How Google crawls and indexes: a non-technical explanation [Video]

[embedded content]

The crawling and indexing processes lay the groundwork for search engines to rank results. Despite being fundamental aspects of how search engines operate, crawling and indexing are often overlooked or misunderstood. During our crawling and indexing session of Live with Search Engine Land, Martin Splitt, search developer advocate at Google, explained these two processes using a simple analogy about librarians.

“Imagine a librarian: If you are writing a new book, the librarian has to actually take the book and figure out what the book is about and also what it relates to, if there’s other books that might be source material for this book or might be referenced from this book,” Splitt said. In his example, the librarian is Google’s web crawler (referred to as Googlebot) and the book is a website or webpage.

“Then you . . . have to read through [the book], you have to understand what it is about, you have to understand how it relates to the other books, and then you can sort it into the catalog,” he said, explaining the indexing process. The content of your webpage is then stored in the “catalog” (i.e., the search engine’s index), where it can be ranked and served as a result for relevant queries.

To bring the analogy full circle, Splitt also described the process in technical terms: “We have a list of URLs . . . and we take each of these URLs, we make a network request to them, then we look at the server response and then we also render it (we basically open it in a browser to run the JavaScript) and then we look at the content again, and then we put it in the

index where it belongs, similar to what the librarian does.”

Why we care. For content to be eligible to appear in search results, it must first be crawled and indexed. Understanding how crawling and indexing work can help you resolve technical SEO issues and ensure your pages are accessible to search engines.

Want more Live with Search Engine Land? Get it here:

About The Author

George Nguyen is an Associate Editor at Third Door Media. His background is in content marketing, journalism, and storytelling.
http://feeds.searchengineland.com/~r/searchengineland/~3/Avu2sPxNwiw/how-google-crawls-and-indexes-a-non-technical-explanation-video-335245