**CS572 Information Retrieval and Web Search Engines**

**Assignment 5**

**Submission date: 11/29/2017**

**Monika Devanga Ravi**

**3448881178**

**Part 1: Steps followed to complete this assignment**

This HW is built on top of HW4, which already had the facility of searching and displaying top 10 results from Solr Lucene and Page rank algorithms.

**Tools used:** Ubuntu, Eclipse IDE, Apache, Solr 7.0.1.

**Languages used:** JAVA, PHP, AJAX

**Steps 1: For “SPELL CHECK” implementation**

* I used the Norwigs’s Spell check algorithm in PHP which was suggested in homework description named “SpellCorrector.php” (this file is included in the submission)
* This algorithm uses external dictionary of words i.e., big.txt file for calculating edit distance and appropriately correct the entered misspelled word.
* **“Big.txt”** file was created using JAVA program and the Apache Tika Parser external JAR that was suggested in the homework description.
* This JAVA program works by scanning and parsing all the html files which was given to us in HW4 and creates a lexicon.
* I integrated Spell check feature into the client side PHP code(main.php) which was created during HW4.
* When “Spell corrector” algorithm was ran for the first time, “Serialized dictionary” file was created. This Serialized dictionary file is used by SpellCorrector.php program.
* Ex: if u misspell the word in the query. The correct spelling based on Norwig’s spell check algorithm was displayed.
* The output contains 2 links, along with the result. First link shows the result for correct spelling i.e., Showing results for facebook. Second link showing link to search the entered query i.e., Search instead for facbok. Below is the screenshot of how spell check works for the word “facebook”.

Query : facbok



Figure 1: Spell check example

**Step 2: Implementing autosuggest feature**

* Solr includes the feature for suggesting, but by default configuration is not done.
* To configure suggesting feature in Solr, I modified SolrConfig.xml by adding solr.SuggestComponent and its attributes. To handle suggest functionality along with other request, a request handler named solr.SearchHandler was also added to the same xml as mentioned in the homework description.
* AJAX was used in main.php file in order to show the suggestions to the user at real time.
* JAVASCRIPT along with AJAX gets the query entered in the search box at real time and sends the /suggest request to Solr.
* Fuzzy lookup factory and prefix matching is used by Solr to lookup for the words to suggest and list of suggestions is returned to the user in JSON format.
* JAVASCRIPT that is included in main.php file will parse the JSON suggestion result and suggestion is shown to the user as a dropdown with the same prefix entered by the user.
* Eg: when “ph” is typed in the search box, Solr suggests photos, photo, photo.jpg, photographer as the suggestion. Below is the screenshot for how auto suggest works.

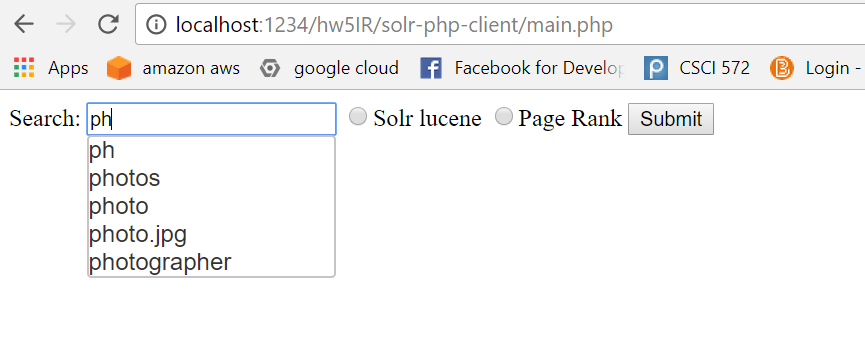


Figure 2: auto suggest example

**Step 3: Implementing snippet.**

* To generate the snippet of 160 characters, the entered query term was searched in the title and description first, if the entered query(multiple or single) term was found then, the word was highlighted and the sentence was displayed as the snippet.
* Suppose query term was not present in the title or description, then the whole body content for <p>, <h1>, <h3>, <a> are used to find the appropriate snippet containing the query word.
* This matching of query term is done for both single word or multiple words.
* If the html page doesn’t contain any word that belongs to the query, then No Snippet is shown.
* Eg: when the query word “Donald trump” was entered in the search box, below is the snippet screenshot for the entered query.

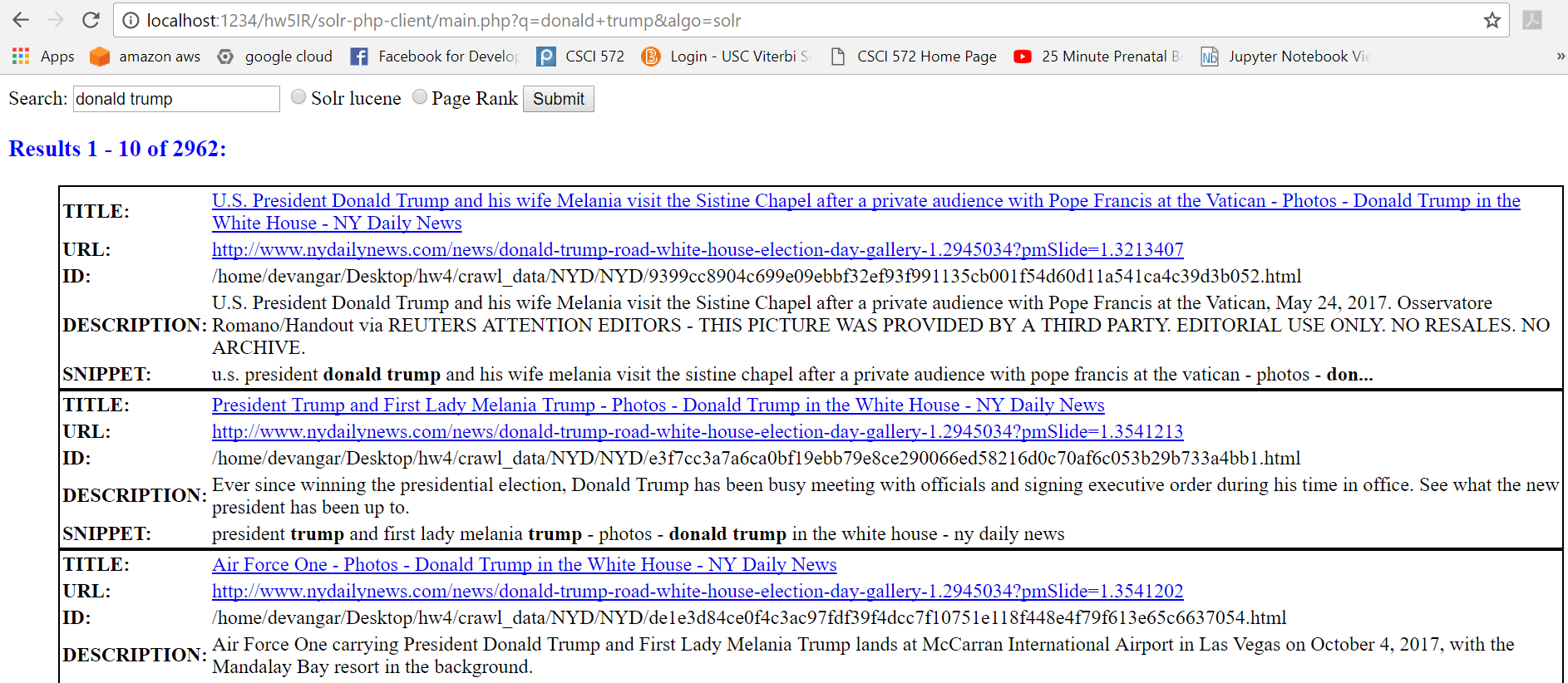


Figure 3: snippet example

**Part 2: Analysis of results**

**Spell Correction example:**

|  |  |  |
| --- | --- | --- |
| No | Query | Corrected Spelling |
| 1 | facbok | facebook |
| 2 | snpcht | snapchat |
| 3 | iphne | iphone |
| 4 | spig | spin |
| 5 | sylvst | stylist |

**Auto complete examples:**

|  |  |  |
| --- | --- | --- |
| No | Query | Suggested spelling |
| 1 | face | face, facebook. fact, facebook.com, faces |
| 2 | north | north, northing, northern, northeast, northwestern |
| 3 | Elon mu | Elon mu, elon music, elon much, elon must, elon murder |
| 4 | hurric | Hurricane, hurricanes, hurrican, hurricanes.jpg, hurricane.jpg |
| 5 | stree | Stree, stream\_content\_type, steam\_size, street, stream |