

Cómo sacar partido a los Informes de Cobertura de Google Search Console | SEO Mallorca

Todo SEO que se precie debería estar actualmente deseoso de disponer de cada vez más y más proyectos migrados al nuevo Google Search Console. ¿No es tu caso? Pues debería serlo... En SEO siempre hemos ido muy faltos de datos claros sobre como nos trata el buscador y estos nuevos añadidos si bien tampoco es que nos resuelvan la vida completamente vienen a suponer una gran cantidad de insights accionables con los que sacar un claro provecho a nuestros proyectos.

El nuevo Search Console, por un motivo que creo que nadie entiende, se está habilitando muy poco a poco en las cuentas. No es niquiera que un usuario tenga acceso y otros no, es que proyecto a proyecto y bloque datos a bloque de datos poco a poco se nos va a ir sumando información a este informe. Actualmente si manejas un buen conjunto de proyectos lo normal es que al menos 2 o 3 ya te ofrezcan sus informes de rendimiento y cobertura. Esto se encargan de notificártelo por email pero si quieres probar a ver si se te ha escapado alguno puedes entrar en el nuevo search console con la siguiente URL: <https://search.google.com/search-console/index>

Dentro de esta nueva interfaz podremos ir al famoso **informe de Cobertura**: el que para muchos es la gran revolución de información de rastreo e indexación. En él se nos dan suficientes datos para diagnosticar y luego accionar problemas y optimizaciones de indexación en un proyecto. Básicamente se nos está hablando de cómo esta tratando Google las distintas URLs que va encontrando de nuestro dominio al rastrear. Eso, si trabajas en sites medianos o grandes es sencillamente

brutal. Vamos a ver como sacarle partido...

Cómo es el informe de cobertura

No me extenderé mucho con este punto. Simplemente es un informe donde podemos ver catalogadas 4 grandes bloques:

Esto ya supone una gran mejora de lo que teníamos antes: **se nos habla de páginas excluidas**, páginas que no pasan los filtros de Google y con las que podremos trabajar al detalle.

Pero hay además 4 grandes puntos o perspectivas desde las que ahora si podemos investigar nuestro estado de indexación y que marcan un antes y un después en nuestro día a día como SEOs.

Control de tendencias los últimos 3 meses y comparado con impresiones

Un gran añadido es el poder ver la evolución de de estos 4 grandes datos en los últimos 3 meses (aun que lo normal es que aun no disfrutes de tanto tiempo pues ha empezado ahora a recopilar la información). Esto en la práctica supone que podemos ver cuando se producen los problemas o las mejoras de optimización pues podemos ver los cambios de tendencia de cada tipo de valor y como un tipo de páginas se transforma en otro con los cambios de la web.

Es decir, que si unimos está información a nuestro histórico de acciones que hemos realizado en los sites que gestionamos o con las subidas de nuevas versiones al servidor tendremos una nueva puerta abierta al aprendizaje y a creación de hipótesis.

El hecho de que además estas gráficas se complementen con las impresiones de resultados en el buscador nos permite descubrir aun de forma más rápida todas las relaciones de causa efecto entre indexación y visibilidad.

Os dejo solo una gráfica sobre esto donde vemos claramente

como el hecho de ir aumentando la cantidad de URLs excluidas del índice (páginas y mas páginas que google pierde el tiempo en analizar para que finalmente no formen parte del índice) nos afectó negativamente en impresiones mientras duraba su rastreo. Tambien vemos como cuando este rastreo inutil terminó recuperamos impresiones.

El control de Enviadas y de Sitemaps

De primeras puede pasar un poco desapercibido pero este informe queda muy muy conectado a tu subida de sitemaps haciendo aún más útiles todas las prácticas que solemos realizar a día de hoy de trocear los sitemaps para sacar información por grupos.

En definitiva, podemos seleccionar ver todo el rastreo por libre de Google (“Todas las páginas conocidas”), ver solo lo que le enviamos por sitemaps y que idealmente es lo que queremos posicionar (“Todas las páginas enviadas”) o escoger cada uno de nuestros sitemaps.xml por separado y ver como es el estado de indexación de cada conjunto por separado.

Si unimos esto a la información de tendencias de la que hablábamos en el punto anterior veremos para cada sitemap como ha ido evolucionando su indexación. Vamos que si nuestros sistemaps contienen cada uno tipologías de páginas distintas este filtrado acaba siendo realmente práctico.

Aquí la pena es que la gráfica de impresiones no se filtre por solo las impresiones de las URLs contenidas en el sitemap que estamos viendo pero el resto de valores (Error, Advertencias, Válidas y Excluidas) si lo hacen así que si que supone una gran ayuda para el seguimiento del proyecto.

Los motivos de cada estado

Y aquí es donde todos casi lloramos de alegría cuando empezamos a verlo en algún proyecto: no solo nos dicen el estado de las URLs sino el motivo por el que Google las mete en un estado u otro. Al principio tanta información es abrumadora pero en pocos minutos empezamos a entender lo que supone disponer de toda esta información.

Google Search console tan solo nos muestra en su listado los Motivos que nos afectan, es decir, podemos ver una lista más o menos larga dependiendo de todas las casuísticas que se sucedan en nuestro proyecto. Por suerte nos ha documentado un poco todos los motivos de los que disponen y podemos tirar de su documentación para saber todo lo que no vemos (que también es importante saber todo lo que no nos sucede): <https://support.google.com/webmasters/answer/7440203?hl=es>

Al final esta lista de motivos es mucho más importante de lo que parece pues prácticamente nos indica si lo que vemos es un problema o no y nos ayuda a saber que podemos hacer para solucionar aquellas cosas que no nos acaban de gustar.

Tratamiento de motivos de Error

- Error del servidor (5xx)
- Error de redirección
- Robots.txt ha bloqueado la URL enviada
- La URL enviada contiene la etiqueta "noindex"
- La URL enviada devuelve un soft 404
- La URL enviada devuelve una solicitud no autorizada (401)
- No se ha podido encontrar la URL enviada (404)
- La URL enviada tiene un problema de rastreo

Muchos de los motivos aquí citados son puramente técnicos y deben remediarse hablando con IT o con sistemas. Sin embargo hay problemas que si nos atañen directamente como redirecciones que son demasiado largas o terminan formando un

loops, marcados de noindex, 404 que podríamos remediar con 301, links que le hemos ofrecido a páginas que piden password, etc. Por ultimo tenemos el “problema de rastreo” que viene a decirnos que ni Google sabe exactamente porque ha fallado esa URL... en fin. Nadie es perfecto.

Tratamiento de los motivos de advertencia

Aqui de momento solo se marca un tipo de problema: “Se ha indexado aunque un archivo robots.txt la ha bloqueado”

No hace falta explicar mucho, podemos validar que realmente se bloquea lo que esperamos bloquear. Sobre este punto tenemos que recordar que Google no solo indexa mediante el rastreo de tu site por lo que aunque le bloqueemos con robots.txt una URL el puede encontrarla con señales externas y si quiere indexarla de todas formas.

Este resultado nos puede alarmar un poco al principio: por un lado GSC al darse de alta parece que nos marca muchas URLs de este tipo pero ellas solas acaban desapareciendo de las gráficas sin que cambiemos nada (parece que le lleva tiempo saber que realmente no las ha indexado). De las que finalmente nos quedan muchas terminarán desapareciendo también solas y acabaremos prácticamente con URLs que reciben enlaces externos. Ahi si que tenemos cosas que arreglar:

- Podemos Abrirlas de nuevo (quitando el disallow) del robots.txt y aprovecharlas dandoles contenido y enlaces.
- o bien redirigirlas (tambien abriendolas) hacia un punto más interesante con un 301 (o canonical si no queda más remedio).

Forzar desindexarlas porque si (porque nos molestan) a mi al menos no me parece una buena idea. Si no perdemos crawl budget (porque no se rastrean) y no nos están haciendo daño (canibalizando otras páginas) ¿para que vamos a molestarnos en ellas?

Tratamiento de las validas indexadas

- Enviada e indexada
- Indexada, no enviada en sitemap
- Indexada; recomendamos marcarla como canónica

Se nos indican las páginas que realmente va a ofrecer el buscador pero además se nos diferencian por dos criterios:

Por un lado el tema de los sitemaps vuelve a estar aquí presente, podemos comprobar que URLs google ha decidido meter en el buscador y que nosotros no le habíamos dado en sitemaps. Una validación de este grupo nos puede avisar de que cosas se nos habían pasado por alto o que páginas que realmente no queríamos ofrecer están ahí luchando por las posiciones.

El segundo criterio es la canonización, google nos indica que páginas han tenido problemas de duplicidad pero que aun así ha escogido para indexar. La nota que nos hace google es que deberíamos marcarlas como canónicas pero nosotros deberíamos leer “escogidas por google” y ver si nos gusta su decisión o no antes de obedecerle. Para mí resulta más práctico plantearse las y quizás marcar a otras que google no había escogido como canónicas más que obedecerle porque sí.

Tratamiento de las excluidas

La tipología de páginas excluidas es tan grande que creo que es mejor clasificarlas un poco antes de presentarlas...

1. Excluidas, por sacarse del índice por directrices SEO

- Página con redirección
- Excluida por una etiqueta “noindex”
- Bloqueada por la herramienta de borrado de URLs
- Bloqueada por robots.txt
- Bloqueada por una solicitud no autorizada (401)
- No se ha encontrado (404)
- Soft 404

Estos motivos indican un cambio en el estado de la URL. Anteriormente la URL si perteneció al índice pero por indicaciones de nuestra página se ha excluido del mismo. Una buena oportunidad para validar nuestras acciones o posibles problemas inesperados del proyecto.

2. Excluidas, por duplicidades y etiquetas canónicas

- Página alternativa con etiqueta canónica adecuada
- Google eligió una página canónica diferente a la del usuario
- Página duplicada sin etiqueta canónica
- Página duplicada por una página noHTML

Por fin información clara sobre como nos trata Google con las duplicidades. Si unimos todo esto a las validas a las que nos recomienda poner etiqueta con url canónica tendríamos la foto completa de todo el tratamiento de duplicidadesw internas del contenido.

Es interesante ver esta clasificación donde existen matches claros entre contenido:

- “Pagina duplicada sin etiqueta canonica” y “pagina duplicada por una página noHTML” pierden su contenido en favor de las válidas donde se nos recomienda incluir etiqueta canónica.
- Las “paginas alternativas con etiqueta canónica adecuada” lo pierden por una canonica a la que google ha hecho caso.
- Y las mejores, las paginas a las que pusimos etiqueta canónica pero que Google ha pensado que no tenían sentido y ha preferido llevar a resultados otra URL, una que tenía en otro sitio y que responde (según Google) mejor a los intereses de los usuarios. Vamos, ¡problema gordo! No le da la gana hacernos caso y nos lo dice a la cara.

En cualquier caso y aunque nada nos sorprenda y la foto sea la que esperabamos según el etiquetado que nosotros mismos hemos definido, todo esto nunca nos debe llevar a pensar que todo esta bien. Estas paginas son siempre rastreos innecesarios a evitar. Siempre va a ser mejor que Google no rastree urls

inútiles así que estos avisos nunca son del todo positivos.

3. Excluidas, a pesar de que las hemos enviado intencionadamente

- La URL enviada no se ha seleccionado como canónica
- La URL enviada se ha caído del índice

Páginas que hemos enviado intencionadamente (sitepaps, solicitudes de indexación) pero que Google no va a contemplar. De entre los motivos solo nos detalle si no la quiere por contenido duplicado, el otro motivo básicamente es "cualquier otro motivo".

4. Información sobre el proceso de indexación

- Rastreada: actualmente sin indexar
- En cola de rastreo
- Descubierta: actualmente sin indexar

3 estados con los que hay que tener mucho cuidado en su interpretación. La primera vez que los leemos creemos que tienen que ver con que google aun no ha tenido tiempo de interpretar los contenidos, pero en realidad nos hablan de bloqueos que se producen en el proceso de indexación porque las URLs no cumplen ciertos criterios.

Para mi el más importante de ellos es "Rastreada: actualmente sin indexar", este estado nos habla de URLs que se han analizado pero que google ha decidido no meter en el índice: No ha asociado la URL a keywords comerciales porque no cumplen los criterios mínimos de calidad. Aquí encontramos un poco de todo pero sobretodo: Thin Content, Urls de listados, URLs que han perdido autoridad (ya tienen mucha distancia de rastreo o pocos o ningún link externo), formatos de archivo poco claros, etc. Vamos es el cajón para la baja calidad.

En cola de rastreo parece que no tiene demasiados problemas: Nos habla de URLs que aun no ha rastreado pero nos dice que terminará haciéndolo.

Por ultimo descubierta: actualmente sin rastrear, nos indica que Google conoce la URL pero que por el motivo que sea (normalmente técnico) no ha querido rastrearla e indexarla. Poco podemos saber sobre esto y gran parte se soluciona solo. Lo que si hemos notado es que ante caídas del servidor y errores 500 este tipo de URLs se disparan y cuesta bastante más pasarlas a válidas que las que ya estaban en cola de rastreo.

¿Y eso es todo? ¿No nos faltan cosas?

El tema es que si que nos faltan cosas. ¿Y el thin content? ¿Forma parte de las URLs válidas o cae dentro de las urls que se caen del indice? Parece que forme parte de las “Rastreadas pero no indexadas” pero en realidad no lo tenemos todo claro... lo iremos viendo seguro en los proximos meses a medida que todos hagamos las pruebas pertinentes.

Sea como sea todo esto es información valiosísima con la que hay que trabajar desde ya mismo.

Ver URLs exactas de cada estado y motivo o filtrar por ellas

Y la guinda para el final. Todo lo que os he contado hasta ahora podemos verlo URL a URL. No veremos todas nuestras URLs, Google Search Console ya nos deja claro que solo nos dan ejemplos (solo una muestra del total y solo hasta 1000 resultados de cada tipo) pero para descubrir patrones en cada tipología de URL estos “ejemplos” son más que suficientes. En sites pequeños podremos hacer correcciones URL a URL hasta que Google las trate como deseamos que lo haga, en sites más grandes esta información nos ayudará a descubrir patrones con los que trabajar.

Y este es solo uno de los nuevos informes...

Quizas el más llamativo, quizás el más deseado, pero solo uno. Esta claro que este nuevo Search Console solo está empezando y yo creo que incluso en este informe iremos viendo como los motivos cambian con el tiempo (por desgracia eso no significa que siempre a mejor). Nos toca aprender a lidiar con todo esto y sacarle el máximo partido posible a todo.

Nueva información significa nuevos aprendizajes y ya solo el ver como google lo está estructurando todo te da muchas ideas. Más de uno al verlo descubrirá que Google puede pasar de los canonicals y elegir por su cuenta y se le abrirá un mundo por delante. No será porque no nos lo hubiesen dicho antes pero no todo el mundo era consciente de esto.

Bueno, esperemos que no quede mucho para que esta nueva versión esté activa en todos nuestros proyectos.

Comparte este artículo:

[Twittear](#)

This content was originally published [here](#).