Homework 5: Adding Spell Checking, AutoComplete and Snippets to Your Search Engine

To implement Auto Complete

1. Modified the solrconfig.xml file to add search component - SuggestComponent
2. Next added request handler for the "suggest" search component
3. jQuery ajax() Method is used to call the autocomplete function as the user types in query box and results were limited to 5
4. The results returned in JSON are parsed and displayed in dropdown box

To implement Spell correct

1. Used Apache TIKA , a java library to parse all html files and extract text content
2. The extracted words were stored in "big.txt'
3. For Spell correction, I used Peter Norvig's code in PHP. When the code is first run, it create serialized_dictionary.txt file that store every word with it count
4. Peter Norvig's code returns all the correct words that are within one and two edit distance of incorrect word.
5. For every incorrect query keyword I displayed the correct word as returned by peter Norvig's code by calling correct() function
6. On clicking the corrected word the result for that word will be displayed

To implement Snippets

1. Parsed all the html files using Apache TIKA Java library
2. For every search result, the document id is used to locate the parsed file
3. get_file_contents() is used to extract the content of file
4. TextSnippet.php file is included to generate snippets. The createSnippet method of TextSnippet class object generates snippet
5. The keyword is highlighted and length of snippets is limited to around 160 character
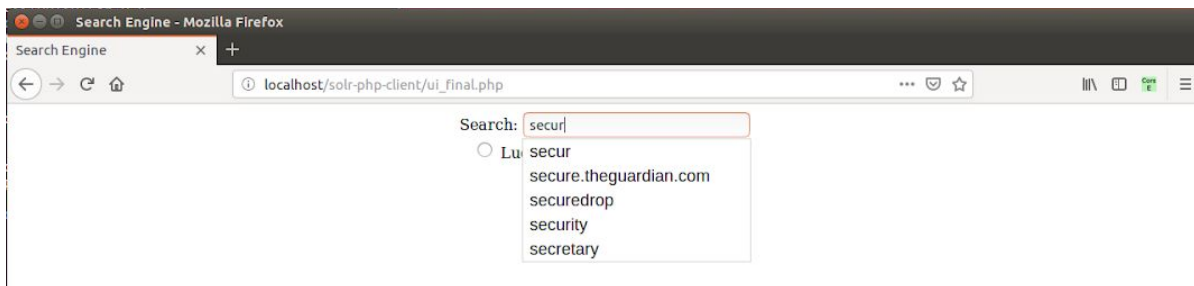6. If snippet was not found, then I displayed "No snippets"

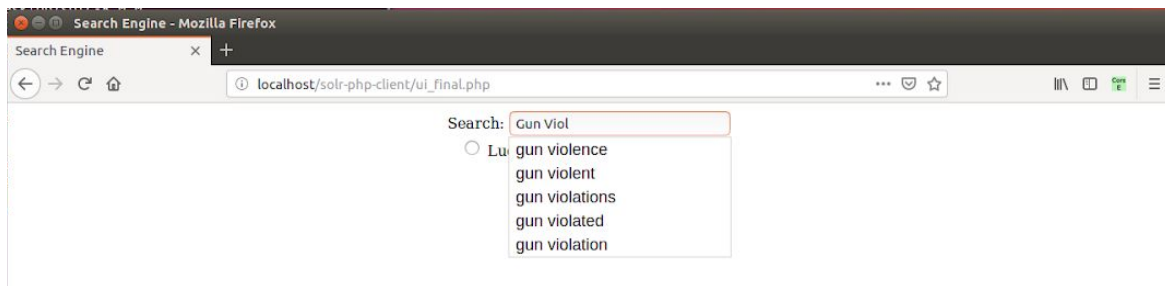## Result 1 - Autocomplete for "presi__"
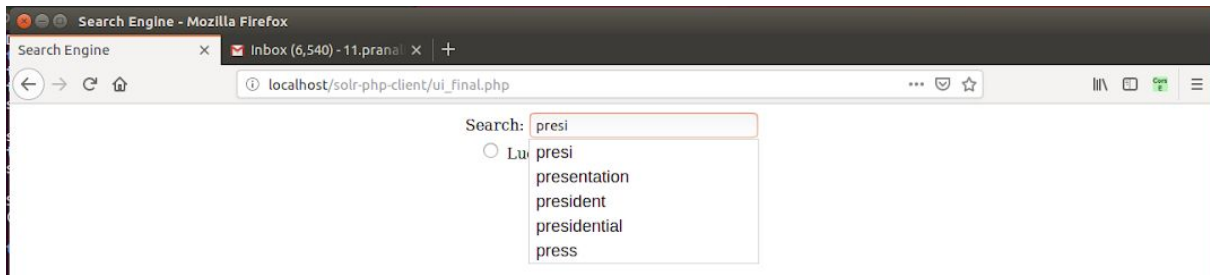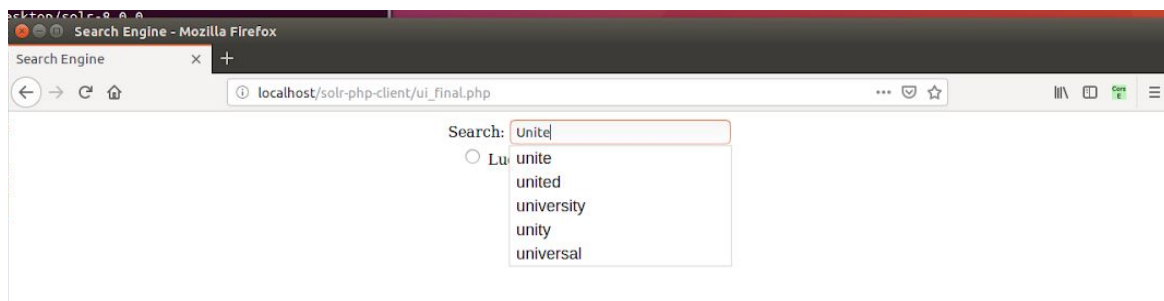

Result 2 - Autocomplete for "mental healt_ "
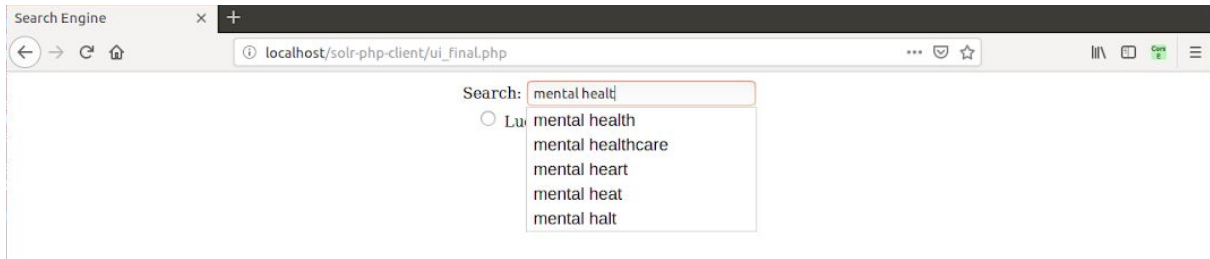
Result 3 - Autocomplete for secui__"

Result 4 - Autocomplete for "Gun Vio__"

Result 5 - Autocomplete for "Unite__"

Result 6 - Snippets for "Security"

Result 7 - Snippets for "Gun Violence"

Result 8 - Spell correction for " winezuela"

Search: winezuela
○ Lucene (Default) ○ Page Rank
Submit Query

Did you mean: venezuela
Your search winezuela did not match any documents
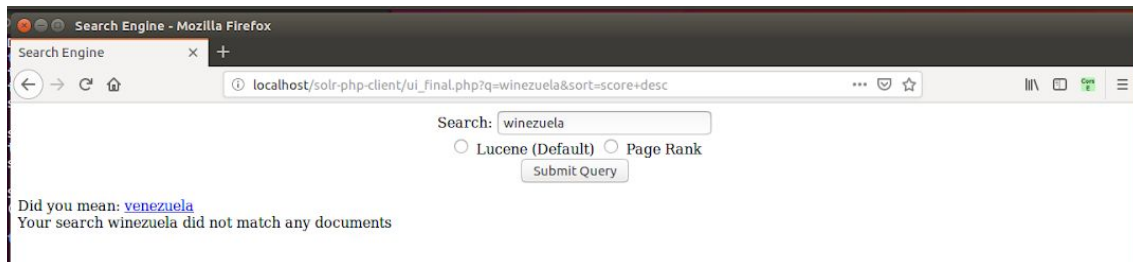
Result 9 - Spell correction for " ceenate"

Search: ceenate
○ Lucene (Default) ○ Page Rank
Submit Query

Did you mean: senate
Your search ceenate did not match any documents
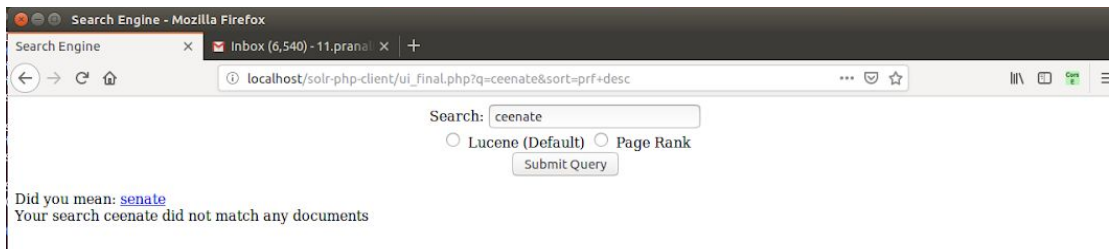
Result 10 - Spell correction for " democratz"

Search: democratz
● Lucene (Default) ○ Page Rank
Submit Query

Did you mean: democrats
Your search democratz did not match any documents
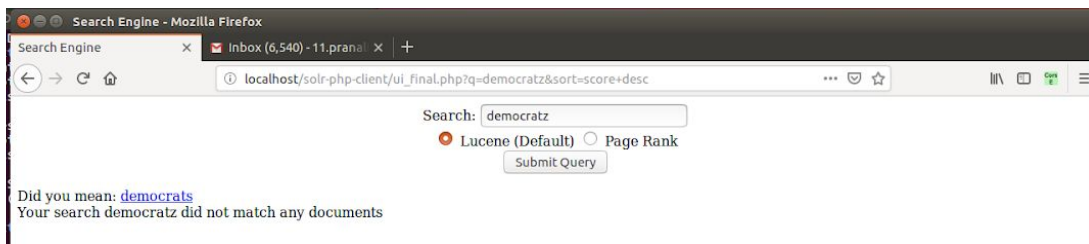
Result 11 - Spell correction for " inda"

Search: inda
○ Lucene (Default) ○ Page Rank
Submit Query

Did you mean: india
About 1 results

1. Iowa: will America's strictest abortion law drive female voters to the left? | US news | The Guardian
http://www.theguardian.com/us-news/2018/may/23/iowa-fetal-heartbeat-law-push-voters-left
/home/pranali/Desktop/solr-8.0.0/guardiannews/f1666736-a27b-4074-9909-9f55fecb8cca.html
No Snippets

Result 12 - Spell correction for " elction"
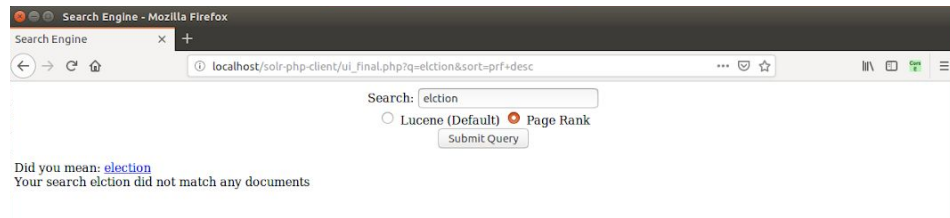
Result 13 - Clicking on corrected term "election"