meu Web site que eu não quero que estejam no índice do Google. Como faço para que elas não apareçam nos resultados de pesquisa do Google?

Normalmente, a maioria dos proprietários quer que o Googlebot aceda aos seus Web sites para que as suas páginas da Web sejam encontradas pelos utilizadores que pesquisam no Google. No entanto, talvez não queira indexar determinadas páginas: por exemplo, logs internos ou artigos de notícias que estejam associadas a um acesso pago.

Poderá excluir páginas do índice do Google ao criar um arquivo robots.txt e posicioná-lo no directório raiz no seu servidor da Web. O arquivo robots.txt relaciona as páginas que os mecanismos de pesquisa não devem indexar. A criação de um arquivo robots.txt é simples e proporciona aos editores um nível sofisticado de controlo em relação à forma pela qual os mecanismos de pesquisa irão aceder aos seus Web sites.

Por exemplo, se um webmaster quiser impedir a indexação dos seus logs internos, o arquivo robots.txt deve conter:

`User-Agent:Googlebot` – a linha User-Agent especifica que a próxima secção é um conjunto de instruções específicas para o Googlebot.

`Disallow: /logs/` – a linha Disallow instrui o Googlebot a não aceder a arquivos no subdirectório logs do seu Web site.

O proprietário do Web site especificou que nenhuma das páginas no directório de logs deve ser exibida nos resultados de pesquisa do Google.

Todos os principais mecanismos de pesquisa irão verificar e seguir as instruções inseridas no robots.txt; além disso, se desejar, poderá especificar diferentes regras para diferentes mecanismos de pesquisa.

Meta tags: controlo mais específico

Além do arquivo *robots.txt* – que permite que especifique instruções precisas para um grande número de arquivos no seu Web site – poderá utilizar a meta tag de robô para ter um controlo mais específico sobre cada uma das páginas do seu Web site. Para implementar, basta adicionar meta tags específicas a uma página HTML, a fim de controlar a forma como essa página deverá ser indexada. Juntos, o *robots.txt* e as meta tags, proporcionam flexibilidade para expressar de maneira relativamente fácil as políticas de acesso complexas.

> Tenho um artigo específico de notícias no meu Web site que está acessível apenas para utilizadores registrados. Como impeço que ele seja incluído nos resultados de pesquisa do Google?

Para fazer isto, basta adicionar a meta tag NOINDEX à primeira <head> secção do artigo. Deverá ter o seguinte formato:

```
<html>
<head>
<meta   name="googlebot"   content="noindex">
[...]
```

Isto irá impedir que o Google indexe o respectivo arquivo.

No entanto, note que nalguns casos, poderá querer que o Google indexe este tipo de páginas – por exemplo, um artigo de notícias em arquivo que os utilizadores pagam para a leitura on-line. Embora este tipo de conteúdo "premium" não seja exibido nos resultados de pesquisa do Google, determinados serviços do Google, como a Pesquisa no arquivo do Google Notícias, incluirão o artigo nos seus índices, com as informações de pagamento claramente visíveis para os utilizadores.

Robots.txt versus meta tags

Em geral, o *robots.txt* é um bom modo de obter o controlo sobre todo o Web site e que as meta tags de fornecer o controlo mais específico sobre arquivos individuais. As meta tags são especialmente úteis se tiver permissão para editar arquivos individuais mas não todo o Web site. As meta tags também permitem que especifique políticas complexas de controlo de acesso em cada uma das páginas.

Às vezes, qualquer uma das duas ferramentas pode resolver o mesmo problema.

> Como posso ter certeza de que o texto de uma página será indexado, mas não as imagens?

Uma opção seria bloquear o acesso às imagens por extensão de arquivo em todo o seu Web site utilizando o *robots.txt*. As seguintes linhas num arquivo robots.txt indicam ao Google que não deverá indexar quaisquer arquivos com a extensão **.jpg* ou **.jpeg*:

```
User-agent:   Googlebot
Disallow:      /*.jpg$
Disallow:      /*.jpeg$
```

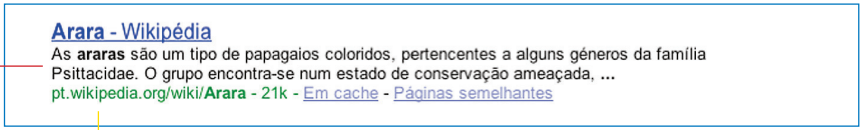
Em alternativa, se o seu Sistema de gestão de conteúdo (CMS, Content Management System) armazenar imagens num directório separado, poderá excluir o directório completo. Se as suas imagens estiverem num directório intitulado /imagens, poderá excluir esse directório de todos os mecanismos de pesquisa utilizando:

```
User-agent:   *
Disallow:   /images/
```

Outra opção seria adicionar uma tag NOINDEX a cada arquivo que incluir uma imagem. Todas estas abordagens irão impedir que as suas imagens sejam indexadas; a única questão é o quão ampla gostaria que fosse esta exclusão de imagens.

Controlar o armazenamento em cache e snippets

Os resultados de pesquisa geralmente mostram um link de página em cache e um snippet. Veja a seguir, por exemplo, um dos primeiros resultados visualizados quando pesquisamos por "arara":



Snippet – um excerto de texto da página da Web

Link em cache – este link direcciona os utilizadores para uma cópia da página armazenada nos servidores do Google

Porquê ter um snippet? É mais provável que os utilizadores acedam um Web site se os resultados de pesquisa mostrarem um snippet desse Web site. Isto acontece pelo facto de os snippets facilitarem a identificação da relevância do resultado da sua pesquisa por parte dos utilizadores. Se um utilizador não puder ter esta percepção rapidamente,haverá uma maior probabilidade de passar para o resultado de pesquisa seguinte.

Porquê ter um link em cache? O link em cache é útil em vários casos como, por exemplo: quando os Web sites ficam temporariamente indisponíveis, quando os Web sites de notícias ficam sobrecarregados devido a um evento importante; ou quando os Web sites são acidentalmente excluídos. Outra vantagem é que a cópia em cache do Google destaca as palavras utilizadas pelo utilizador na sua pesquisa, permitindo uma rápida avaliação da relevância da página.

A maioria dos editores da Web prefere que o Google exiba o snippet e o link em cache. No entanto, há alguns casos em que um proprietário do Web site pode desejar desabilitar uma ou ambas as opções:

> O conteúdo do meu jornal é alterado várias vezes ao dia. Parece que o Googlebot não está a indexar este conteúdo tão rapidamente quanto nós o actualizamos, e o link em cache está a apontar para uma página que não é a mais recente. Como posso impedir que o Google crie um link em cache?

O proprietário do Web site de notícias poderá impedir que este link em cache seja exibido nos resultados de pesquisa adicionando a tag NOARCHIVE à sua página:

```
<META   NAME="GOOGLEBOT"   CONTENT="NOARCHIVE">
```

Também poderá indicar ao Google que não deverá exibir um snippet para uma página utilizando a tag NOSNIPPET:

```
<META   NAME="GOOGLEBOT"   CONTENT="NOSNIPPET">
```

Observação: Adicionar a tag NOSNIPPET também tem o efeito de impedir que um link em cache seja exibido, pelo que, se especificar NOSNIPPET, obterá automaticamente também a tag NOARCHIVE.

O seu Web site apresenta conteúdo exclusivo e útil?

Assim que o Web site puder ser detectado e indexado, a pergunta final é se o conteúdo das páginas da Web é exclusivo e útil.

Antes de mais, olhe para o seu texto como um todo. O seu título e os links de texto são descritivos? O texto flui naturalmente, de forma clara e intuitiva?

Da mesma forma que o capítulo de um livro é organizado por áreas e temas específicos, cada página da Web deve estar centralizada numa área ou num tópico específico. As palavras-chave e frases surgem naturalmente deste tipo de texto, e é muito mais provável que os utilizadores permaneçam numa página da Web que disponibilize conteúdo e links relevantes.

No entanto, certifique-se de que as frases que utilizar incluem as frases que os visitantes provavelmente irão pesquisar. Por exemplo, se o seu Web site for direccionado para um clube de fãs dos carros da MG, certifique-se de que as palavras "MG" e "carros" aparecem de facto no texto, em vez de apenas em termos como "automóveis britânicos".

Aumentar a visibilidade: práticas recomendadas

Os proprietários de Web sites perguntam-nos com frequência quais as melhores formas de aumentar a visibilidade e o posicionamento dos seus Web sites nos nossos resultados de pesquisa. A nossa resposta é simples: "Pense como um utilizador, pois é assim que tentamos pensar".

O que é que isto significa na prática? Acima de tudo, certifique-se de que fornece aos visitantes o conteúdo que eles estão à procura, pois é a relevância que irá direccionar o tráfego para o seu Web site e que o irá ajudar a mantê-lo.

Muitos proprietários de Web sites concentram-se na forma como as suas páginas da Web irão ser posicionadas. Porém, a posição é determinada por mais de 200 critérios além do PageRank. É muito melhor investir o seu tempo na qualidade do seu conteúdo e na respectiva acessibilidade do que tentar descobrir formas de "influenciar" um algoritmo do mecanismo de pesquisa. Se um Web site não atender às nossas directrizes de qualidade, ele poderá ser bloqueado do índice.

O que fazer:

1. Crie conteúdo relevante e atraente: os visitantes irão aceder às suas páginas através de vários links; portanto, certifique-se de que cada página chama a atenção.
2. Envolva os utilizadores: pode adicionar uma secção de comentários ou um blog no seu Web site? A criação de uma comunidade ajuda a estimular a utilização regular do seu Web site. Envolva os seus visitantes para acelerar a visibilidade e a lealdade dos utilizadores.
3. Acompanhe o seu Web site: utilize o Centro para webmasters (consulte a página 13) para ver quais os termos de pesquisa que estão a direccionar os visitantes para o seu Web site, ou para controlar as alterações no posicionamento nos resultados de pesquisa em relação a alterações efectuadas no Web site.
4. Tenha links de alta qualidade incorporados.
5. Forneça links de texto claros: posicione os links de texto adequadamente no seu Web site e certifique-se de que eles incluem termos que descrevem o tópico.

O que evitar:

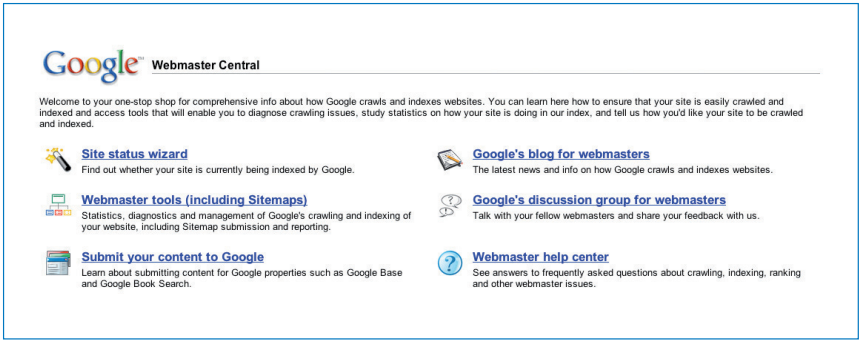
1. Não encha a sua página com listas de palavras-chave.
2. Não tente "camuflar" páginas com texto que possa ser visto pelos mecanismos de pesquisa mas não pelos utilizadores.
3. Não insira páginas "apenas para rastreio" ao definir páginas ou links cujo único objectivo seja "enganar" os mecanismos de pesquisa.
4. Não utilize imagens para exibir nomes, conteúdo ou links importantes – os mecanismos de pesquisa não "lêem" imagens.
5. Não crie várias cópias de uma página em URLs diferentes com a intenção de enganar os mecanismos de pesquisa.

Quando tiver dúvidas, consulte as directrizes para os nossos webmasters, disponíveis em: *www.google.com/webmasters/guidelines.html*

Centro para webmasters

Como uma empresa que pretende fornecer os resultados de pesquisa mais relevantes e úteis na Web, empenhamo-nos em fornecer um apoio generalizado e justo para todos os webmasters e todos os Web sites, independentemente da sua dimensão. É por isso que criámos o Centro para webmasters, localizado em google.pt/webmasters.

O Centro para webmasters é um excelente recurso para todos os editores da Web. Este responde a questões sobre rastreio, indexação e posicionamento; dispõe de um canal para trocar comentários e alertar para problemas; e oferece ferramentas de diagnóstico que apoiam os webmasters na detecção de potenciais problemas de rastreio.



Veja uma exemplo do que pode encontrar no Centro para webmasters.

- Diagnosticar potenciais problemas no acesso a páginas e fornecimento de soluções
- Solicitar a exclusão de páginas específicas do nosso índice
- Garantir que o arquivo robots.txt esteja a permitir e bloquear as páginas que pretende.
- Consultar as estatísticas de termos de pesquisa e de páginas relacionadas com o seu Web site:
- Estatísticas de termos de pesquisa: verifique quais os termos de pesquisa que direccionam mais visitantes para o seu Web site, e quais os tópicos que poderiam ser ampliados no seu Web site para conseguir mais tráfego.
- Análise da página: veja a sua página da Web como o Google a vê. Veja as palavras mais comuns no seu Web site, os links incorporados no Web site e como os outros descrevem o seu Web site quando o visitam.
- Taxa de rastreio : veja com que frequência o seu Web site está a ser rastreado pelo Googlebot e dê indicações ao Google para rastrear com menor ou maior rapidez.

Sitemaps

O Centro para webmasters também disponibiliza o Sitemaps aos editores para resultados de pesquisa na Web, por telemovel e de notícias.

O Sitemaps é um protocolo compatível com outros mecanismos de pesquisa, a fim de ajudar os webmasters a fornecer mais informações sobre as suas páginas. O Sitemaps complementa os mecanismos padrão de rastreio na Web; os webmasters podem utilizá-lo para dar informações ao Google sobre as páginas de seu Web site, a fim de melhorar o rastreio e a visibilidade das suas páginas nos resultados de pesquisa do Google.

Além do Sitemaps para pesquisa na Web, também oferecemos o Google Mobile Sitemaps, permitindo aos editores enviar URLs com conteúdo para dispositivos móveis para o nosso índice para telemoveis.

Além disso, para estes editores cujos Web sites de notícias estejam incluídos no Google Notícias, o Sitemaps do Google Notícias pode ajudar a fornecer estatísticas sobre os artigos do editor, desde as consultas à frequência de visualização. Se utilizado com as ferramentas de diagnóstico do Centro para webmasters, o Sitemaps do Google Notícias também poderá disponibilizar relatórios de erros que ajudam a explicar quaisquer problemas que o Google possa ter ao rastrear ou extrair artigos de notícias do Web site de um editor. Além disso, um editor pode enviar um Sitemap do Google Notícias contendo URLs que gostaria que fossem considerados para inclusão no Google Notícias. Os Sitemaps do Google Notícias, ao contrário dos Sitemaps da Web e móvel, estão actualmente disponíveis apenas em inglês, embora esperemos disponibilizá-los em breve noutros idiomas.

Perguntas frequentes

Por que motivo não é possível ter um apoio individualizado para o meu Web site?
De acordo com algumas estimativas, existem aproximadamente 100 milhões de Web sites na Web. Cada um destes Web sites é importante para nós, pois sem eles, não importa se pequenos ou grandes, o nosso índice seria menos abrangente e, por fim, menos útil para os nossos utilizadores.

O Centro para webmasters é uma excelente fonte de apoio para todos os tipos de Web sites. Publicamos e respondemos às dúvidas dos editores, de forma a que todos possam beneficiar das informações. No Centro para webmasters, também poderá encontrar uma comunidade amigável e útil de webmasters com quem pode partilhar sugestões e obter ajuda para a solução de diversos problemas.

Os anúncios exibidos influenciam o posicionamento dos resultados? As listagens de anúncios são totalmente separadas dos resultados de pesquisa?
Os posicionamentos dos anúncios e dos resultados de pesquisa não estão minimamente relacionados; Na realidade, temos equipas totalmente separadas que trabalham em cada um dos serviços, de forma a que não haja qualquer interferência. Acreditamos que a objectividade dos nossos resultados de pesquisa é crucial para fornecer a melhor experiência para os nossos utilizadores.

Como adiciono um Web site ao índice de pesquisa do Google?
A inclusão nos resultados de pesquisa do Google é gratuita e fácil; além disso, não requer um envio manual do Web site para o Google. O Google é um mecanismo de pesquisa totalmente automatizado; ele rastreia a Web regularmente e detecta Web sites para adicionar ao nosso índice. Assim, a grande maioria dos Web sites relacionados nos nossos resultados de pesquisa não são enviados manualmente para inclusão, mas localizados e adicionados automaticamente quando os nossos robôs rastreiam a Web.

Além disso, as Ferramentas para webmasters do Google (no Centro para webmasters) fornecem um método fácil para que os webmasters enviem um sitemap dos seus URLs ao índice do Google e obtenham relatórios detalhados sobre a visibilidade das suas páginas no Google. Com as Ferramentas para webmasters do Google, os proprietários de Web sites podem manter o Google automaticamente informado sobre todas as páginas actuais e sobre quaisquer actualizações feitas nas mesmas.

Quanto tempo demora, em média, a que o Google detecte um Web site recém-criado, e com que frequência o Google rastreia a Web em geral?
Não existe um período de tempo definido para que o Google detecte um novo Web site.

O Googlebot rastreia regularmente a Web para recriar o nosso índice. Ao utilizar o Centro para webmasters, um webmaster pode ver com que frequência o seu Web site está a ser rastreado pelo Googlebot e informá-lo se deve rastrear de forma mais rápida ou mais lenta.

E se eu quiser que o meu Web site seja exibido nos resultados de pesquisa da Web, mas não em serviços separados, como o Google Notícias ou no Google Pesquisa de Imagens?
O Google permite que os editores da Web escolham participar ou não dos serviços; além disso, um editor pode entrar em contacto com a equipa de apoio de um produto específico e fazer esta solicitação.

Como abordado anteriormente neste folheto, o Protocolo de exclusão de robôs pode ser utilizado para bloquear a indexação das páginas de imagens e da Web. O recurso de "remoção de URL" no Centro para webmasters também pode ser utilizado com este objectivo e abrange a pesquisa na Web e por imagens.

Além disso, como o Googlebot conta com vários robôs diferentes, poderá segmentar o que deseja bloquear:

- Googlebot: rastreia páginas do nosso índice da Web e de nosso índice de notícias
- Googlebot-Mobile: rastreia páginas do nosso índice para telemóvel
- Googlebot-Image: rastreia páginas do nosso índice de imagens
- Mediapartners-Google: rastreia as páginas para definir o conteúdo do AdSense. Utilizamos este robô apenas para rastrear o seu Web site se exibir anúncios do AdSense no seu Web site.
- rastreia páginas para avaliar a qualidade da página de destino do AdWords. Utilizamos este robô apenas se utilizar o Google AdWords para anunciar o seu Web site.

Por exemplo, para bloquear totalmente o Googlebot, poderá usar a seguinte sintaxe:

```
User-agent:   Googlebot
Disallow:   /
```

Posso escolher que texto desejo que seja especificado como um snippet?
Não. Não é uma boa ideia, tanto para o utilizador como para o próprio criador do conteúdo. Escolhemos um snippet de texto a partir do Web site que mostra o termo de pesquisa do utilizador no contexto, o que por sua vez demonstra a relevância do resultado.

Os estudos mostram que é mais provável que os utilizadores acedam a um Web site se os resultados de pesquisa mostrarem o snippet. Isto acontece pelo facto de os snippets facilitarem muito a identificação por parte dos utilizadores da relevância do resultado para a sua pesquisa. Se um utilizador não puder definir isto rapidamente, a probabilidade de passar para o resultado de pesquisa seguinte é maior.

Os editores da Web podem incluir uma meta tag nas suas páginas, a fim de fornecer ao Google informações adicionais nos casos em que não puderem gerar algoritmicamente um snippet útil a partir do conteúdo na página. Para fazer isso, basta adicionar o seguinte à secção <head> da página:

```
<meta   name="description"   content="Por que Ana não
gosta de coelhos? Estamos prestes a descobrir.">
```

Qualquer editor da Web que não quiser que um snippet das suas páginas seja criado, pode utilizar a tag NOSNIPPET, com a seguinte descrição:

```
<meta name="robots" content="nosnippet">
```

Finalmente, às vezes, utilizamos uma descrição do Web site a partir do Open Directory Project do snippet do resultado de pesquisa. Se não desejar que esta descrição seja utilizada, basta adicionar a seguinte meta tag:

```
<meta name="robots" content="noodp">
```

Os artigos com notícias mais recentes no meu Web site são exibidos apenas durante algumas horas antes de serem actualizados e movidos para uma secção de artigos padrão. Quero que o artigo completo apareça no índice do Google, não as últimas notícias.
Uma opção é colocar todos os artigos com as últimas notícias num directório e utilizar o robots.txt para não permitir que o Googlebot aceda a esse diretório.

Outra opção é adicionar a tag NOFOLLOW à secção <HEAD> do html da sua secção de últimas notícias. Isto irá dar indicação ao Googlebot de que não deverá seguir quaisquer links que detecte nessa página. Note, porém, que a tag NOFOLLOW impede apenas que o Googlebot siga os links de uma página para outra. Se outra página da Web estiver associada a esse artigo, o Google irá detectará o artigo quando a indexar.

Se eu tiver vários nomes de domínio e divulgar o mesmo conteúdo nesses diferentes domínios serei excluído dos resultados de pesquisa?
Embora alguns editores possam tentar enganar os mecanismos de pesquisa ao duplicar o conteúdo e divulgar Web sites espelhados, também há um conteúdo legítimo que pode ser duplicado por bons motivos. O Google não quer penalizar estes sites. Por exemplo, não tratamos conteúdo semelhante expresso em diferentes idiomas (digamos, inglês num Web site e francês noutro) como conteúdo duplicado.

Ter o mesmo conteúdo em diversos Web sites (p. ex., distribuição de artigos) não resulta necessariamente na remoção total de um ou mais Web sites dos resultados de pesquisa. No entanto, lembre-se de que é provável que cada ocorrência do artigo apareça mais abaixo no posicionamento, dado que tem apenas uma parte dos links provenientes do que teria uma única cópia. Em geral, uma única cópia de um artigo terá um posicionamento mais elevado e, portanto, poderá ser visto por mais utilizadores do que várias cópias do mesmo conteúdo.

Além disso, para garantir a qualidade da pesquisa, o Google não inclui várias cópias de uma página nos resultados de pesquisa. Em vez disso, é habitual escolher uma versão da página a ser visualizada. No entanto, os webmasters podem indicar ao Google a sua versão preferida utilizando o robots.txt ou uma meta tag para bloquear quaisquer cópias que eles não queiram mostrar nos resultados de pesquisa.

Porque é que o meu Web site está a ser bloqueado do índice do Google?

Antes de mais, é possível que o seu Web site não tenha sido bloqueado. Há vários motivos pelos quais um Web site pode não aparecer nos resultados de pesquisa (consulte as páginas 5 a 11).

Se o seu Web site não tiver obstáculos para a descoberta ou indexação, pode ter sido bloqueado. Os Web sites podem ser bloqueados do índice por não atenderem aos padrões de qualidade descritos nas diretrizes para webmasters (disponíveis no Centro para webmasters). Isto ocorre com maior frequência quando um Web site está a utilizar métodos desonestos para tentar aparecer em posições superiores na pesquisa. As violações comuns das directrizes incluem "camuflagem" (texto escrito de tal forma que possa ser visto pelos mecanismos de pesquisa mas não pelos utilizadores) ou configuração de páginas/links com o único objectivo de enganar os mecanismos de pesquisa e manipular os resultados desses mecanismos.

Quando os webmasters suspeitarem que estes Web sites violam as directrizes de qualidade, poderão modificar os Web sites para que respeitem estas directrizes e, em seguida, clicar no link "solicitar reinclusão" no interface de Ferramentas para webmasters para solicitar a reavaliação do Web site.

Glossário

A indexar
O processo de adicionar o conteúdo do seu Web site a um mecanismo de pesquisa.

Camuflagem
Mostrar aos mecanismos de pesquisa conteúdo diferente do que mostra aos utilizadores.

Conteúdo dinâmico
Conteúdo, como imagens, animações ou vídeos, que conta com o Flash, JavaScript, frames ou URLs criados dinamicamente.

Directório raiz
O principal directório ou directório superior num sistema de arquivos de computador.

Extensão de arquivo
Nome de um arquivo de computador (.doc, .txt, .pdf, etc.) frequentemente utilizado para indicar o tipo de dados armazenados.

HTML (Hypertext Markup Language)
Linguagem de marcação utilizada na Web para estruturar o texto.

Link em cache
Um instantâneo de como uma página foi exibida aquando a última vez que o Google a acedeu. Uma cópia em cache permite que os utilizadores visualizem uma página quando a versão dinâmica não estiver disponível, embora o conteúdo possa ser ligeiramente diferente. Para visualizar uma cópia em cache, clique no link "em cache" que aparece abaixo de um resultado de pesquisa.

Meta tags
Uma tag no HTML que descreve o conteúdo de uma página da Web. As meta tags podem ser utilizadas para controlar a indexação de páginas individuais num Web site.

Page Rank
Um recurso do Google que ajuda a definir o posicionamento de um Web site nos nossos resultados de pesquisa. O PageRank conta com a natureza exclusivamente democrática da Web ao utilizar a sua vasta estrutura de links como um indicador do valor de uma página individual. Ainda mais importante é que os Web sites de alta qualidade recebem um PageRank superior, que é tido em conta sempre que o Google realiza uma pesquisa. O Google combina o PageRank com técnicas sofisticadas de correspondência de texto, a fim de encontrar páginas que sejam importantes e relevantes para as pesquisas.

Palavra-chave
Um termo que é inserido na caixa de pesquisa de um mecanismo de pesquisa. O mecanismo de pesquisa então pesquisa páginas que incluam a palavra-chave ou frase.

Protocolo de exclusão de robôs
Uma especificação técnica que dá indicação aos mecanismos de pesquisa de que Web site ou partes de um Web site não devem ser pesquisáveis, e quais as partes que devem permanecer visíveis nos resultados de pesquisa.

Rastreador
Software utilizado para detectar e indexar URLs na Web ou numa intranet.

Rastreio
Processo utilizado pelos mecanismos de pesquisa para aceder a páginas da Web.

Robots.txt
Um arquivo de texto que permite que um editor da Web controle o acesso ao seu Web site em vários níveis, desde o Web site completo a directórios individuais, páginas de um tipo específico ou até mesmo páginas individuais. Este arquivo mostra aos rastreadores quais os directórios que eles podem ou não rastrear.

Sistema de gestão de conteúdo (CMS, Content Management System)
Sistema de software utilizado para gerir o conteúdo a partir de arquivos do computador, de imagem e som para o conteúdo da Web.

Site espelhado
Uma página da Web duplicada; às vezes, utilizada para enganar os mecanismos de pesquisa e tentar optimizar a indexação e a classificação na Web de um Web site.

URL (Uniform Resource Locator)
O endereço de um Web site na Internet, o qual consiste no protocolo de acesso (http), nome de domínio (www.google.pt) e, em alguns casos, a localização de outro arquivo (www.google.com/webmasters).

Para mais informações sobre o Centro para webmasters, visite

www.google.com/webmasters

Google™

© Copyright 2008. A Google e o Google Maps são marcas registadas da Google Inc. Todos os outros nomes de empresas e produtos podem ser marcas registadas das respectivas empresas às quais estão associados.