17   5   help

How does google return so many search results?

When googling for some term we're told that we see the k's 10 results out of thousands. I.e 1-10th results out of 10000. We can skip right to 50-60th result and google will return them in less than a second. If google would cache all of the 1-40, I guess it would waste RAM. If it would return all of the results at once, it would waste traffic. How does google return so many search results while it sorts them?

architecture | search-engine

asked Jul 20 '13 at 10:48
Vitali Pom
169   1   12

This question appears to be off-topic because it is not about programming – John Conde Jul 20 '13 at 12:37

1   Saying that this question is not about programming is as saying that a question on recipe is not about food. – Vitali Pom Jul 20 '13 at 14:13

@JohnConde then why is the architecture tag on SO? – rahulserver Jul 20 '13 at 14:19
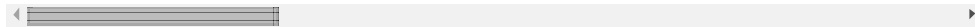
add a comment

1 Answer

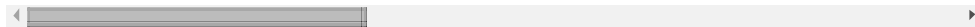Here is something I tried: For the google search:

How google search engine works

Here is the url sent to the server:

```
https://www.google.co.in/?gws_rd=cr#gs_rn=20&gs_ri=psy-ab&tok=v1e-J_ynDKaPw_oDlFH-
```

And if I select the page number 5 link on the bottom of the page, here is the url which is being sent:

```
https://www.google.co.in/?gws_rd=cr#q=how+google+search+engine+works&ei=qm3qUeXSA4
```

Note the additional `start=40&sa=N` parameter here.

So what I feel after going thru google search enging working step 2 of 3 that google has its own index for over **30 TRILLION web pages** and for each query they rank the result using over 200 factors. So I feel that the `start` and `sa` parameters form one of those factors while fetching the result.

Hence my opinion is that google is not caching the results, but carries out a new search for each queries using different url params.

EDIT:

To further support my observation, here is a list of time taken for navigating to the different pages of the above search query:

1. For first page: `157,000,000 results (0.22 seconds)`
2. For page 5: `Page 5 of about 157,000,000 results (0.26 seconds)`
3. For page 10: `Page 10 of about 157,000,000 results (0.30 seconds)`

4. For page 14: `Page 14 of about 157,000,000 results (0.33 seconds)`

5. For page 18: `Page 18 of about 157,000,000 results (0.38 seconds)`
6. For page 22: `Page 22 of about 157,000,000 results (0.35 seconds)`
7. For page 26: `Page 26 of about 157,000,000 results (0.43 seconds)`
8. For page 30: `Page 30 of about 157,000,000 results (0.45 seconds)`
9. For page 34: `Page 34 of about 157,000,000 results (0.40 seconds)`
10. For Page 38: `Page 38 of about 371 results (0.50 seconds)`

So the time is almost increasing for increasing page numbers of queries. If it were caching the results, above would not have been the trend.

edited Jul 20 '13 at 11:33   answered Jul 20 '13 at 11:08

                                                      rahulserver
                                                      1,347 1   11   30

For more information, Google for DAAT and TAAT algorithms which retrieve the top k's results. (I found it out later). –   Vitali Pom   Jul 21 '13 at 10:21

----

add a comment

Not the answer you're looking for? Browse other questions tagged   architecture

search-engine   **or ask your own question.**