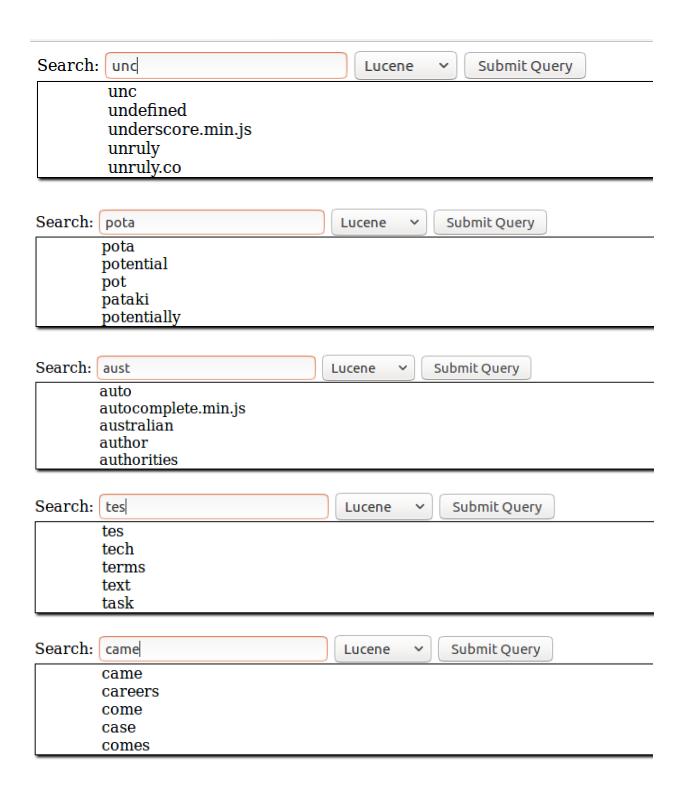
IR HW 4 - Adding Spell Checking, AutoComplete and Snippets to Your Search Engine - Report

Name: Aayush Khanna; USC ID: 8406082994; Site Indexed: NYPost

Steps followed for this assignment -

- 1. **Spellcheck:** Used Peter Norvig's Spell Corrector in php (SpellCorrector.php) to implement the spellcheck feature.
 - a. First, I installed Apache Tika and wrote a Java Program configured over IntelliJ and Maven to interact with Tika.
 - b. This program parsed the HTML files to generate only the contents between the html tags.
 - c. I named this file big.txt and passed it through Spellcheck.php to generate the correct spelling suggestions.
 - d. I then wrote JavaScript to check if the input text needed any corrections and if it did I made it display the correct word using the Spellcheck.php file.
- 2. **Autocomplete:** Configured Solr to give suggestions using the /suggest endpoint using its FuzzyLookupFactory as instructed.
 - a. Wrote custom JavaScript from scratch which displayed the results based on the query results acquired from Solr and displayed it in a drop-down menu with respect to the input box.
- **3. Snippets:** Wrote a JS program to first check if the description had the exact match of the words or any of the word combinations of the query.
 - **a.** If there was no description for the document, I proceeded to make an Ajax call to the respective documented HTML and then proceeded to parse the HTML and removed the script tags to get the content.
 - **b.** I further used regex to get cleaner content, and then parsed the content into an array of sentences.
 - **c.** I then search each element of this array for instances of the query words and display it using Vanilla JS.
 - **d.** I also trim it to 160 characters before displaying.



Search: star wasr Lucene V Submit Query
Did you mean: star was?
Results 1 - 10 of 10679:
1. Title: The kid from E.T. is all grown up
Link: https://nypost.com/2013/11/08/betrayal-star-henry-thomas-sheds-child-star-roots/
Search: norh korae Lucene Submit Query
Did you mean: <u>norh korea</u> ?
Results 1 - 1 of 1:
1. Title: April 17, 2010 New York Post
Search: [Ibron james Lucene V Submit Query
Did you mean: <u>lebron james</u> ?
Results 1 - 10 of 4595:
1. Title: James Harden involved in alleged phone-throwing nightclub fight
Search: dnolad tramp Lucene V Submit Query
Did you mean: <u>donald tramp</u> ?
Results 1 - 10 of 40:
1. Title: April 16, 2017 New York Post
Search: potat Lucene V Submit Query
Did you mean: potato?
Results 0 - 0 of 0: