Come lavora Google?

di Paolo Franzese

Tutti i contenuti sono sotto Licenza Creative Commons
Contatti +39 388 1020417 - info@imaginepaolo.com

Google(graphic)

How google works

Google Search | I'm Feeling Lucky

You write a blog post, tweet, update a site, or otherwise add content to the web.

Google bots crawling the web come across your post.

Google bots follow links. If there are no links to your site, typically it will not get crawled deeply or regularly.

Google bots won't crawl your site if you tell them not to with a robots.txt

If links to your site have a nofollow tag, the google bots won't arrive from those links.

Google may also find your site by being pinged by blog software or xml sitemaps.

Page titles and link data are stored in one index used for broad and competitive searches.

On page content is stored in another index used for obscure and long tail searches.

Page content is stored in a reverse index.

Once crawled, the page is indexed within seconds.

When you search google you are not searching the active web, but Google's cache of it, which is constantly being updated.

Google estimates the domain and page's overall authority based on links.

The more links you have from higher authority pages, the greater your own pages' authority.

As long as they are not tagged 'nofollow'.

Google's Search Quality team and Webspam Team review and refine algorithms.

10,000+ remote testers rate the quality of their searches.

Google solicits spam reports from users.

Google gets DMCA notifications to take down pirated work.

Pages are checked against editorial policies.

Penalties are applied and each page now has many pieces of data attached to it that help in user searches.

"On most Google queries, you're actually in multiple control or experimental groups simultaneously... Essentially, all the queries are involved in some test." -Patrick Riley, Google search quality engineer

Search Query: 1033034356021

User queries Google

Google suggests keywords based on what has been typed so far.

Google uses synonyms to look for similar words to include in the search query.

Google may claim millions of results but only 1,000 or less are ever displayed.

Initial result set is created.

Result localization: Local websites are promoted in the search results.

Result set is sorted by authority and PageRank and duplicate pages are removed.

With universal search, if Google thinks news results, shopping results, video results, book results, local results, or any other form of vertical search are relevant then they may mix these directly into the search results.

User Personalization: Past websites the user visits are often promoted.

Excessive anchor text manipulation can cause websites to be removed from the results.

Local interconnectivity of the result set: If pages are well linked among other high ranking sites, then their ranking may be boosted.

Filters applied.

The official word

"The software behind our search technology conducts a series of simultaneous calculations requiring only a fraction of a second. Traditional search engines rely heavily on how often a word appears on a web page. We use more than 100 signals, including our patented PageRank algorithm, to examine the entire link structure of the web and determine which pages are most important. We then conduct hypertext-matching analysis to determine which pages are relevant to the specific search being conducted. By combining overall importance and query-specific relevance, we're able to put the most relevant and reliable results first." -Google

Google is very secretive about its data centers. This is what we know so far.



There are over 19 data centers in the US, with at least 17 more around the world.



They can be up to 500,000 square feet and $600 million each.



Google data centers highly efficient & Google aims to become carbon neutral.



Data centers use between 50-100 megawatts of power and often found near water (cooling).



Google servers are housed in standard shipping containers that hold 1,160 servers each.

Google finds relevant ads based on keywords, ad match type and user location.

Advertisers operating outside the guidelines may have their ad accounts banned.

Ads are subject to editorial policy

If keyword has low search volume or ad generates too few clicks, ads might be automatically disabled.

Relevant ads are ordered based on earnings potential. (bid x ad quality score)

Favored businesses, perhaps like Amazon.com, maybe be given a discount.

Dynamic keyword content is used to make the ad appear more relevant.

Some ads also have extensions available, like site links, phone numbers, product links, location, etc.

Tutti i contenuti sono sotto Licenza Creative Commons
Contatti +39 388 1020417 - info@imaginepaolo.com

Infographic by the Pay Per Click Blog

Da un articolo scritto da Paolo Franzese il 11 Luglio 2012

Tutti i contenuti sono sotto Licenza Creative Commons