

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

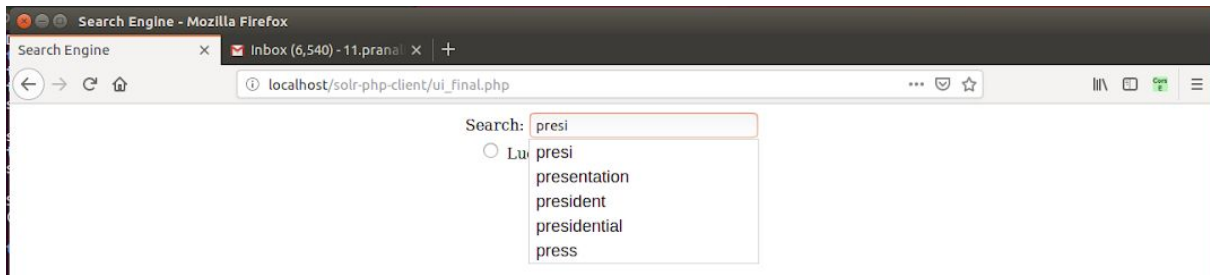
To implement Spell correct

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

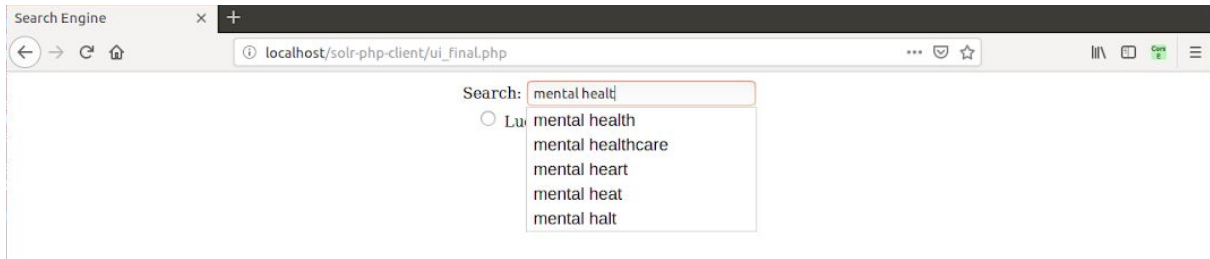
To implement Snippets

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"

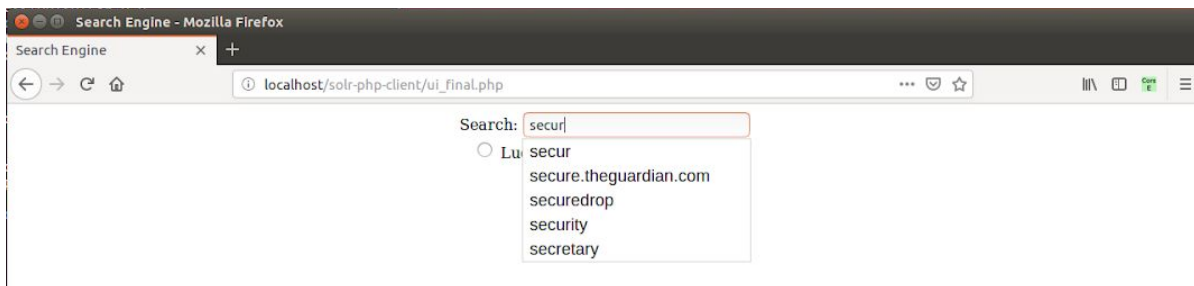
Result 1 - Autocomplete for “presi__”



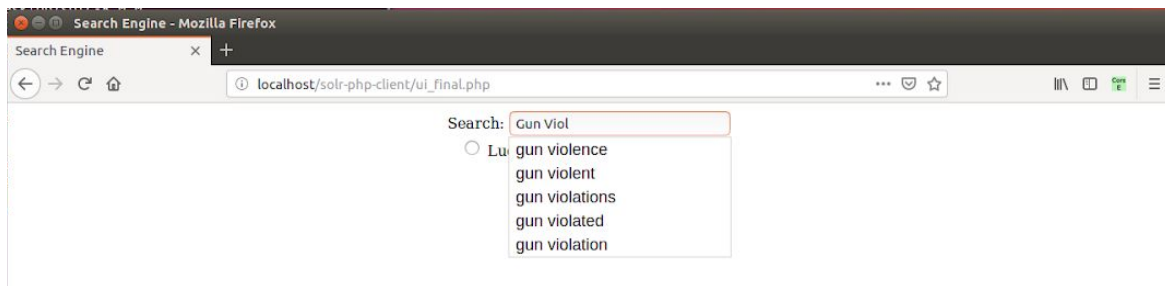
Result 2 - Autocomplete for “mental healt_”



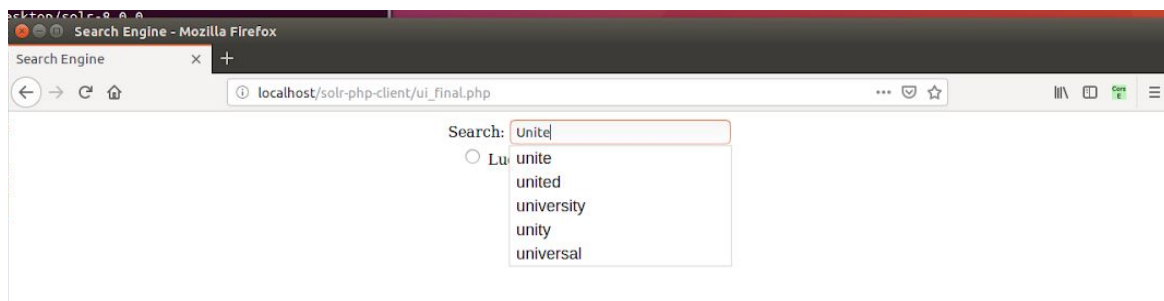
Result 3 - Autocomplete for securi__



Result 4 - Autocomplete for “Gun Vio__”



Result 5 - Autocomplete for “Unite__”



Result 6 - Snippets for “Security”

The screenshot shows a web browser window titled "Search Engine - Mozilla Firefox". The address bar displays "localhost/solr-php-client/ui_final.php?q=Security&sort=score+desc". The search bar contains the text "Security". Below the search bar, there are radio buttons for "Lucene (Default)" and "Page Rank", with "Lucene (Default)" selected. A "Submit Query" button is located below the radio buttons. The results section indicates "About 4627 results". The first four results are listed, each with a number, a title, a URL, and a snippet. The titles are: 1. "Michael Flynn resigns: Trump's national security adviser quits over Russia links - as it happened | US news | The Guardian", 2. "Michael Flynn resigns: Trump's national security adviser quits over Russia links - as it happened | US news | The Guardian", 3. "Trump dares 'worst CIA director' Brennan to sue as security row deepens | US news | The Guardian", and 4. "Democrats launch inquiry into White House's use of security clearances | US news | The Guardian". The snippets provide additional context for each result.

Search: Security

☒ Lucene (Default) ☐ Page Rank

Submit Query

About 4627 results

1. Michael Flynn resigns: Trump's national security adviser quits over Russia links - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2017/feb/14/flynn-resigns-donald-trump-national-security-adviser-russia-links-live>
/home/pranali/Desktop/solr-8.0.0/guardiannews/e73481ce-0a20-491b-a694-b2304163aec7.html
Trump's national **security** adviser quits over Russia links as it happened. ... who is Trump's acting national **security** adviser?.
2. Michael Flynn resigns: Trump's national security adviser quits over Russia links - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2017/feb/14/flynn-resigns-donald-trump-national-security-adviser-russia-links-live>
/home/pranali/Desktop/solr-8.0.0/guardiannews/08737cd5-71cb-4afd-9a0a-bd1edda3bf5d.html
Trump's national **security** adviser quits over Russia links as it happened. ... who is Trump's acting national **security** adviser?.
3. Trump dares 'worst CIA director' Brennan to sue as security row deepens | US news | The Guardian
<http://www.theguardian.com/us-news/2018/aug/20/trump-brennan-security-clearance-cia-director>
/home/pranali/Desktop/solr-8.0.0/guardiannews/75f2dde3-0385-454e-bf0a-5cbded744109.html
Trump dares 'worst CIA director' Brennan to sue as **security** row deepens.
4. Democrats launch inquiry into White House's use of security clearances | US news | The Guardian
<http://www.theguardian.com/us-news/2019/jan/23/trump-white-house-security-clearances-investigation-democrats-house>
/home/pranali/Desktop/solr-8.0.0/guardiannews/1f102a06-f32a-4904-8553-5b0ed4c6245b.html
Democrats launch inquiry into White House's use of **security** clearances. Trump officials accused of grave breaches of national **security**.

Missile crisis by candlelight: Donald Trump's use of Mar-a-Lago raises security questions | US

www.theguardian.com/us-news/2018/aug/20/trump-brennan-security-clearance-cia-director

Result 7 - Snippets for “Gun Violence”

The screenshot shows a web browser window titled "Search Engine - Mozilla Firefox". The address bar displays "localhost/solr-php-client/ui_final.php?q=gun+violence&sort=score+desc". The search bar contains the text "gun violence". Below the search bar, there are radio buttons for "Lucene (Default)" and "Page Rank", with "Lucene (Default)" selected. A "Submit Query" button is located below the radio buttons. The results section indicates "About 1200 results". The first four results are listed, each with a number, a title, a URL, and a snippet. The titles are: 1. "March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian", 2. "March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian", 3. "March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian", and 4. "March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian". The snippets provide additional context for each result.

Search: gun violence

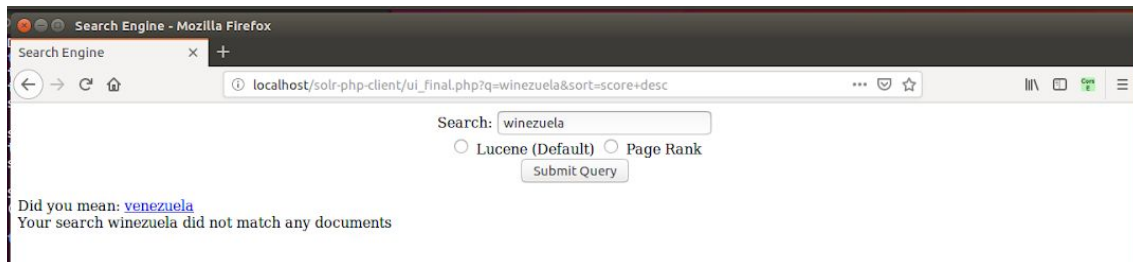
☒ Lucene (Default) ☐ Page Rank

Submit Query

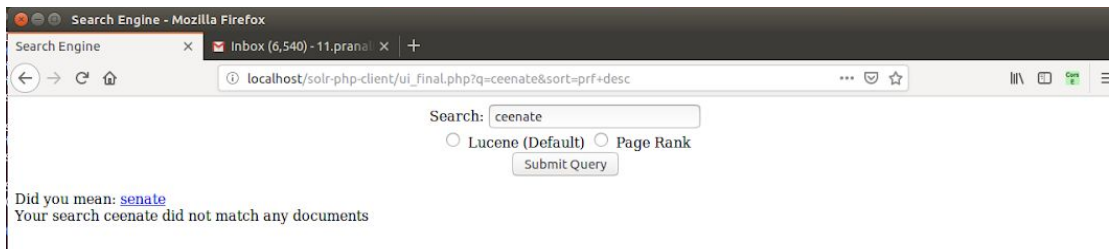
About 1200 results

1. March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2018/mar/24/march-for-our-lives-protest-gun-violence-washington>
/home/pranali/Desktop/solr-8.0.0/guardiannews/62da3d13-7467-4afa-ba2a-559d7882fafb.html
hundreds of thousands demand end to **gun violence** as it happened. ... Thousands join anti-**gun** protests around the world. ... a rally for **gun** control.
2. March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2018/mar/24/march-for-our-lives-protest-gun-violence-washington>
/home/pranali/Desktop/solr-8.0.0/guardiannews/accc8ba8-d004-4b25-905d-86a1bd1ed558.html
hundreds of thousands demand end to **gun violence** as it happened. ... Thousands join anti-**gun** protests around the world. ... a rally for **gun** control.
3. March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2018/mar/24/march-for-our-lives-protest-gun-violence-washington>
/home/pranali/Desktop/solr-8.0.0/guardiannews/45b7dd69-c4ee-4055-a7ae-2b21b4aff467.html
hundreds of thousands demand end to **gun violence** as it happened. ... Thousands join anti-**gun** protests around the world. ... a rally for **gun** control.
4. March for Our Lives: hundreds of thousands demand end to gun violence - as it happened | US news | The Guardian
<http://www.theguardian.com/us-news/live/2018/mar/24/march-for-our-lives-protest-gun-violence-washington>
/home/pranali/Desktop/solr-8.0.0/guardiannews/e673c03a-e40f-4ae2-bb8f-41532637fa89.html

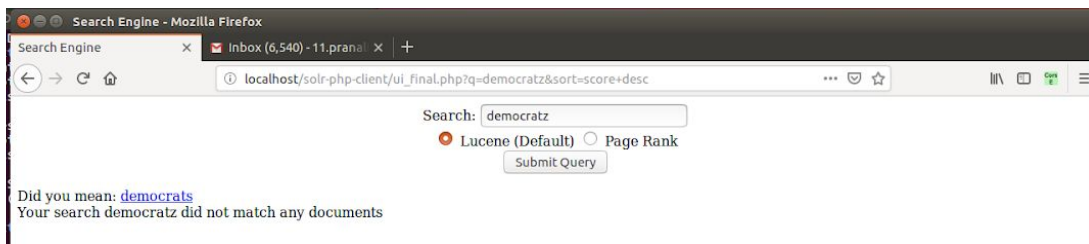
Result 8 - Spell correction for “winezuela”



Result 9 - Spell correction for “ceenate”



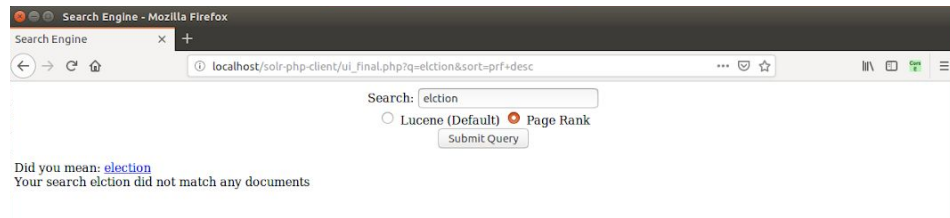
Result 10 - Spell correction for “democratz”



Result 11 - Spell correction for “inda”



Result 12 - Spell correction for “ elction”



Result 13 - Clicking on corrected term “election”

