How search works

- How crawlers browse the web
- Building an index
- Serving the most useful results
- Next steps: how to optimize for search engines

Search engines are the digital version of a librarian. They use a comprehensive index to help find the right information for a query. Understanding the basics of search prepares you to make your content **discoverable** for users.

How crawlers browse the web

Crawling is like reading through all the books in the library. Before search engines can bring any search results, they need to have as much information from the web as possible. To do this, search engines use a crawler—a program that travels from site to site and acts like a browser.

If a book or document is missing or damaged, the crawler can't read it. Crawlers try to fetch each URL to determine the state of the document. If a document returns an error status code, crawlers cannot use any of its content, and might retry the URL at a later time. This ensures only publicly accessible documents get into the index.

If crawlers discover a redirection status code (like 301 or 302), they follow the redirection to a new URL and continue there. Once they get a successful response, meaning they've found a document accessible to users, they check if it's allowed to be crawled and then download the content.

This check includes the HTML and all content mentioned in the HTML, such as images, videos, or JavaScript. Crawlers also extract the links from HTML documents so that the crawler can visit the linked URLs as well. Following links is how crawlers find new pages on the web.

Crawlers don't actively click links or buttons, but instead send URLs to a queue to crawl them later. When accessing a new URL, no cookies, service workers or local storage (like IndexedDB) are available.

Building an index

After retrieving a document, the crawler hands the content to the search engine to add it to the index. The search engine now renders and analyzes the content to understand it. Rendering means displaying the page as a browser would (with some limitations).

Search engines look at keywords, the title, links, headings, text, and many other things. These are called **signals** which describe the content and context of the page. Signals allow search engines to answer any given query with the best possible page.

Search engines might find the same content at different URLs. For example a recipe for "apple pie" might live under `/recipes/apple-pie` and under `/recipes/1234`. To avoid indexing and showing the recipe twice, search engines determine what the main URL should be and discard the alternative URLs showing the same content.

Serving the most useful results

Search engines do more work than just matching the query to keywords in the index. To give useful results, they might consider context, alternative wording, location of the user, and more. For example, "silicon valley" might refer to the geographic region or the TV show. But if the query is "silicon valley cast", results on the region aren't very helpful.

Some queries can be indirect, like "the song from pulp fiction", and search engines need to interpret that and show results for the music in the film. When a user searches for something, search engines determine the most useful results and then show them to the user. Ranking, or ordering, the pages happens based on the query. The order can often change over time if better information becomes available.

Next steps: how to optimize for search engines

Now that you understand the basics of how search engines work, you may see the value in optimizing for search engines. This is called SEO, or 'Search Engine Optimization.' By making sure search engines can find and automatically understand your content, you are improving the visibility of your site for relevant searches. This can result in more interested users coming to your site. Audit your site with Lighthouse and check the SEO results to see how well search engines can make users aware of your content.

□□□ #2
□ □□□ □□□ 2026-10-31 22:32:46 UTC
□ □□□ □□□ 2026-10-31 22:32:47 UTC