Assignment 5 Report

Jing Peng 9802767543

1. Steps to complete the assignment

- Preparation
  Write java program to generate **big.txt** from FOX_NEWS html files using Tika parser package. Save it for later use as vocabulary for spelling correction and auto-completion.
- Spelling correction
  Spelling correction is implemented by Norvig's spell correction program php version.
  Download file **SpellCorrector.php** from given link and included it into my **index.php** file. Every time when user input a query, the program will check it with previously generated vocabulary to see if the query is a valid term. If not, corrector will return correction. Then my code will return message "Did you mean" + correction to user.
- Auto-completion
  Add prefix and suffix to user input to form query for auto-completion.
  Call solr back-end with query, then get its return result, which includes auto-completion suggestions. Display suggestions in a dropdown list under input box.
- Snippet
  Split html content by dot and query by whitespace. Include **simple_html_dom.php** to ensure only plaintext will be extracted from html files.
  Check every sentence in html file to see if it contains one or more queries. If so, replace matched string by add <b></b> at its two sides to implement highlight.
  Save modified matched sentence as snippet and later print it to user.

2. Examples

- Five spelling correction examples
  1) Input: informatiom, correction: information
  2) Input: lunkh, correction: lunch
  3) Input: computre, correction: computer

    4) Input: helllo, correction: hello
    5) Input: asdignment, correction: assignment

- Five auto-completion examples
    1) Input: charac, auto-completion: character, character’s, characterisation, characteristic, characteristically
    2) Input: bitco, auto-completion: bitcoin, bitcoin’s, bitcoin.amp.html, bitcoin.html, bitcoin.print.html
    3) Input: lunc, auto-completion: lunch, lunchbox, luncheon, lunches, lunchroom
    4) Input: happ, auto-completion: happen, happen.amp.html, happen.html, happen.print.html, happened
    5) Input: schoo, auto-completion: school, school’s, school.amp.html, school.html, school.jpg