STEPS USED FOR SPELLCORRECTOR, AUTOCOMPLETE AND SNIPPET:

To implement **spellCorrector** I used peter norvig's spellCorrector. Norvig's spell checker creates a serialized dictionary based on the data input given to it "big.txt". In order to create the data set for "latimes" site, big.txt was generated using the html files provided for latimes. Big.txt will now have all the words corresponding to the content of html files given in latimes. A java maven project was created which imported Apache Tika library and uses HTML Parser for parsing content through the html files provided. The client code is ir_hw5_1_latest.php accepts the query requests. Norvig's spellcorrector file "SpellCorrector.php" was included in the client php file which uses the big.txt that has been created. Big.txt generated using the java code was copied in the same root folder as the client php file and SpellCorrector.php, so that SpellCorrector could directly read it from same path. If a query or a term is wrongly typed, SpellCorrector will take the wrong word as input and suggest the correct word based on the serialized dictionary it created using the data available for latimes. The correct word is shown as the result with link which redirects to the top 10 Lucene results when query for correct word is made.

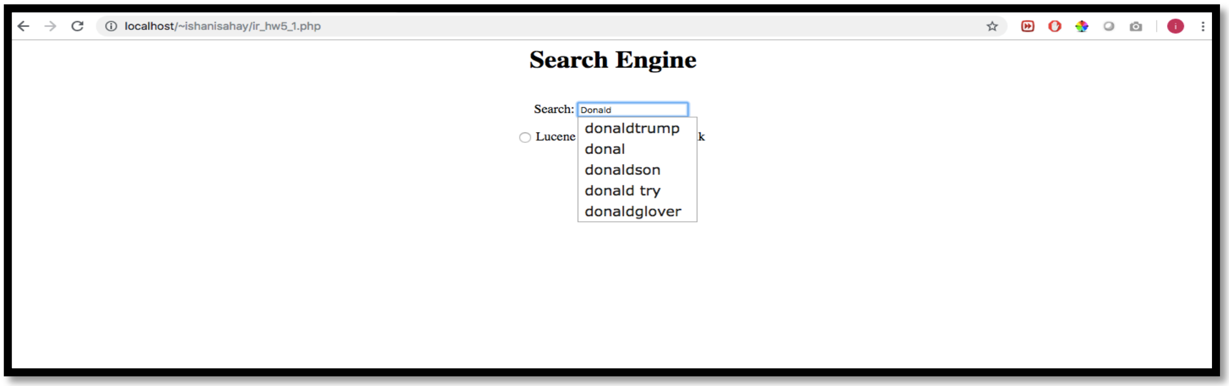
Autocomplete feature has been implemented using the solr setup we had done for assignment 4. For this I had modified the solrconfig.xml file. Solr.SuggestComponent was added in the solr.SearchHandler. With this new config of solr, if a query is fired with "/suggest", the new requestHandler which uses the suggestComponent is called. In order to make the query with suggest function, I added the new solr url(the localhost solr url for solr search engine) which would be called with the query term "q" and "suggest". The new url would make a request to solr for the query term using the suggest handler and return the json output. The url that fires the request has parameters for solr to return the response as json. The top five results returned by solr.SuggestComponent is displayed in the drop down of the query text box. When user selects one of the terms from the suggested list and hits submit, the top ten results for that term is returned based on the algorithm (Lucene / pagerank) selected.
If there is a spelling error in the query, autocomplete list should suggest the correct spell for the query.

Snippet is shown for the top ten results for the query term. For each result the contents of the page/url is retrieved using the file_get_contents() function. Special characters are removed from the content returned. A string corresponding to all the query terms is created. Using a regex match function, this string containing all the query terms is matched with each sentence of the file content. Since snippets are 156 characters in length and ellipses, the sentence match with 160 character count was generated as the snippet for the query term. In snippet, the words corresponding to the query terms have been highlighted. If for a specific file there is no snippet that matches the query terms, we display "N/A".
Config Note: As mentioned in the notes, the requestHandler of select in solrconfig.xml has been modified to add "<str name="q.op">AND</str>"so that the search results contain all the query terms.

As done in assignment4, the display results show Id, Title, outgoing url, description and snippet corresponding to each result.

AutoComplete Examples:

1) Donald:



localhost/~ishanisahay/ir_hw5_1.php

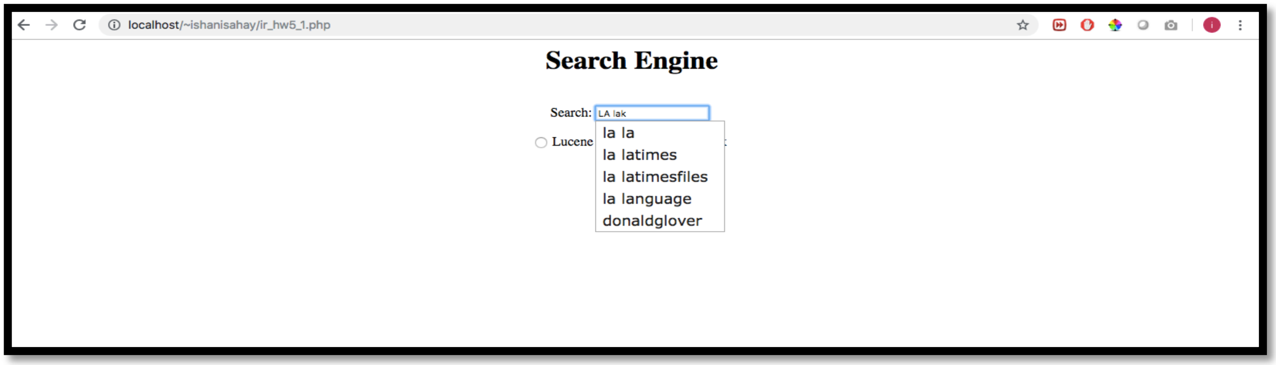
Search Engine

Search: Donald

○ Lucene

donaldtrump
donal
donaldson
donald try
donaldglover

2) LA lak:



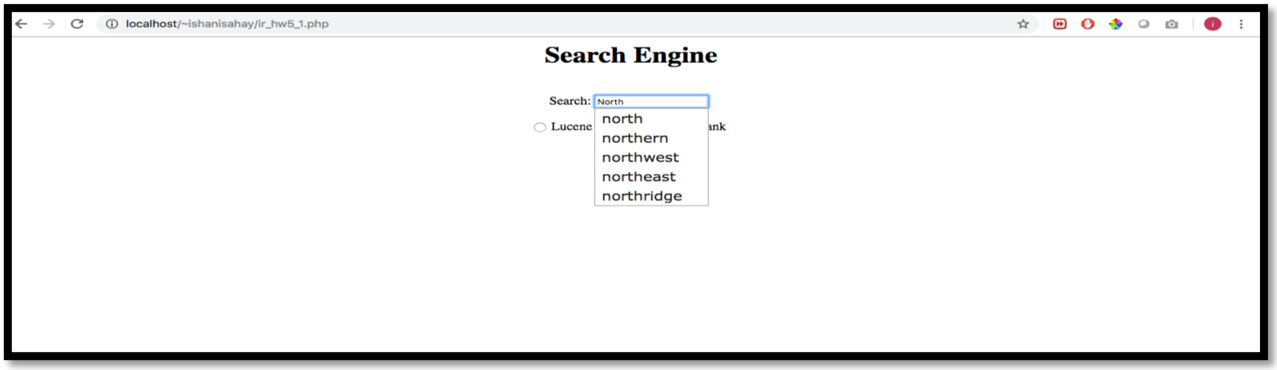
Search Engine

Search: LA lak

○ Lucene

- la la
- la latimes
- la latimesfiles
- la language
- donaldglover

3) North:



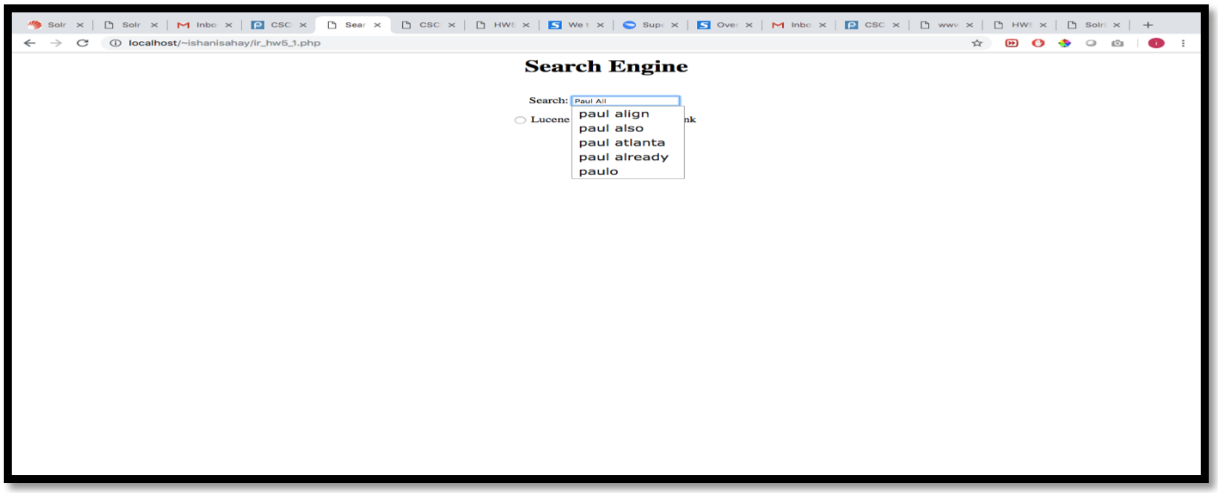
Search Engine

Search: North

○ Lucene ank

- north
- northern
- northwest
- northeast
- northridge

4) Paul Al:



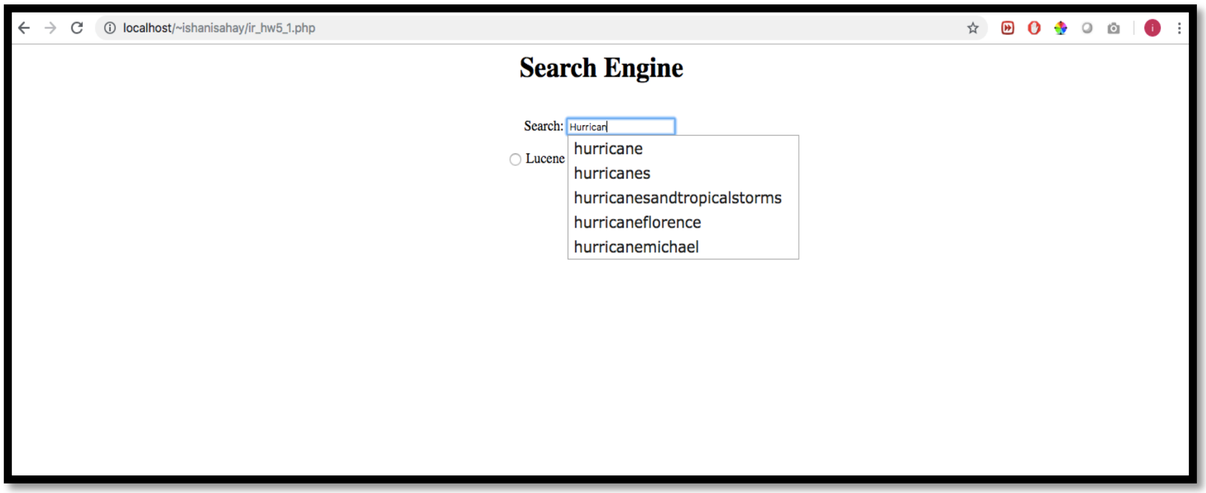
Search Engine

Search: Paul All

○ Lucene nk

- paul align
- paul also
- paul atlanta
- paul already
- paulo

5) Hurrican:



localhost/~ishanisahay/ir_hw5_1.php

Search Engine

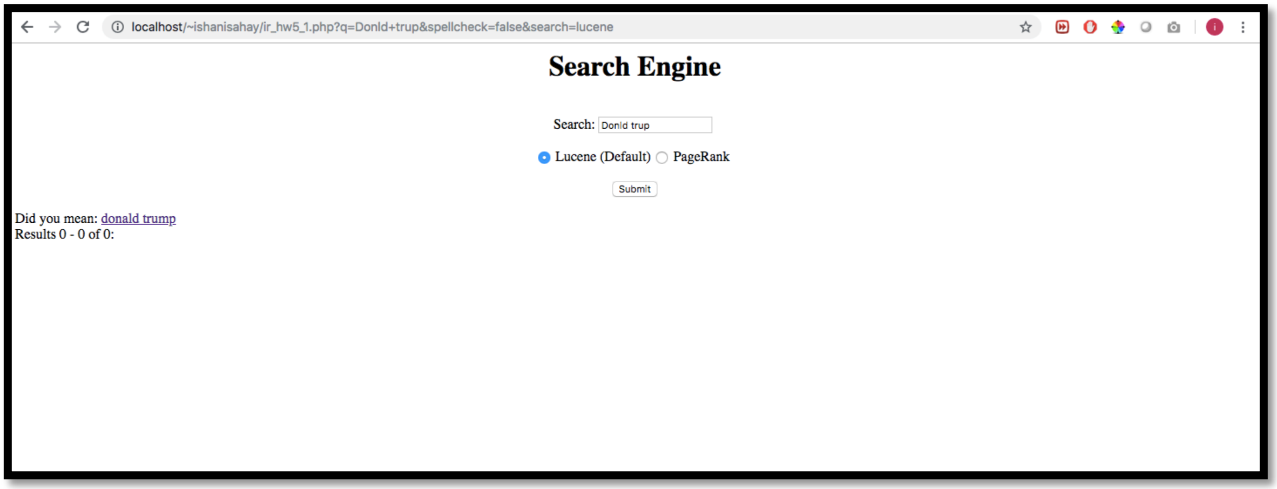
Search: Hurrican

```
hurricane
hurricanes
hurricanesandtropicalstorms
hurricaneflorence
hurricanemichael
```

○ Lucene

Spell Correct:

1) Donld trup:



localhost/~ishanisahay/ir_hw5_1.php?q=Donld+trup&spellcheck=false&search=lucene

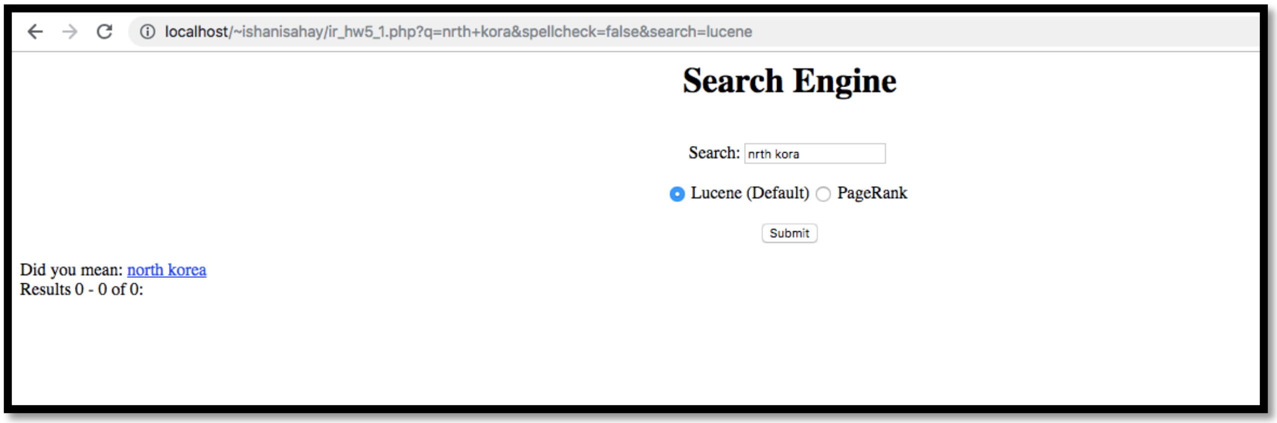
Search Engine

Search: Donld trup

● Lucene (Default) ○ PageRank

Did you mean: donald trump
Results 0 - 0 of 0:

2) Nrth kora:



localhost/~ishanisahay/ir_hw5_1.php?q=nrth+kora&spellcheck=false&search=lucene

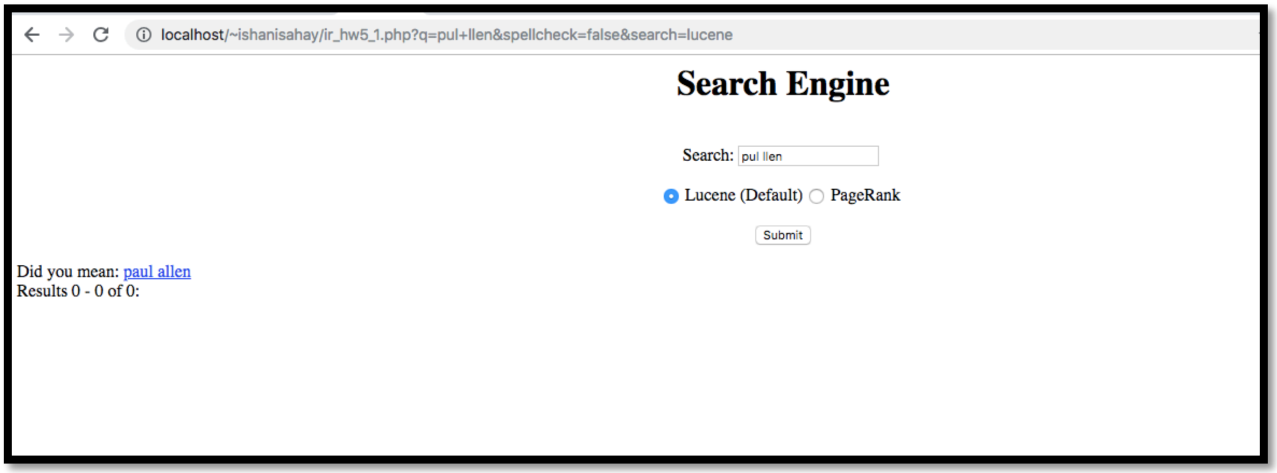
Search Engine

Search: nrth kora

● Lucene (Default) ○ PageRank

Did you mean: north korea
Results 0 - 0 of 0:

3) Pul llen:



localhost/~ishanisahay/ir_hw5_1.php?q=pul+llen&spellcheck=false&search=lucene

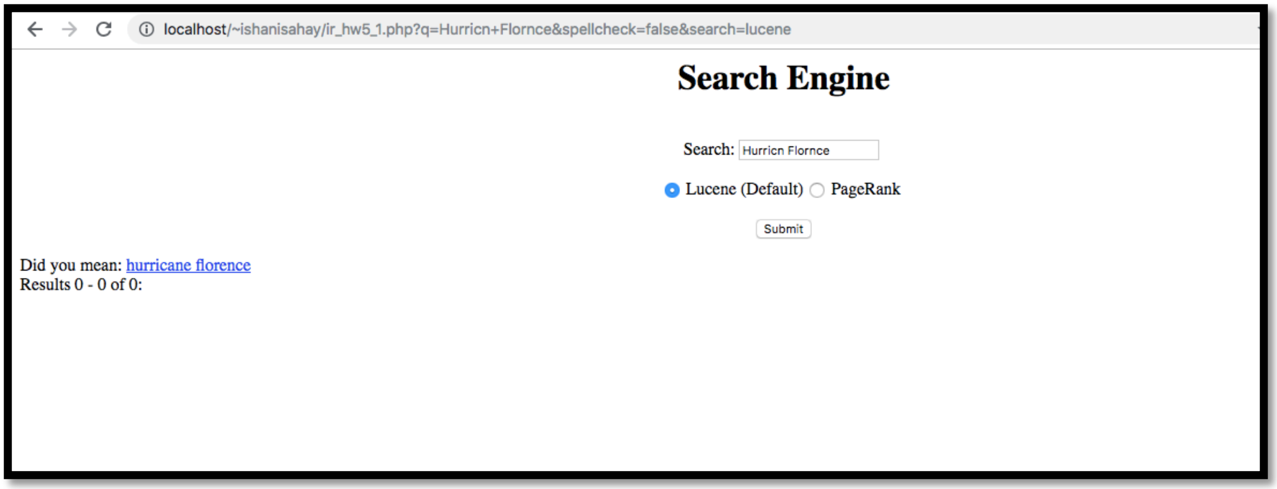
Search Engine

Search: pul llen

● Lucene (Default) ○ PageRank

Did you mean: paul allen
Results 0 - 0 of 0:

4) Hurrican Flornce:



localhost/~ishanisahay/ir_hw5_1.php?q=Hurricn+Flornce&spellcheck=false&search=lucene

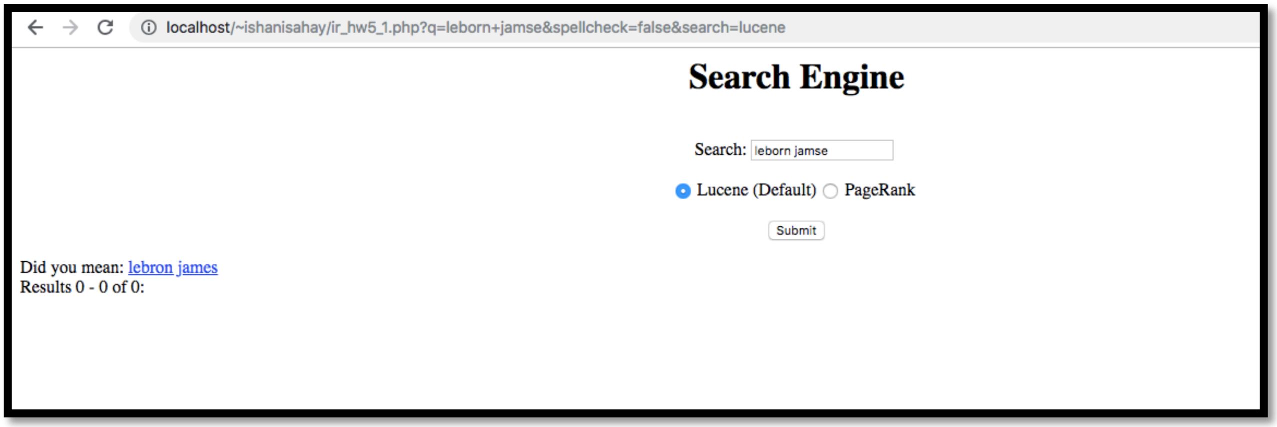
Search Engine

Search: Hurricn Flornce

● Lucene (Default) ○ PageRank

Did you mean: hurricane florence
Results 0 - 0 of 0:

5) Leborn jamse



localhost/~ishanisahay/ir_hw5_1.php?q=leborn+jamse&spellcheck=false&search=lucene

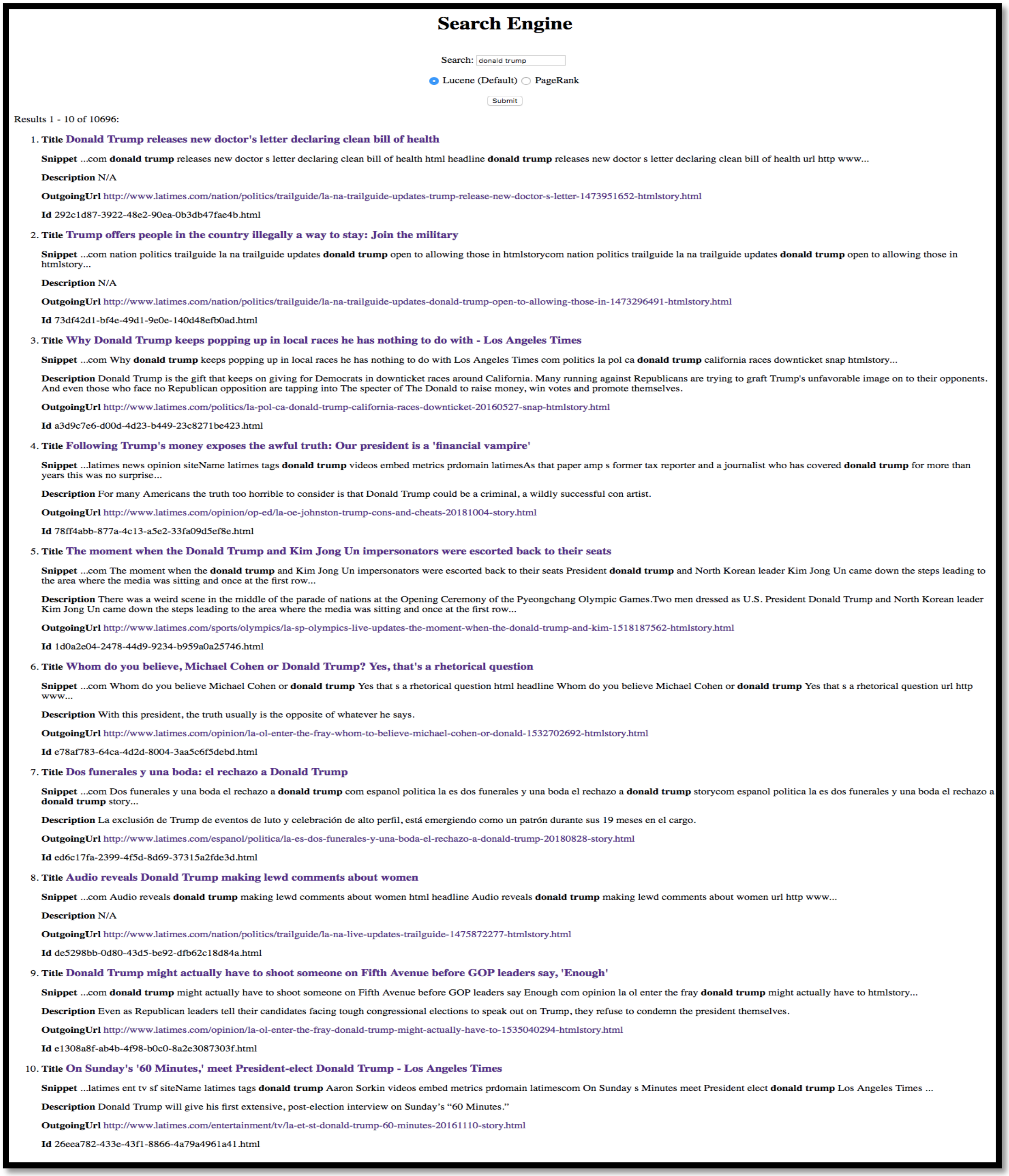
Search Engine

Search: leborn jamse

● Lucene (Default) ○ PageRank

Did you mean: lebron james
Results 0 - 0 of 0:

Below screenshot shows the results displayed when the suggested correct spelling term (Donald Trump) is clicked for the query Donld Trup . The results would be same as search done for correct query term.



Search Engine

Search: donald trump

● Lucene (Default) ○ PageRank

Submit

Results 1 - 10 of 10696:

1. **Title** Donald Trump releases new doctor's letter declaring clean bill of health

   Snippet ...com **donald trump** releases new doctor s letter declaring clean bill of health html headline **donald trump** releases new doctor s letter declaring clean bill of health url http www...

   Description N/A

   OutgoingUrl http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-trump-release-new-doctor-s-letter-1473951652-htmlstory.html

   Id 292c1d87-3922-48e2-90ea-0b3db47fae4b.html

2. **Title** Trump offers people in the country illegally a way to stay: Join the military

   Snippet ...com nation politics trailguide la na trailguide updates **donald trump** open to allowing those in htmlstorycom nation politics trailguide la na trailguide updates **donald trump** open to allowing those in htmlstory...

   Description N/A

   OutgoingUrl http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-donald-trump-open-to-allowing-those-in-1473296491-htmlstory.html

   Id 73df42d1-bf4e-49d1-9e0e-140d48efb0ad.html

3. **Title** Why Donald Trump keeps popping up in local races he has nothing to do with - Los Angeles Times

   Snippet ...com Why **donald trump** keeps popping up in local races he has nothing to do with Los Angeles Times com politics la pol ca **donald trump** california races downticket snap htmlstory...

   Description Donald Trump is the gift that keeps on giving for Democrats in downticket races around California. Many running against Republicans are trying to graft Trump's unfavorable image on to their opponents. And even those who face no Republican opposition are tapping into The specter of The Donald to raise money, win votes and promote themselves.

   OutgoingUrl http://www.latimes.com/politics/la-pol-ca-donald-trump-california-races-downticket-20160527-snap-htmlstory.html

   Id a3d9c7e6-d00d-4d23-b449-23c8271be423.html

4. **Title** Following Trump's money exposes the awful truth: Our president is a 'financial vampire'

   Snippet ...latimes news opinion siteName latimes tags **donald trump** videos embed metrics prdomain latimesAs that paper amp s former tax reporter and a journalist who has covered **donald trump** for more than years this was no surprise...

   Description For many Americans the truth too horrible to consider is that Donald Trump could be a criminal, a wildly successful con artist.

   OutgoingUrl http://www.latimes.com/opinion/op-ed/la-oe-johnston-trump-cons-and-cheats-20181004-story.html

   Id 78ff4abb-877a-4c13-a5e2-33fa09d5ef8e.html

5. **Title** The moment when the Donald Trump and Kim Jong Un impersonators were escorted back to their seats

   Snippet ...com The moment when the **donald trump** and Kim Jong Un impersonators were escorted back to their seats President **donald trump** and North Korean leader Kim Jong Un came down the steps leading to the area where the media was sitting and once at the first row...

   Description There was a weird scene in the middle of the parade of nations at the Opening Ceremony of the Pyeongchang Olympic Games.Two men dressed as U.S. President Donald Trump and North Korean leader Kim Jong Un came down the steps leading to the area where the media was sitting and once at the first row...

   OutgoingUrl http://www.latimes.com/sports/olympics/la-sp-olympics-live-updates-the-moment-when-the-donald-trump-and-kim-1518187562-htmlstory.html

   Id 1d0a2e04-2478-44d9-9234-b959a0a25746.html

6. **Title** Whom do you believe, Michael Cohen or Donald Trump? Yes, that's a rhetorical question

   Snippet ...com Whom do you believe Michael Cohen or **donald trump** Yes that s a rhetorical question html headline Whom do you believe Michael Cohen or **donald trump** Yes that s a rhetorical question url http www...

   Description With this president, the truth usually is the opposite of whatever he says.

   OutgoingUrl http://www.latimes.com/opinion/la-ol-enter-the-fray-whom-to-believe-michael-cohen-or-donald-1532702692-htmlstory.html

   Id e78af783-64ca-4d2d-8004-3aa5c6f5debd.html

7. **Title** Dos funerales y una boda: el rechazo a Donald Trump

   Snippet ...com Dos funerales y una boda el rechazo a **donald trump** com espanol politica la es dos funerales y una boda el rechazo a **donald trump** storycom espanol politica la es dos funerales y una boda el rechazo a **donald trump** story...

   Description La exclusión de Trump de eventos de luto y celebración de alto perfil, está emergiendo como un patrón durante sus 19 meses en el cargo.

   OutgoingUrl http://www.latimes.com/espanol/politica/la-es-dos-funerales-y-una-boda-el-rechazo-a-donald-trump-20180828-story.html

   Id ed6c17fa-2399-4f5d-8d69-37315a2fde3d.html

8. **Title** Audio reveals Donald Trump making lewd comments about women

   Snippet ...com Audio reveals **donald trump** making lewd comments about women html headline Audio reveals **donald trump** making lewd comments about women url http www...

   Description N/A

   OutgoingUrl http://www.latimes.com/nation/politics/trailguide/la-na-live-updates-trailguide-1475872277-htmlstory.html

   Id de5298bb-0d80-43d5-be92-dfb62c18d84a.html

9. **Title** Donald Trump might actually have to shoot someone on Fifth Avenue before GOP leaders say, 'Enough'

   Snippet ...com **donald trump** might actually have to shoot someone on Fifth Avenue before GOP leaders say Enough com opinion la ol enter the fray **donald trump** might actually have to htmlstory...

   Description Even as Republican leaders tell their candidates facing tough congressional elections to speak out on Trump, they refuse to condemn the president themselves.

   OutgoingUrl http://www.latimes.com/opinion/la-ol-enter-the-fray-donald-trump-might-actually-have-to-1535040294-htmlstory.html

   Id e1308a8f-ab4b-4f98-b0c0-8a2e3087303f.html

10. **Title** On Sunday's '60 Minutes,' meet President-elect Donald Trump - Los Angeles Times

    Snippet ...latimes ent tv sf siteName latimes tags **donald trump** Aaron Sorkin videos embed metrics prdomain latimescom On Sunday s Minutes meet President elect **donald trump** Los Angeles Times ...

    Description Donald Trump will give his first extensive, post-election interview on Sunday’s “60 Minutes.”

    OutgoingUrl http://www.latimes.com/entertainment/tv/la-et-st-donald-trump-60-minutes-20161110-story.html

    Id 26eea782-433e-43f1-8866-4a79a4961a41.html