

CSCI 572 Information Retrieval and Web Search Engines

Homework 5: Enhancing Your Search Engine

Name : Anubhav Jindal

USC ID : 5963113610

Steps followed to complete the Assignment:

1. Spelling Correction:

- a. Downloaded Apache Tika Jar and included it in the build path of my Java project.
- b. Wrote Java code to extract the Text from the HTML files(bigtxt.java) and write to big.txt
- c. Downloaded Peter Norvig's PHP code for Spelling Correction after signing up (SpellCorrector.php)
- d. I provided the path of my big.txt file to Peter's code.
- e. To execute the SpellCorrector.php, I included the file in my index.php file and made a call to the correct () function.
- f. After the execution, a file serialized_dictionary.txt is generated which is used for spelling correction.

2. Autocomplete:

- a. Changes were made to Solrconfig.xml as specified in the document provided as a part of the assignment.
- b. I reused the code from Assignment 4 (index.php) and made modifications to it to include Autocomplete function.
- c. The tutorial and code mention on <http://api.jqueryui.com/autocomplete/> was used for reference.
- d. The suggested terms, for any Query term were parsed (json) and stored in a variable.
- e. All the suggestions are iterated through and shown to the user as a drop-down list.

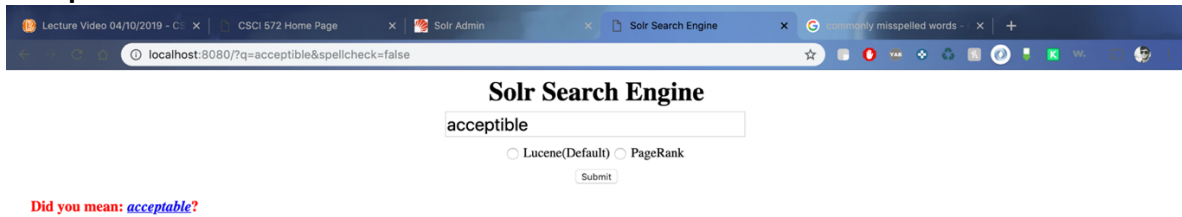
3. Snippet:

- a. In order to parse the HTML pages in PHP, I used simple_html_dom.php and included it in my index.php file.
- b. The retrieved text was divided into sentences and words.
- c. The rule followed to obtain the snippet was:
 - i. For a multi-term query, obtain a sentence in which all terms occur together if possible.
 - ii. Else the sentence containing the words is obtained if possible.
 - iii. Else the first sentence containing one of the terms is obtained if possible
 - iv. Else look into the description associated for the query words.
 - v. Else N/A is returned to the user.
- d. The query terms were put in bold within the snippet for easy identification.

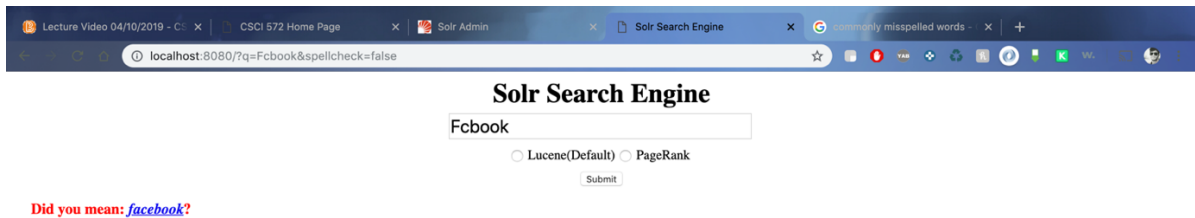
Analysis of the Results:

1. Examples of Misspelled words:

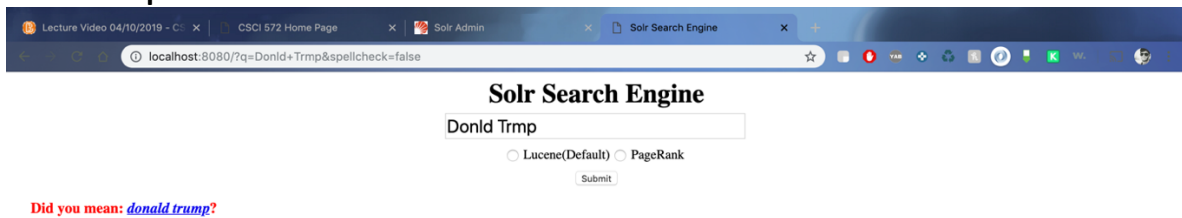
a. Acceptible



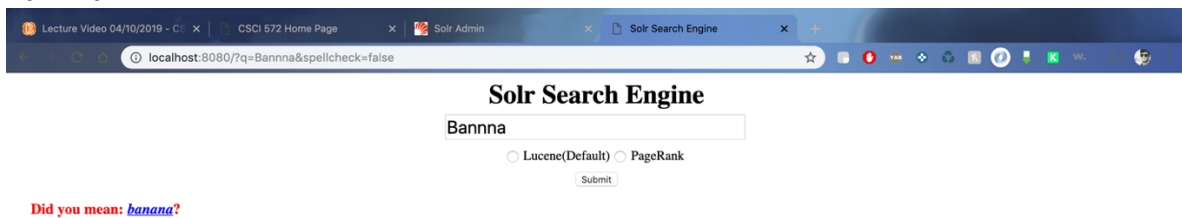
b. Fcbook



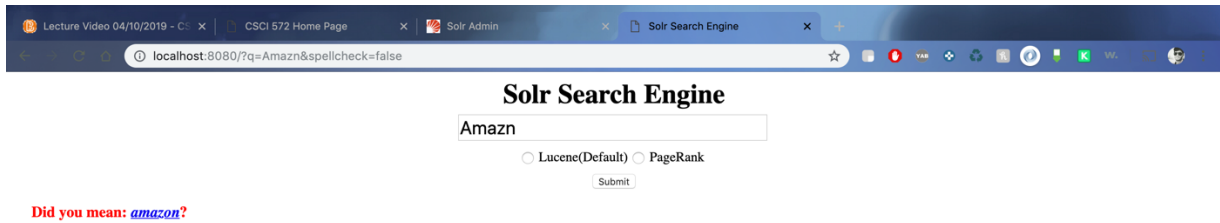
c. Donld Trmp



d. Bannna

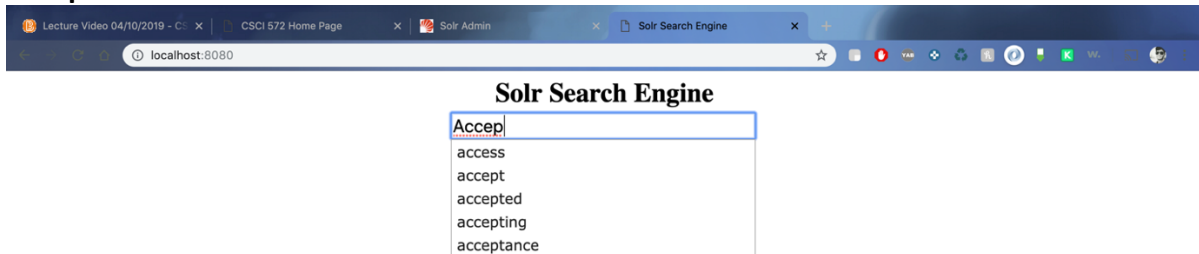


e. Amzn

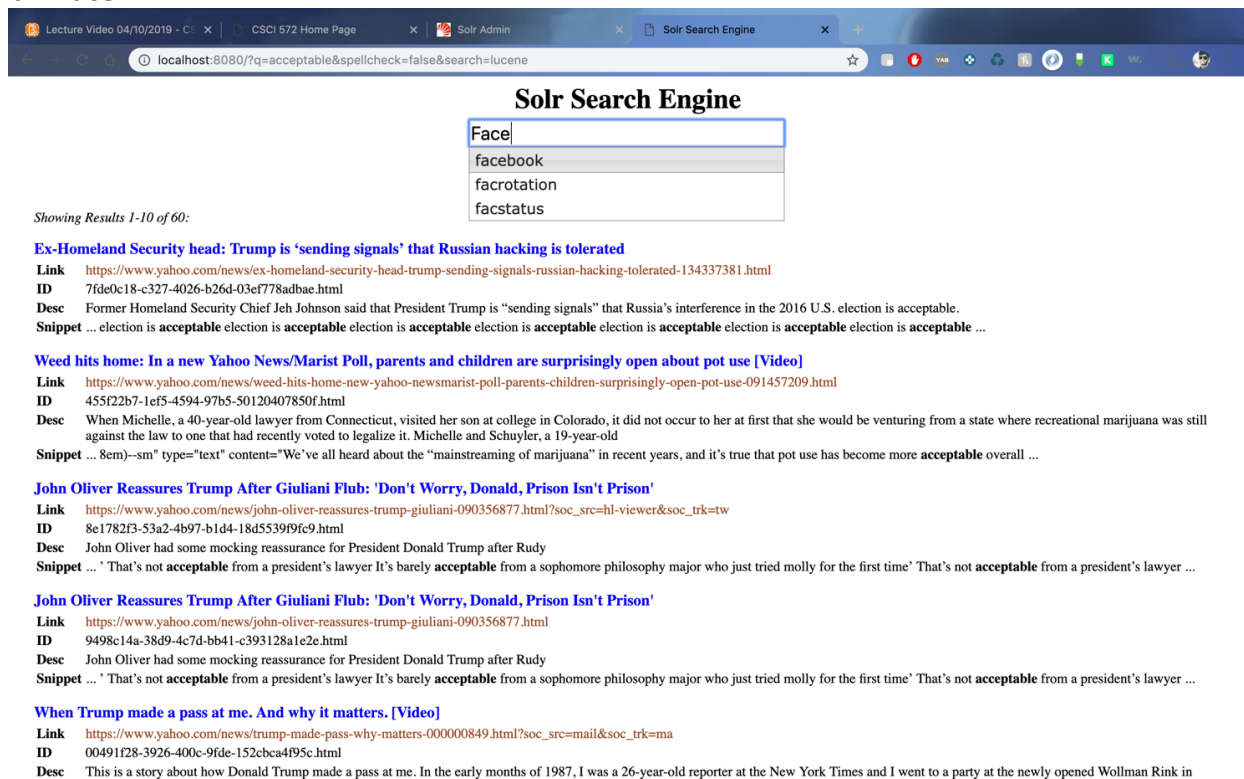


2. Examples of Autocomplete:

a. Accep



b. Face



c. Amaz

Solr Search Engine

Amaz

amazing
amazon
amanda
against

Showing Results 1-10 of 8761:

Mark Zuckerberg speaks at the Facebook F8 developer conference and more: April 18 in photos

Link <https://www.yahoo.com/news/mark-zuckerberg-speaks-facebook-f8-slideshow-wp-203001167/photo-p-facebook-founder-ceo-mark-photo-203001635.html>

ID 1ee2b21d-6497-4117-a816-41d4e764299a.html

Desc Facebook founder and CEO Mark Zuckerberg speaks during the annual Facebook F8 developer conference in San Jose, Calif.; President Trump speaks during a visit to the headquarters of tool manufacturer Snap-On in Kenosha, Wis.; Pennsylvania state police investigate the scene where Steve Stephens, the suspect in the random killing of a Cleveland retiree posted on Facebook, died in Erie, Pa. These are some of the photos of the day. (AP/EPA/Getty/Reuters)

Snippet ...; Pennsylvania state police investigate the scene where Steve Stephens, the suspect in the random killing of a Cleveland retiree posted on **facebook**, died in Erie, Pa ...

Zuckerberg hedges on how Facebook tracks you

Link <https://www.yahoo.com/news/zuckerberg-hedges-facebook-tracks-214550221.html>

ID 8bb90e42-ccf8-432e-8c05-5c81e8a72bc2.html

Desc Some questions about Facebook are too tricky for its chief executive to answer.

Snippet ... 8em)--sm" type="text" content="Some questions about **facebook** are too tricky for its chief executive to answer, asked **facebook**'s founder and CEO whether the company collected "user data through ...

Facebook data-sharing deals with major tech companies under investigation in criminal inquiry

Link <https://www.yahoo.com/news/facebook-data-sharing-deals-major-131611255.html>

ID a189d0a9-efb5-44de-9a0e-7fe1c4b98f61.html

Desc Federal prosecutors are conducting a criminal investigation into data deals Facebook struck with some of the world's largest technology companies, intensifying scrutiny of the social media giant's business practices as it seeks to rebound from a year of scandal and setbacks. Both companies had entered

Snippet ... 8em)--sm" type="text" content="Both companies had entered into partnerships with **facebook**, gaining broad access to the personal information of hundreds of millions of its users ...

Facebook vs. ISIS: Inside the tech giant's antiterror strategy [Video]

Link <https://www.yahoo.com/news/facebook-vs-isis-inside-the-1352650033922102.html>

ID b9f74d66-5523-4c28-a2a7-c98bddcd824.html

Desc Facebook, under pressure to crack down on Islamic State militants using its network, has quietly ramped up its efforts to block terrorist messages and videos in what some experts say is a potentially significant move by the Silicon Valley giant in the U.S. government's battle against the terror group

Snippet ... Bickert, **facebook**'s head of global policy management, says the company is doing everything it can to cooperate with the U8em)--sm" type="text" content="**facebook**, under pressure to crack down on Islamic State militants using its network, has quietly ramped up its efforts to block terrorist messages and videos in what some experts say is a potentially significant move by the Silicon Valley giant in the U ...

d. Bara

Solr Search Engine

Bara

bar
barack
brand
balance

Showing Results 1-10 of 26:

Microsoft Says It Stopped Cyberattacks on Three 2018 Congressional Candidates

Link <https://www.yahoo.com/news/microsoft-says-stopped-cyberattacks-three-195333274.html>

ID e050771b-947a-48bd-a510-99b9c8e0ce35.html

Desc Microsoft VP says the same tactic was used by hackers in 2016 election

Snippet ... 8em)--sm" type="text" content="(Bloomberg) — **microsoft** Corp8em)--sm" type="text" content="The targets, who it didn't identify, were "all people who because of their positions might have been interesting from an espionage standpoint, as well as an election disruption standpoint," Tom Burt, **microsoft**'s corporate vice president for customer security and trust, said at the Aspen Security Forum in Colorado on Thursday ...

Don't Break Up Big Tech

Link <https://www.yahoo.com/news/don-t-break-big-tech-103041413.html>

ID bebddb21-fa7e-4241-b2c1-524ab3a58236.html

Desc The worst ideas in Washington are often bipartisan. Big Tech is about to learn this lesson, if it hasn't already.Elizabeth Warren is out with a headline-grabbing proposal to break up Big Tech companies, the sort of overly ambitious government plan that once would have engendered knee-jerk Republican

Snippet ... 8em)--sm" type="text" content="It's not true, as Warren asserts, that the antitrust suit against **microsoft** in the 1990s opened up the space for Google and Facebook to thrive ...

U.S. prosecutors probing Facebook's data deals: New York Times

Link <https://www.yahoo.com/news/u-prosecutors-probing-facebooks-data-deals-nyt-0131446617--finance.html>

ID fb18a0e0-c6a9-44ab-8ebb-6b981bd13473.html

Desc U.S. federal prosecutors are conducting a criminal investigation into data deals Facebook Inc struck with some of the world's largest technology companies, the New York Times reported on Wednesday. A grand jury in New York has subpoenaed records from at least two prominent makers of smartphones

Snippet ... com Inc, Apple Inc and **microsoft** Corp, that have entered into partnerships with Facebook for access to the personal information of hundreds of millions of its users, according to the report ...

Starbucks

Link <https://www.yahoo.com/news/topics/starbucks-1490333697>

ID 8faf1700-0bca-420f-8ef4-1d930ff5bc60.html

Desc

Snippet ... King County Executive Dow Constantine will kick off Changemakers' Rally, followed by a program with representatives from Starbucks Coffee Company, Seattle Mariners, **microsoft** and others ...

Facebook data-sharing deals with major tech companies under investigation in criminal inquiry

Link <https://www.yahoo.com/news/facebook-data-sharing-deals-major-131611255.html>

e. Dona

Lecture Video 04/10/2019 - CSCI 572 Home Page Solr Admin Solr Search Engine

localhost:8080/?q=barack&spellcheck=false&search=lucene

Solr Search Engine

done

don't

donald

don't

Showing Results 1-10 of 1040:

Barack Obama, Hillary Clinton, and other world leaders are posting messages in the wake of the New Zealand mosque shootings [Video]

Link <https://www.yahoo.com/news/barack-obama-hillary-clinton-other-170447931.html>

ID 489564e2-6533-4757-8075-f6d7a4754dab.html

Desc Today, March 15th, two mosques in the city of Christchurch, New Zealand, were targeted in two mass shootings. According to CNN, at least 49 people have died in the New Zealand mosque shootings so far, and

Snippet ... html%3Fsoc_src%3Dhl-viewer%26soc_trk%3Dfb&name=barack%20Obama%2C%20Hillary%20Clinton%2C%20and%20other%20world%20leaders%20are%20posting%20messages%20in%20the%20wake%20of%20the%20New% ...

'American values are at stake': Barack Obama 'heartened' by resistance to Donald Trump

Link <https://www.yahoo.com/news/american-values-are-at-stake-barack-obama-heartened-by-resistance-to-donald-trump-195026090.html>

ID c3f9bbd4-7585-44e3-9844-e14620d60533.html

Desc Former President Barack Obama broke his postpresidential silence on Monday, issuing a cautious statement about President Trump's travel ban and the protests against it. Before leaving the White House, Obama repeatedly emphasized how important the role of citizen is in American democracy. Now, he says

Snippet ... 8em)--sm" type="text" content="Former President barack Obama broke his postpresidential silence on Monday, issuing a cautious statement about President Trump's travel ban and the protests against it ...

Barack Obama: Going for broke in the bin Laden raid

Link <https://www.yahoo.com/news/barack-obama-going-broke-bin-laden-raid-090035949.html>

ID cc3906ee-7451-443d-8609-09b8e375751a.html

Desc Author Peter Bergen discusses Barack Obama's defining moment of leadership.

Snippet ... html%3Fsoc_src%3Dhl-viewer%26soc_trk%3Dfb&name=barack%20Obama%3A%20Going%20for%20broke%20in%20the%20bin%20Laden%20raid&description=Author%20Peter%20Bergen%20discusses%20barack%20Obama% ...

Barack, Michelle Obama sign bumper book deal

Link <https://www.yahoo.com/news/barack-michelle-obama-sign-bumper-book-deal-publisher-003153519.html>

ID ccfl1ddab-b729-4e88-9fbd-879f1f78a295.html

Desc Barack and Michelle Obama have signed a deal to publish their memoirs with New York-based Penguin Random House, in a coveted contract reportedly worth tens of millions of dollars. America's first African American president is already the author of two memoirs and a children's book. Michelle

Snippet ... 8em)--sm" type="text" content="New York (AFP) - barack and Michelle Obama have signed a deal to publish their memoirs with New York-based Penguin Random House, in a coveted contract ...