**1. Spell Check:**

i. Generate big.txt file to give input to Norvig’s spell correct php program

This file is generated using a java code (Jsoup library) which reads the html file and does a simple regex matching allowing alphabets only followed by a “.”.

ii. Give this file as a input to Norvigs spell correct code.

iii. In my main php, I call this function of Norvig to tell me whether the input provided by the user is of a valid spelling from the corpus or not. If not, it provides me with the corrected spelling.

**2. AutoComplete/ Suggest:**

i. Make the necessary changes in solrconfig.xml file to add suggest component, thus telling solr that we will be using the suggest component.

ii. Once the config is in place, reload the cores config

iii. Add a js file to php (have to enable CORS) for the same.

iv. Have created a file autosuggest.js which uses ajax to pop up a suggestion box as soon as a word is keyed in, this file calls the suggest query of solr.

**3. Snippet**

i. Once, the results are given back by solr, examine each and every returned html file.

ii. Using DomDocument and file\_get\_contents to read the files and parse the html file,

first look for meta tags to get the description and append the result in a text variable.

iii. Once all tags related to meta are completed, check for paragraph tags and extract their contents.

iv. Append the content to the same text variable and then using map reduce get the count of matching query in a sentence.

v. At any point if we get an exact match for the query, break the loop and display the sentence as the snippet.

vi. Get the sentence with the max match and display this sentence as the snippet.

**5 Queries with spell error**

|  |  |
| --- | --- |
| **Keyed In** | **Auto Correct** |
| Donald trmp | Donald Trump |
| Brext | Brexit |
| Gallilio | galliano |
| Spce | Space |
| Anta | Ante |

**5 Queries with Auto Complete**

|  |  |
| --- | --- |
| **Keyed In** | **Suggestions** |
| Sna | Snarled, Snapchat, Snap, snakes,snake |
| Nas | Nasa, nasdaq, nasty, nasa’s |
| Anthr | Anterior, Anthony, Anteriories, Anthem, Anthems |
| Basi | Baseball, baskeball, base |
| Cric | Crisis, crime, crimen, critisized, crimes |