IR ASSIGNMENT 5 REPORT

Name : SHREYA SUDHIR WAGLE

USC ID : 3298824698

Website : Washington Post

Steps followed to implement spelling correction:

- Used a Java code that functioned as an HTML parser. Its basic aim was to parse the large amount of data under Washington Post website and to generate a big.txt that contains all the content of the HTML files (by the name HtmlParse.java)
- Downloaded the spell corrector program from an online source (by the name SpellCorrector.php) which is an implementation based on Peter Norvig's algorithm
- Included this code in the PHP code for the UI generation
- Tools used – Eclipse (IDE for Java) to parse HTML files, Peter Norvig's SpellCorrector program for handling misspelt words and a final UI.php which was carried forward from Assignment 4 based on PageRank algorithm

Steps followed to implement autosuggest:

- Updated the solrconfig.xml file to include changes related to search component and request handler
- Reloaded core in Solr for the changes to get reflected
- Included JQuery in the PHP code that elaborated the UI
- Auto-suggest feature was added using JQuery such that suggestions are displayed in a drop-down fashion as soon as the user begins typing and the suggestions get more streamlined as the user gives in more input
- Excluded stopwords to improve the auto-suggest feature

Steps followed to implement snippet generation:

- Included simple_html_dom.php to extract HTML content
- Generated snippets for websites (if present) and also ensured that the snippets would highlight the query terms from the user's search
- Added regex to remove special characters

Analysis of the results:

- Spelling correction examples:
    - Misspelt word: dinnre

      Word after spelling correction: dinner
    - Misspelt word: dicsuss

      Word after spelling correction: discuss
    - Misspelt word: lugage

      Word after spelling correction: luggage
    - Misspelt word: swich

      Word after spelling correction: switch
    - Misspelt word: tommorrow

      Word after spelling correction: tomorrow

- Auto-completion examples:
    - Characters typed: su

      Suggestions prompted: support, supporting, summary, subscribe, state
    - Characters typed: cali

      Suggestions prompted: California, click, client, call, called
    - Characters typed: phot

      Suggestions prompted: photo, photos, phone, prototype, post
    - Characters typed: din

      Suggestions prompted: dine, dinner, dies, display, din
    - Characters typed: bow

      Suggestions prompted: bow, body, bottom, border, boolean