Yeun-Yuan Kuo

CSCI572 – hw5

April 24, 2019

**Steps I took to complete this assignment:**

Spelling Correction:

1. Used Apache Tika library to implement a Java program that parses all the HTML files in the reuters news file given to us. Using the Tika library I am able to extract contents and metadata from the HTML files and generate my own big.txt that is tailord to reauter news.
2. After creating my own big.txt, I used Peter Norvig’s spelling correction program written in php form and feed my reuters news big.txt to that program.
3. Then in my original index.php file I included the SpellCorrector.php file that I got from Peter Norvig’s website so I can call functions within that file and implement the spelling correction codes.

Autocomplete:

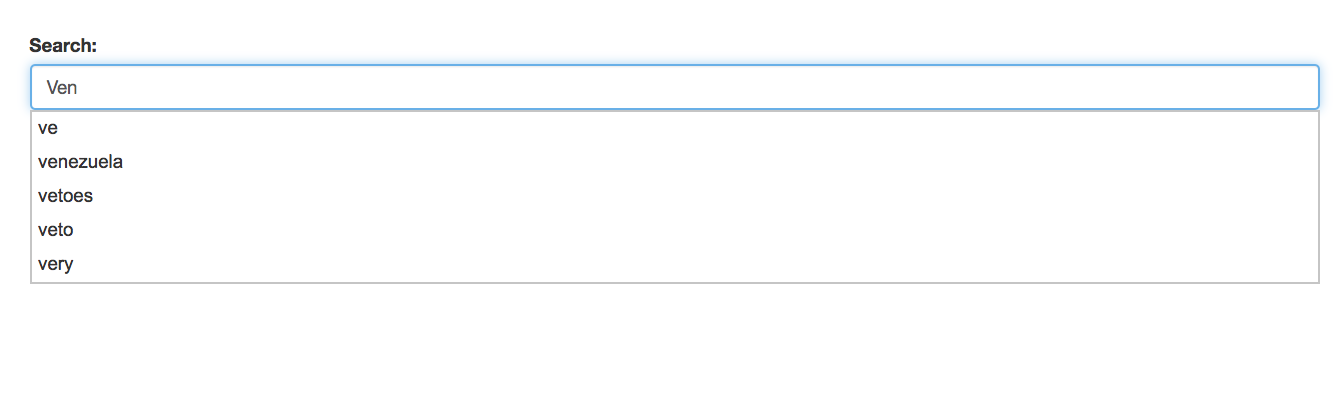
1. I used the Solr’s FuzzyLookupFactory function for autocomplete function.
2. First, I modified my solrconfig.xml file adding the new solr.SuggestComponent in my code.
3. Then, I added the corresponding handler for that SuggestionComponent so I can call the suggest component.
4. Lastly, in my original index.php file add in the calling functions through AJAX that would automatically display a list of auto complete words when a new character is typed in the search box.

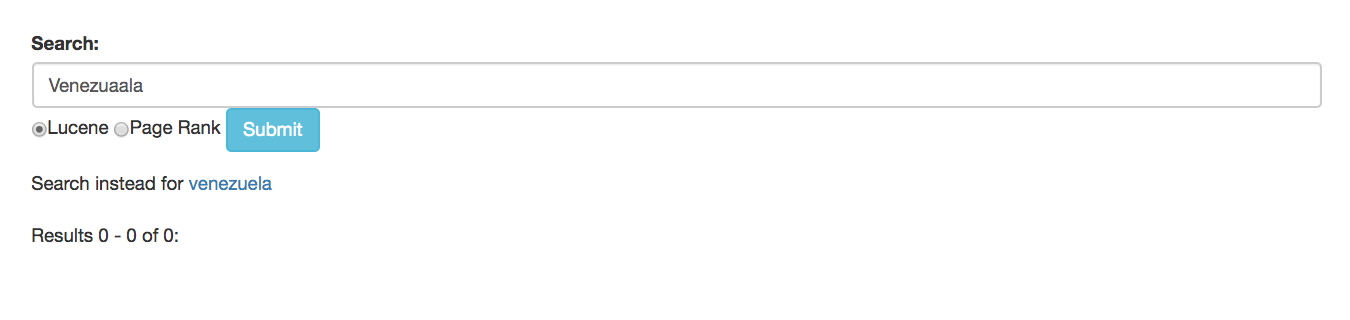
Snippets:

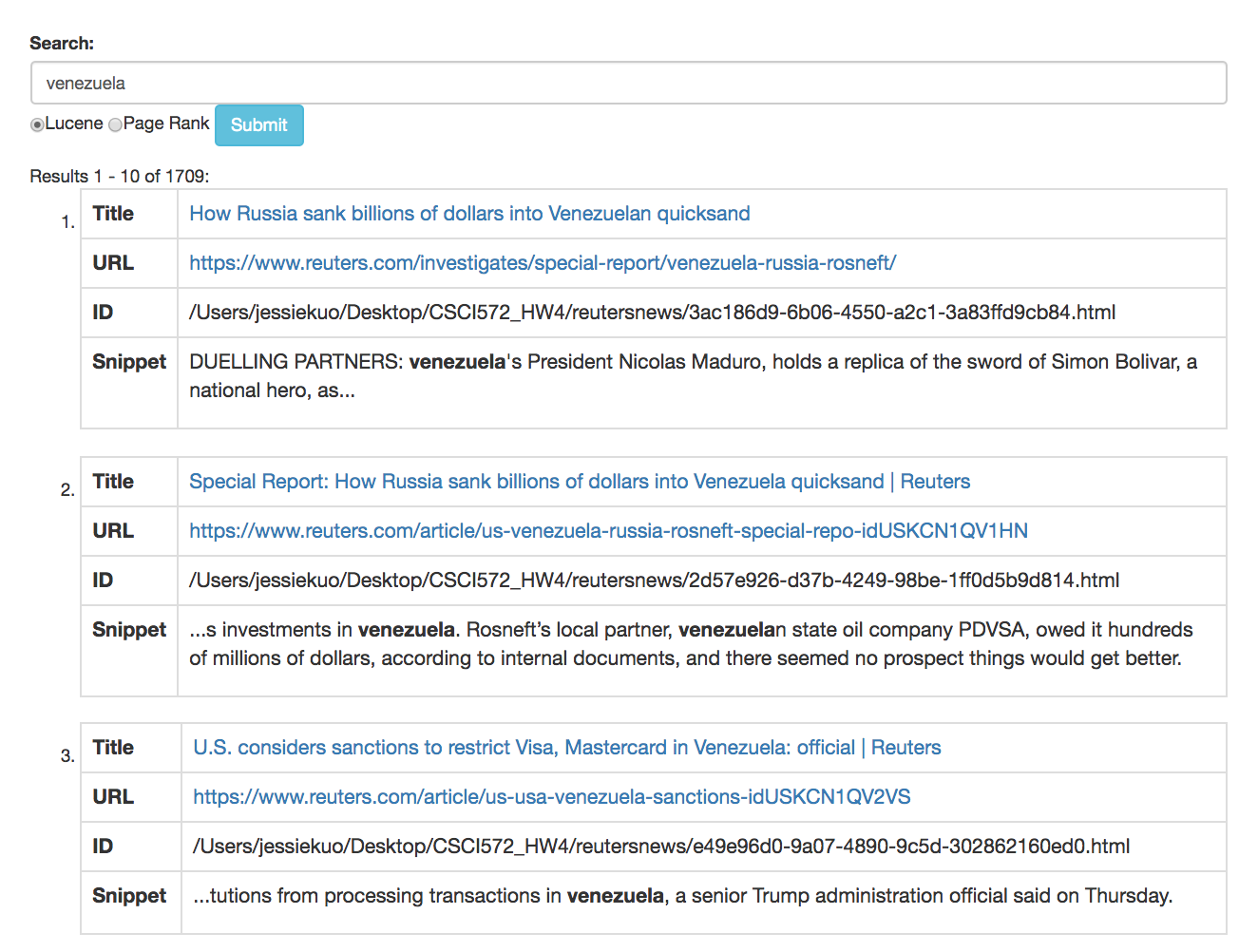
1. For snippets I used the 'simple\_html\_dom.php' and the internal function of parsing a html page.
2. I wrote my own code for parsing a webpage and finding the first sentence that contains the query word in it.
3. Then I wrote another function custom\_echo to process the sentence and trim it so that the sentence is within 160 characters and that the query word would be in that 160 characters bolded.

**Analysis of my results:**

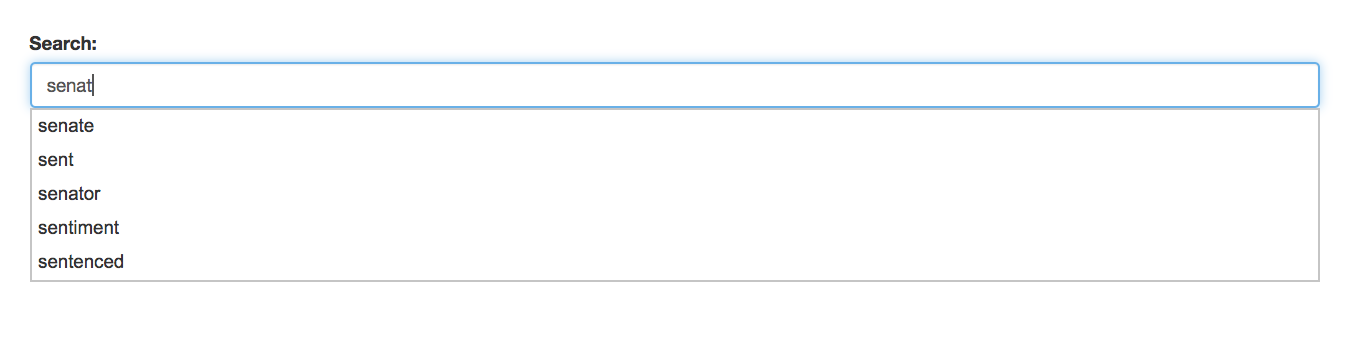
**Example 1: Venezuela**

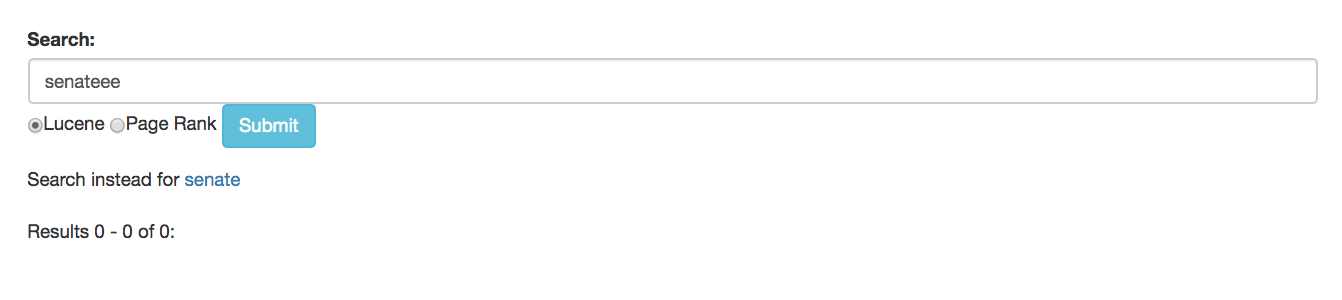


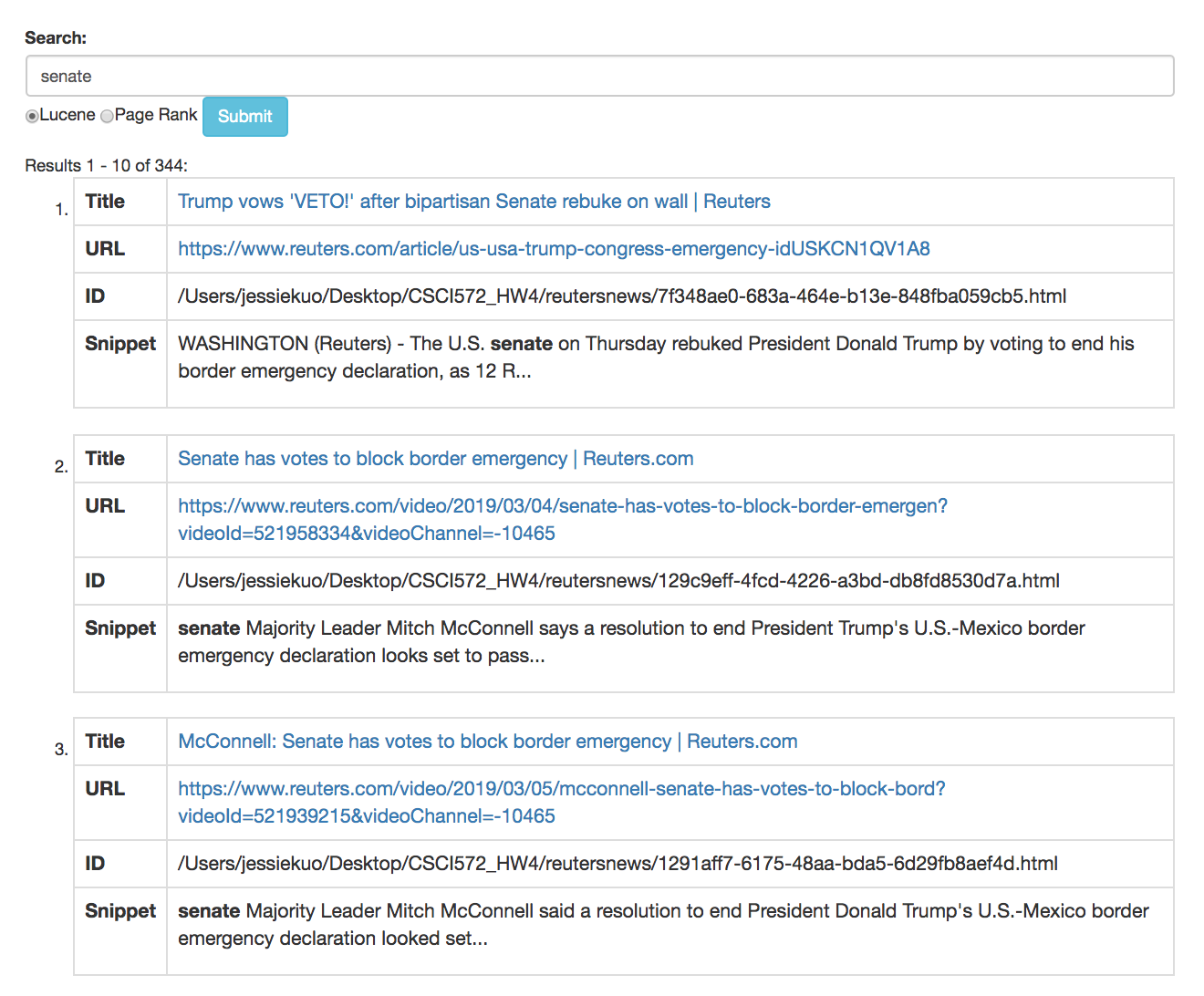




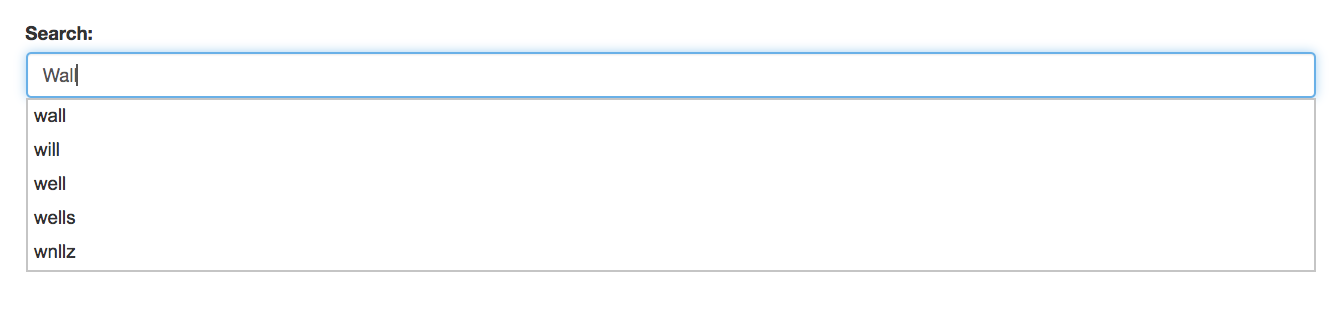
**Example 2: Senate**

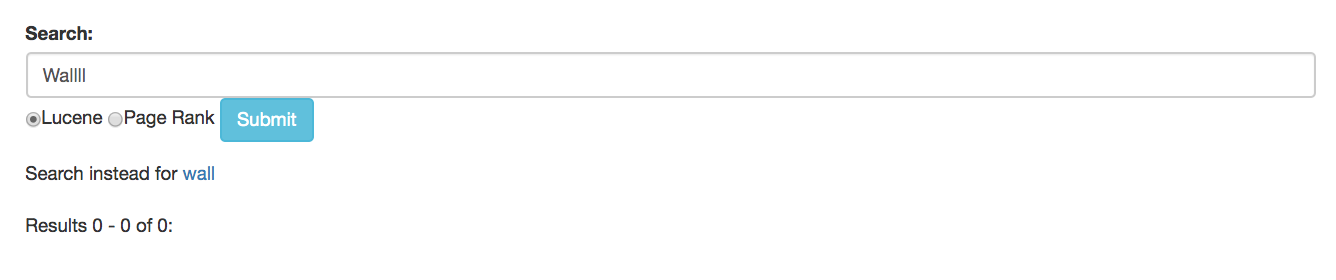


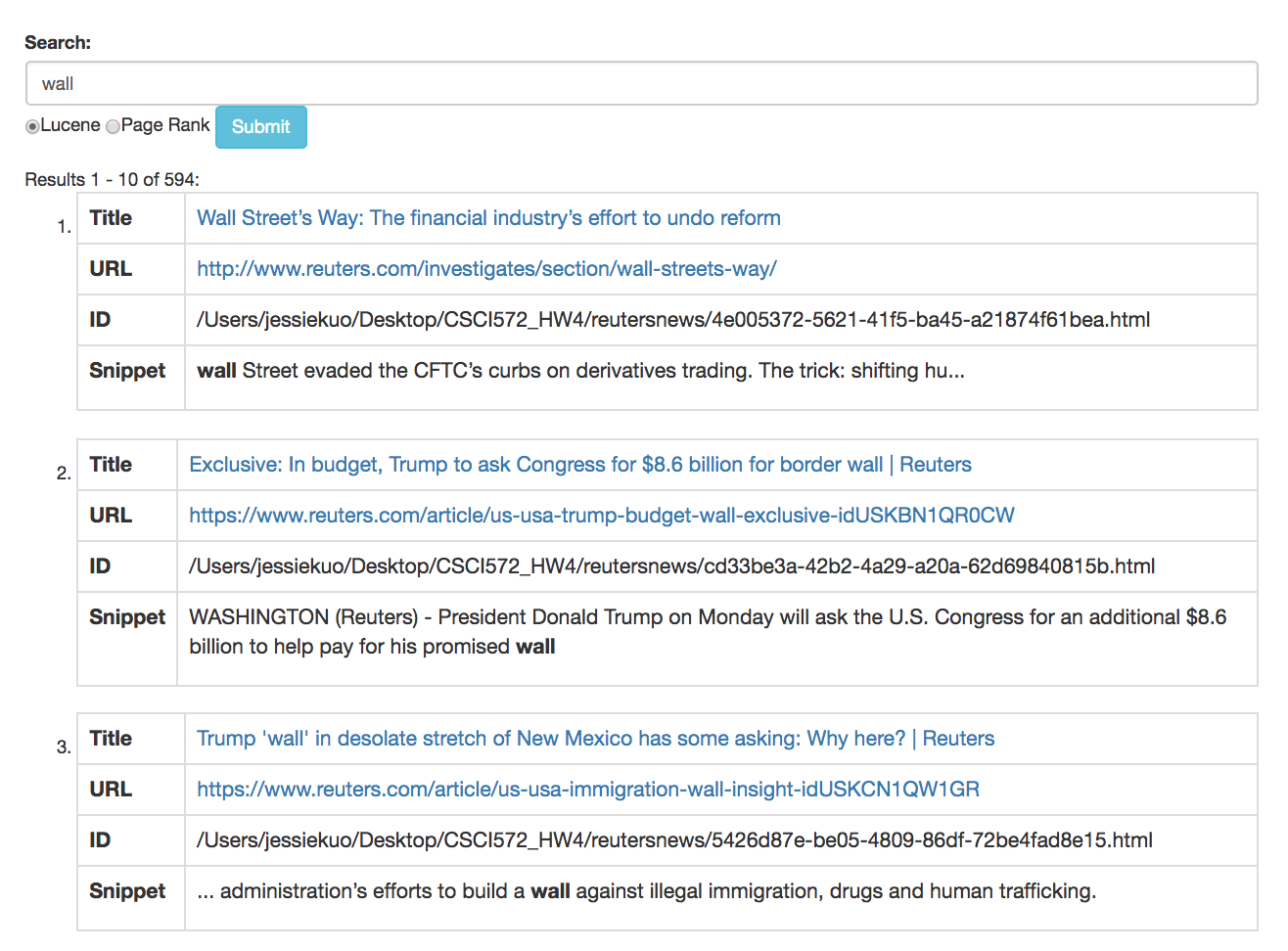




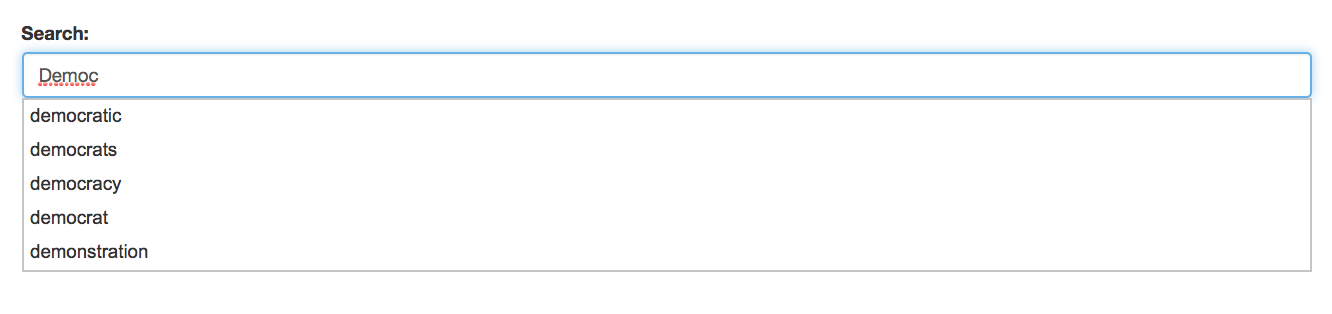
**Example 3: Wall**

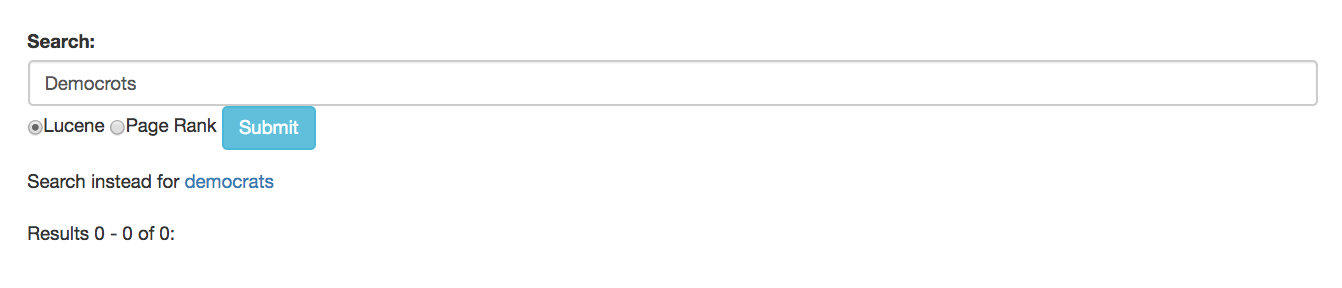


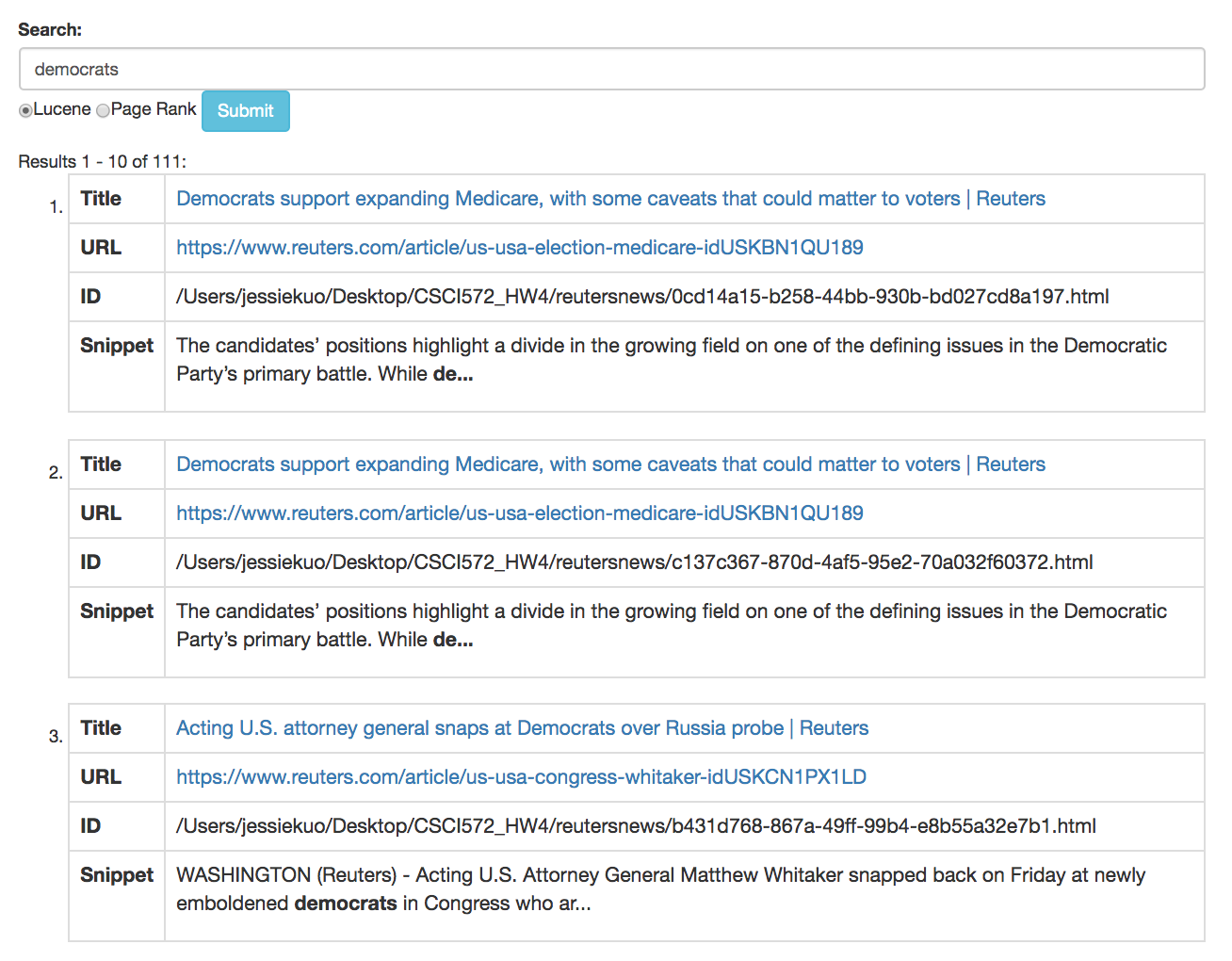




**Example 4: Democrats**







**Example 5: Republicans**

