Technical SEO for Beginners. PDF Download

Ultimate Guide to Technical SEO for Beginners – PDF Download

Technical SEO

As the name suggests, it is the technical aspect of SEO.
Technical SEO refers to website and server optimizations that help search engine spiders crawl and index your site more effectively (to help improve organic rankings).

Google helps you to monitor and optimise these technical aspects with Google Search Console to which you must have access. You add your Property/Properties (website/s) to Google Search Console to better handle the Technical SEO.

Why Technical SEO should be carried out?

Search wants to present the users (every search intent) with the best possible results for their query/searches. There in the case of Google, Google bots crawl and evaluate web pages on a wide range of parameters. Some of these parameters are base on how fast a page loads, performance, core web vitals, user experience, mobile usability..etc., Other parameters help crawlers understand better and fast what your content is all about, this is what Structured Data does. So if you take care of these aspects then search engines will reward you with good rankings.

It also works the other way around: if you make serious technical mistakes on your site, they can cost you.

But it's a misconception you should focus on the technical details of a website just to please search engines. A website should work well – be fast, clear, and easy to use in the first place. Fortunately, creating a strong technical foundation often coincides with a better experience for both users and search engines.

8 Technical aspects taken care-off by Ultimate Technical SEO.

The characteristics of a well-optimised Website with excellent Technical SEO are as follows.

1. **Website (pages) speed to load**
2. **Broken links are next to nil (no dead links)**
3. **Secure with HTTPS**

4. **Crawling is optimised**
   a. **Meta Robots Tag**
   b. **Robots.txt**
5. **Structured data is used**
6. **XML Sitemap is used**
7. **Should not confuse Crawlers with Duplicate Content**
8. **Hreflang (for International Websites)**

What is Robots.txt?

Use robots.txt rules to prevent crawling, and sitemaps to encourage crawling. Block crawling of duplicate content on your site, or unimportant resources (such as small, frequently used graphics such as icons or logos) that might overload your server with requests. Don't use robots.txt as a mechanism to prevent indexing; use the noindex tag or login requirements for that.

Structured data and Technical SEO

Google Search works hard to understand the content of a page. You can help google crawlers by providing explicit clues about the content of a page by including structured data on the page.

Structured data is a standardized format for providing information about a page and classifying the page content; for example, on a recipe page, what are the ingredients, the cooking time and temperature, the calories, and so on. Similarly, about a course page with details of the content of the course, fees structure, timing, timetable and so on.

Google uses structured data that it finds on the web to understand the content of the page, as well as to gather information about the web and the world in general. For example, here is a JSON-LD structured data snippet that might appear on a recipe page, describing the title of the recipe, the author of the recipe, and other details:

```
<html>
  <head>
    <title>Hyderabadi Chicken Biryani</title>
    <script type="application/ld+json">
    {
      "@context": "https://schema.org/",
      "@type": "Recipe",
      "name": "Hyderabadi Chicken Biryani",
      "author": {
        "@type": "Person",
        "name": "Kiran Kumar"
      },
      "datePublished": "2021-04-28",
      "description": "Best Hyderabadi Chicken Biryani in Town.",
      "prepTime": "PT50M"
    }
    </script>
  </head>
```

```
<body>
  <h2>Hyderabadi Chicken Biryani recipe</h2>
  <p>
    <em>by Kiran Kumar, 2021-04-28</em>
  </p>
  <p>
    Best Hyderabadi Chicken Biryani in Town.
  </p>
  <p>
    Preparation time: 50 minutes
  </p>
 </body>
</html>
```

Google Search also uses structured data to enable special search result features and enhancements. For example, a recipe page with valid structured data is eligible to appear in a graphical search result.

Structured data is coded using in-page markup on the page that the information applies to. The structured data on the page describes the content of that page. Don't create blank or empty pages just to hold structured data, and don't add structured data about information that is not visible to the user, even if the information is accurate.

For more technical and quality guidelines, see the Structured data general guidelines.

Structured data format

Most Search structured data uses **schema.org** vocabulary, but you should rely on the Google Search Central documentation as definitive for Google Search behaviour, rather than the schema.org documentation. There are more attributes and objects on schema.org that aren't required by Google Search; they may be useful for other services, tools, and platforms.

Be sure to test your structured data using the **Rich Results** Test during development, and the **Rich result status reports** after deployment, to monitor the health of your pages, which might break after deployment due to templating or serving issues.

1. JSON-LD* (Recommended): A JavaScript notation embedded in a <script> tag in the page head or body. The markup is not interleaved with the user-visible text, which makes nested data items easier to express, such as the Country of a PostalAddress of a MusicVenue of an Event. Also, Google can read JSON-LD data when it is dynamically injected into the page's contents, such as by JavaScript code or embedded widgets in your content management system.

2. Microdata: An open-community HTML specification used to nest structured data within HTML content. Like RDFa, it uses HTML tag attributes to name the properties you want to expose as structured data. It is typically used in the page body but can be used in the head.

3. RDFa: An HTML5 extension that supports linked data by introducing HTML tag attributes that correspond to the user-visible content that you want to describe for search engines. RDFa is commonly used in both the head and body sections of the HTML page.

Structured data guidelines

Be sure to follow the general structured data guidelines, as well as any guidelines specific to your structured data type; otherwise, your structured data might be ineligible for rich results display in Google Search.

Please read and understand all guidelines, it helps in a great Technical SEO.

Google Search Console (GSC) and Monitor, Optimise Technical SEO

Google Search Console is a free tool from Google that can help anyone with a website to understand how they are performing on Google Search, and what they can do to improve their appearance on the search to bring more relevant traffic to their websites.

Search Console provides information on how Google crawls, indexes, and serves websites. This can help website owners to monitor and optimize Search performance.

There is no need to log in to the tool every day. If new issues are found by Google on your site, you'll receive an email from Search Console alerting you. But you might want to check your account around once every month, or when you make changes to the site's content, to make sure the data is stable. Learn more about managing your site with Search Console.

To get started, follow these steps:

1. **Verify site ownership**. Get access to all of the information Search Console makes available.
2. **Make sure Google can find and read your pages**. The Index Coverage report gives you an overview of all the pages Google indexed or tried to index in your website. Review the list available and try to fix page errors and warnings.
3. **Review mobile usability errors Google found on your site**. The Mobile usability report shows issues that might affect your users experience while browsing your site on a mobile device.
4. **Consider submitting a sitemap to Search Console**. Pages from your site can be discovered by Google without this step. However, submitting a sitemap via Search Console might speed up your site's discovery. If you decide to submit it through the tool, you'll be able to monitor information related to it.
5. **Monitor your site's performance**. The Search performance report shows how much traffic you're getting from Google Search, including breakdowns by queries, pages, and countries. For each of those breakdowns, you can see trends for impressions, clicks, and other metrics.

How to setup Google Search Console for Technical SEO

Now, let us learn how to set up GSC for Technical SEO in steps. Note: It's a free tool from Google.

Why use Search Console?

Google Search Console (or GSC) is one of the most — if not most — powerful SEO tools out there.

But why is it important? At its core, Search Console helps you, monitor, maintain and optimize your website's organic search presence. Most primarily use GSC to view clicks and impressions. While that's cool, it has much more to offer. For example, it can:

- Find search queries that drive traffic
- Find how well all your pages rank
- Identify and leverage backlinks to boost link juice
- Add sitemaps
- Locate errors that need fixing
- Ensure the eligibility for rich snippets and schema
- Make your site more mobile-friendly
- Monitor your Core Web Vitals
- Show if your site has been hacked

Step 1: Sign in to Search Console With Your Google account

You'll need to have a Google account for this method to work when setting up Search Console. Don't worry, that's free, too. If you already have Google Analytics, Adwords or Gmail, you can use the same login.

Step 2: Enter Your Website's Domain (or URL-Prefix) to Add a Property

After you sign in, you have the option to add a property type via your domain or a URL prefix.

Here's why starting with some key definitions:

- **Property** – a catch-all term for a single website, URL, mobile app, or device with a uniroot directory (aka public_html) of your siteque tracking ID
- **Domain** – the name of your website (without *http(s)://* and *www.*). Our domain is *digitalmarketacademy.com*
- **Subdomain** – an extension added to a domain, like ***www**.digitalmarketacademy.com* or ***blog**.digitalmarketacademy.com*
- **URL** – an address for a web page. (Domain is the name of a site; URL leads to a page within that site)
- **URL Prefix** – the protocol that appears before your domain. For example, *http://* or *https://*

Setting up Search Console via the "Domain" option sets up your account as a domain-level property.

This means you're creating a single property that includes all subdomains and protocol prefixes associated with your domain. In other words, this option connects Google Search Console to every aspect of your site.

So here's the next step for how to set up Google Search Console with a domain-level property. Enter your site's root domain in the entry field and hit "Continue."

Selecting "URL-prefix" sets up a URL-prefix property**.**

This means you're creating a single property for only one URL prefix for your site. As such, Search Console will only be connected to one version of your site – not the whole thing with all protocols/subdomains – and so it may not provide accurate data. But sometimes you have no choice but to use a URL prefix.

To set up a URL prefix property, enter a URL with a prefix in the field, and hit "Continue."

NOTE**:** To ensure Search Console provides accurate data with URL-prefix properties, create a GSC property for each of the following four URLs:

1. *https://yourdomain.com*
2. *http://yourdomain.com*
3. *https://www.yourdomain.com*
4. *http://www.yourdomain.com*

If you use other subdomains, like *blog.yourdomain.com* or *shop.yourdomain.com*, then you'll want to create a property for each of those too. All told, you will have to repeat the entire Google Search Console set-up process for each of these URLs.

Step 3: Verify Your Website

To implement Google Search Console and start gathering data, you need to verify that you own your site. The verification process varies depending on which option you chose in the previous step.

Jump to the instructions that apply to you:

- Verification for a Domain Property
- Verification for a URL-prefix Property

Verification for a Domain Property

There is only one way to verify a domain-level property, and that's through your DNS provider (or domain name system provider). Here's the screen you'll start with.

First, see if you can find your DNS provider (the company you pay to use your domain) in the dropdown:

This will display detailed instructions specific to your provider. If you want to know how to implement Google Search Console for your digital marketing strategy, using this method, you might want to work with either your developer or DNS provider. If you don't see your provider, you can leave it as "*Any DNS service provider.*"

Next, hit the "Copy" button to copy the TXT record provided to you by Google.

Once you've copied the TXT record, open your domain registrar's site in a new tab (for example GoDaddy, BlueHost, Hostgator, etc…) and log into your account with them.

Navigate to the list of domains you own and select the one you wish to configure. Find the option to manage your DNS records. This will be located in different places, depending on your provider's site. Look for any mention of "DNS" and click it.

For example, on GoDaddy, you would go to "*My Account > My Products*" and select "DNS" next to your domain.

You'll then be brought to a Domain Management screen where you'll find a list of your DNS Records. Select "Add" to create a new one.

Select "Type" choose TXT. Under "Host" type in the @ symbol. Leave "TTL" at 1 hour. And, most importantly, paste the TXT record you got from Google into the field for "TXT Value." Then hit "Save."

This will add a new TXT record for Google Search Console. (In case you're wondering, a TXT record is used to provide info about your domain to an outside source — e.g. show Google you own a domain.)

The process we just outlined above for GoDaddy is very similar for all domain providers. You can even use the same entries for "Type," "Hostname," and "TTL." Some providers will ask for "TXT Record" instead of "TXT Value."

With your TXT record added, return to the Google Search Console set up and select "Verify."

If everything went according to plan, you should see a message like this:

Keep in mind, updating DNS records can take up to 72 hours. If your ownership isn't verified immediately, come back in a few hours or the next day, and check again.

Verification for a URL-prefix Property

If you don't have access to your registrar or would rather not mess with your DNS records, you can set up Google Search Console using a URL prefix property. This provides several alternative options for verification.

Google recommends verifying via an HTML file. (But remember, this is just for URL prefix properties. Ultimately, they recommend you create a Domain property, if possible.)

Notice there's an option to verify via domain name provider. We showed how to do that above. But if you're considering using that method, you might as well create a domain-level property.

Here's how to verify via the other methods.

HTML File

For this method, you need to upload an HTML file to the root folder of your website. It's easy to do, but the downside is you will need to have access to your server, either via FTP or a cPanel File Manager. If you're not familiar with either don't attempt to verify via this method.

If you are comfortable working with your site's server, here's how you verify using an HTML file:

1. Download the file provided by Google. (By clicking the download box shown in the image above.)
2. Access the root directory (aka public_html) of your site.
3. Upload the file. (Like the image example of how the file will appear shown below)
4. Return to Search Console, and select "Verify"

HTML Tag

To verify using the HTML tag you need to add a meta tag to your site's <head> section. To do so you'll need to have developer access to your site's CMS. We'll use WordPress as an example.

Using WordPress, there are two ways you can do this:

1. Adding the meta tag directly to your header.php file
2. Using a plugin to add to the header

More than likely you'll go with option two. Because if you're comfortable working with the header.php, you're better off verifying via HTML file.

Here are the steps to add the GSC HTML tag to WordPress using a plugin:

1. Copy the tag.
2. Log into your site's WordPress admin in a new tab.
3. Install the Insert Headers and Footers plugin on your site.
4. Go to *Settings>Insert Headers and Footers*.
5. Past the Search Console meta tag in the "Scripts in Header**"** field.
6. Return to Search Console, and select "Verify"

Implement GSC on WordPress

We showed above how to verify Search Console on WordPress using an HTML tag and header plugin. But you can also verify using your WordPress SEO plugin. The two most popular SEO plugins are Yoast SEO and All-in-One SEO.

Here's how to set up Google Search Console with each:

1. Create a GSC account and add a URL prefix property
2. Copy the HTML tag provided.
3. For Yoast SEO: In your WordPress Dashboard go to *SEO>General*. Select the "Webmaster Tools" tab. Paste your HTML tag into the field for Google verification code. Hit "Save changes." (As shown in the first image below.)
4. For All-in-One SEO: In your WordPress Dashboard go to *All in One SEO>General Settings.* Paste your HTML tag into the field for Google Webmaster Tools (what Search Console used to be called). Then hit "Update." (As shown in the second image below.)

   - Note: Yoast and All-in-One SEO will automatically strip the code so only the tag ID shows after saving.

5. Go back to Google Search Console and select "Verify."

Sitemaps and how it helps in Technical SEO

A *sitemap* is a file where you provide information about the pages, videos, and other files on your site, and the relationships between them. Search engines like Google read this file to crawl your site more efficiently. A sitemap tells Google which pages and files you think are important in your site and also provides valuable information about these files. For example, when the page was last updated and any alternate language versions of the page.

You can use a sitemap to provide information about specific types of content on your pages, including video, image, and news content. For example:

- A sitemap *video entry* can specify the video running time, category, and age-appropriateness rating.

- A sitemap *image entry* can include the image subject matter, type, and license.

- A sitemap *news entry* can include the article title and publication date.

For Better understanding, https://www.youtube.com/watch?v=JlamLfyFjTA

Do I need a sitemap?

If your site's pages are properly linked, Google can usually discover most of your site. Proper linking means that all pages that you deem important can be reached through some form of navigation, be that your site's menu or links that you placed on pages.

Even so, a sitemap can improve the crawling of larger or more complex sites or more specialized files.

You might need a sitemap if:

1. **Your site is really large.** As a result, it's more likely Google web crawlers might overlook crawling some of your new or recently updated pages.

2. **Your site has a large archive of content pages that are isolated or not well linked to each other.** If your site pages don't naturally reference each other, you can list them in a sitemap to ensure that Google doesn't overlook some of your pages.

3. **Your site is new and has few external links to it.** Googlebot and other web crawlers crawl the web by following links from one page to another. As a result, Google might not discover your pages if no other sites link to them.

4. **Your site has a lot of rich media content (video, images) or is shown in Google News.** If provided, Google can take additional information from sitemaps into account for search, where appropriate.

You might *not* need a sitemap if:

1. **Your site is "small".** By small, we mean about 500 pages or fewer on your site. (Only pages that you think need to be in search results count toward this total.)

2. **Your site is comprehensively linked internally.** This means that Google can find all the important pages on your site by following links starting from the homepage.

3. **You don't have many media files** (video, image) **or news pages** that you want to show in search results. Sitemaps can help Google find and understand video and image files, or news articles, on your site. If you don't need these results to appear in image, video, or news results, you might not need a sitemap.

Security and Manual Actions Report

What is a Manual Action?
According to Google, manual action is taken against a site when Google has determined that pages on the site are not compliant with Google's **webmaster quality guidelines.** Most manual actions attempt to manipulate our search index. Most issues reported here will result in pages or sites being ranked lower or omitted from search results without any visual indication to the user.
If you violate Google's guidelines by following black hat practices such as Cloaking, Doorway Pages, Copied Content, etc, Google will penalize you by either demoting your web page or in some cases such as having obscene content, completely removing you from the SERP.

If Google has taken manual action against your site, you will be notified about it under the Manual Actions report. To fix the manual actions taken against your site, follow these steps:

1. Understand which pages of your site are affected. Know the type and status of these actions.

2. Google will suggest some steps to fix these issues. Follow these steps carefully and rectify the changes on the affected page.

3. Make sure that Google can reach your pages. After you fix all the issues with the page, ask Google to review it using the **Request Review** option which will be shown to you along with the report.

4. In this report, describe all changes have you made to the pages and show the affected and rectified page for Google to understand that the page does not violate any guidelines. Finally, ask Google to reconsider.

5. Google can take up to a week to reconsider your changes and put your pages back on its indexed source.

What are Security Issues?

The Security Issues report lists indications that your site was hacked, or behaviour on your site that could potentially harm a visitor or their computer: for example, phishing attacks or installing malware or unwanted software on the user's computer. These pages can appear with a warning label in search results, or a browser can display a warning page when a user tries to visit them.

What is a Canonical Url?

If you have a single page accessible by multiple URLs, or different pages with similar content (for example, a page with both a mobile and a desktop version), Google sees these as duplicate versions of the same page. Google will choose one URL as the *canonical* version and crawl that, and all other URLs will be considered *duplicate* URLs and crawled less often.
Note that all the websites that are created now have a **rel=canonical** tag by default for all the web pages.
If you don't define your canonical URL then Googlebot will choose one for you and crawl that URL more often.

Links Report on GSC

In this report, you can see who links to your website the most, both internally and externally, and you can also check your top-linked pages.
You can find the following information in this report.

HTTP Status Code Error (Important ones)

200 OK

The request has succeeded. The information returned with the response is dependent on the method used in the request, for example:

- GET an entity corresponding to the requested resource is sent in the response;
- HEAD the entity-header fields corresponding to the requested resource are sent in the response without any message-body;
- POST an entity describing or containing the result of the action;
- TRACE an entity containing the request message as received by the end server.

300 Multiple Choices

- The requested resource corresponds to any one of a set of representations, each with its specific location, and agent-driven negotiation information (section 12) is being provided so that the user (or user agent) can select a preferred representation and redirect its request to that location.
- Unless it was a HEAD request, the response SHOULD include an entity containing a list of resource characteristics and location(s) from which the user or user agent can choose the one most appropriate. The entity format is specified by the media type given in the Content-Type header field. Depending upon the format and the capabilities of the user agent, selection of the most appropriate choice MAY be performed automatically. However, this specification does not define any standard for such automatic selection.
- If the server has a preferred choice of representation, it SHOULD include the specific URI for that representation in the Location field; user agents MAY use the Location field value for automatic redirection. This response is cacheable unless indicated otherwise.

301 Moved Permanently

- The requested resource has been assigned a new permanent URI and any future references to this resource SHOULD use one of the returned URIs. Clients with link editing capabilities ought to automatically re-link references to the Request-URI to one or more of the new references returned by the server, where possible. This response is cacheable unless indicated otherwise.
- The new permanent URI SHOULD be given by the Location field in the response. Unless the request method was HEAD, the entity of the response SHOULD contain a short hypertext note with a hyperlink to the new URI(s).
- If the 301 status code is received in response to a request other than GET or HEAD, the user agent MUST NOT automatically redirect the request unless it can be confirmed by the user, since this might change the conditions under which the request was issued.
- **Note:** When automatically redirecting a POST request after receiving a 301 status code, some existing HTTP/1.0 user agents will erroneously change it into a GET request.

302 Found
The requested resource resides temporarily under a different URI. Since the redirection might be altered on occasion, the client SHOULD continue to use the Request-URI for future requests. This response is only cacheable if indicated by a Cache-Control or Expires header field.

The temporary URI SHOULD be given by the Location field in the response. Unless the request method was HEAD, the entity of the response SHOULD contain a short hypertext note with a hyperlink to the new URI(s).

If the 302 status code is received in response to a request other than GET or HEAD, the user agent MUST NOT automatically redirect the request unless it can be confirmed by the user, since this might change the conditions under which the request was issued.

Note: RFC 1945 and RFC 2068 specify that the client is not allowed to change the method on the redirected request. However, most existing user agent implementations treat 302 as if it were a 303 response, performing a GET on the Location field-value regardless of the original request method. The status codes 303 and 307 have been added for servers that wish to make unambiguously clear which kind of reaction is expected of the client.

400 Bad Request
The request could not be understood by the server due to malformed syntax. The client SHOULD NOT repeat the request without modifications.

401 Unauthorized
The request requires user authentication. The response MUST include a WWW-Authenticate header field (section 14.47) containing a challenge applicable to the requested resource. The client MAY repeat the request with a suitable Authorization header field (section 14.8). If the request already included Authorization credentials, then the 401 response indicates that authorization has been refused for those credentials. If the 401 response contains the same challenge as the prior response, and the user agent has already attempted authentication at least once, then the user SHOULD be presented the entity that was given in the response, since that entity might include relevant diagnostic information. HTTP access authentication is explained in "HTTP Authentication: Basic and Digest Access Authentication".

403 Forbidden
The server understood the request but is refusing to fulfil it. The authorization will not help and the request SHOULD NOT be repeated. If the request method was not HEAD and the server wishes to make public why the request has not been fulfilled, it SHOULD describe the reason for the refusal in the entity. If the server does not wish to make this information available to the client, the status code 404 (Not Found) can be used instead.

404 Not Found
The server has not found anything matching the Request-URI. No indication is given of whether the condition is temporary or permanent. The 410 (Gone) status code SHOULD be used if the server knows, through some internally configurable mechanism, that an old resource is permanently unavailable and has no forwarding

address. This status code is commonly used when the server does not wish to reveal exactly why the request has been refused, or when no other response is applicable.

500 Internal Server Error
The server encountered an unexpected condition that prevented it from fulfilling the request.

501 Not Implemented
The server does not support the functionality required to fulfil the request. This is the appropriate response when the server does not recognize the request method and is not capable of supporting it for any resource.

502 Bad Gateway
The server, while acting as a gateway or proxy, received an invalid response from the upstream server it accessed in attempting to fulfil the request.

503 Service Unavailable
The server is currently unable to handle the request due to a temporary overloading or maintenance of the server. The implication is that this is a temporary condition that will be alleviated after some delay. If known, the length of the delay MAY be indicated in a Retry-After header. If no Retry-After is given, the client SHOULD handle the response as it would for a 500 response.

Note: The existence of the 503 status code does not imply that a server must use it when becoming overloaded. Some servers may wish to simply refuse the connection.

HTTP Status Codes & Details