

CSCI 572 - Assignment 5
Spring 2019
Name – Karishma Chadha (7853208046)

Steps Followed to complete the Assignment

1. Enabled auto-suggest in Solr as described in the document provided.
2. Made changes to the solrconfig.xml by adding a search component and a request handler for suggest.
3. Reloaded the core and checked the suggest functionality by changing the requesthandler from “/select” to “/suggest” in Query View.
4. Added spell-correction functionality, by writing a java code which parsed the html webpages and created a big.txt file, using the tika jar – tika-app-1.2.jar.
5. Downloaded Peter Norvig’s spell correction program in PHP from the link provided. It uses the edit distance from the words in big.txt and returns the most probable word from the corpus that could be a correct word. Included this file in Main.php.
6. For autocomplete feature, modified php client with javascript and jquery. To make sure that the solr ‘suggest’ works for n-term queries, the query is split and the suggestions are then stored in an array which is returned as a responseText to the javascript. When the user starts typing even one character in the query box, the ajax code gets activated.
7. File simple_html_dom parses the text from html pages in order to create snippets. Using the method file_get_contents on each link, I matched the strings in the content to find the index of the words in the query.
8. Determined the snippet to be displayed with the help of this by finding a sentence with all the query terms together or finding the sentence that has all the terms in it, even if they are not together or in same order or finding the first sentence with at least one query term in it.
9. I then highlighted the query terms if they appear in the snippet.
10. All this is returned on the php front end page.

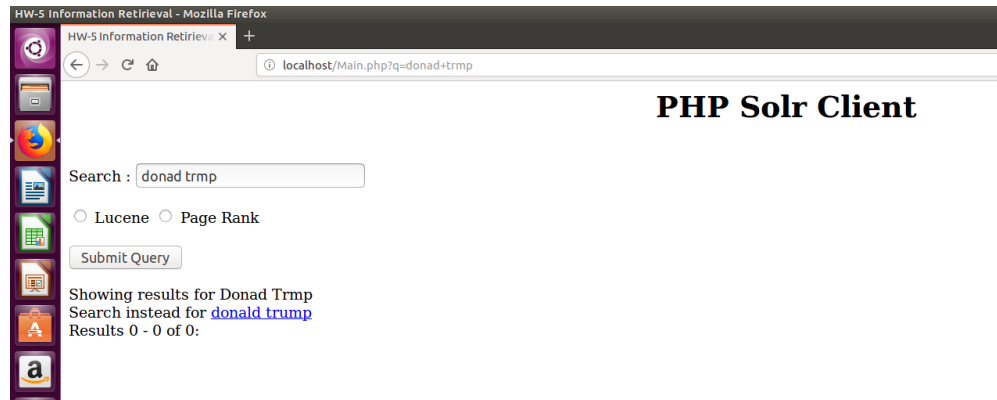
Analysis of Results

FIVE examples of misspelled terms

Query 1 - Incorrect Spelling - califnia
Correct Spelling - california



Query 2-
Incorrect Spelling - Donad Trmp
Correct Spelling - Donald Trump



PHP Solr Client

Query 3-

Incorrect Spelling - hurricane

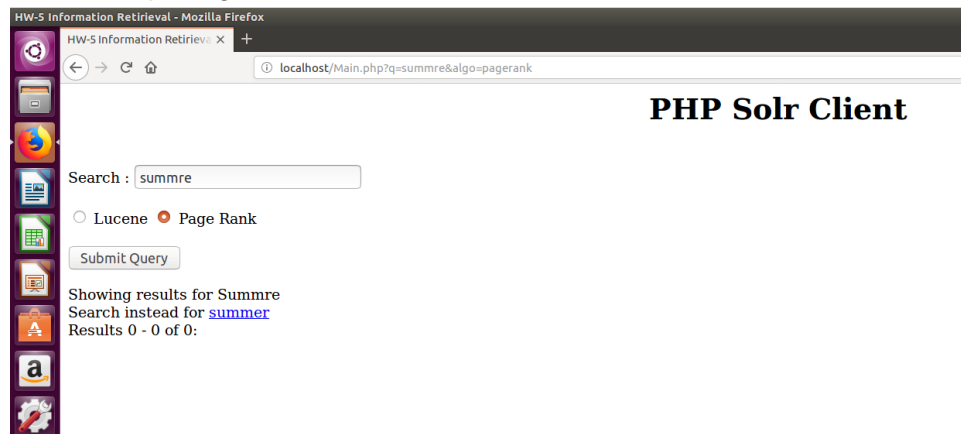
Correct Spelling - hurricane



Query 4-

Incorrect Spelling - sumer

Correct Spelling - summer



Query 5-

Incorrect Spelling - iphoen

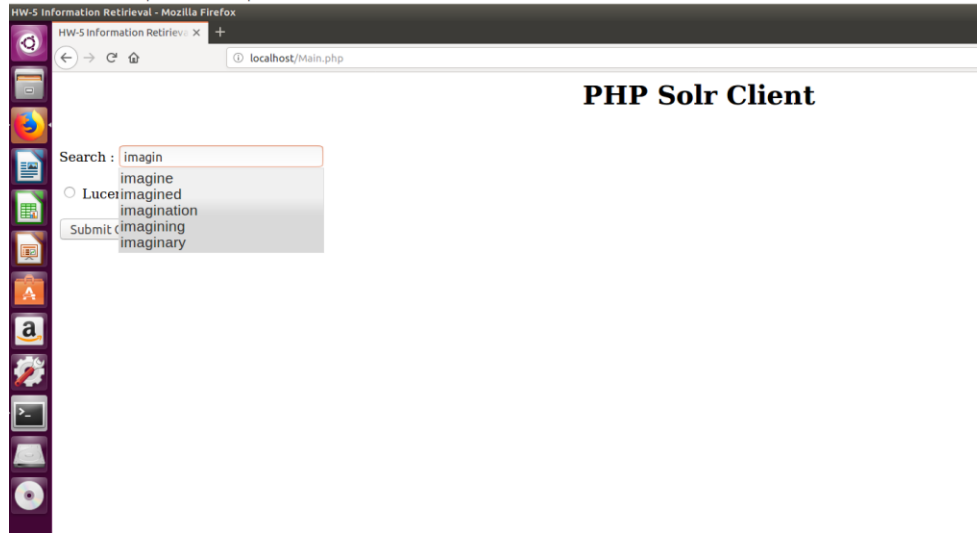
Correct Spelling - iphone



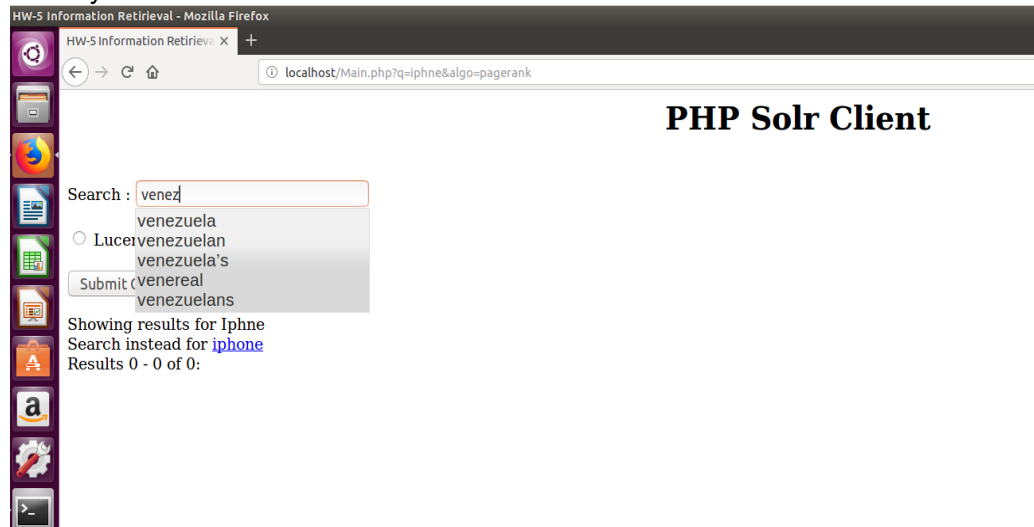
Analysis of Results

FIVE examples of auto-completion

Query 1 – imagine



Query 2 - venezuela



Query 3 – vice president



Query 4 – bomb attack



Query 5 – How Are You

