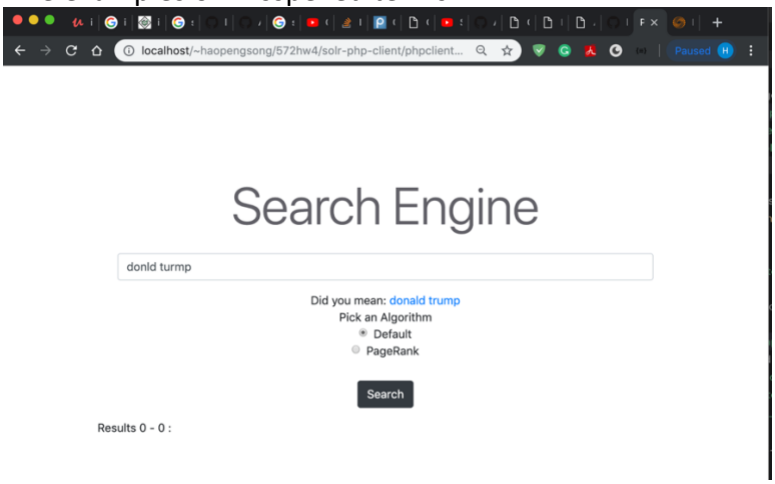
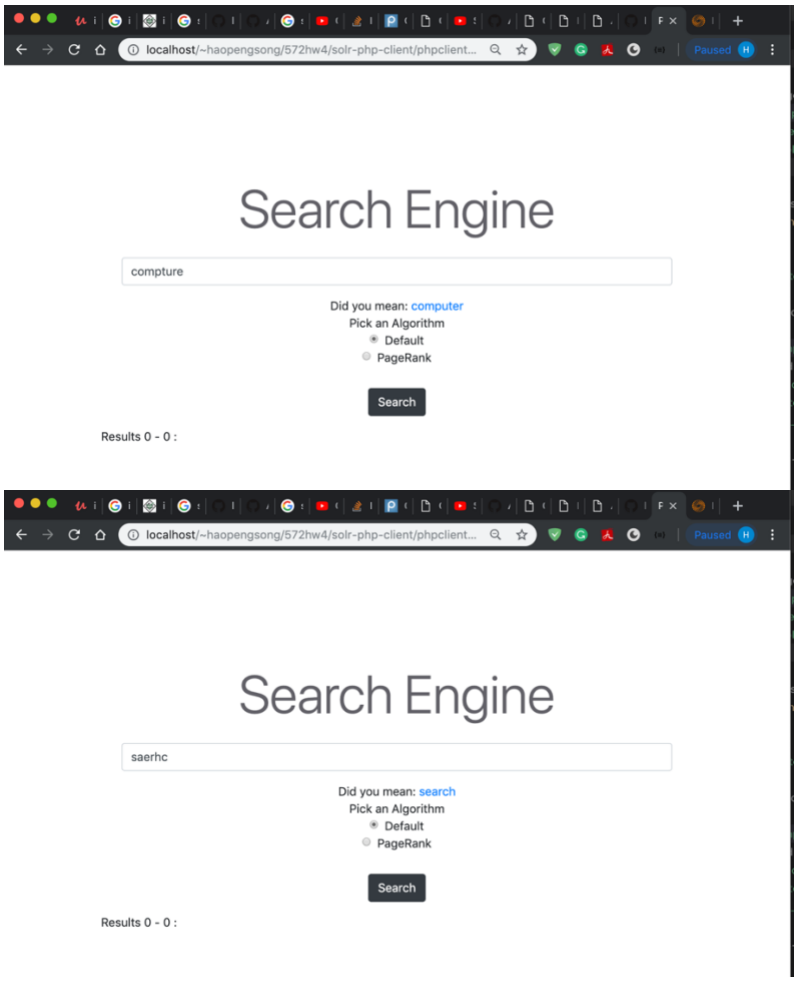
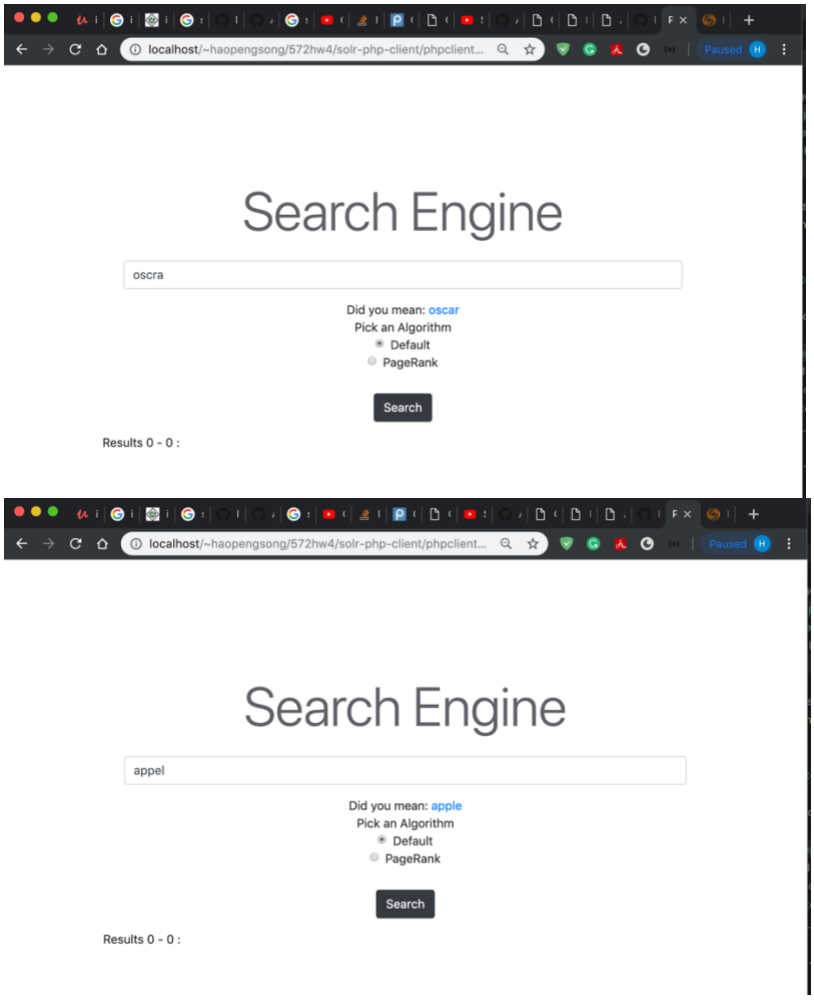
CSCI 572 HW5 Report
Haopeng Song
haopengs@usc.edu

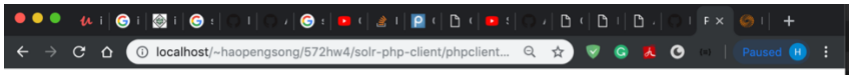
1. Steps followed to complete this assignment.
    a. For **autocompletion**. I first configured Solr to enable auto suggests by editing the solrconfig.xml. I added a searchComponent as well as a requestHandler corresponding to the searchComponent.
    b. After setting up Solr, I used JQueryUI to implement the autocomplete feature on the frontend. I used the autocomplete UI module to display the suggested words returned from Solr. The module has an asynchronous REST function to handle the response from the server. In my implementation, each time the function is triggered it makes a request to Solr. The response will be parsed using JSON parser and returned to the frontend as an array of words. I also choose the minimum trigger characters length to be one in order to achieve the desired performance.
    c. For **spelling correction**, I used the php version of Norvig's spell checker. I included the php file within the frontend.
    d. I used Apache Tika to parse the html file and generated a dictionary of words (big.txt) provided to the spell checker as training data.
    e. Each time the user makes a search request, the query will first go through the spell checker to make sure it only contains valid words. If the query contains misspelled words, the correction will be shown as clickable texts under the input field. Also, the results for misspelled query will still be displayed. Upon clicking the correct word, the page will refresh and display the new results.
    f. For generating **snippets**, I used regular expression to match the text content within the response html file. If there is a match, the snippet field will show sentences with the matched query term in a highlighted/italic manner.
2. Analysis of the results
    a. Five examples of misspelled terms:




localhost/~haopengsong/572hw4/solr-php-client/phpclient...
Paused
Search Engine
compture
Did you mean: computer
Pick an Algorithm
Default
PageRank
Search
Results 0 - 0 :
localhost/~haopengsong/572hw4/solr-php-client/phpclient...
Paused
Search Engine
saerhc
Did you mean: search
Pick an Algorithm
Default
PageRank
Search
Results 0 - 0 :



b. Five examples of autocomplete:



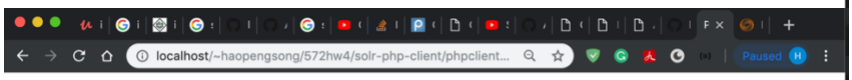
Search Engine

fruit

- fruit
- frustrated
- frqito
- fruitful
- frustration

Search

Results 0 - 0 :



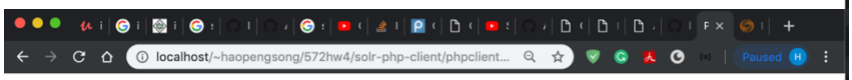
Search Engine

tiger woods

- tiger woods
- tigers wood
- tier words
- tigres wooden
- tierpark woodside

Search

Results 0 - 0 :



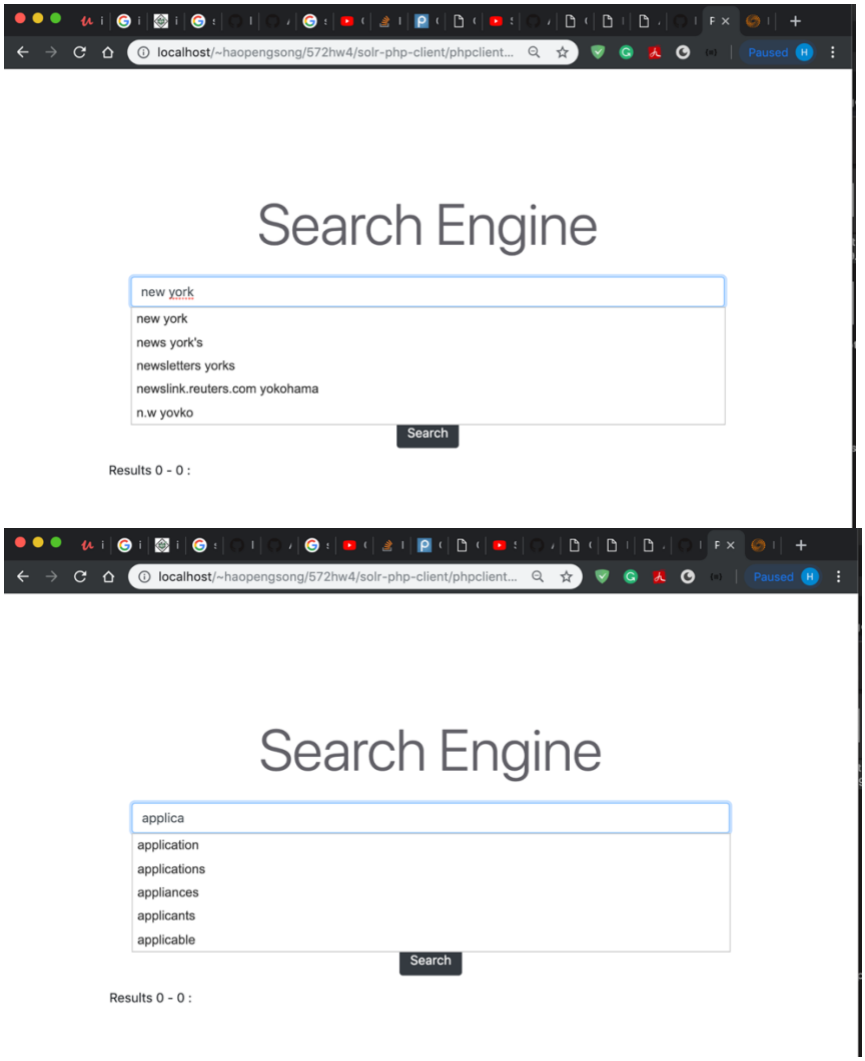
Search Engine

sport

- sport
- shortcut
- sports
- short
- spotify

Search

Results 0 - 0 :


localhost/~haopengsong/572hw4/solr-php-client/phpclient...
Paused
Search Engine
new york
new york
news york's
newsletters yorks
newslink.reuters.com yokohama
n.w yovko
Search
Results 0 - 0 :

localhost/~haopengsong/572hw4/solr-php-client/phpclient...
Paused
Search Engine
applica
application
applications
appliances
applicants
applicable
Search
Results 0 - 0 :