

In-depth guide to how Google Search works

Google Search is a fully-automated search engine that uses software known as web crawlers that explore the web regularly to find pages to add to our index. In fact, the vast majority of pages listed in our results aren't manually submitted for inclusion, but are found and added automatically when our web crawlers explore the web. This document explains the stages of how Search works in the context of your website. Having this base knowledge can help you fix crawling issues, get your pages indexed, and learn how to optimize how your site appears in Google Search.

Looking for something less technical? Check out our [How Search Works site](https://www.google.com/search/howsearchworks/) (<https://www.google.com/search/howsearchworks/>), which explains how Search works from a searcher's perspective.

A few notes before we get started

Before we get into the details of how Search works, it's important to note that Google doesn't accept payment to crawl a site more frequently, or rank it higher. If anyone tells you otherwise, they're wrong.

Google doesn't guarantee that it will crawl, index, or serve your page, even if your page follows the [Google Search Essentials](/search/docs/essentials) (</search/docs/essentials>).

Introducing the three stages of Google Search

Google Search works in three stages, and not all pages make it through each stage:

1. **Crawling:** (#crawling) Google downloads text, images, and videos from pages it found on the internet with automated programs called crawlers.
2. **Indexing:** (#indexing) Google analyzes the text, images, and video files on the page, and stores the information in the Google index, which is a large database.
3. **Serving search results:** (#serving) When a user searches on Google, Google returns information that's relevant to the user's query.

Crawling

The first stage is finding out what pages exist on the web. There isn't a central registry of all web pages, so Google must constantly look for new and updated pages and add them to its list of known pages. This process is called "URL discovery". Some pages are known because Google has already visited them. Other pages are discovered when Google follows a link from a known page to a new page: for example, a hub page, such as a category page, links to a new blog post. Still other pages are discovered when you submit a list of pages (a [sitemap](/search/docs/crawling-indexing/sitemaps/overview)) for Google to crawl.

Once Google discovers a page's URL, it may visit (or "crawl") the page to find out what's on it. We use a huge set of computers to crawl billions of pages on the web. The program that does the fetching is called [Googlebot](/search/docs/crawling-indexing/googlebot) (also known as a crawler, robot, bot, or spider). Googlebot uses an algorithmic process to determine which sites to crawl, how often, and how many pages to fetch from each site. Google's [crawlers](/search/docs/crawling-indexing/overview-google-crawlers) are also programmed such that they try not to crawl the site too fast to avoid overloading it. This mechanism is based on the responses of the site (for example, [HTTP 500 errors mean "slow down"](/search/docs/crawling-indexing/http-network-errors#http-status-codes)) and [settings in Search Console](https://support.google.com/webmasters/answer/48620).

However, Googlebot doesn't crawl all the pages it discovered. Some pages may be [disallowed for crawling](/search/docs/crawling-indexing/robots/robots_txt#disallow) by the site owner, other pages may not be accessible without logging in to the site.

During the crawl, Google renders the page and [runs any JavaScript it finds](/search/docs/crawling-indexing/javascript/javascript-seo-basics#how-googlebot-processes-javascript) using a recent version of [Chrome](https://www.google.com/chrome/), similar to how your browser renders pages you visit. Rendering is important because websites often rely on JavaScript to bring content to the page, and without rendering Google might not see that content.

Crawling depends on whether Google's crawlers can access the site. Some common issues with Googlebot accessing sites include:

- [Problems with the server handling the site](/search/docs/crawling-indexing/http-network-errors#http-status-codes)
- [Network issues](/search/docs/crawling-indexing/http-network-errors#network-and-dns-errors)
- [robots.txt rules preventing Googlebot's access to the page](/search/docs/crawling-indexing/robots/intro)

Indexing

After a page is crawled, Google tries to understand what the page is about. This stage is called indexing and it includes processing and analyzing the textual content and key content tags and attributes, such as <title> elements (/search/docs/appearance/title-link) and alt attributes, images (/search/docs/appearance/google-images), videos (/search/docs/appearance/video), and more.

During the indexing process, Google determines if a page is a duplicate of another page on the internet or canonical (/search/docs/crawling-indexing/consolidate-duplicate-urls). The canonical is the page that may be shown in search results. To select the canonical, we first group together (also known as clustering) the pages that we found on the internet that have similar content, and then we select the one that's most representative of the group. The other pages in the group are alternate versions that may be served in different contexts, like if the user is searching from a mobile device or they're looking for a very specific page from that cluster.

Google also collects signals about the canonical page and its contents, which may be used in the next stage, where we serve the page in search results. Some signals include the language of the page, the country the content is local to, the usability of the page, and so on.

The collected information about the canonical page and its cluster may be stored in the Google index, a large database hosted on thousands of computers. Indexing isn't guaranteed; not every page that Google processes will be indexed.

Indexing also depends on the content of the page and its metadata. Some common indexing issues can include:

- The quality of the content on page is low (/search/docs/essentials)
- Robots meta rules disallow indexing (/search/docs/crawling-indexing/block-indexing)
- The design of the website might make indexing difficult
(/search/docs/crawling-indexing/javascript/javascript-seo-basics)

Serving search results

Google doesn't accept payment to rank pages higher, and ranking is done programmatically. Learn more about ads on Google Search

(<https://www.google.com/search/howsearchworks/our-approach/ads-on-search/>).

When a **user enters a query, our machines search the index for matching pages and return the results we believe are the highest quality and most relevant to the user's query.**

Relevancy is determined by hundreds of factors, which could include information such as the **user's location, language, and device (desktop or phone).** For example, searching for "bicycle repair shops" would show different results to a user in Paris than it would to a user in Hong Kong.

Based on the user's query the search features that appear on the search results page also change. For example, searching for "bicycle repair shops" will likely show local results and no image results (</search/docs/appearance/visual-elements-gallery#image-result>), however searching for "modern bicycle" is more likely to show image results, but not local results. You can explore the most common UI elements of Google web search in our Visual Element gallery (</search/docs/appearance/visual-elements-gallery>).

Search Console might tell you that a page is indexed, but you don't see it in search results. This might be because:

- The content on the page is irrelevant to users' queries (</search/docs/fundamentals/seo-starter-guide>)
- The quality of the content is low (</search/docs/essentials>)
- Robots meta rules prevent serving (</search/docs/crawling-indexing/block-indexing>)

While this guide explains how Search works, we are always working on improving our algorithms. You can keep track of these changes by following the Google Search Central blog (</search/blog>).

Except as otherwise noted, the content of this page is licensed under the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the Apache 2.0 License (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the Google Developers Site Policies (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2023-12-01 UTC.