

Problemas de contenido duplicado y sitios web múltiples

Greg Grothaus

Google Search Quality Engineer

12 de Agosto de 2009 - SES San Jose



Foto del usuario de flickr popilop.

Información sobre Greg Grothaus

Soy:

- Un geek (antes que nada)
- Antiguo estudiante de bioinformática en Virginia Tech
- Un apasionado empleado de Google
- No muy experto en Powerpoint

En Google:

- Ingeniero de Software durante 4 años
- Enfoque en Calidad de búsqueda
 - encontrar información y posicionarla correctamente
- Educar e informar a la comunidad Webmaster

El mito de la penalización por contenido duplicado

Para mostrar los resultados más pertinentes, omitimos ciertas entradas similares a las que ya hemos encontrado.

Puede [repetir la búsqueda e incluir los resultados omitidos](#).

"Mi sitio no aparece en los resultados a no ser que haga clic en "repetir la búsqueda e incluir los resultados omitidos."

"Parece como si Google filtrara tu sitio web pensando que es de baja calidad."

"Recibo este mensaje cuando uso el comando site: ... Me preocupa que Google haya omitido mis páginas por contenido duplicado o que me pueda penalizar."

El mito de la penalización por contenido duplicado

El contenido duplicado se refiere a contenido **sin intención de manipular** nuestras clasificaciones.

Ejemplos de contenido duplicado:

- Foros de discusión que pueden generar páginas normales y páginas para móviles.
- Páginas archivadas o enlazadas desde varias URL distintas.
- Versiones para imprimir de páginas web.

Google intenta indexar y mostrar páginas con información variada. Este *filtro* supone, por ejemplo, que si tu sitio web tiene una versión normal y otra para imprimir de cada artículo, y ninguna de éstas está bloqueada por robots.txt o una **metaetiqueta noindex**, **mostraremos sólo una de ellas.***

*dependiendo de la búsqueda

<http://www.google.com/support/webmasters/bin/answer.py?>



El mito de la penalización por contenido duplicado

Sin embargo, en algunos casos el contenido duplicado se usa de modo intencionado para **manipular el posicionamiento en los motores de búsqueda** o ganar tráfico.

En casos **excepcionales** en que el contenido duplicado se usa con la **intención de manipular nuestras clasificaciones y engañar a los usuarios**, Google tomará medidas al indexar y posicionar dichos sitios.



<http://www.google.com/support/webmasters/bin/answer.py?answer=66359>

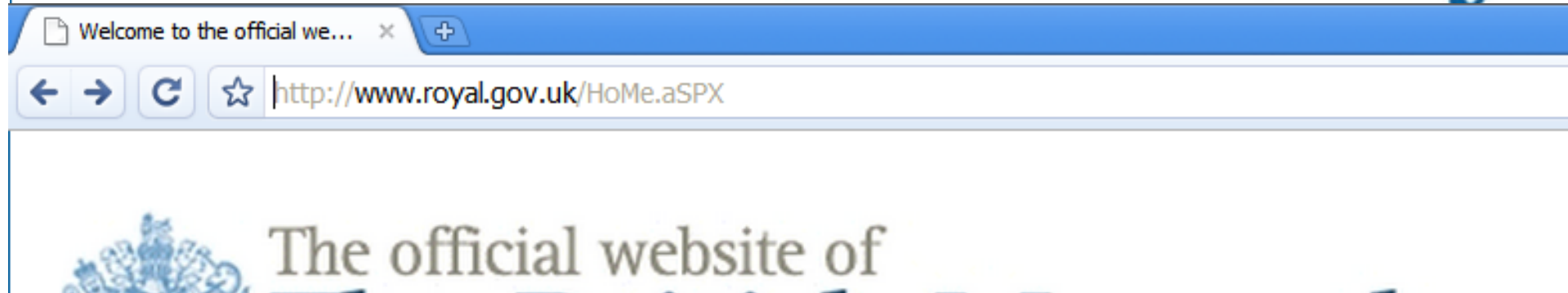


Contenido duplicado

Estas URL son diferentes:

- example.com/
- example.com/?
- example.com/index.html
- example.com/Home.aspx
- www.example.com/
- www.example.com/?
- www.example.com/index.html
- www.example.com/Home.aspx

¿Es un problema común? ¡Sí!



Posibles situaciones: Sin penalización

- **Menor popularidad de enlaces**

Un sitio web que recibe enlaces a distintas URL con el mismo contenido.

- **URL útiles para el usuario en los resultados de búsqueda**

URL con parámetros poco útiles pueden resultar poco productivas y empeorar la usabilidad

- **Rastreo ineficaz**

Más tiempo rastreando el mismo contenido y menos tiempo para descubrir contenido nuevo.

¿Cómo solucionar problemas de contenido duplicado?

URL canónica : reducida a la forma más simple y significativa sin perder generalidad.

- Escoge una URL canónica para cada página y asegúrate que de que los enlaces sean consistentes en tu sitio web.
- Haz que las URL no canónicas redirijan permanentemente (301) a la URL canónica o preferida.
- Herramientas para Webmasters de Google: especifica con www. o sin www.
- Etiqueta HTML rel=canonical

Uso correcto de los códigos de respuesta

- Código 301 para redirecciones permanentes:
 - Indica al motor de búsqueda que transfiera propiedades, como la popularidad de enlaces, a otra URL.
 - Se aplica a situaciones tales como:
 - Mover un sitio a otro domain/host
 - <http://googlewebmaster-es.blogspot.com/2009/07/vas-mover-tu-sitio-practicas.html>
 - Modificar la estructura de una URL



prácticas recomendadas al migrar tu sitio

Buscar

Contenido duplicado - nivel avanzado



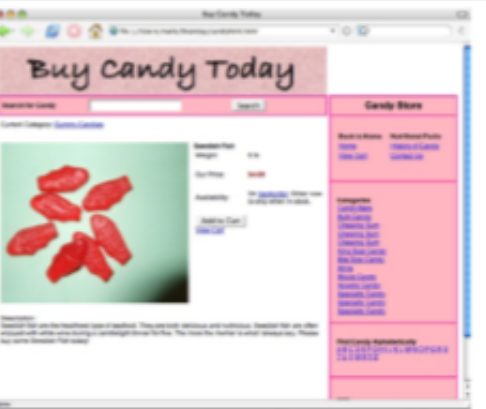
This screenshot shows a web browser window with the address bar containing 'http://en.wikipedia.org/wiki/Canonical'. The page features the Wikipedia logo on the left and navigation tabs for 'article', 'discussion', 'edit this page', and 'history'. The main content area displays the title 'Canonical' and a sub-header 'From Wikipedia, the free encyclopedia'. Below this, there is a note: 'For the free software company, see [Canonical Ltd.](#)'. The main text states: '**Canonical** is an adjective derived from [canon](#). *Canon* comes from the Greek word *kanon*, "rule"'. A Wiktionary widget is visible on the right side of the page.

This screenshot shows a second web browser window with the address bar containing 'http://en.wikipedia.org/w/index.php?title=Canonical&printable=yes'. The page displays the title 'Canonical' in a large font, followed by the sub-header 'From Wikipedia, the free encyclopedia'. The main text reads: '**Canonical** is an adjective derived from *canon*. *Canon* comes from the Greek word *kanon*, "rule" (perhaps originally from *kanna* "reed", cognate to *cane*), and is used in various meanings. **Basic, canonic, canonical:** reduced to the simplest and most significant form possible without loss of generality, e.g., "a basic story line"; "a canonical syllable pattern."'. A navigation menu is visible on the left side of the page.

Contenido duplicado - nivel avanzado

http://stuff.com/breadcrumbs/tents/bags/red/tent_bag.html

http://stuff.com/breadcrumbs/bags/tents/red/tent_bag.html

URL que desea mostrar el webmaster	Duplicado	Duplicado
		
<p>http://www.example.com/product.php?item=swedish-fish</p>	<p>http://www.example.com/product.php?category=gummy-candy&item=swedish-fish&affiliateid=ABCD</p>	<p>http://www.example.com/product.php?item=swedish-fish&trackingid=1234&sort=price&sessionid=5678</p>

Nueva opción para contenido duplicado

Elemento de enlace canónico a nivel de página.

En <http://www.example.com/page.html?sid=asdf314159265>

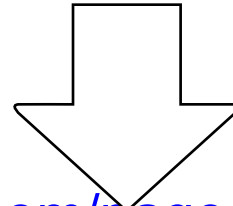
<head>

...

<link rel="canonical" href="http://example.com/page.html"/>

...

</head>



<http://googlewebmaster-es.blogspot.com/2009/02/especifica-tu-url-canonica.html>



Preguntas frecuentes sobre rel=canonical

P: ¿Funciona para **subdominios** diferentes?

R: **Sí**. zeta.zappos.com podría sugerir www.zappos.com como URL canónica.

P: ¿Funciona para **dominios** diferentes?

R: **No**, sólo para el mismo dominio.

P: ¿Lo puedo usar para sugerir que http://example.com/ sea la URL canónica, en lugar de http**s**://example.com/?

R: **Sí**, claro.

P: ¿Debería usar rel=canonical o una redirección 301 permanente?

R: Depende, ahora tienes más opciones a tu disponibilidad.

P: ¿Las páginas tienen que ser exactamente idénticas?

R: No, pero deberían ser parecidas. No pasa nada si hay pequeñas diferencias.

Dominios múltiples

Los dominios múltiples **no** suponen un problema. Por ejemplo, contenido para un público distinto (país, idioma, etc).

Posibles áreas problemáticas:

- Reputación distribuida en varios dominios.
- La "mejor" página para una búsqueda no es necesariamente todos tus dominios.
- Perder las ventajas de una IU tabulada:

Google

Google. [Advanced Search](#) · [Preferences](#) · [Language Tools](#) · [Advertising Programs](#) - [Business Solutions](#) - [About Google](#). ©2009 - [Privacy](#). [+](#) [Show stock quote for GOOG](#)

[www.google.com/](#) - [Cached](#) - [Similar](#) -   

[Add your URL to Google](#) - 4 visits - Jul 30

For example: <http://www.google.com/> . You may also add comments or keywords that describe the content of your page. These are used only for our information ...

[www.google.com/addurl/](#) - [Cached](#) - [Similar](#) -   

[+](#) [Show more results from www.google.com](#)

Preguntas y respuestas

¡Gracias!

- Enviad vuestras preguntas al Foro para Webmasters de Google
 - <http://www.google.com/support/forum/p/webmasters?hl=es>