# Homework - 5
# Enhancing Your Search Engine

**Name: Anusha Mysore Swamy**
**USC ID: 5562476545**

## Steps followed for:

### 1. Autocomplete

Used Solr/Lucene's FuzzyLookupFactory to implement this feature, which automatically creates suggestions for misspelt words in fields. For this feature to work, I have added a search component and a request handler to solarconfig.xml, as suggested in the homework pdf "AutoCompleteInSolr". JQuery is used in the front end. An ajax request is made with the URL to Solr, whose response is the list of keywords matching the string input, and is later displayed to the user using the HTML5 dataList component.

### 2. Spell Correction

- Created a java program to parse all the HTML files for Fox news website, to create big.txt. Used Apache Tika library to extract the HTML content from each of the webpages
- Downloaded Norvig's SpellCorrector.php file from the link http://www.phpclasses.org/package/4859-PHP-Suggest-corrected-spelling-text-in-pure-PHP.html#download
- The input to Novig program's correct function is the big.txt. When the user types a word in the input, the program reads this from the php and sends the word to the SpellCorrector program. This program tries to check if the word exists in the dictionary(i.e. big.txt). If not, then it tries to find the nearest possible match and returns this as the corrected suggestion to the user and displays the results for the corrected word
- A link is also provided with the user's actual input to search for (if the user intends to continue with this input itself) and on clicking it, the obtained results are returned
- If the word does not exist in the webpages then the program returns an appropriate message to the user and asks him to try again with a different spelling

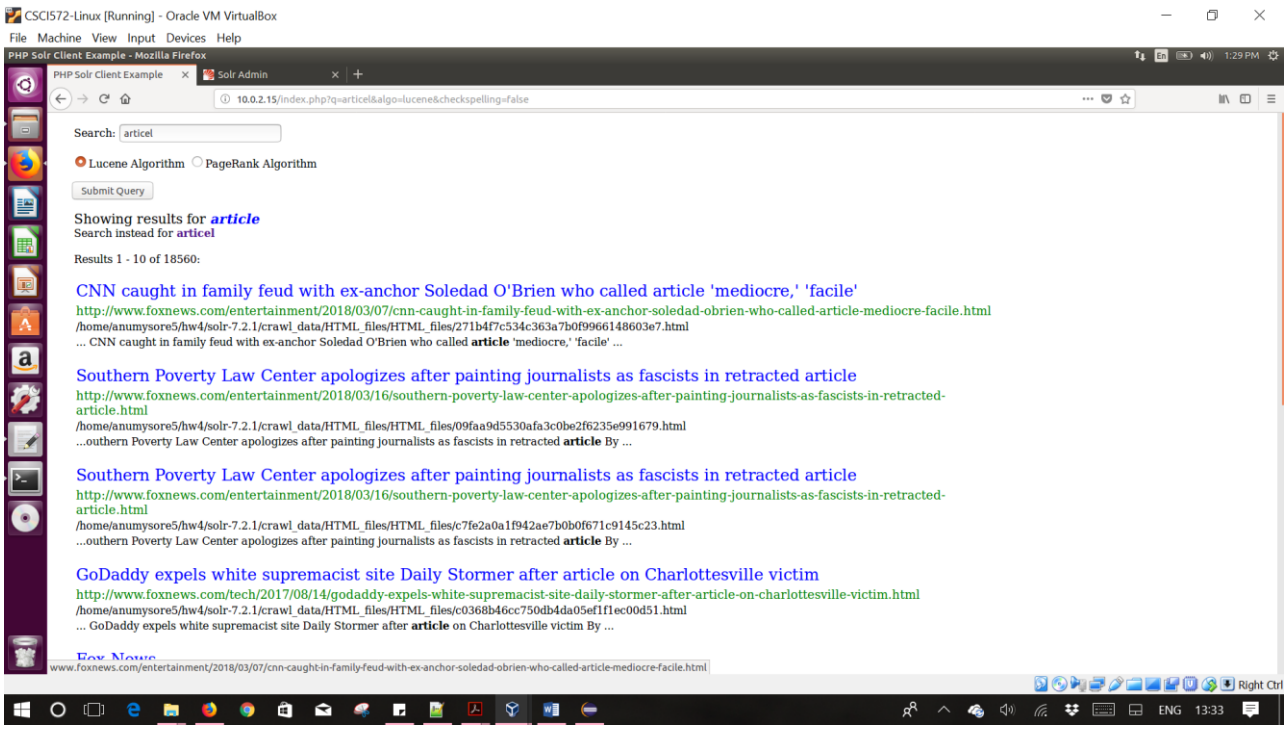
### 3. Snippets

The program tries to find the query term(s) in the html page contents and highlights them. If there are multiple terms, then it tries to find the first occurrence where all the terms occur together. If this is not found, then it tries to find the first sentence where at least one of the terms occurs. Once this is done, we extract the required amount of text and display to the user as a snippet.

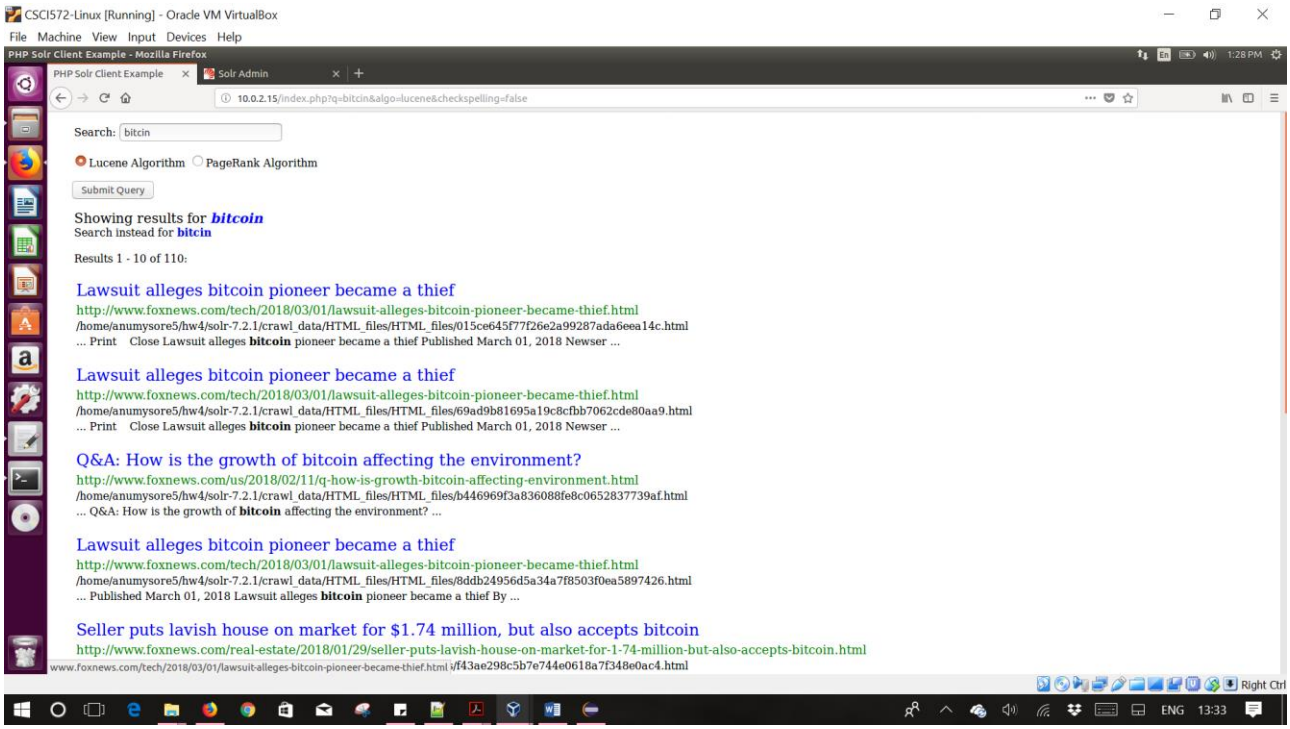
# Analysis of results:

## Five examples of spell correction:

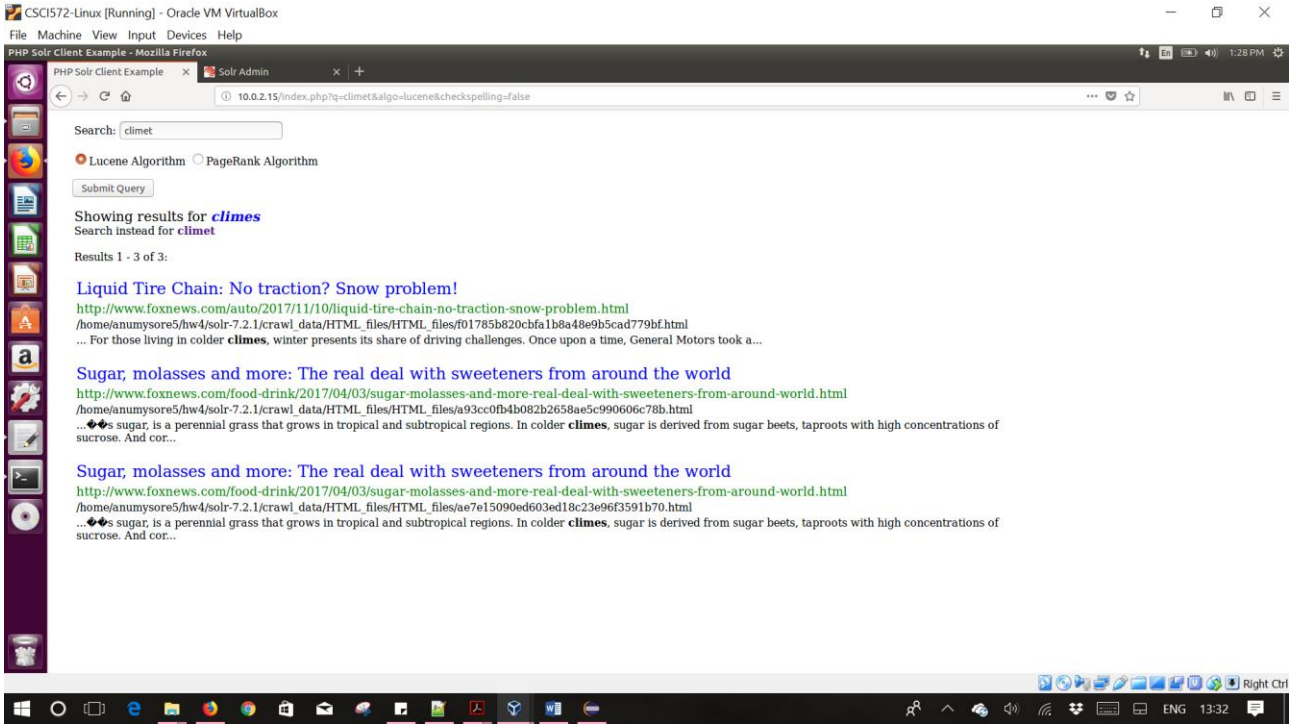
### 1.



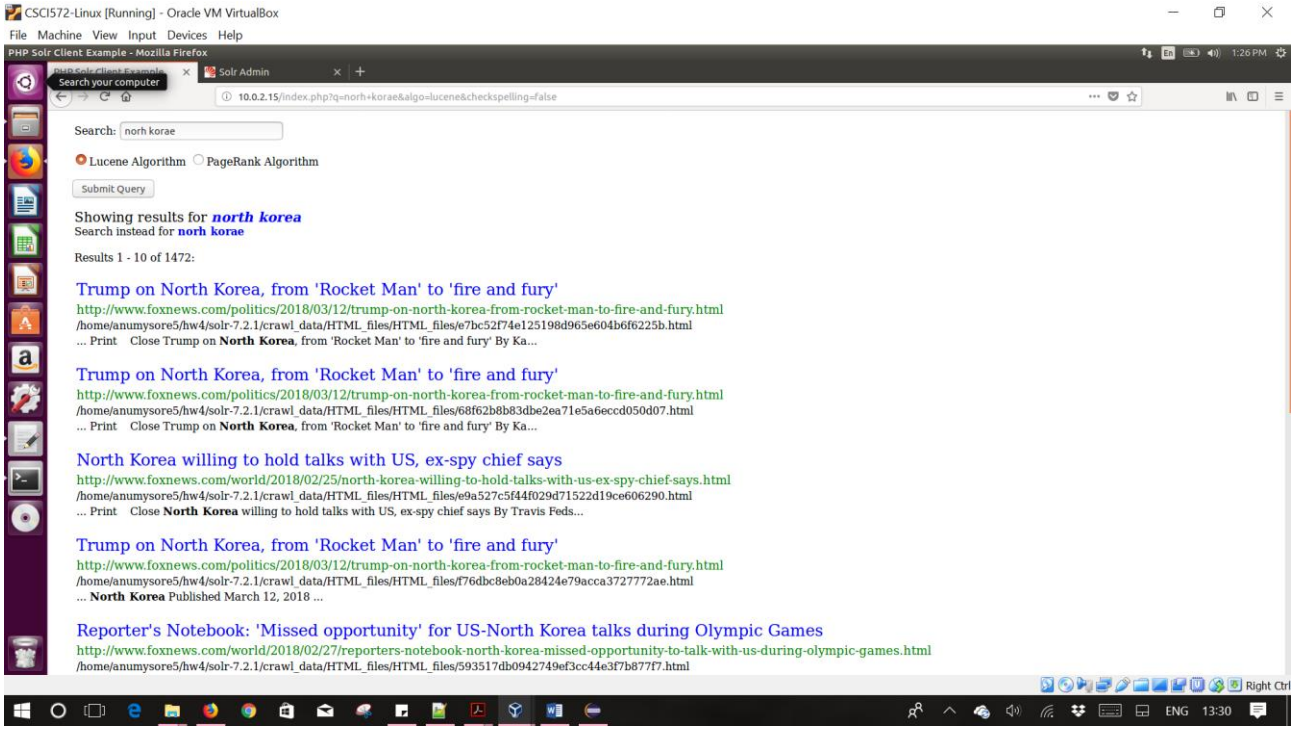

### 2.



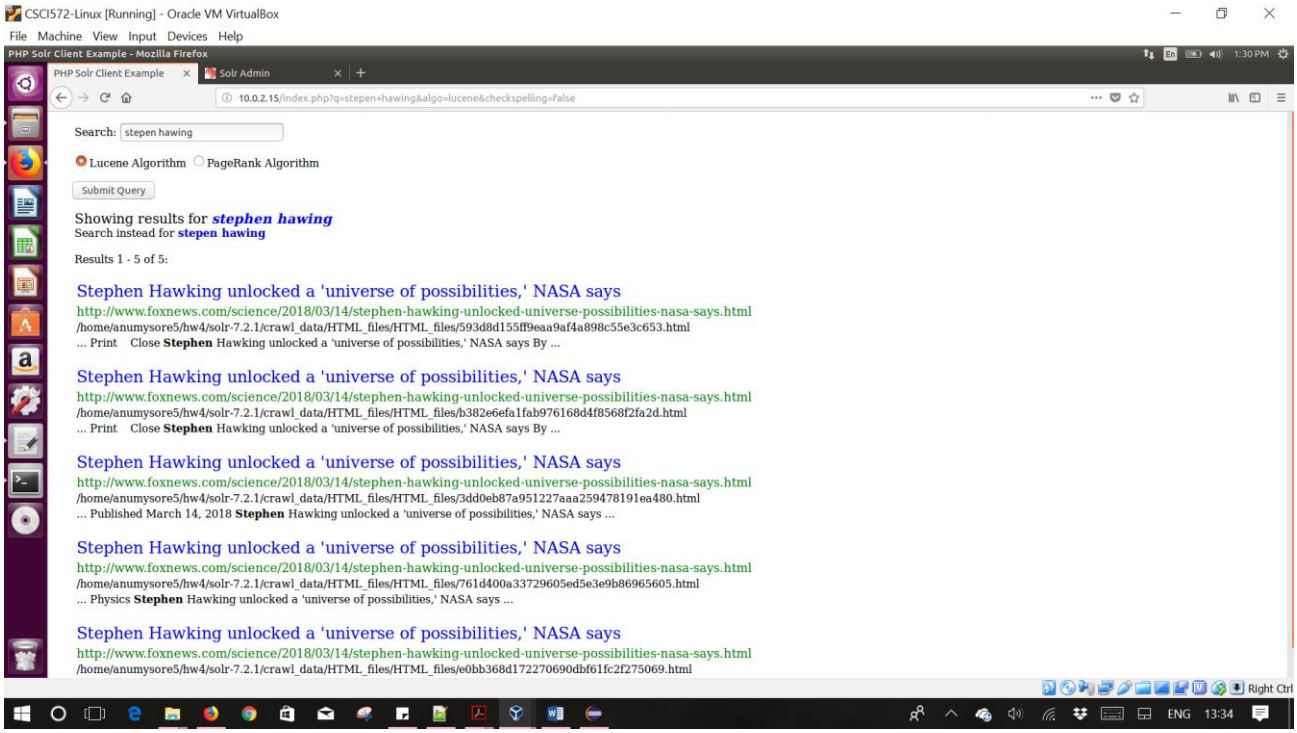
**3.**


CSCI572-Linux [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
PHP Solr Client Example - Mozilla Firefox
PHP Solr Client Example
Solr Admin
10.0.2.15/index.php?q=climet&algo=lucene&checkspelling=false
Search: climet
Lucene Algorithm
PageRank Algorithm
Submit Query
Showing results for climes
Search instead for climet
Results 1 - 3 of 3:
Liquid Tire Chain: No traction? Snow problem!
http://www.foxnews.com/auto/2017/11/10/liquid-tire-chain-no-traction-snow-problem.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/f01785b820cbfa1b8a48e9b5cad779bf.html
... For those living in colder climes, winter presents its share of driving challenges. Once upon a time, General Motors took a...
Sugar, molasses and more: The real deal with sweeteners from around the world
http://www.foxnews.com/food-drink/2017/04/03/sugar-molasses-and-more-real-deal-with-sweeteners-from-around-world.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/a93cc0fb4b082b2658ae5c990606c78b.html
...s sugar, is a perennial grass that grows in tropical and subtropical regions. In colder climes, sugar is derived from sugar beets, taproots with high concentrations of sucrose. And cor...
Sugar, molasses and more: The real deal with sweeteners from around the world
http://www.foxnews.com/food-drink/2017/04/03/sugar-molasses-and-more-real-deal-with-sweeteners-from-around-world.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/ae7e15090ed603ed18c23e96f3591b70.html
...s sugar, is a perennial grass that grows in tropical and subtropical regions. In colder climes, sugar is derived from sugar beets, taproots with high concentrations of sucrose. And cor...
Right Ctrl
ENG 13:32

**4.**

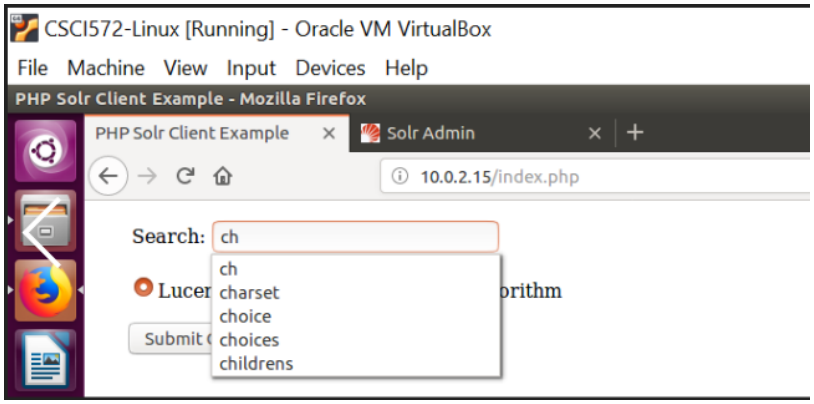

CSCI572-Linux [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
PHP Solr Client Example - Mozilla Firefox
Search your computer
Solr Admin
10.0.2.15/index.php?q=norh+korae&algo=lucene&checkspelling=false
Search: norh korae
Lucene Algorithm
PageRank Algorithm
Submit Query
Showing results for north korea
Search instead for norh korae
Results 1 - 10 of 1472:
Trump on North Korea, from 'Rocket Man' to 'fire and fury'
http://www.foxnews.com/politics/2018/03/12/trump-on-north-korea-from-rocket-man-to-fire-and-fury.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/e7bc52f74e125198d965e604b6f6225b.html
... Print Close Trump on North Korea, from 'Rocket Man' to 'fire and fury' By Ka...
Trump on North Korea, from 'Rocket Man' to 'fire and fury'
http://www.foxnews.com/politics/2018/03/12/trump-on-north-korea-from-rocket-man-to-fire-and-fury.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/68f62b8b83dbe2ea71e5a6eccd050d07.html
... Print Close Trump on North Korea, from 'Rocket Man' to 'fire and fury' By Ka...
North Korea willing to hold talks with US, ex-spy chief says
http://www.foxnews.com/world/2018/02/25/north-korea-willing-to-hold-talks-with-us-ex-spy-chief-says.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/e9a527c5f44f029d71522d19ce606290.html
... Print Close North Korea willing to hold talks with US, ex-spy chief says By Travis Feds...
Trump on North Korea, from 'Rocket Man' to 'fire and fury'
http://www.foxnews.com/politics/2018/03/12/trump-on-north-korea-from-rocket-man-to-fire-and-fury.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/f76dbc8eb0a28424e79acca3727772ae.html
... North Korea Published March 12, 2018 ...
Reporter's Notebook: 'Missed opportunity' for US-North Korea talks during Olympic Games
http://www.foxnews.com/world/2018/02/27/reporters-notebook-north-korea-missed-opportunity-to-talk-with-us-during-olympic-games.html
/home/anumysore5/hw4/solr-7.2.1/crawl_data/HTML_files/HTML_files/593517db0942749ef3cc44e3f7b877f7.html
Right Ctrl
ENG 13:30

**5.**

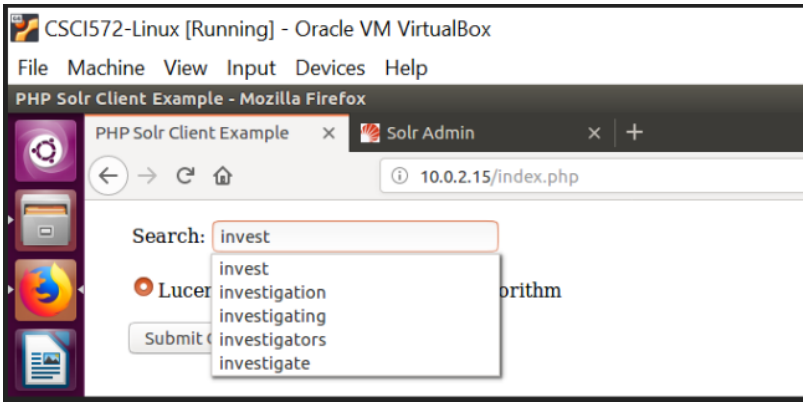



## Five examples of auto-completion:

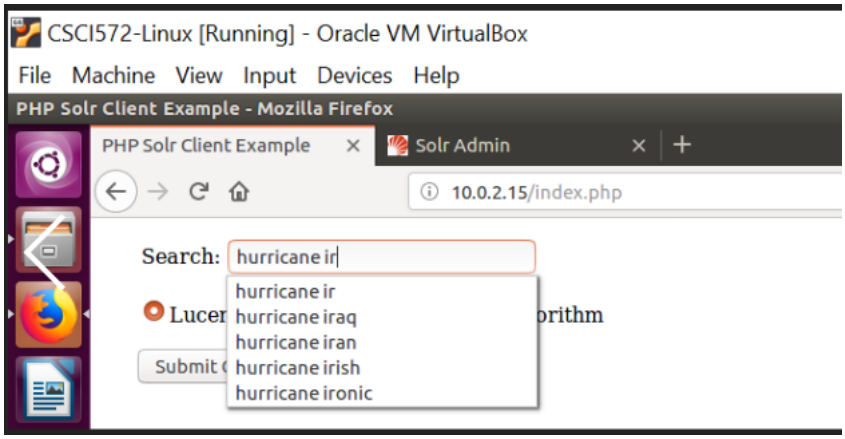
**1.**



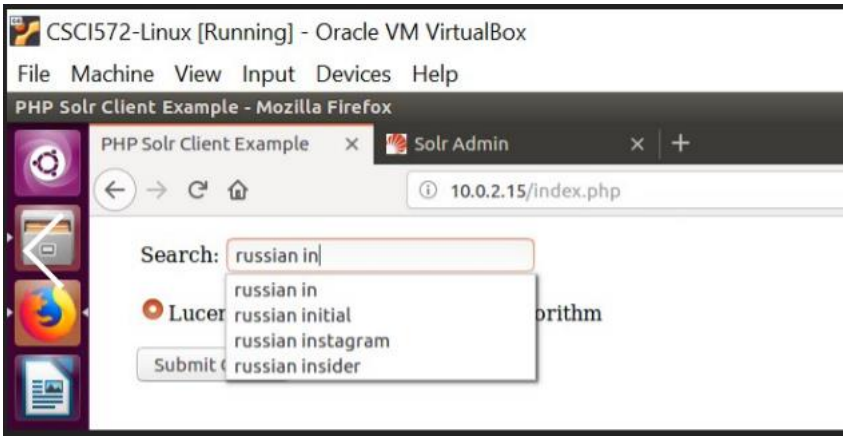

**2.**



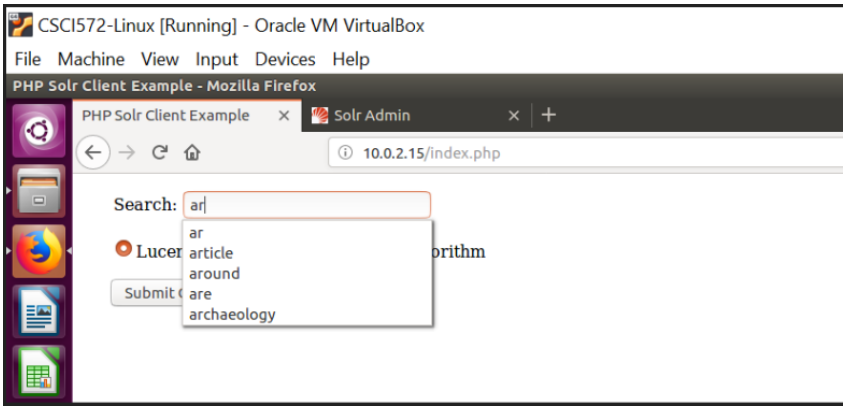
**3.**


CSCI572-Linux [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
PHP Solr Client Example - Mozilla Firefox
PHP Solr Client Example
Solr Admin
10.0.2.15/index.php
Search: hurricane ir
hurricane ir
hurricane iraq
hurricane iran
hurricane irish
hurricane ironic
Submit

**4.**


CSCI572-Linux [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
PHP Solr Client Example - Mozilla Firefox
PHP Solr Client Example
Solr Admin
10.0.2.15/index.php
Search: russian in
russian in
russian initial
russian instagram
russian insider
Submit

**5.**


CSCI572-Linux [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
PHP Solr Client Example - Mozilla Firefox
PHP Solr Client Example
Solr Admin
10.0.2.15/index.php
Search: ar
ar
article
around
are
archaeology
Submit