Lesson 1.3: How Search Works Contents:

- Crawling the web
- Understanding how Google ranks results
- Identifying advertisements

Access Lesson 1.3 slides here

This lesson features a video by Matt Cutts, an engineer at Google. It demonstrates how spiders work: how they crawl the web, collect information, and pull it all together to provide an index Google uses every time you do a search.

Here's Matt Cutts explaining how search actually works:

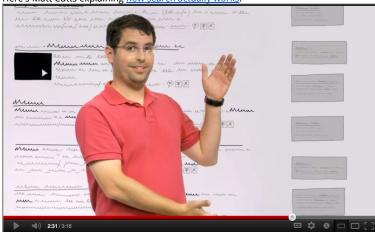


Figure: Matt Cutts talking in the video How Search Works

Matt Cutts: Hi. My name is Matt Cutts. I am an engineer in the Quality group at Google, and I'd like to talk today about what happens when you do a web search. The first thing to understand is, when you do a Google search, you aren't actually searching the web. You're searching Google's index of the web, or at least as much of it as we can find.

We do this with software programs called spiders. Spiders start by fetching a few web pages, then they follow the links on those pages and fetch the pages they point to; and follow all the links on those pages, and fetch the pages they link to, and so on, until we've indexed a pretty big chunk of the web; many billions of pages stored on thousands of machines.

Now, suppose I want to know how fast a cheetah can run. I type in my search, say [cheetah, running, speed] and hit return.

Our software searches our index to find every page that includes those search terms. In this case, there are hundreds of thousands of possible results.

How does Google decide which few documents I really want? By asking questions; more than two hundred of them. Like:

- How many times does this page contain your key words?
- Do the words appear in the title? In the URL (web address)?
- Do the words appear directly adjacent?
- Does the page include synonyms for those words?
- Is this page from a quality website? Or is it low quality, even spammy? What is this page's PageRank? That's a formula invented by our founders, Larry Page and Sergey Brin, that rates a web page's importance by looking at how many outside links point to it, and how important those links are.

Finally, we combine all those factors together to produce each page's overall score and send you back your search results, about half a second after you submit your search. At Google, we take our commitment to delivering useful and impartial search results VERY seriously. We don't ever accept a payment to add a site to our index, update it more often, or improve its ranking.

Sometimes, along the right, and at the top, you'll see ads.

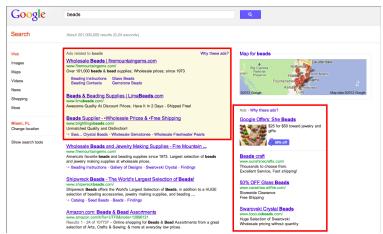


Figure: A search results page, with highlighted ads at the top and on the right side.

We take our advertising business very seriously as well. Both, our commitment to delivering the best possible audience for our advertisers, and to strive to only show ads that you really want to see.

We're very careful to distinguish your ads from your search results. And we won't show you any ads at all if we can't find any we think will help you find the information you're looking for. Which in this case, a cheetah's top running speed, is more than sixty miles an hour.

Thanks for watching. I hope this makes Google a little more understandable. Please give the activity a try!

Power Searching with Google © 2015 Google, Inc. (DMR 7-18-15)