**ASSIGNMENT-5 REPORT**

The following files have been submitted. The main file in which the code resides is the ‘code.php’ file. Further SpellCorrector.php and a.php have been submitted.

**STEP 1:**

**-SPELLING CORRECTION**

1. Started with downloading Norvig’s SpellCorrector.php
2. Further used the Apache Tika to extract data from the html pages
3. HTML pages were from LATimes website.
4. Created ‘big.txt’ and placed it along with SpellCorrector.php in apache htdocs folder.
5. ‘Big.txt’ serves as the dictionary of words used by SpellCorrector.php. When SpellCorrecter.php is used for the first time, it creates serialized\_dictionary.txt file.
6. Created an object of the ‘SpellCorrector’ class and ‘correct’ function in code.php

**-AUTO-COMPLETE**

1. Started with modifying the ‘solrconfig.xml’ file and added SuggestComponent and RequestHandler.
2. Used HTML 5’s datalist tags and XHTTP requests to fetch autocomplete list.
3. Made an Xhttp request and the response was later stored in datalist tag which produced auto suggestion list in search box.

**-SNIPPETS**

1. Used Simple HTML DOM Parser to get html contents of a webpage
2. Matched user entered queries with the extracted contents to display the first occurrence of the string as a snippet. Snippet was printed as first occurrence with 160 characters in it.

**STEP 2: ANALYSIS OF RESULTS**

**-Examples of misspelled terms**

1. Misspelled term- immigartion

Corrected term- immigration

1. Misspelled term- donld trump

Corrected term- donald trump

1. Misspelled term- connectavity

Corrected term- connectivity

1. Misspelled term- californai

Corrected term- California

1. Misspelled term- Twotter

Corrected term- Twitter

**-Examples of autocomplete**

1. Term – a

Auto complete- author,about,apple,advertisement

1. Term- w

Auto complete- Wednesday, width, weather

1. Term- donal

Auto complete-donal, donalsons, Donald

1. Term- e

Auto complete- entertainment, encoding, element, elect

1. Term- pol

Auto complete- politics, policy, possible

**Overall**: The assignment involved testing both, auto-complete and spelling-correction for the search engine.