1. Steps followed to complete the assignment:

Spelling Correction –

I have used **Norvig's Spell Corrector PHP version**.

For the word list, I used big.txt file available on Norvig's website containing many common words. Then I extracted words from the files downloaded by crawling to get specific terms for my school. I used TIKA tool and java in eclipse to parse the contents of file and extract words. I appended these words to the big.txt file.

Instead of sending user enter query directly to solr I was sending the (single or multiple words) query to the SpellCorrector tool (the Norvig's PHP program). This program checks each term of the query for the most probable correct word, by choosing the word of highest frequency from the dictionary.

Then I used the spell corrected version of all query terms to send the search request to the Solr and displayed the results received.

Autocomplete –

I used **Solr's 'suggest' feature** to get a list of suggestions for a term. I did the required changes in the configuration files of my core in Solr. I also did changes in Service.php by adding a function so that it handles suggest requests the same way it does for search requests because without these changes I was getting zero suggestions.

For each term sent to solr it returns a list of words being arranged in descending order by weight.

In the User Interface, as soon as a character is typed in the textbox, the onkeyup event makes a call to a javascript function.

Inside the function I am sending an Ajax call to re-direct to an intermediate server side php code which is communicating with solr in order to send a suggest request. For each sequence of characters typed, a suggest request is sent to solr to get a list of suggestions.

The set of words returned are extracted by parsing the response received in xml form from solr. Then added these set of words to a json object and returned it to the User Interface (UI).

The UI code takes each term from the json object after decoding it then two other techniques are applied – (1) stop word removal and (2) stemming.

I downloaded a list of more than 450 stop words from internet and then I have removed the stop words from suggestions by comparing the suggestions received with the stop words list.

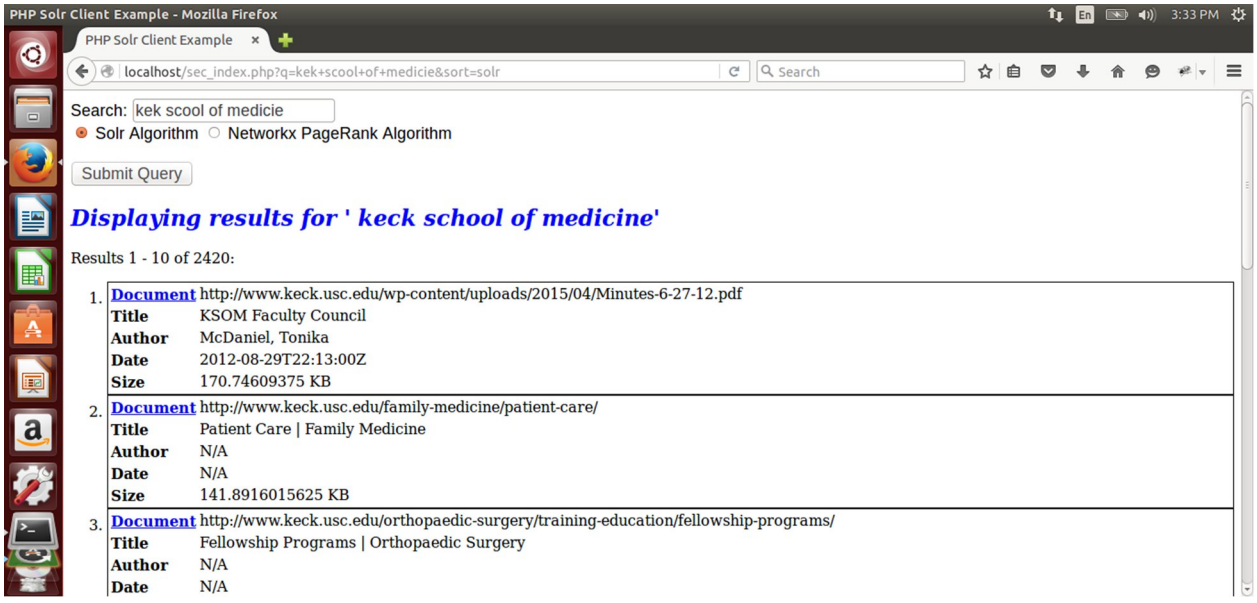
Then I applied stemming on the list of suggestions received. I used Porter's Algorithm available online in javascript version. This algorithm takes a word and returns root form of the word. If the root form is not already present in the final list of suggestions which I'm using to display then only that word was added to the list of final suggestions otherwise not. The Porter's stemming was returning some weird words like cacul for calculations, so I have removed such words from suggestions.

Once these two techniques are applied and a final list of suggestions is ready, I added a jquery based tool to show the suggestions and give users the option to select one of the suggestions and see the search results.

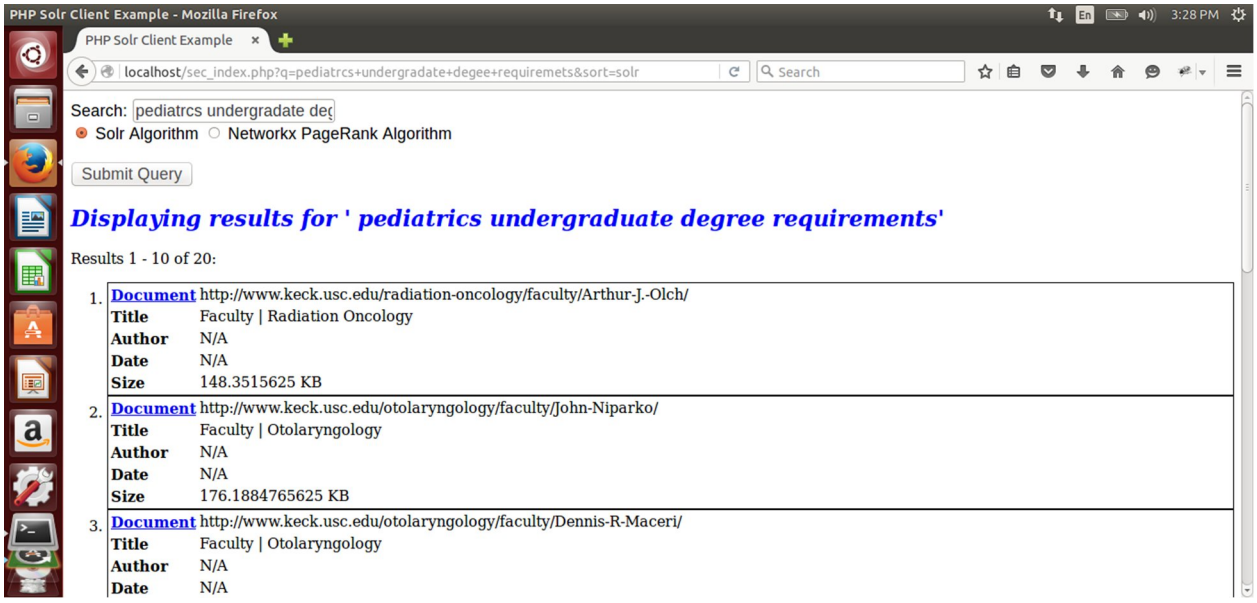
2. Examples:

Spelling Correction –

1. When I type the query ***wrk***, it shows results for ***work.***
2. When I type the query ***neurolgy resarch***, it shows results for ***neurology research.***
3. When I type the query ***kek scool of medicie***, it shows results for ***keck school of medicine.***

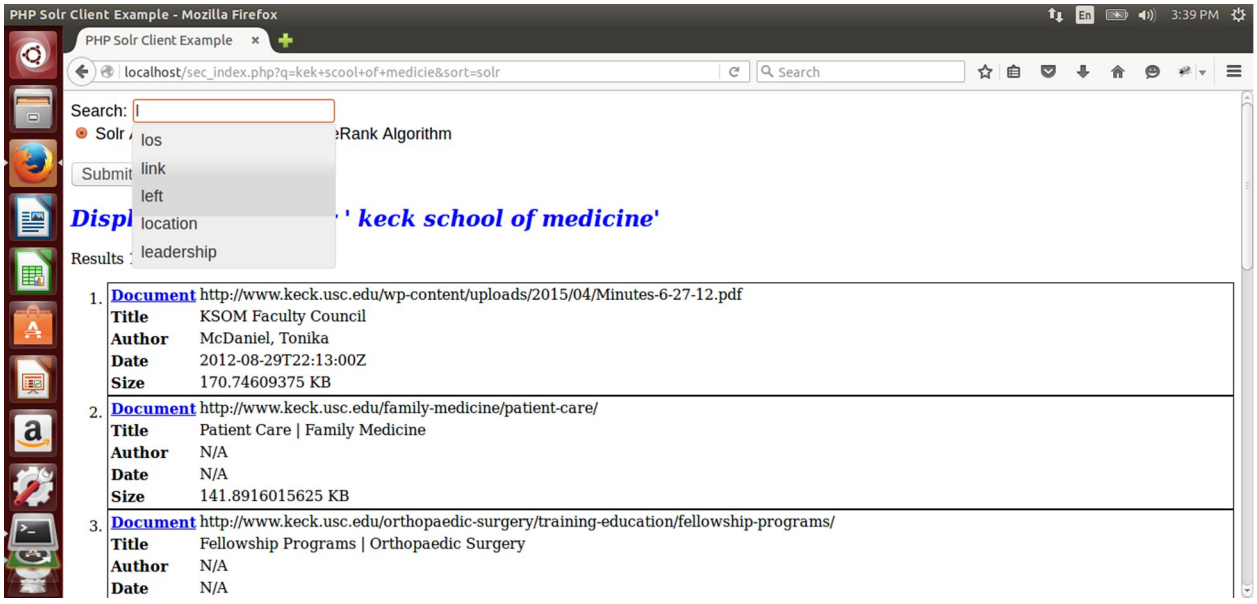


4. When I type the query ***radiolgy deprtment ophthalmoloy faculy***, it shows results for ***radiology department ophthalmology faculty.***
5. When I type the query ***pediatrcs undergradate degee requiremets***, it shows results for ***pediatrics undergraduate degree requirements.***



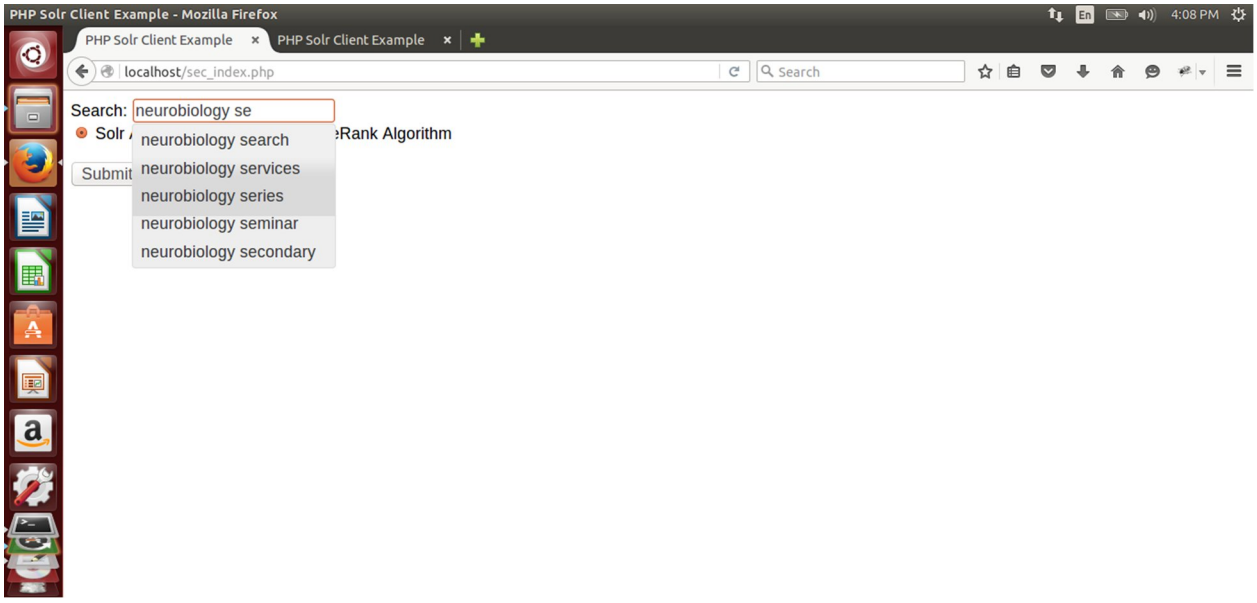
Autocomplete –

1. When I type ***l*** , I get auto suggestions as ***los, link, left, location, leadership.***



2. When I type ***me*** , I get auto suggestions as ***medicine, medical, media, med, meet.***
3. When I type ***p*** , I get auto suggestions as ***parsed, page, program, post, patient.***

4. When I type ***com*** , I get auto suggestions as ***community, comprehensive, compliance, commitee, company.***
5. When I type ***neur*** , I get auto suggestions as ***neurology, neurobiology, neuroimaging, neurogenetic, neurosurgery.***
6. When I type ***neurobiology***, I get auto suggestions as ***neurobiology search, neurobiology services, neurobiology series, neurobiology seminar, neurobiology secondary.***



The search results after selecting the suggestion ***neurobiology seminar*** are: