**Homework 5 – Adding Spell Checking, AutoComplete and Snippets to Your Search Engine**

**1. Steps you followed to complete this assignment. Include the details of**

**what tools and techniques used to implement spelling correction and**

**autocomplete:**

Step 1: Created Big.txt:

• Extracted text from Latimes crawled data sets using Tika parser written in Java.

and added the text into Big.txt, which is the vocabulary for spell check

and auto suggest.

Step 2: Spell Correction:

• Used Norvig’s spell check php version code to implement spell

correction.

• When a user inputs an incorrect query and submits the page, the

program checks to see if it’s a valid word or not. If it’s a valid word

corresponding results are displayed else a phrase with “Did you

mean?” and the keyword gets displayed. On click of new keyword the

corresponding results are displayed.

• When it’s an incorrectly spelt word it will be sent to the Norvig spell

corrector and the correct word will be returned as a suggestion to continue with.

Step 3: Auto complete:

• Called the autocomplete php from the main page and sent every character individually to have it implemented.

• This php called the url to extract suggestions

http://localhost:8983/solr/myexample/suggest?wt=json&indent=on&q="

and sent it back to the home page.

• The results of auto complete are thus displayed in a dropdown list.

• On select of a key word the corresponding results are displayed.

**2. Analysis of the results: In this you should provide examples of misspelled**

**terms that are correctly handled by your spelling correction program. You**

**should also provide some examples of auto completion.**

**SPELL CORRECTION:**

|  |  |
| --- | --- |
| **Incorrect Word** | **Correct Word** |
| Rusasi | Russia |
| Donal Trmp | Donald Trump |
| Snapcht | Snapchat |
| Brext | Brexit |
| Illegal immgraton | Illegal immigration |

**AUTO-COMPLETE:**

|  |  |
| --- | --- |
| **Prefix** | **Autocompletion** |
| **Cali** | **Cali**  **Calc**  **Click**  **Called**  **California** |
| **Trum** | **Trum**  **True**  **Trump**  **Trump’s**  **Thumbnail** |
| **Immi** | **Immigration**  **Immigrants**  **Immediately**  **Immigrant** |
| **We** | **We**  **Were**  **Week**  **Weather**  **Wednesday** |
| **Rus** | **Rus**  **Rss**  **Russia**  **Reskin**  **Resourcename** |