Homework Assignment 5

I have used same environment for assignment 5 as I have used in the assignment 4. Following is the environment setup.

Environment Setup:

- 1. Installed ubuntu operating system and all the required software.
- 2. Downloaded apache solr from http://www.apache.org/dyn/closer.lua/lucene/solr/7.5.0 and installed it.
- 3. Apache Solr is launched using the command bin/solr.cmd start. Verified the installation of Solr from browser using http://localhost:8983/solr
- 4. Created a core called irassignment4_version1 Using bin/solr create -c irassignment4_version1.
- 5. Given my usc-id, I am working on LA times. Downloaded webpage dump from drive and extracted it.
- 7. At the end of this step, we have environment with apache solr installed and LA times webpages are indexed.

Auto-complete:

To create auto complete functionality used **java** rest api's are used. As soon as there is change detected in the input box(using angular ng-change), /suggest rest api is invoked. through rest calls.

Following changes are made in solr-config.xml to enable suggestions in apache solr.

```
<searchComponent class="solr.SuggestComponent" name="suggest">
<|st name="suggester">
   <str name="name">suggest</str>
   <str name="lookuplmpl">FuzzyLookupFactory</str>
   <str name="field">_text_</str>
   <str name="suggestAnalyzerFieldType">string</str>
 </searchComponent>
<requestHandler class="solr.SearchHandler" name="/suggest">
  <|st name="defaults">
   <str name="suggest">true</str>
   <str name="suggest.count">5</str>
   <str name="suggest.dictionary">suggest</str>
  </lst>
  <arr name="components">
   <str>suggest</str>
  </arr>
 </requestHandler>
```

Snapshots of Auto completion: x 🗋 localhost:8080/search-en x 🖪 Facebook x | 💀 Microsoft Word Online x | 🕟 IR Assignment 4.docx x | 💽 YouTube × Netflix ← → C ☆ ① localhost:8080/search-engine/# G Google 💆 Twilter 🏿 To-Do-Platform 🖍 youtube 🖍 Stuff 🦍 TECH 🖍 MS 🐚 DL 🐚 Courses 🐚 Preperation 🕒 CSCI 572 🗋 CSCI 567 📳 USC connectsc. 📳 Viterbi Gateway Enter the query Hillar Hide dropdown Select Algorithm hillary hillaryclinton hilarious hilary hilariously Activities 🧐 Google Chrome ▼ × 🕒 localhost:8080/search-en × 📑 Facebook x | 📳 Microsoft Word Online x | 📳 IR Assignment4.docx x | 💿 YouTube × Netflix × | + \leftarrow \rightarrow C \bigcirc localhost:8080/search-engine/# G Google 💆 Twitter 🏿 To-Do-Platform 🗈 youtube 🗈 stuff 🛍 TECH 🛍 MS 🛍 DL 🛍 Courses 🛍 Preperation 🖰 CSCI 572 🖰 CSCI 567 📳 USC connectsc 📳 Viterbi Gateway Enter the query calif Hide dropdown Select Algorithm calif california california's california's californians x 🗋 localhost:8080/search-en x 📳 Facebook x | 😰 Microsoft Word Online x | 😨 IR Assignment4.docx x | 💿 YouTube G Google \leftarrow \rightarrow C \bigcirc (i) localhost:8080/search-engine/# G Google 📝 Twitter 🖫 To-Do-Platform 🛍 youtube 🛍 stuff 🛍 TECH 🛍 MS 🚵 DL 🛍 Courses 🛍 Preperation 🖰 CSCI 572. 🗋 CSCI 567 📳 USC connectsc 📳 Viterbi Gateway Enter the query Donald Hide dropdown Select Algorithm donald donaldtrump donaldtrumpjr donaldson donal G Google X 🗋 localhost:8080/search-en X 📳 Facebook X | 🕡 Microsoft Word Online X | 🕡 IR Assignment4.docx X | 💌 YouTube X | 📉 Netflix ← → C ① ① localhost:8080/search-engine/# ☆ () (e) G Google 💆 Twitter 🛄 To-Do-Platform 🛍 youtube 🛍 stuff 🛍 TECH 🛍 MS 🛍 DL 🛍 Courses 🛍 Preperation 🖰 CSCI 572 🖰 CSCI 567 👢 USC connectsc 👢 Viterbi Gateway Enter the query Micro Hide dropdown Select Algorithm micro microsoft microscopic mirror microbes



Misspelled Term correction:

Snapshots:

1. Used Norvig's Spell Correction Java Version program.

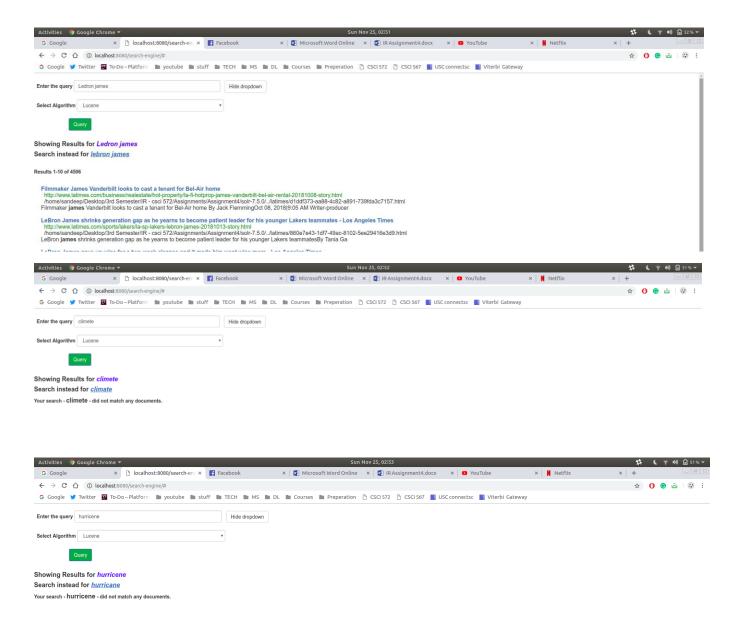
Trump finally comes to California — to see donors and border wall prototypes, keeping protesters at a distance - Los Angeles Times http://www.latimes.com/politics/la-na-pol-flurmp-california-20180313-storys html
//home/sandeep/Desklop/dro Semester/IR - cost of 572/Assignments/Assignment4/soir-7.5.0/_/latimes/800c2bbb-2b8d-4d8a-af09-e8ca3a778eb5.html
Politics trump finally comes to California — to see donors and border wall prototypes, keeping protesters at a distance

Trump's complicated relationship with California, explained by his own tweets

http://www.latimes.com/politics/la-pol-ca-frump-california-border-201803-trump-s-complicated-relationship-with-1520952968-htmlstory.html
//home/sandeep/Desklop/drd Semester/IR - cost 572/Assignments/Assignment4/soir-7.5.0/_/latimes/d583410f-752a-4bda-b4d3-d576arf2c836.html
and ad-Politics trump spent 22 hours in California

- 2. Generate big.txt using Apache Tika and iterating over all the html files of LA times.
- 3. Invoked the spell correction from the UI when user request for search, showed the suggestion and when suggestion is clicked, corresponding results are showed.

Activities 🤴 Google Chrome ▼ x 🗋 localhost:8080/search-en x 📑 Facebook x 📳 Microsoft Word Online x 📳 IR Assignment-1.docx x 💌 YouTube ← → C ① ① localhost:8080/search-engine/# ☆ O @ @ @ : G Google 🐓 Twitter 🖫 To-Do-Platform 🛍 youtube 🛍 stuff 🛍 TECH 🛍 MS 🛍 DL 🛍 Courses 🛍 Preperation 🖰 CSCI 572 🖰 CSCI 567 📳 USC connectsc 📳 Viterbi Gateway Hide dropdown Select Algorithm Lucene Showing Results for North korae Search instead for north korea Trump becomes first U.S. president to meet a North Korean head of state http://www.latimes.com/worldla-fig-trump-kim-north-korea-summil-trump-becomes-firsi-u-s-president-to-1528766878-htmistory.htmome/sandeep/Desktop/3rd Semester/iR - csci 572/Assignments/Assignment4/solr-7.5.0/_.latimes/bco3e78b-3348-43dh-82ce-cTrump and Morth Korea's Kim Jong Un32 postsExplainersSceneClear 32 postsfilter ExplainersSceneAdvertisementgo to new pos cehce863e3ca html will i trump pring up form Acrea's preadur lumain rights rector when ne meets Kint / http://www.latimes.com/worldia-fg-trump-kinn-north-korea-summit-will-president-trump-bring-up-north-1528706649-htmlstory.html /home/sandeep/Deshop/dr Semesterfill R-csci 572/painers/Sene(Deshop-dr S-20, /latimas/6a39967-8a2e-4661-a934-4264037968.html Trump and North Korea's Kin Jong Un32 postSeptioners/Sene(Desa'2 SpostSiffer Explainers/Sene(Park)-deshop-dr S-20, /latimas/Basil-grade-deshop-dr S-20, /latimas/Basil-grade-dr S-20 Activities 7 Google Chrome * Sun Nov 25, 02:50 x 🖰 localhost:8080/search-en x 📳 Facebook x 📳 Microsoft Word Online x 📳 IR Assignment 4.docx x o YouTube ← → C ① ① localhost:8080/search-engine/# # O @ @ E G Google 🐓 Twitter 🔛 To-Do-Platform 🛍 youtube 🛍 stuff 🛍 TECH 🛍 MS 🛍 DL 🛍 Courses 🛍 Preperation 🐧 CSCI572 🐧 CSCI567 📓 USC connectsc 🕵 Viterbi Gateway Enter the query Dobald trump Hide dropdown Select Algorithm Lucene Showing Results for Dobald trump Search instead for donald trump Results 1-10 of 10530 Trump has arrived in Los Angeles. Here's what happens next http://www.latimes.com/politics/la-pol-a-frump-california-border-201803-frump-has-arrived-in-tos-angeles-here-s-1520980807-htmlstory.html/home/sandeep/Desktop/3rd Semester/IR - csci 572/Assignments/Assignment4/soir-7.5.0/./latimes/czd22544-0f54-43da-b348-3aa0519680da.html an adPolitics trump spent 22 hours in California



Steps to run the code:

Front-end part of the application built using AngularJS, HTML and CSS and Bootstrap.

Backend part of the application is built using java and communication to the front-end is over rest api.

The code is written in java plus angular js and maintained through maven. So, for the ease I have created a .war file called search-engine.war

Download apache-tomee-plus-8.0.0-M1 and copy search-engine.war to the Webapps folder.

Make sure that apache tomee http is running on 8080 port (check in conf/server.xml)

Run bin/catalina.sh run or bin/startup.sh

Go to any browser and run http://localhost:8080/search-engine

It pops with the search form, you can give queries and check the results.

PRECAUTIONS:

Make sure the folder where you copied apache tomee won't have any spaces in its absolute path