

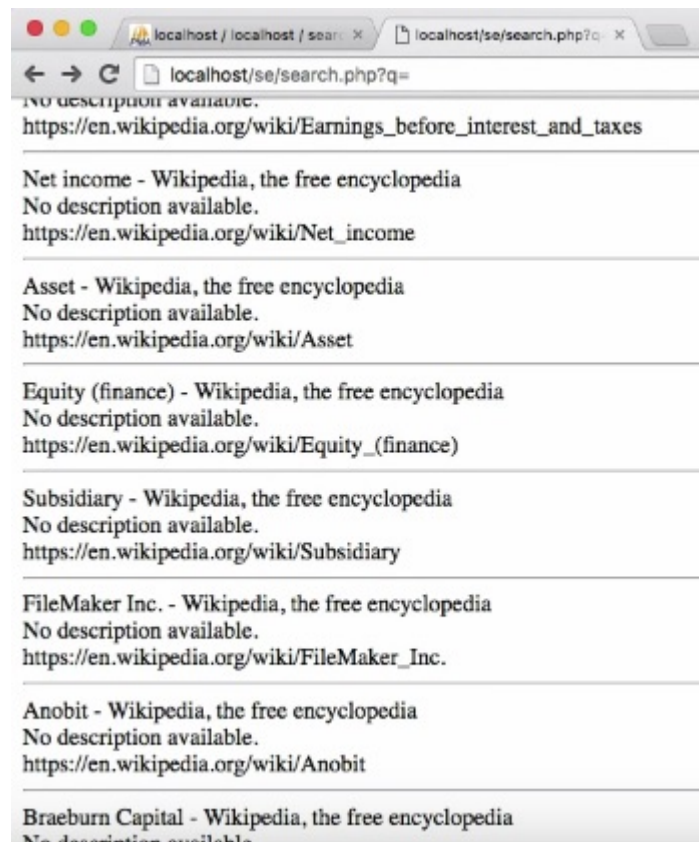
Creating a simple search engine

Aninda Maulik

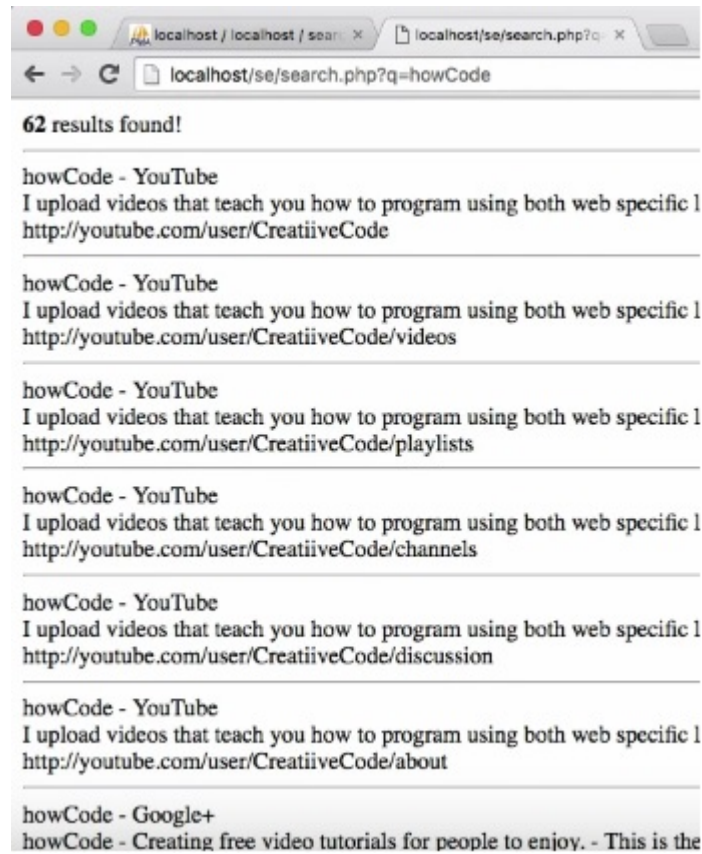
January 2020

1 Introduction

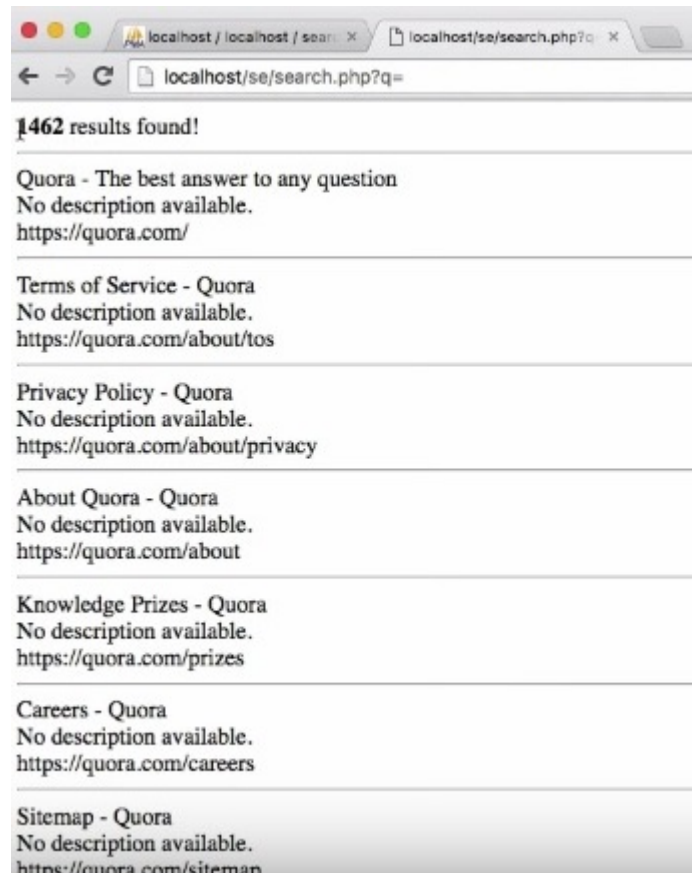
We'll be creating a simple search engine like below:



We would now try to query something, in our case, "howCode". Please note that we could have chosen to query anything for eg: Natural Disaster which is our topic but then we have got to have an associated database at the background:- and that's not available as of yet.



After passing a query sample, we get all the results available in the database. Here's how it would look like:



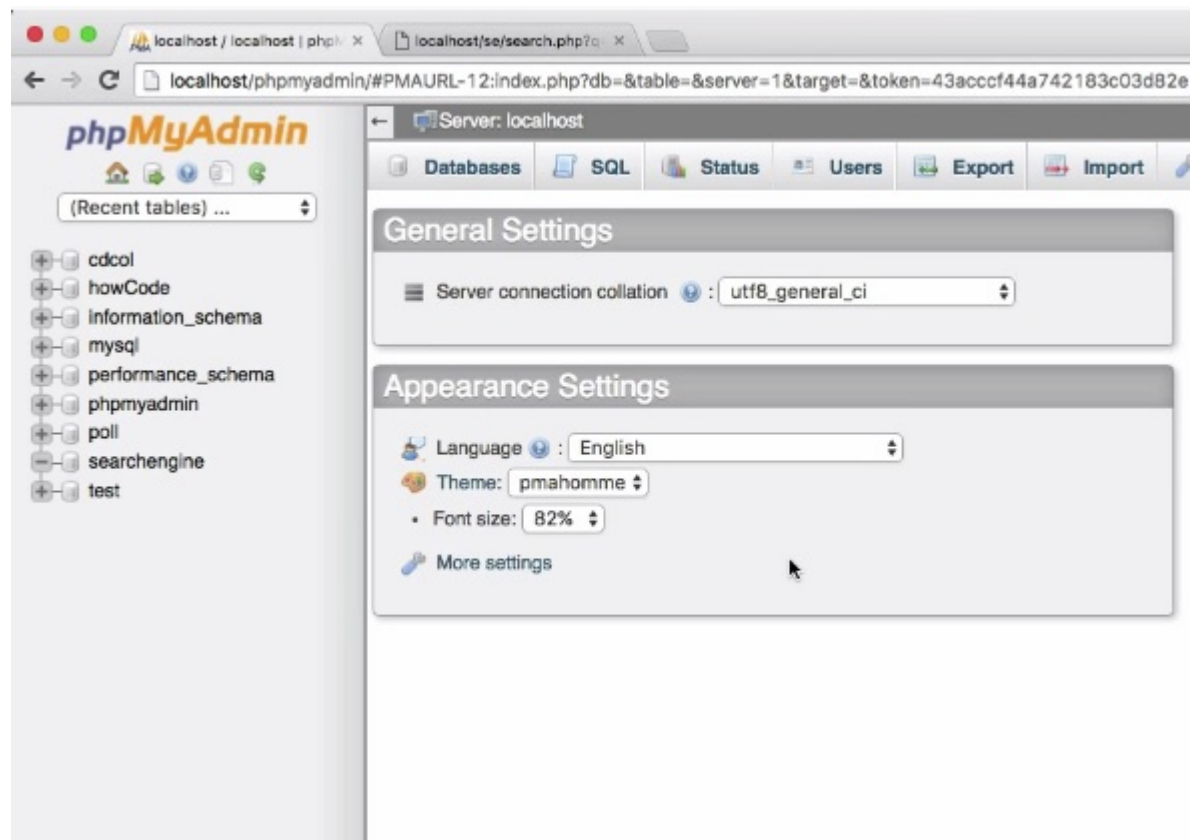
Search engine would be comprising of two pages. There's an index page which is going to be our webcrawler and the second page would be our search page. We could move the index page to a different page altogether and call that a crawler to make a home page like how Google does on their index page. Here's a look:

```

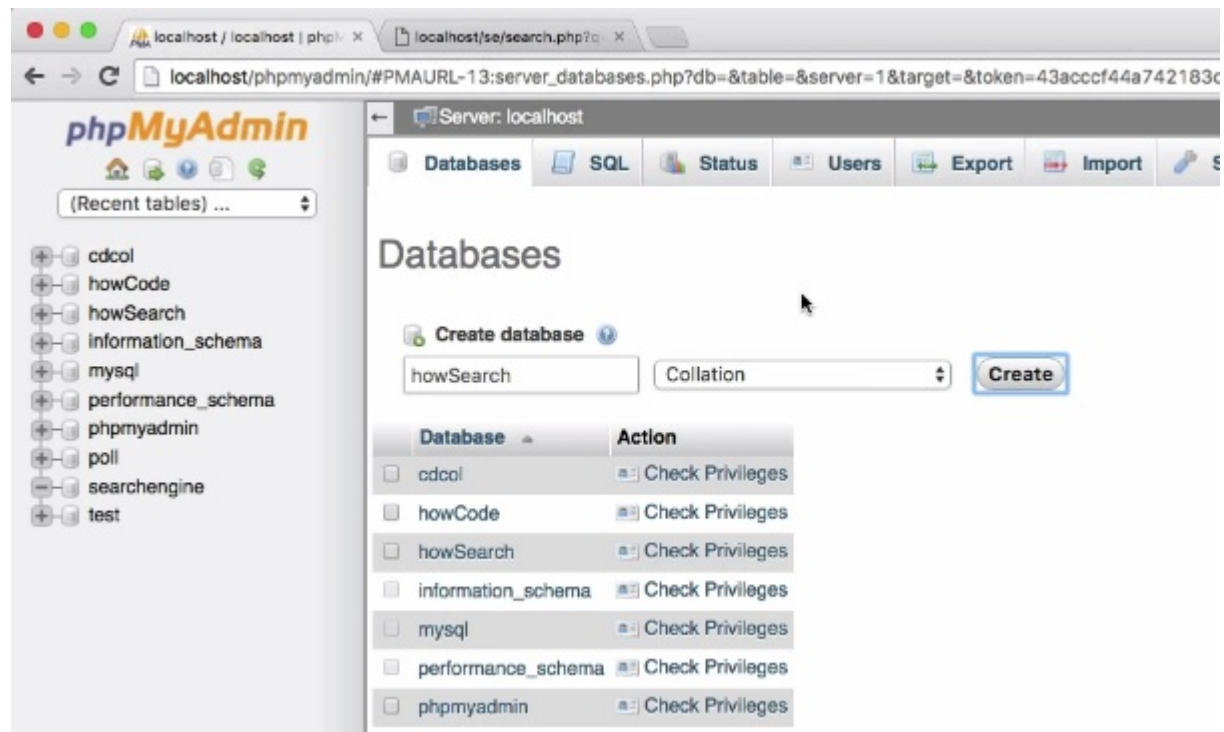
index.php search.php: 1x
76     } else if (substr($l, 0, 5) != "https" && substr($l, 0, 4) != "http") {
77         $l = parse_url($url)["scheme"]."://".parse_url($url)["host"]."/". $l;
78     }
79     // If the link isn't already in our crawl array add it, otherwise ignore it.
80     if (!in_array($l, $already_crawled)) {
81         $already_crawled[] = $l;
82         $crawling[] = $l;
83         // Output the page title, descriptions, keywords and URL. This output is
84         // piped off to an external file using the command line.
85         echo get_details($l)."\n";
86     }
87
88 }
89 // Remove an item from the array after we have crawled it.
90 // This prevents infinitely crawling the same page.
91 array_shift($crawling);
92 // Follow each link in the crawling array.
93 foreach ($crawling as $site) {
94     follow_links($site);
95 }
96
97 }
98 // Begin the crawling process by crawling the starting link first.
99 follow_links($start);
100

```

Creating a Database: Here's the first look

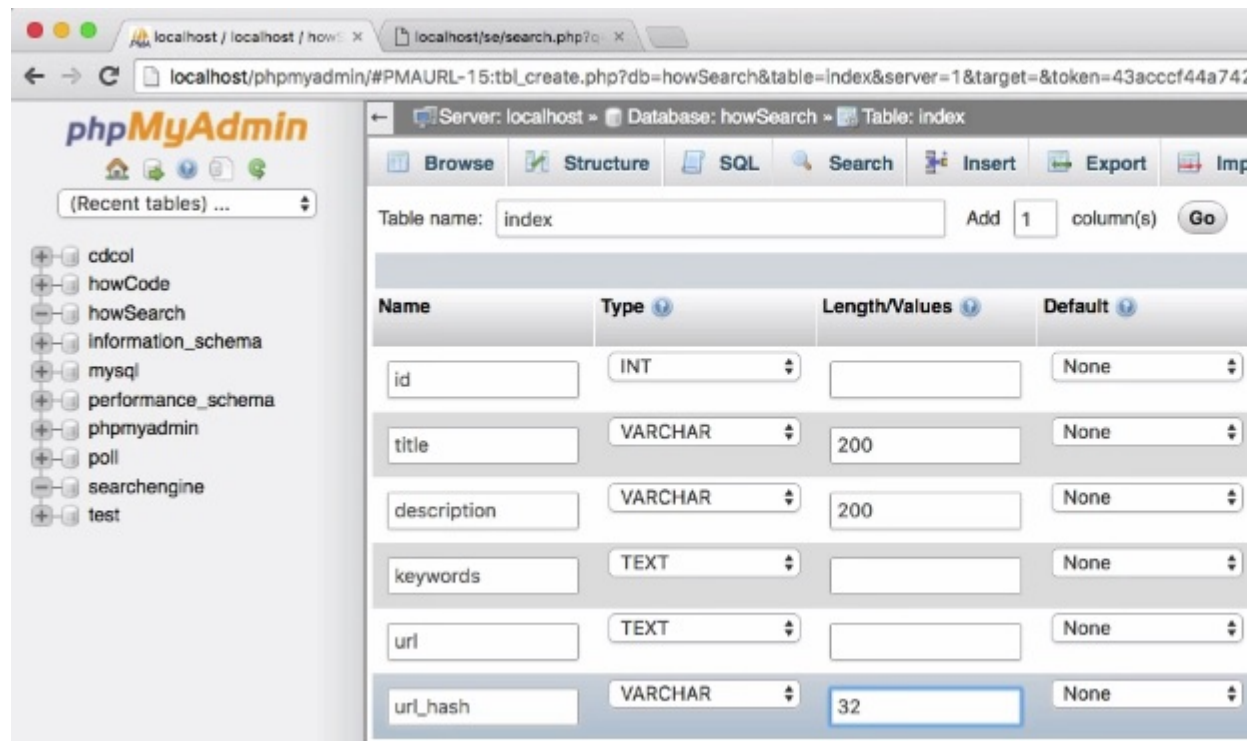


Then we attempt to create a database and call it "HowSearch", click on it which would lead to creating of the database.



Then we create a table and call it index, and the reason we are calling it an index is because when the search engine crawls through the internet, it stores all of it's pages in something called index. We would have 6 columns for the table namely: ID, Title,Description,Keywords,URL,URL-hash

Here's a look:



In the next session, we'll probably modify our crawler to insert it's results into our database and after that we'll be creating the search engine.