**CSCI 572 Assignment 5**

**Neekita Salvankar**

**USC ID : 8591-3366-93**

These are a few steps I followed to do this assignment:

1. **To implement Spell Check**

1. To use Peter Norvig’s code for Spelling corrector, I first needed the big.txt file which consists of set of words to calculate edit distance.
2. I wrote a java program for the same using the Eclipse IDE, imported required jar files and the program created the big.txt file. To download and parse the html files to big.txt file, I used htmlparser
3. I used Peter Norvig code , gave it the big.txt file as input to generate the serailized\_dictionary.txt file
4. I felt the most important part of Peter Norvig code is the correct() function, it generates a set of candidates which are at a minimum edit distance from the query term. Out of all these, the term closest to the main query term would get selected to be displayed.

2. **To implement auto complete**

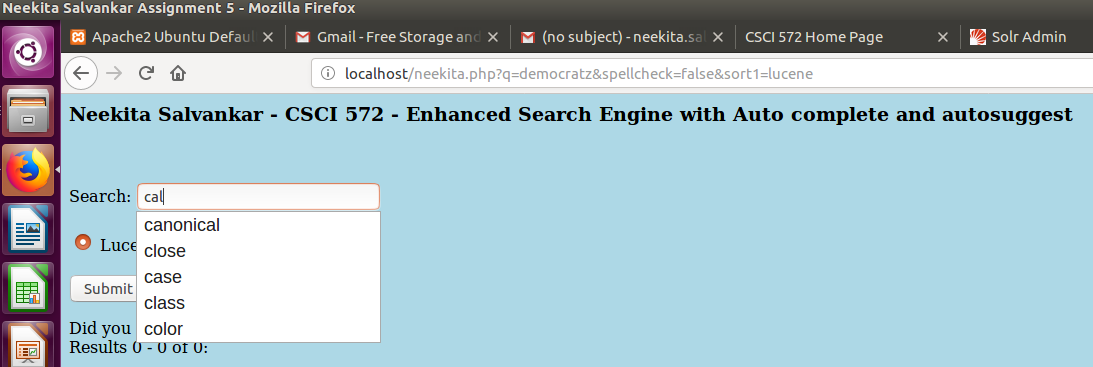
1. Firstly, as mentioned in the documentation, I made changes to the solrconfig file, that is, added the search component. I also added a requesthandler to the same file.
2. I added a JQuery snippet to enable autocomplete in my previous assignment of Lucene. 2 concepts.
3. The main part of implementation were these 2 : a variable to store term suggested by solr which will be closest to query term (as prefix) and a variable to specify the output is json (wt=JSON)
4. Thus the nearest terms are retrieved in JSON format and can be stored. On looping on the maximum count, we are able to view all the suggestions

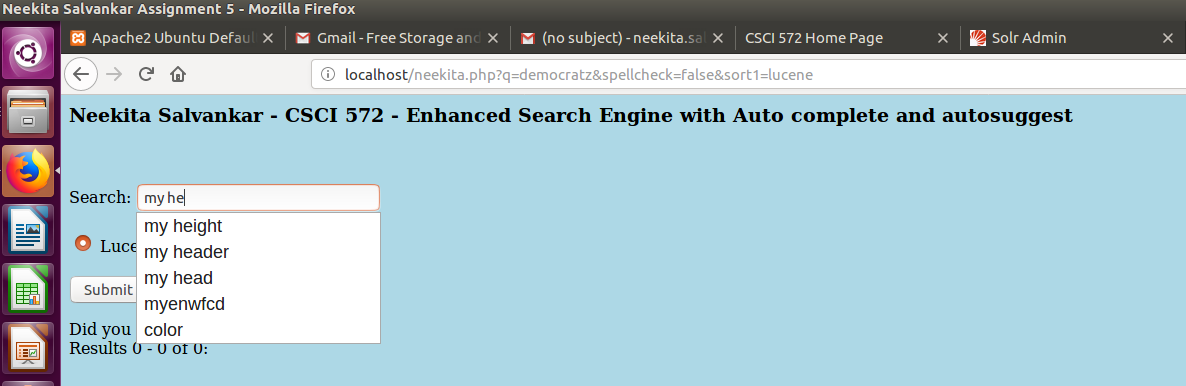
3. **To implement snippets**

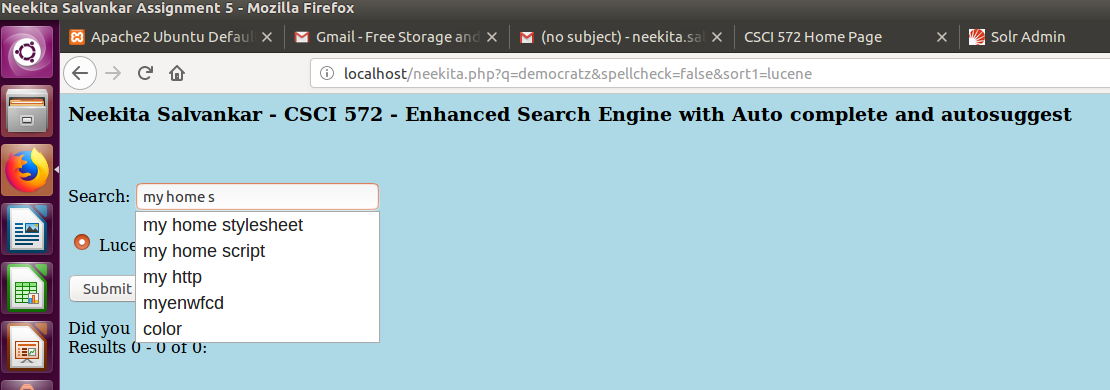
1. I used the simple\_htm\_dom.php file for this feature
2. I extracted contents of each and every file. The content was divided into 2 parts - sentences and words. Special character were removed from these.
3. Then I created loops to extract those parts in the content where our query term occurs and added ellipses in front and behind it. Also bolded the query.
4. If no snippet found, returned 0.

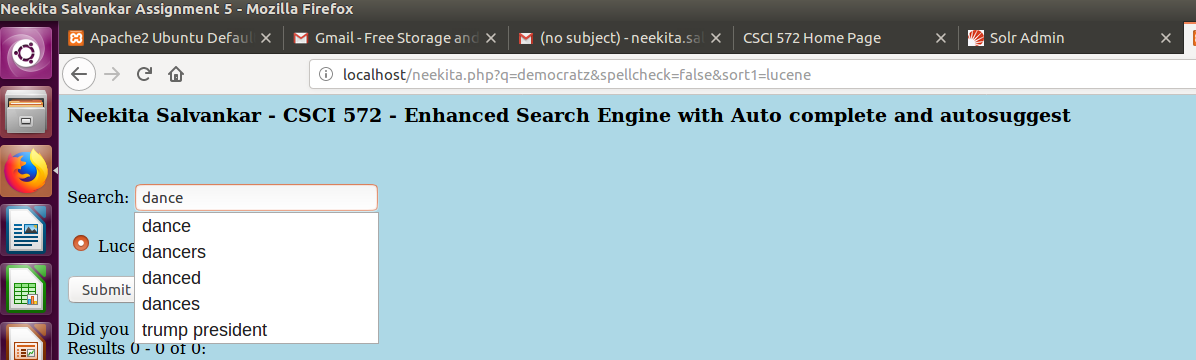
**4. Analysis( Screenshots cropped for better visibility)**

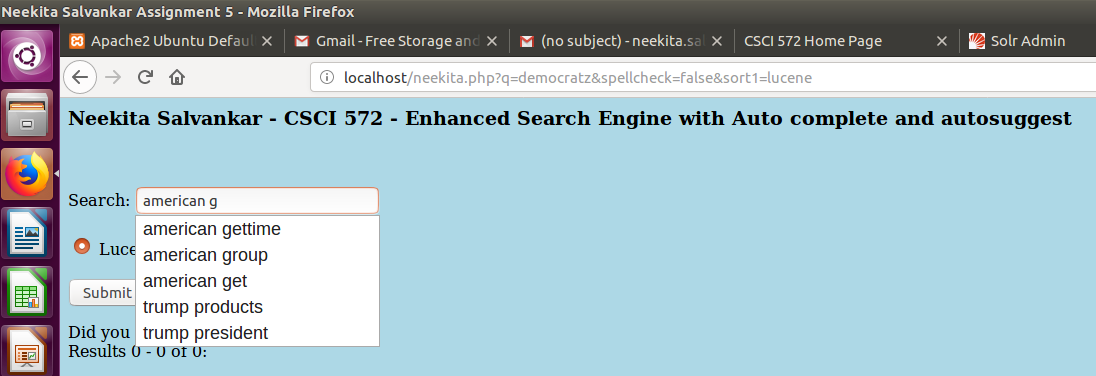
**Autocomplete examples**



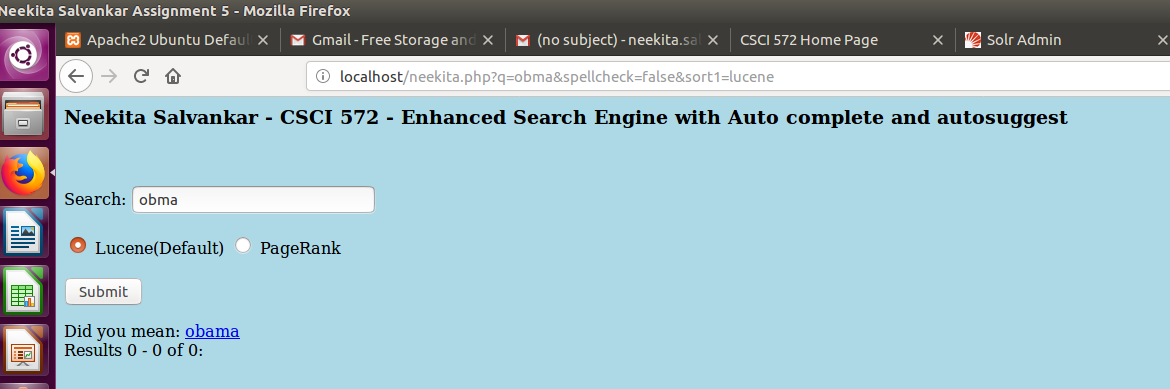


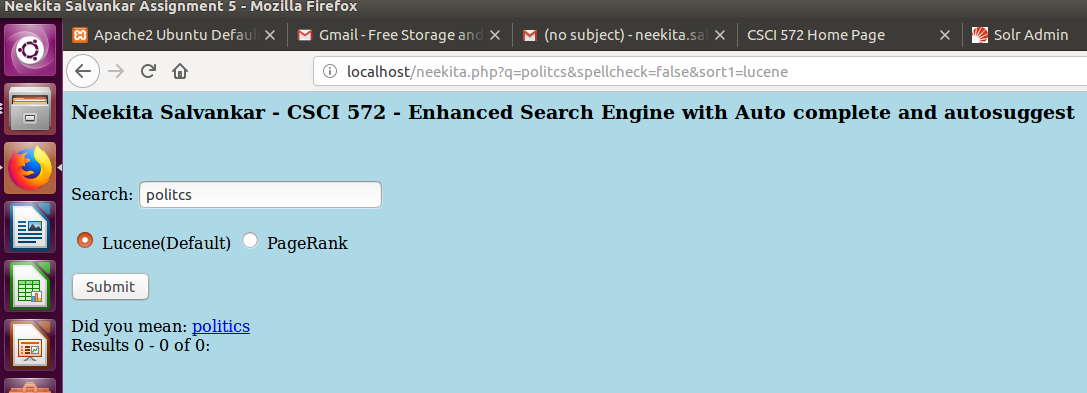


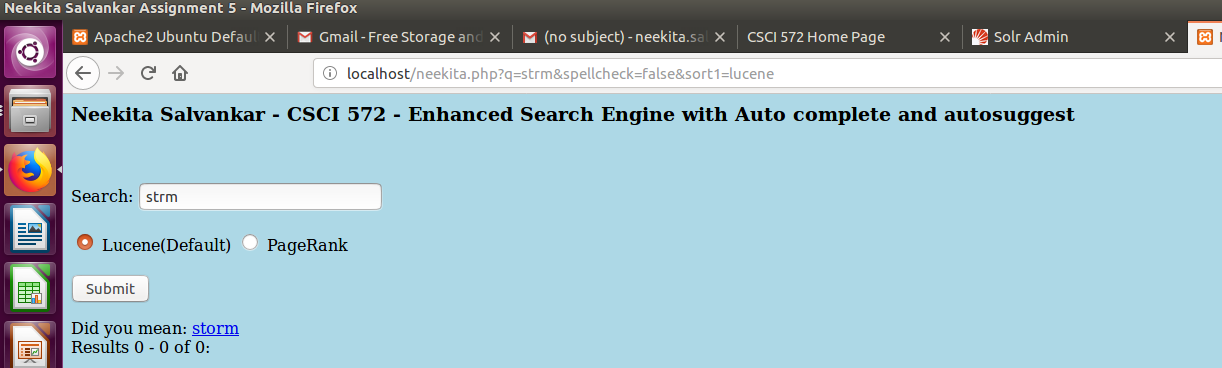


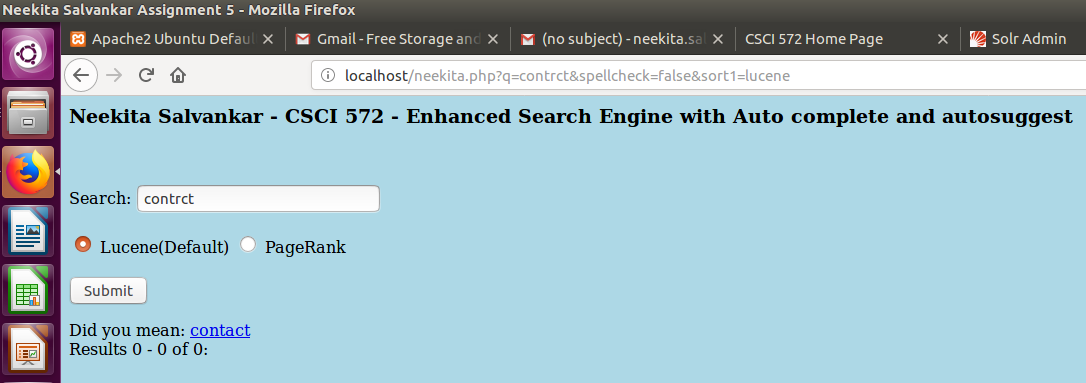


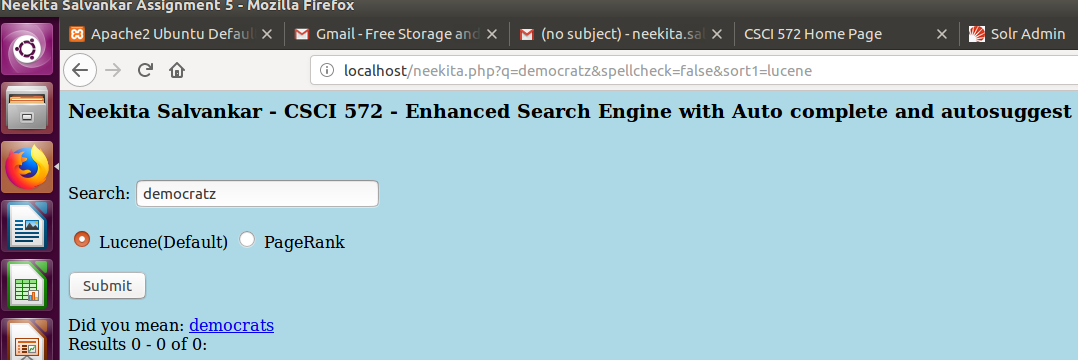
**Suggest examples**:











5. Final output:

