

Arquitectura de la Información

URL, Arquitectura, robots.txt, Sitemaps XML...

IMPORTANTE

- ▶ Las direcciones URL

NO DEBEN CAMBIAR

- ▶ Así que vigila muy bien lo que decides escoger, porque cualquier cambio implicará una pérdida de visitas desde buscadores si no se hace correctamente.

Máquinas que entiendan como personas

- ▶ Los humanos somos capaces de crear e interpretar elementos y grupos.
- ▶ Las máquinas son capaces de «clusterizar» contenidos de un mismo tipo (color, tamaño, datos, URL...)
- ▶ Los sitios web son escalables, mediante carpetas, ficheros, subdominios, dominios...
- ▶ ¿Cuál es la mejor forma de organizar contenidos para un sitio web?

Contenidos por país vs. idioma

- ▶ Lo primero que hay que saber es quién es nuestro público: gente de un país o gente de un idioma
- ▶ *Venta de coches vs. Diario de información*
- ▶ Si vamos a vender en un país, lo mejor es usar el ccTLD de ese país (.es, .fr...)
- ▶ Si vamos a vender a un idioma, lo mejor es usar los subdominios de ese idioma (es., fr....)

Subdominios vs. Carpetas vs. Ficheros

- ▶ Cada buscador trata la arquitectura de las URL de una forma distinta, y le da peso según le parece más o menos importante.
- ▶ Es importante que un sitio tenga forma de árbol, con más o menos hojas y ramas. No podemos dejar elementos huérfanos.
- ▶ No tiene la misma importancia un contenido en la carpeta «raíz» del sitio que en una carpeta: ¡pero no podemos poner todo en la raíz!

Estructura básica de un sitio

- ▶ Hay 3 formas de hacer consultas en los buscadores, por tanto hay 3 tipos de contenidos (o cómo ha de organizarse la información).
- ▶ Hay que tener en cuenta que no podemos aparecer por todas las palabras en la página principal, por lo que hay que incidir y dar peso a las ramas, no al tronco.
- ▶ Marca: búsquedas del tipo «example.com» o «example» que hagan referencia a una marca/dominio. Estos resultados han de ir a la página principal del sitio.

Estructura básica de un sitio

- ▶ Cosas generales: en principio este tipo de consulta debería ir a una categoría o elemento navegacional. Es algo «tan genérico» que no tiene un contenido concreto.
- ▶ Cosas concretas: debería ir directamente a un contenido final del sitio web.
- ▶ **Ejemplo:**
 - ▶ [ford]
<http://www.ford.com/>
 - ▶ [ford fiesta]
<http://www.ford.com/cars/fiesta/>
 - ▶ [ford fiesta sel sedan]
<http://www.ford.com/cars/fiesta/trim/?trim=selsedan>

El árbol

ford.es

coches

comercial

concesionario

equipamiento

fusion

focus

fiesta

fiesta-van

transit-connect

tourneo

barcelona

madrid

badalona

3-puertas

gasolina

automatico

control-de-crucero

gasolina

3-puertas

La idea

- ▶ El objetivo es tener una arquitectura que agrupe contenidos de mismo tipo.
- ▶ Coches = carpeta con las fichas
- ▶ Equipamientos = carpeta de navegación
- ▶ Separamos y agrupamos los contenidos. Contenidos finales en una carpeta, árbol de navegación en otra.

¿Con barra o sin barra?

- ▶ Comparemos:

<http://example.com/noticias>

<http://example.com/noticias/>

- ▶ Comparemos:

<http://example.com/info>

<http://example.com/info/>

<http://example.com/info/cualquier-cosa>

<http://example.com/info/cualquier-cosa/>

Evitar URL duplicada

- ▶ Utilicemos la etiqueta *rel='canonical'* en toda la web, siempre y cuando no haya un *robots-noindex*.
- ▶ Esta URL evitará ataques por URL, duplicidad de URL parecidas, limitará problemas del *si-www* y *no-www*.
- ▶ Nos aseguramos de que la URL sea única, ya que se calcula al 100% para evitar que haya multiplicidad de resultados.
- ▶ Es algo que debería ir de serie en cualquier sitio web. Ayuda a los buscadores a tener una estructura del sitio correcta.

Contenidos de alta calidad

- ▶ Utilizando el *robots-noindex* podremos decidir si una página / contenido concreto se ha de indexar o no.
- ▶ Google tiene ansia por conocer, así que le dejamos que sepa de todo nuestro sitio, pero que muestre sólo lo que queramos en los resultados.
- ▶ Esto no evita que tengamos una arquitectura de URL correcta.
- ▶ Páginas como el «aviso legal», «términos y condiciones», «contacto» y similares no tiene sentido que se enlacen en el 100% de la página y aparezcan en los buscadores.

Usando el robots.txt

- ▶ Hay que indicar el Sitemap XML para que todos los robots que lo conozcan lo utilicen (sin necesidad de panel de control).
- ▶ Por norma general es mejor no hacer uso de sus filtros, al menos para contenidos «textuales».
- ▶ Podemos filtrar carpetas de contenidos multimedia, como PDF, imágenes u otros elementos que no queramos indexar.

Sitemap: <http://www.example.com/sitemap.xml>

User-Agent: *

Sitemap XML

- ▶ Es un sistema similar a los RSS, pero específicos para los motores de búsqueda.
- ▶ Sólo hay que enviar las direcciones URL nuevas.
 - ▶ En un sitio nuevo podemos lanzar los primeros 3 meses el 100% de los contenidos
 - ▶ En un sitio que ya funciona, sólo mandemos un porcentaje de URL nuevas o actualizadas.
 - ▶ Si se publican 100 contenidos/día, mandemos los últimos 3.000 contenidos (aprox. Un mes)
- ▶ No hace falta mandar toda la navegación del sitio (categorías, por ejemplo) sino sólo los contenidos finales.

Paneles donde controlar qué ocurre

▶ **Bing:**

<http://www.bing.com/webmasters/>

▶ **Google:**

<http://www.google.com/webmasters/tools/>

▶ **Yahoo!:**

<http://siteexplorer.search.yahoo.com/>

▶ **Yandex:**

<http://webmaster.yandex.ru/>



Gracias

Javier Casares

javier.casares@kisslab.com - <http://javiercasares.com>