**CSCI 572 Information Retrieval and Web Search Engines**

**Homework 5: Enhancing Your Search Engine**

**Name : Anubhav Jindal**

**USC ID : 5963113610**

**Steps followed to complete the Assignment:**

1. **Spelling Correction:**
   1. Downloaded Apache Tika Jar and included it in the build path of my Java project.
   2. Wrote Java code to extract the Text from the HTML files(bigtxt.java) and write to big.txt
   3. Downloaded Peter Norvig’s PHP code for Spelling Correction after signing up (SpellCorrector.php)
   4. I provided the path of my big.txt file to Peter’s code.
   5. To execute the SpellCorrector.php, I included the file in my index.php file and made a call to the correct () function.
   6. After the execution, a file serialized\_dictionary.txt is generated which is used for spelling correction.
2. **Autocomplete:**
   1. Changes were made to Solrconfig.xml as specified in the document provided as a part of the assignment.
   2. I reused the code from Assignment 4 (index.php) and made modifications to it to include Autocomplete function.
   3. The tutorial and code mention on <http://api.jqueryui.com/autocomplete/> was used for reference.
   4. The suggested terms, for any Query term were parsed (json) and stored in a variable.
   5. All the suggestions are iterated through and shown to the user as a drop-down list.
3. **Snippet:**
   1. In order to parse the HTML pages in PHP, I used simple\_html\_dom.php and included it in my index.php file.
   2. The retrieved text was divided into sentences and words.
   3. The rule followed to obtain the snippet was:
      1. For a multi-term query, obtain a sentence in which all terms occur together if possible.
      2. Else the sentence containing the words is obtained if possible.
      3. Else the first sentence containing one of the terms is obtained if possible
      4. Else look into the description associated for the query words.
      5. Else N/A is returned to the user.
   4. The query terms were put in bold within the snipped for easy identification.

**Analysis of the Results:**

1. **Examples of Misspelled words:**
   1. **Acceptible**   
      A screenshot of a cell phone

      Description automatically generated
   2. **Fcbook**  
      A screenshot of a social media post

      Description automatically generated
2. **Donld TrmpA screenshot of a cell phone

   Description automatically generated**
3. **Bannna**

A screenshot of a social media post

Description automatically generated

1. **Amazn**

**A screenshot of a social media post

Description automatically generated**

1. **Examples of Autocomplete:**
   1. **Accep**

**A screenshot of a social media post

Description automatically generated**

* 1. **Face**

**A screenshot of a social media post

Description automatically generated**

* 1. **Amaz**

A screenshot of a social media post

Description automatically generated

* 1. **Bara**

**A screenshot of a social media post

Description automatically generated**

* 1. **Dona**

**A screenshot of a social media post

Description automatically generated**