

IR HW5 REPORT

Adding Autocomplete, Spell Check, Snippets to our Search Engine

Steps:

Spell Check

1. Downloaded Peter Norvig's spelling corrector's PHP version
2. Added Apache Tika jar files to a new java project.
3. Wrote java program to parse all the html files and create a list of words in "big.txt" file. This big.txt file is used by SpellCorrector.php to correct misspellings.
4. Implemented the following in my reuters search engine :

When a user inputs an incorrect query and submits the page, the program checks to see if it's a valid word or not. If it's a valid word corresponding results are displayed.

When it's an incorrectly spelt word it will be sent to the Norvig spell corrector and the correct word will be returned as a suggestion to continue with a phrase with "Did you mean?" and the keyword gets displayed.

On click of new keyword the corresponding results are displayed.

Auto Suggest:

1. Added the suggest component to Solr as given in the exercise- Added search component and request handler in solrconfig.xml
2. Checked on Solr UI- using /suggest in Request Handler to verify that auto-suggest was working
3. Made changes in php code so that it can give suggestion when user types the query. Used JQuery to implement this functionality. JQuery has an AutoComplete widget <https://jqueryui.com/autocomplete/> for the same.
4. Splitted the query terms based on space, queried solr to get autocomplete suggestions for the last term and added it to all the initial query terms.
5. Then used Ajax to request solr to get suggestions, simultaneously not blocking the website's front end, fetched results and displayed it as suggestions.
6. Made changes in solrconfig.xml to make "AND" as default instead of "OR" for multi-word queries. Added <str name="q.op">AND</str> line in solr.Search Handler in list defaults tag.

Snippets

1. Used simple HTML Doc Parser to parse the html files.
2. Used file_get_contents function to extract contents from the html file.
3. Divided the contents into sentences and words and removed special characters.
4. Added the best matching sentences till the snippet's length is less than 160 characters. If no matching sentence found, snippet is set to N/A.

Analysis of Results

Spell Check:

1. incorrect word : winezuela
corrected word : venezuela

Reuters Search Engine

localhost/searchEngine.php?q=winezuela&algo=l

Search:

☒ Lucene(Default) ☐ PageRank

Did u mean [venezuela](#)

Showing results for Winezuela

Results 0 - 0 of 0:

Clicking on venezuela, we get

Reuters Search Engine

localhost/searchEngine.php?q=venezuela&algo=l

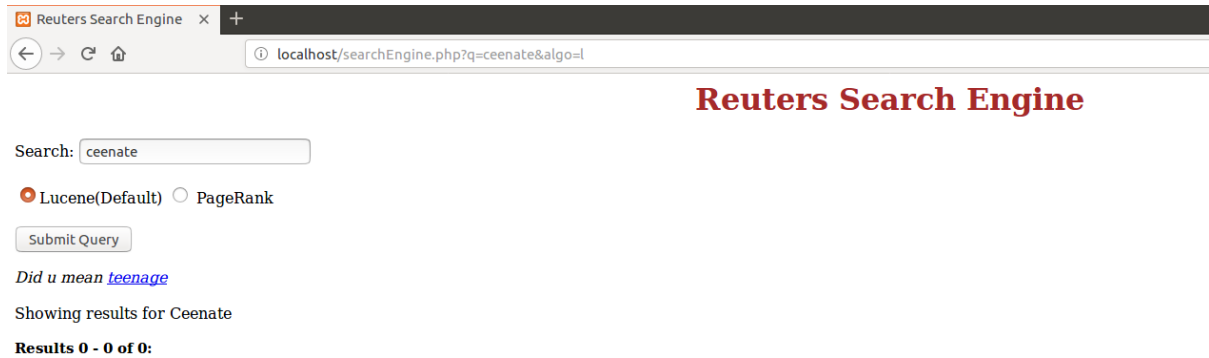
Search:

☒ Lucene(Default) ☐ PageRank

Results 1 - 10 of 1709:

1.	Title: How Russia sank billions of dollars into Venezuelan quicksand Reuters URL: https://www.reuters.com/investigates/special-report/venezuela-russia-rosneft/ Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/3ac186d9-6b06-4550-a2c1-3a83fd9cb84.html Snippet: Joint ventures between Rosneft and venezuela 's state oil company, PDVSA, have not always gone smoothly. GMT MOSCOW - At the end of 2015, managers at Rosneft, the Russian state-controlled oil firm, sounded the alarm to their bosses about the company's investments in venezuela
2.	Title: Special Report: How Russia sank billions of dollars into Venezuela quicksand Reuters URL: https://www.reuters.com/article/us-venezuela-russia-rosneft-special-report-idUSKCN1QV1HJN Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/2d57e926-d37b-4249-98be-1ff0d5b9d814.html Snippet: ... They cover the firm's operations in venezuela between 2012 and 2015. We all had to contribute," said an executive at a Russian oil firm that partnered with Rosneft in venezuela
3.	Title: U.S. considers sanctions to restrict Visa, Mastercard in Venezuela; official Reuters URL: https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HJM Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/e49e96d0-9a07-4890-9c5d-302862160ed0.html Snippet: ... President Donald Trump's administration has hit Maduro's government, his political allies and venezuela 's state-owned oil company PDVSA with a series of sanctions since recognizing opposition leader Juan Guaido as the country's interim president at the end of January. ...
4.	Title: U.S. envoy says Venezuela oil production dropping steadily Reuters URL: https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HJM Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/c725f3e2-eda2-47a6-97b2-8840c3ad3e89.html Snippet: ... draft resolution calling for free and fair presidential elections in venezuela at U. special representative for venezuela said on Friday that venezuela 's oil exports have been dropping steadily by roughly 50,000 barrels per month and production is likely to dip below a million barrels a day within a "month or two. ...
5.	Title: Shipbroker McQuillan ends tanker contracts with Venezuela's PDVSA Reuters URL: https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1QW2L6 Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/f5434ebb-7882-4ae9-9962-83ed792b2306.html Snippet: ...-based McQuillan Partners Inc, which provided PDVSA with four contract tankers, joined German tanker operator Bernhard Schulte Shipmanagement (BSM) in withdrawing from providing oil-shipping services to venezuela
6.	Title: Billions spent, little to show Reuters URL: https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1QV1HJN Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/8930da40-6b0b-4276-9c99-83217309951f.html

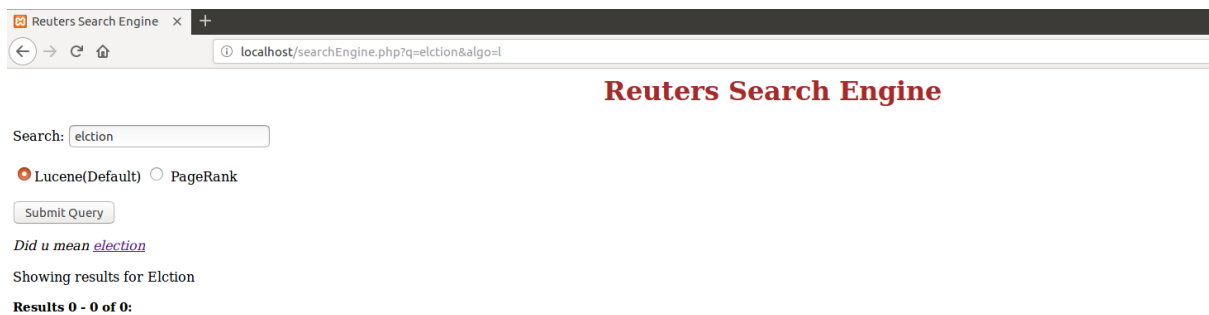
2. incorrect word : ceenate
corrected word : teenage



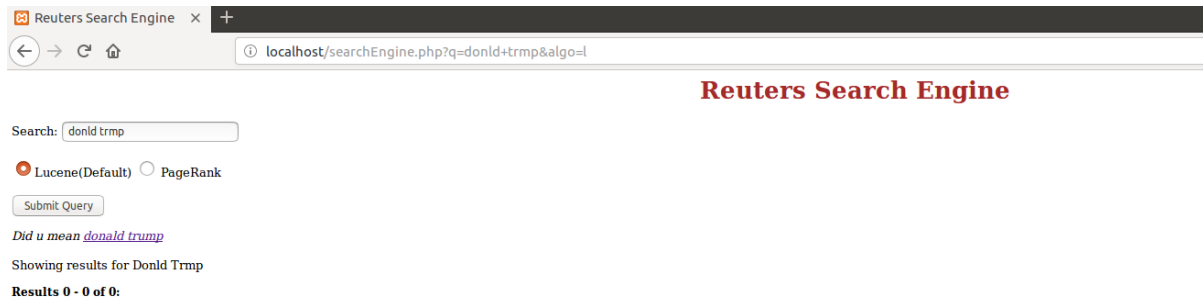
3. incorrect word : democratrz
corrected word : democrats



4. incorrect word : elction
corrected word : election

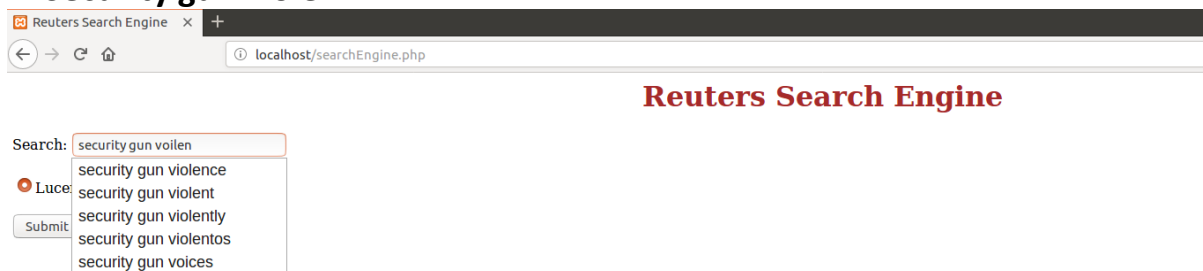


5. incorrect words: donld trmp
corrected word: donald trump

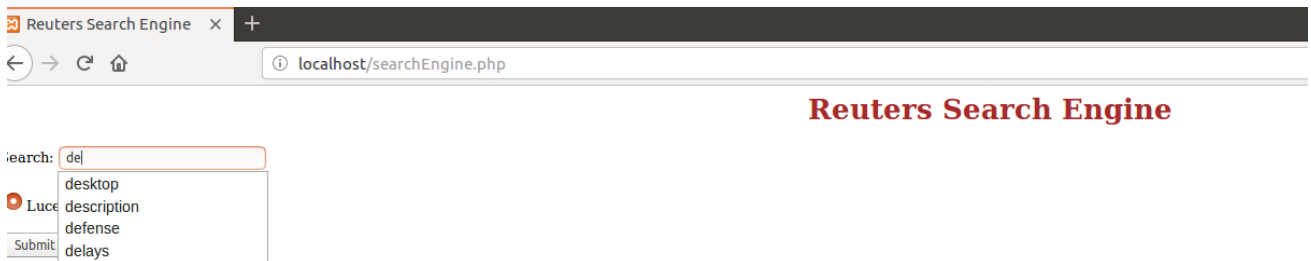


Examples of AutoComplete

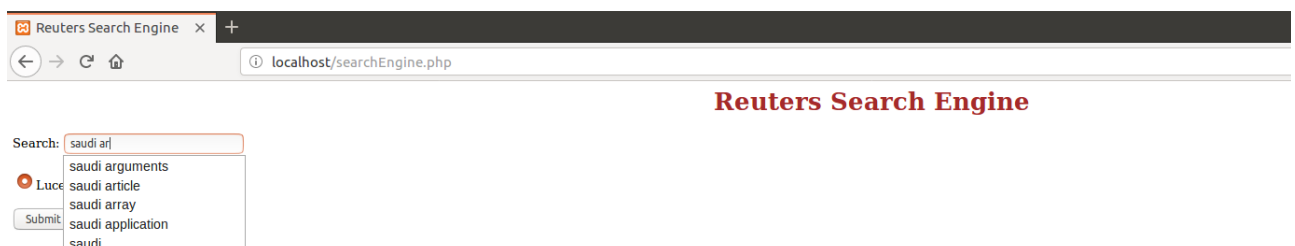
1. Security gun violen



2. de



3. saudi ar



4. democra

Reuters Search Engine x +

localhost/searchEngine.php

Reuters Search Engine

Search:

☒ Lucene ☐ PageRank

democratic

democr

democracy

democratizing

5. senat

Reuters Search Engine x +

localhost/searchEngine.php

Reuters Search Engine

Search:

☒ Lucene ☐ PageRank

senate

sent

senator

sentiment

sentenced

Snippets

Snippets for query “Venezuela”

Reuters Search Engine x +

localhost/searchEngine.php?q=venezuela&algo=l

Reuters Search Engine

Search:

☒ Lucene(Default) ☐ PageRank

Results 1 - 10 of 1709:

1.

Title: [How Russia sank billions of dollars into Venezuelan quicksand](#)

URL: <https://www.reuters.com/investments/special-report/venezuela-russia-rosneft/>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/3ac186d9-6b06-4550-a2c1-3a83f8d9cb84.html

Snippet: Joint ventures between Rosneft and **venezuela**'s state oil company, PDVSA, have not always gone smoothly. GMT MOSCOW - At the end of 2015, managers at Rosneft, the Russian state-controlled oil firm, sounded the alarm to their bosses about the company's investments in **venezuela** ...

2.

Title: [Special Report: How Russia sank billions of dollars into Venezuela quicksand | Reuters](#)

URL: <https://www.reuters.com/article/us-venezuela-russia-rosneft-special-repo-idUSKCN1OV1HN>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/2d57e7d6-437b-4249-98be-1ff45b9d814.html

Snippet: ... They cover the firm's operations in **venezuela** between 2012 and 2013. We all had to contribute," said an executive at a Russian oil firm that partnered with Rosneft in **venezuela** ...

3.

Title: [U.S. considers sanctions to restrict Visa, Mastercard in Venezuela, official | Reuters](#)

URL: <https://www.reuters.com/article/us-usa-venezuela-sanctions-idUSKCN1OV2VS>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/e49e96d0-9a07-4890-0c5d-302862160ed0.html

Snippet: ... President Donald Trump's administration has hit Maduro's government, his political allies and **venezuela**'s state-owned oil company PDVSA with a series of sanctions since recognizing opposition leader Juan Guaido as the country's interim president at the end of January ...

4.

Title: [U.S. envoy says Venezuela oil production dropping steadily | Reuters](#)

URL: <https://www.reuters.com/article/us-venezuela-politics-usa-sbrams-idUSKCN1OW2HM>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/c7253e2-eda2-47a6-97b2-8840c3ad3e89.html

Snippet: ... draft resolution calling for free and fair presidential elections in **venezuela** at U. special representative for **venezuela** said on Friday that **venezuela**'s oil exports have been dropping steadily by roughly 50,000 barrels per month and production is likely to dip below a million barrels a day within a "month or two."

5.

Title: [Shipbroker, McQuillan ends tanker contracts with Venezuela's PDVSA | Reuters](#)

URL: <https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1OW2L6>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/f5434ebb-7882-4ae9-9962-83ed792b2306.html

Snippet: ...-based McQuillan Partners Inc, which provided PDVSA with four contract tankers, joined German tanker operator Bernhard Schulte Shipmanagement (BSM) in withdrawing from providing oil-shipping services to **venezuela** ...

6.

Title: [Billions spent, little to show | Reuters](#)

URL: <https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1OV1HS>

Doc Id: /home/shakshi/Desktop/solr-8.0.0/crawl_data/reutersnews/8930da40-660b-4276-9c99-83217309951f.html

Snippet: ...5 billion more in **venezuela** than it had earned in the form of oil allocated to it as dividends, according to Reuters' calculations. This figure is reached by calculating the value of oil Rosneft received from its joint venture projects and subtracting from that outstanding loans Rosneft issued to PDVSA, official payments Rosneft made to **venezuela** for access to oil fields, Rosneft capital expenditure on the ventures, and the cost of extracting Rosneft's share of the oil ...