

SEO WEB

Un breve recorrido sobre la terminología, conceptos y herramientas.



Oscar Casado Lorenzo
Victor Clemente Nieto
Raúl Beltrán Marco

14/01/2022

1. Introducción

¿Qué es SEO? Básicamente, es el proceso de mejorar un sitio web para que los buscadores puedan comprenderlo mejor y así poder **mejorar su posicionamiento en buscadores**, es decir, que si un usuario busca palabras claves relacionadas con tu sitio web este aparezca de los primeros resultados gracias a este proceso.

El objetivo principal de este proceso es “controlar” cómo **rastrea** o **indexa** (clasificar las palabras y **crear un índice**[index] con ellas) los buscadores, como podría ser el de Google, los sitios web.

Lo básico para aplicar esta técnica es tener claro cómo afectan al rastreo y a la indexación las **páginas canónicas** de tu sitio web; si tienes varias URL para tu sitio web **los buscadores elegirán una URL como “canónica”** y considerará el **resto como duplicados** y las rastreará con menos frecuencia; esta URL canónica se puede indicar explícitamente, en caso contrario será el buscador quien elija una, la clave es que todas tus URL sigan una estructura lo más parecido a la URL canónica y así poder mejorar los resultado de búsqueda.

Y **retirar o gestionar el contenido duplicado** de tu sitio web cuando sea necesario, como por ejemplo unificar URL similares o duplicadas; algunos buscadores como **Google** permite a los sitios web establecer **reglas de robots.txt** para impedir que se rastreen páginas y sitemaps, por ejemplo, deberías bloquear el rastreo de contenido duplicado o de recursos que no son importantes de tu sitio web, como iconos, logotipos u otros gráficos pequeños y de uso frecuente.

A lo largo de este documento profundizaremos mucho más en esta técnica mostrando distintas **directrices, más formas de controlar la indexación** de los buscadores, incluso como **optimizar** la experiencia de la página y mucho más.

2. Directrices

En este apartado se proporcionan **pautas sobre cómo debe ser el código** de tu sitio web para que pueda aparecer en los buscadores. Las directrices se dividen en las siguientes categorías:

2.1. Directrices para webmasters

- Es recomendable comprobar que se pueda acceder a todas las páginas de tu sitio mediante un enlace desde otra página que también se pueda encontrar, así como que cada enlace tenga un atributo **"alt"** que describa correctamente la página destino, los enlaces que se pueden rastrear son etiquetas **<a>** que tienen un atributo **href**.
- Facilitar un **archivo de sitemap** con enlaces que dirijan a las páginas importantes de tu sitio junto a una página con una lista de enlaces legibles (**<a>** con href) en ocasiones llamada **"página de índice del sitio"** o **"página de mapa del sitio"**.
- Una cantidad razonable de enlaces en las páginas (unos pocos miles como mucho)
- Asegúrate de que tu servidor web admita la cabecera HTTP **If-Modified-Since**, con esta función puedes indicar a los buscadores mediante el servidor web si tu contenido ha **cambiado** desde la última vez que se rastreó tu sitio.
- Impide con el **archivo robots.txt** (comentado anteriormente) de tu servidor web que se rastreen **espacios infinitos** como podrían ser páginas de resultado de búsqueda, es importante ir actualizando este archivo.

2.2. Directrices generales


La mayoría de los buscadores recomiendan el uso de las siguientes prácticas:

- Simplificar la estructura de direcciones URL: Siempre que sea posible, utiliza palabras en vez de números de ID largos

 **Recomendado:** URLs descriptivas y sencillas

`http://en.wikipedia.org/wiki/Aviation` 

 **No recomendado:** URLs complejas y no descriptivas

`http://www.example.com/index.php?id_sezione=360&sid=3a5ebc944f41daa6f849f730f1` 

- Etiquetar un sitio web como contenido dirigido a niños
- Compatibilidad del navegador: Cada navegador interpreta el código del sitio web de manera ligeramente distinta, por lo que es posible que se muestre de forma diferente en función del navegador que se use. Por lo general, no dependas de funciones específicas de los navegadores
- Evitar que se cree contenido duplicado
- Permitir el rastreo de enlaces: Usando la etiqueta **<a>** tal y como se ha explicado en el anterior apartado.

2.3. Directrices de contenido

Las directrices siguientes son consejos adicionales sobre determinados tipos de contenido, como imágenes, vídeos, AMP, AJAX y sitios optimizados para móviles:

- **Prácticas recomendadas para vídeos**
 - Comprobar que **todos** los vídeos están disponibles en una web pública en las que los usuarios puedan verlos. Por ejemplo, en el caso de Google asegurarse que la página no esté bloqueada por el robots.txt
 - Crear una **página específica** para cada uno en la que el vídeo sea lo más destacado.
 - Incluir una **etiqueta HTML adecuada** a su alrededor como <video>, <embed>, <iframe> u <object>.
- **Prácticas recomendadas sobre archivos de rich media**
 - **Rich media** es un término que se usa en el mundo de la publicidad digital para describir los **anuncios con funciones avanzadas** como vídeo, sonido u otros elementos que atraen a los espectadores y consiguen que estos interactúen con el contenido.
 - Utilizar rich media **SOLO** cuando sea necesario.
 - **Facilitar versiones de texto** de las páginas, si en tu página principal tienes una pantalla de inicio que no es HTML, debes incluir en la primera un enlace HTML normal que dirija a otra página con texto
- **Directrices de AMP**(Accelerated Mobile Pages)
 - Todas las directrices que hemos explicado que se ajustan a un sitio web se pueden aplicar a páginas AMP, por supuesto, siguiendo la especificación AMP HTML.
- **Prácticas recomendadas para sitios con contenido educativo:**
 - Pon el contenido de la página a disposición de tanto profesores como alumnos **sin que tengan que iniciar sesión**. Si hay que iniciar sesión los robots de los buscadores **no podrán acceder a ella**.
 - Si has **añadido páginas** a tu sitio recientemente, has hecho cambios en recursos educativos, puedes solicitar que se **vuelva a rastrear y a indexar tus URL**.
 - Google ofrece **Google Trends** para ver qué términos buscan los usuarios, tenlos en cuenta a la hora de elegir palabras clave.

2.4. Directrices de calidad

Las directrices que se exponen a continuación describen **técnicas que están prohibidas de utilizar** y que, de usarse, pueden hacer que tu página o sitio **no se muestren** en los resultados de búsqueda.

- **Contenido generado automáticamente:** Creados mediante programas informáticos, usados para manipular el posicionamiento en los resultados de búsqueda, por ejemplo: texto que no tiene sentido leerlo pero que contiene palabras clave, texto generado mediante procesos automatizados (cadenas de Markov), texto generado usando sinónimos automáticos,
- **Redirecciones engañosas:** es el acto de enviar a un visitante a una URL distinta de la que ha solicitado inicialmente.
- **Enlaces de pago:** El posicionamiento de un sitio en los resultados de búsqueda se basa, en parte, en el análisis de los sitios que incluyen enlaces a él. Sin embargo, algunos propietarios de sitios web y especialistas en SEO **compran o venden enlaces** para manipular el posicionamiento de las páginas.
- **Encubrimiento:** Hace referencia a la práctica de mostrar URL o contenidos distintos a usuarios y a buscadores, por ejemplo, una página te ofrece un producto o información y luego el usuario recibe un resultado distinto al esperado.
- **Texto y enlaces ocultos:** Se puede ocultar contenido de texto (como un número excesivo de palabras clave) de distintas maneras. Por ejemplo:
 - Usar texto blanco con fondo blanco.
 - Incluir texto detrás de una imagen.
 - Seleccionar el valor 0 como tamaño de fuente. Y más.
- **Páginas puerta:** Son sitios o páginas que se crean para que aparezcan como respuesta a consultas de búsqueda similares y específicas.
- **Spam generado por usuarios:** también es posible que visitantes o usuarios generen spam de manera malintencionada en un sitio de calidad. Este spam se genera normalmente en sitios que permiten a los usuarios crear páginas nuevas o añadir contenido al sitio de alguna otra forma.

A continuación, se explican distintos **consejos** con los que se puede evitar utilizar estas técnicas, ya que algunas pueden llegar a escapar de nuestro control (como la última).

- Formas de evitar los comentarios spam:
 - **Moderando los comentarios,** los comentarios no se muestran en el sitio hasta que se revisan y se aprueban.
 - **Solicitar a los usuarios que validen una dirección de correo electrónico real** al registrarse puede evitar que muchos robots de spam creen cuentas automáticamente.
 - **La solución más eficaz** a este problema: no permitas que los usuarios comenten o creen páginas en tu sitio web (o piénsalo dos veces).
- **Denunciar spam, enlaces de pago y software malicioso:** Como usuarios, si encontramos este tipo de spam o software malicioso, está en nuestra responsabilidad denunciar este tipo de acciones para mejorar la calidad de todos los usuarios.

3. Controlar el rastreo y la indexación

Es importante controlar la forma en la que Google busca y analiza tu contenido para mostrarlo en la Búsqueda o en otras de sus propiedades, además de cómo evitar que Google rastree determinado contenido de tu sitio.

3.1. Sitemaps

Tener un sitemap es importante para tu pagina web ya que los sitemaps son archivos en los que se proporciona información sobre las páginas, los vídeos y otros archivos de tu sitio, así como sobre las relaciones que hay entre ellos.

Un sitemap es necesario si :

Tu sitio es muy grande, Tu sitio tiene muchas páginas de contenido que están aisladas o no están bien enlazadas entre sí, Tu sitio es nuevo y hay pocos enlaces externos que dirigen a él o Tu sitio incluye una gran cantidad de contenido rich media, como vídeo o imágenes, o aparece en Google Noticias

Un sitemap no es necesario si:

Tu sitio es relativamente pequeño, Tu sitio está totalmente enlazado de forma interna, No tienes muchos archivos multimedia (vídeo, imagen) o páginas de noticias que quieras que aparezcan en los resultados de búsqueda

3.2. Archivo robots.txt

Un archivo robots.txt indica a los rastreadores de los buscadores a qué URLs de tu sitio pueden acceder. Principalmente, se utiliza para evitar que las solicitudes que recibe tu sitio lo sobrecarguen; no es un mecanismo para impedir que una página web aparezca en Google.

Los archivos robots.txt sirven principalmente para gestionar el tráfico de los rastreadores a tu sitio, aunque también *suelen* usarse para que Google no rastree determinados archivos, según el tipo de archivo. Por ejemplo:

- **Paginas web:** Puedes usar un archivo robots.txt en páginas web (HTML, PDF y otros formatos no multimedia que Google pueda leer) para gestionar el tráfico de los rastreadores si crees que tu servidor se sobrecargará con solicitudes del rastreador de Google, o para evitar que se rastreen páginas sin importancia o similares de tu sitio
- **Archivo Multimedia:** Con un archivo robots.txt, puedes gestionar el tráfico de los rastreadores y evitar que aparezcan archivos de imagen, vídeo y audio en los resultados de la Búsqueda de Google. Aun así, no impedirá que otras páginas o usuarios incluyan enlaces a tu archivo de imagen, vídeo o audio.

- Archivo de recursos: Con un archivo robots.txt, puedes bloquear archivos de recursos (como los de imagen, secuencias de comandos o estilo que no sean importantes) si crees que prescindir de ellos no perjudicará considerablemente a las páginas. No obstante, si crees que complicaría el análisis del rastreador de Google, no debes bloquearlos, ya que si lo haces, Google no podrá analizar correctamente páginas que dependan de estos recursos

Los archivos robots.txt también tienen limitaciones como:

Es posible que algunos buscadores no sigan determinadas directivas del archivo robots.txt

Cada rastreador interpreta la sintaxis de una forma

Es posible que las páginas bloqueadas mediante un archivo robots.txt se indexen si otros sitios incluyen enlaces a ellas.

3.3. Metaetiquetas

Google admite tanto metaetiquetas a nivel de página como directivas insertadas para controlar cómo aparecerán las páginas de tu sitio en la Búsqueda de Google.

Las metaetiquetas a nivel de página son una forma perfecta con la que los propietarios de sitios web pueden facilitar a los buscadores información sobre sus sitios. Además, con ellas se puede ofrecer información a todo tipo de clientes, y cada sistema procesa solo aquellas que entiende e ignora el resto

Cosas Curiosas:

- Google puede leer metaetiquetas HTML y parecidas a XHTML, independientemente del código que se utilice en la página.
- Salvo en el caso de google-site-verification, en las metaetiquetas no se suele distinguir entre mayúsculas o minúsculas.
- Las metaetiquetas que se incluyen en esta lista no son las únicas disponibles y puedes utilizar otras si son importantes para tu sitio. No obstante, si Google no conoce una metatiqueta, la ignorará.

3.4. Retiradas de contenido

Es posible que quieras ocultar algún contenido a Google por diversas razones:

Para mantener la privacidad de los datos. Es posible que tengas información privada alojada en tu sitio a la que no quieres que accedan otros usuarios

Para ocultar contenido de menos valor a tu audiencia. Es probable que tu sitio web muestre el mismo contenido en distintas páginas, lo que podría perjudicar su posicionamiento en la Búsqueda de Google. Un buen ejemplo de dónde puede surgir contenido duplicado suele ser la función de búsqueda en todo el sitio que ayuda a los usuarios a navegar por el contenido

Para que Google se centre en tu contenido importante. Si tienes un sitio muy grande (con muchas miles de URLs) y páginas con contenido menos importante o si tienes mucho contenido duplicado, puede que te interese que Google no rastree las páginas duplicadas o poco importantes y se centre en el contenido principal.

Estas son las principales formas de bloquear contenido

- Retirar contenido: Valido para todo tipo de contenido
- Proteger tus archivos con una contraseña: Valido para todo tipo de contenido
- Archivos robots.txt o solicitudes de retirada de imágenes urgente: Para imágenes
- Directiva noindex: para paginas web

3.5. URLs duplicadas

Si tienes una página a la que se puede acceder mediante varias URLs, o bien páginas diferentes con contenido similar (por ejemplo, una página para móviles y otra para ordenadores), Google las considerará versiones duplicadas de la misma página. En este caso, elegirá una URL como *canónica*, que es la que rastreará, y considerará que las otras URLs son *duplicados*, por lo que las rastreará con menos frecuencia.

Si no indicas explícitamente qué URL es la canónica, Google la seleccionará por ti, aunque también es posible que acabe considerando que todas tienen la misma importancia, lo que podría generar un comportamiento no deseado

Cuando el robot de Google indexa sitios, intenta determinar cuál es el contenido principal de cada página. Si detecta que en un mismo sitio hay varias páginas con contenido muy similar, marca como canónica la página que considera más completa y útil

Para elegir las páginas canónicas, Google tiene en cuenta diferentes factores:

si la página se sirve mediante HTTP o HTTPS.

la calidad que tiene.

si la URL está en un sitemap y si la página incluye la etiqueta *rel=canonical*.

Razones válidas para mantener páginas similares o duplicadas:

- Ofrecer versiones para varios tipos de dispositivos:
- Habilitar URLs dinámicas para usar parámetros de búsqueda o IDs de sesión:

- Si el sistema de blogs que utilizas guarda automáticamente varias URLs al colocar la misma entrada en diferentes secciones:
- Si tu servidor está configurado para publicar el mismo contenido en las variantes con www, sin www, HTTP o HTTPS de una página:
- Si incluyes contenido en un blog para distribuirlo a otros sitios y este contenido se replica parcial o totalmente en ellos:

Motivos por los que elegir una URL canónica

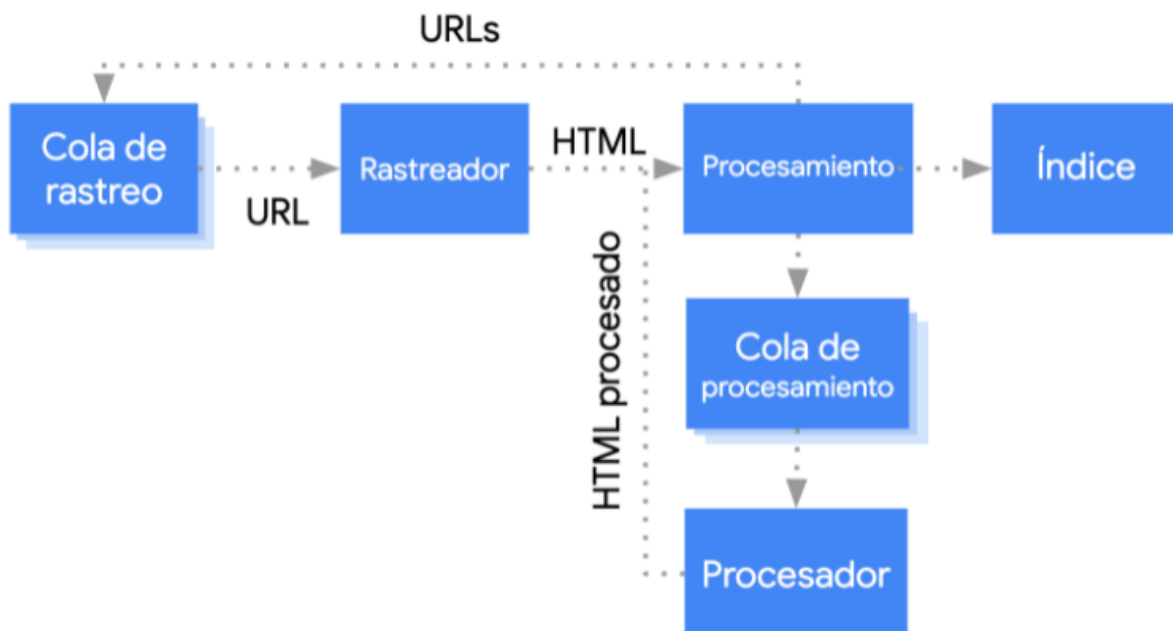
- Indicar la URL que quieres que se muestre en los resultados de búsqueda.
- Unificar las señales de enlaces de páginas similares o duplicadas
- Simplificar las métricas de seguimiento de un mismo producto o tema
- Gestionar el contenido distribuido

Existen varias directrices generales a la hora de tomar en cuenta la URL canónica

3.6. Contenido JavaScript

JavaScript es una parte importante de la plataforma web porque aporta muchas funciones que convierten a la Web en una potente plataforma de aplicaciones. Si permites que la Búsqueda de Google pueda descubrir las aplicaciones web que hayas creado con JavaScript, te resultará más fácil llegar a nuevos usuarios y volver a captar el interés de los que ya tengas cuando busquen el contenido que proporciona tu aplicación web

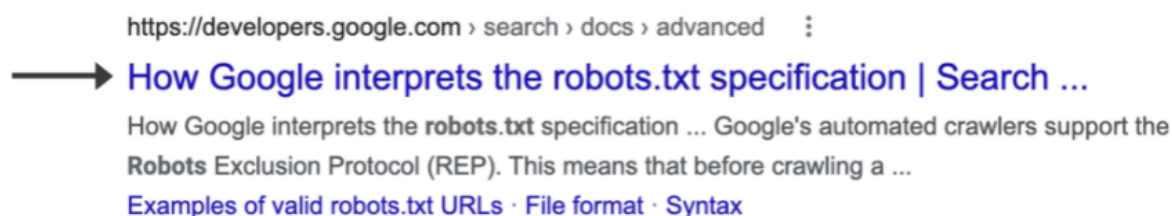
El robot de google Procesa las aplicaciones web JavaScript de la siguiente manera



4. Cambiar la aparición en la Búsqueda

4.1. Controlar enlaces de título

Un enlace de título es el texto de un resultado de búsqueda que enlaza con una página web.

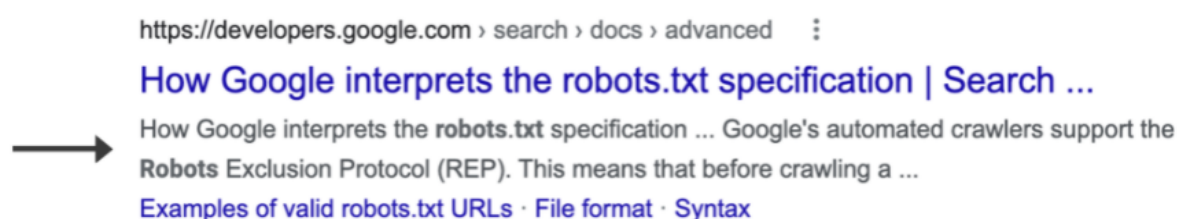


Se pueden indicar las preferencias siguiendo una serie de directrices para escribir elementos <title> descriptivos. De estas directrices las más importantes son: incluir la marca de la web en los títulos, evitar texto repetido o genérico, evitar exceso de palabras clave.

Además, si un <title> contiene una fecha en una página para obtener información periódica la Búsqueda de Google puede detectar esta incoherencia y corregirla.

4.2. Controlar “fragmentos”

Un “fragmento” es la descripción que se muestra en un resultado de búsqueda. Google utiliza diferentes fuentes para determinar este de forma adecuada automáticamente.



Los propietarios de los sitios web pueden sugerir contenido para se incluya en los fragmentos de dos formas:

- **Resultados enriquecidos:** si añades datos estructurados a tu sitio para marcar, por ejemplo, reseñas recetas, empresas o eventos, Google podrá saber de qué tratan tus páginas más fácilmente.
- **Etiquetas de metadescripción:** en algunas ocasiones, Google emplea el contenido de las etiquetas <meta name="description"> para generar fragmentos.

Además, puedes evitar que se creen y se muestren fragmentos de tu sitio. También puedes informar sobre la longitud máxima que quieres que tengan tus fragmentos.

Aplica la metaetiqueta nosnippet para la primera opción o la metaetiqueta max-snippet:[number] para la segunda.


4.3. Datos estructurados

La búsqueda de Google hace todo lo posible por entender el contenido de las páginas. Podemos agilizar este proceso mediante datos estructurados, que son un formato estandarizado con el que proporcionar información sobre páginas y clasificar su contenido.


```
<html>
<head>
  <title>Apple Pie by Grandma</title>
  <script type="application/ld+json">
  {
    "@context": "https://schema.org/",
    "@type": "Recipe",
    "name": "Apple Pie by Grandma",
    "author": "Elaine Smith",
    "image": "http://images.edge-generalmills.com/56459281-6fe6-4d9d-984f-385c9488d824.jpg",
    "description": "A classic apple pie.",
    "aggregateRating": {
      "@type": "AggregateRating",
      "ratingValue": "4.8",
      "reviewCount": "7462",
      "bestRating": "5",
      "worstRating": "1"
    },
    "prepTime": "PT30M",
    "totalTime": "PT1H30M",
    "recipeYield": "8",
    "nutrition": {
      "@type": "NutritionInformation",
      "calories": "512 calories"
    },
    "recipeIngredient": [
      "1 box refrigerated pie crusts, softened as directed on box",
      "6 cups thinly sliced, peeled apples (6 medium)"
    ]
  }
</script>
</head>
<body>
</body>
</html>
```

Recipes

[VIEW ALL](#)



Apple Pie by Grandma
Example.com
4.8 ★★★★★ 7,462 reviews
1 hr 30 min · 512 calories



Grandma's Apple Pie
Example.com
4.8 ★★★★★ 2,055 reviews
1 hr 30 min · 512 calories

5. Optimizar la experiencia en la página

Denominamos “experiencia en la página” a cómo perciben los usuarios su interacción con una página web. Para medir dicha experiencia se utilizan un conjunto de indicadores que evalúan aspectos de las páginas.



5.1. Signed exchanges

SXG permite que la Búsqueda de Google precargue todo tu contenido y, al mismo tiempo, se respete la privacidad del usuario. Cuando alguien hace clic en el resultado, la página web empieza a renderizarse mucho antes porque los recursos clave ya están disponibles (CSS, imágenes, JavaScript, etc).

5.2. Dispositivos móviles

Los sitios que no están optimizados para móviles son difíciles de ver y de utilizar desde uno.



En EE.UU., el 77% de las búsquedas móviles de los usuarios se hacen desde su casa o desde su trabajo, donde normalmente hay ordenadores. Por tanto, este tráfico es fundamental para cualquier negocio electrónico.

El trabajo que hay que dedicar a optimizar un sitio para móviles depende de los desarrolladores, del modelo de negocio y de la experiencia.

El principal consejo para ahorrar esfuerzo, es comenzar desde un diseño adaptable,

siendo el mismo código el que se encargue de realizar los ajustes oportunos (normalmente ajustándose al tamaño de la pantalla).

Por otro lado, tenemos a los usuarios que hacen búsquedas desde clientes móviles lentos, Google les muestra páginas más rápidas y que ocupan menos espacio. Para hacer se transcodifica el código a una versión optimizada. Esta tecnología se llama Web Light.

Las páginas Web Light conservan la mayor parte del contenido relevante e incluyen un enlace para que los usuarios puedan ver la página original, llegando a consumir un 80% menos de bytes.

5.3. Seguridad

HTTPS (HyperText Transfer Protocol Secure) es un protocolo de comunicación de Internet que protege la integridad y la confidencialidad de los datos de los usuarios entre sus ordenadores y el sitio web.

Para poder habilitar el protocolo HTTPS en tu sitio, debes obtener un certificado de seguridad. El certificado lo emite una autoridad de certificación, que toma medidas necesarias para verificar que tu dirección web pertenezca realmente a tu organización. Al configurar el certificado, asegúrate de obtener un nivel de seguridad alto escogiendo una clave de 2048 bits.