

CSCI 572 HW5 Report (Reuters)

Shuming Hao
1547131193

Implement spelling correction:

1. Used HtmlParse.java to parse data from Reuters website. Generate big.txt for future use. Used Eclipse as java IDE.
2. Downloaded SpellCorrector.php which is an implementation based on Peter Norvig's algorithm in php.
3. Included this code in the my index.php for the UI generation

Implement autosuggest:

1. Modify solrconfig.xml file to include changes related to suggest and request handler
2. Reload myexample core for the changes to get reflected
3. Included JQuery in index.php. Suggestions are displayed in a drop-down fashion as soon as the user begins typing and the suggestions get more streamlined as the user gives in more input

Implement snippet generation:

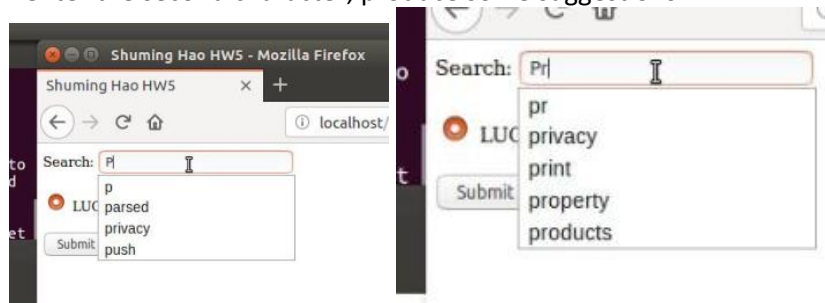
1. Included simple_html_dom.php to extract HTML content
2. Generated snippets for websites.
3. For multiple term queries, try finding a sentence with all the terms together. If not, return the sentence that has all the terms in it, even if they are not together or in same order. If none of the fore-mentioned are found, return the first sentence with at least one query term in it. If no match is found, then no snippet is returned.
4. Highlight the query terms in the snippet

Analysis of the results:

Do steps 1 and 2, first for **President** and then for **Mental Health**

1. TESTING AUTOCOMPLETE

- enter the first character, produce some suggestions
- enter the second character, produce some suggestions

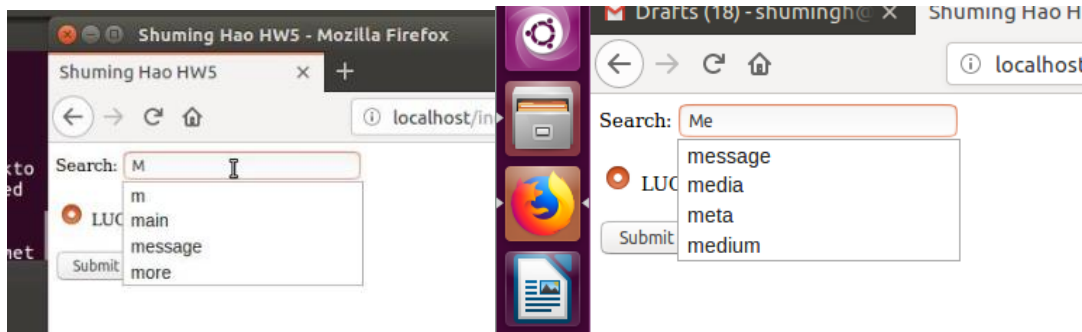


Suggestions prompted: p, parsed, privacy, push

Suggestions prompted: pr, privacy, print, property, products

Suggestions prompted: m, main, message, more

Suggestions prompted: message, media, meta, medium



2. TESTING SEARCH RESULTS WITH SNIPPETS

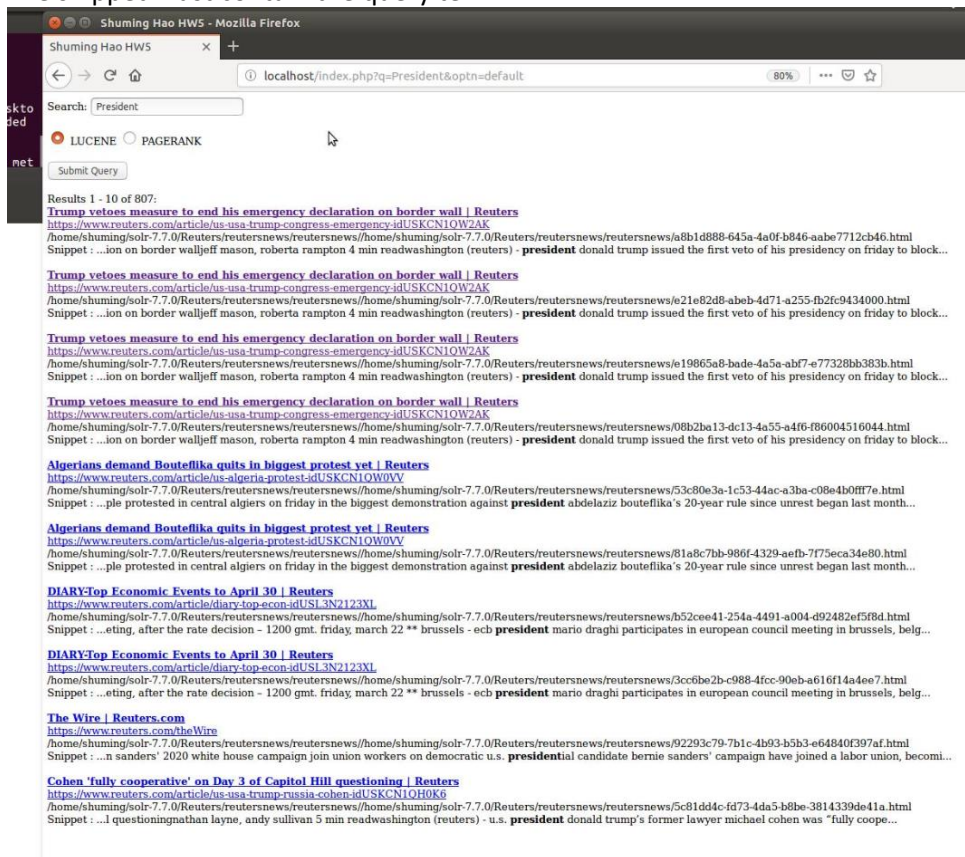
Enter the full Query.

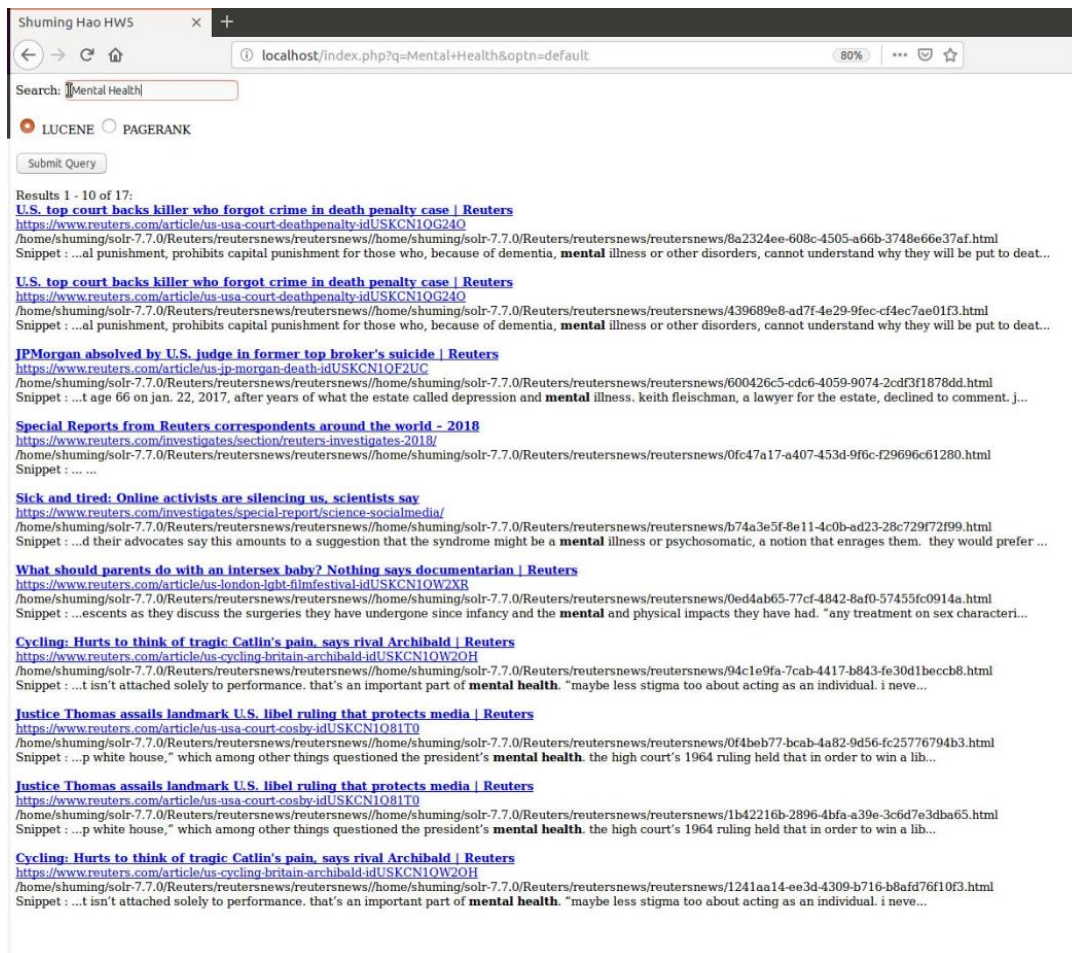
There should be some results

Each result must include a link to the associated web page.

The grader must test the link.

The snippet must contain the query term





3. TESTING SPELL CORRECTION

Results are in the table. Screenshots are in the next page. I did not show all the results to limit the page or this report. All the links and titles are clickable.

	Reuters
winezuela	venezuela
ceenate	teenage
democratrz	democrats
elction	election

Search: winezuela

☒ LUCENE ☐ PAGERANK

Submit Query

Showing results for spell corrected term: [venezuela](#)

Do you want to search instead for the misspelled term: [winezuela](#)

Results 1 - 10 of 1709:

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

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

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/3ac186d9-6b06-4550-a2c1-3a83ffd9cb84.html

Snippet : ... how russia sank billions of dollars into **venezuelan** quicksand ...

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

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

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/2d57e926-d37b-4249-98be-1ff0d5b9d814.html

Snippet : ... special report: how russia sank billions of dollars into **venezuela** quicksand | reuters ...

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

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

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/e49e96d0-9a07-4890-9c5d-302862160ed0.html

Snippet : ... u.s. considers sanctions to restrict visa, mastercard in **venezuela**: official | reuters ...

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

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

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/c725f3e2-eda2-47a6-97b2-8840c3ad3e89.html

Snippet : ... u.s. envoy says **venezuela** oil production dropping steadily | reuters ...

Search: ceenate

☒ LUCENE ☐ PAGERANK

Submit Query

Showing results for spell corrected term: [teenage](#)

Do you want to search instead for the misspelled term: [ceenate](#)

Results 1 - 10 of 1815:

[Beto's secret life as a teenage hacker | Reuters.com](#)

<https://www.reuters.com/video/2019/03/15/betos-secret-life-as-a-teenage-hacker?videoId=526308778&videoChannel=118169>

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/aac1fab9-b07-4520-a010-8962ee03f1a2.html

Snippet : ... beto's secret life as a **teenage** hacker | reuters.com ...

[Beto's secret life as a teenage hacker | Reuters.com](#)

<https://www.reuters.com/video/2019/03/15/betos-secret-life-as-a-teenage-hacker?videoId=526308778&videoChannel=1>

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/561fcdaf-cf35-4401-a525-ff4aaef090da.html

Snippet : ... beto's secret life as a **teenage** hacker | reuters.com ...

[Beto's secret life as a teenage hacker | Reuters.com](#)

<https://www.reuters.com/video/2019/03/15/betos-secret-life-as-a-teenage-hacker?videoId=526308778&videoChannel=1>

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/a2bbcea7-21f7-4b6f-9194-f8aeba5a804e.html

Snippet : ... beto's secret life as a **teenage** hacker | reuters.com ...

[Columns | Reuters.com](#)

<https://www.reuters.com/journalists/josephine-mason>

/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/029c5211-765e-4b4c-ab1b-a5ae95a2bffc.html

Snippet : ... 04:24 beto's secret life as a **teenage** hacker 02:02 ...

Shuming Hao HWS - Mozilla Firefox

Shuming Hao HWS x +

localhost/index.php?q=democratz&optn=default 80% ...

Search:

☒ LUCENE ☐ PAGERANK

Showing results for spell corrected term: [democrats](#)

Do you want to search instead for the misspelled term: [democratz](#)

Results 1 - 10 of 111:

[Democrats support expanding Medicare, with some caveats that could matter to voters | Reuters](#)
<https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QIU189>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/c137c367-870d-4af5-95e2-70a032f60372.html
Snippet : ... **democrats** support expanding medicare, with some caveats that could matter to voters | reuters ...

[Democrats support expanding Medicare, with some caveats that could matter to voters | Reuters](#)
<https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QIU189>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/0cd14a15-b258-44bb-930b-bd027cd8a197.html
Snippet : ... **democrats** support expanding medicare, with some caveats that could matter to voters | reuters ...

[Acting U.S. attorney general snaps at Democrats over Russia probe | Reuters](#)
<https://www.reuters.com/article/us-usa-congress-whitaker-idUSKCN1PX1LD>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/b431d768-867a-49ff-99b4-e8b55a32e7b1.html
Snippet : ... acting u.s. attorney general snaps at **democrats** over russia probe | reuters ...

[Republicans bash Cohen, Democrats at CPAC | Reuters.com](#)
<https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoId=520293110&videoChannel=13976>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/cd8637fe-4d2f-4bfb-b1ee-b497ede9b745.html
Snippet : ... republicans bash cohen, **democrats** at cpac | reuters.com ...

[As Biden weighs 2020 bid, Democrats ask: 'Does he meet the moment?' | Reuters](#)
<https://www.reuters.com/article/us-usa-election-biden-idUSKCN1OH1EX>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/9ad0855d-e270-478a-80ac-dc672b3fdadd.html
Snippet : ... as biden weighs 2020 bid, **democrats** ask: 'does he meet the moment?' | reuters ...

ave

Search:

☒ LUCENE ☐ PAGERANK

Showing results for spell corrected term: [election](#)

Do you want to search instead for the misspelled term: [elction](#)

Results 1 - 10 of 403:

[Beto o'rourke campaigns in iowa idUSRTX6R98N](#)
<https://www.reuters.com/news/picture/beto-orourke-campaigns-in-iowa-idUSRTX6R98N>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/d50e984d-4ca2-4a67-afd7-035902251716.html
Snippet : ... beto o'rourke campaigns in iowa | reuters.com ...

[Zuma-era officials feature on South Africa's ANC election list | Reuters](#)
<https://www.reuters.com/article/us-safrica-politics-idUSKCN1OW2M4>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/045405bc-787b-48c2-9286-b397f0fdac3b.html
Snippet : ... zuma-era officials feature on south africa's anc **election** list | reuters ...

[Trump's top communications aide Shine resigns, moves to re-election campaign | Reuters](#)
<https://www.reuters.com/article/us-usa-election-trump-shine-idUSKCN1OP1YM>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/8ba63d40-82e4-43bb-b59b-faa103d21302.html
Snippet : ... trump's top communications aide shine resigns, moves to **re-election** campaign | reuters ...

[Anti-corruption campaigner set to win Slovakia's presidential vote | Reuters](#)
<https://www.reuters.com/article/us-slovakia-election-president-idUSKCN1OW2Z3>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/a3e7021f-3fcd-46f6-95a6-45a410c5d63c.html
Snippet : ...revious experience of public office is set to win slovakia's presidential **election** on saturday as voters spurn the ruling smer party one year after the mur...

[Anti-corruption campaigner set to win Slovakia's presidential vote | Reuters](#)
<https://www.reuters.com/article/us-slovakia-election-president-idUSKCN1OW2Z3>
/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/home/shuming/solr-7.7.0/Reuters/reutersnews/reutersnews/0dca3a73-de4e-497f-84b4-cc995f85697.html
Snippet : ...revious experience of public office is set to win slovakia's presidential **election** on saturday as voters spurn the ruling smer party one year after the mur...