

## Assignment 5 Report

### Adding Spell Checking, AutoComplete and Snippets to Your Search Engine

#### Steps followed to implement Spell Correction

Used Apache Tika to parse the Newsday HTML files provided and extract content and metadata from the files to create the big.txt file which is used as a vocabulary for Spell checking and Autocomplete suggestion.

Used Peter Norvig Spell correction program to build serialized dictionary of known words which maintains a frequency of each word. Peter Norvig's program uses edit distance to create variation of given search query term.

When the user inputs the query term in search bar the program checks whether the entered query term is correct or not. The results for the correct word is displayed and a link to the original input query term is displayed.

#### Steps followed to implement Auto correction

Made changes to solrconfig.xml file as provided in the assignment instructions document. SuggestComponent was added, the field is used to obtain terms for suggestion. After adding the search component, a request handler must be added to solrconfig.xml.

Once the request handler is defined in solrconfig.xml the core is reloaded from solr UI.

#### Steps followed to implement Snippet Generation

To generate snippet, used html dom parser and utilized function to get the text from a given webpage. Then split the result obtained using explode function to check if any sentence contain full or partial query and return the result.

#### Examples of misspelled terms

Incorrect word	Correct word
Donad Trup	Donald Trump
Chnnai	Chennai
Aprection	Appreciation
Ohnesty	Honesty
Muntain	mountain

#### Examples of Autosuggestion

Query	Autocompletion
Congr	Congratulate Congress Congressional Congressman

	Congrats Congratulatory Congregation
affi	Affiliate Affirms Affidavit Affinity
relia	Reliability Reliance Reliant
Pers	Persian Persisited Person Persona
tr	Trace Tracy Track Traction

## SCREENSHOTS

### Autosuggest

The screenshot shows a web browser window titled "Solr Search - Mozilla Firefox". The address bar displays the URL: `localhost/IR_HW5.php?q=hurricane+irma&hid=no&algo=PageRank+Algorithm`. The search bar contains the text "Search: appl". A dropdown menu is visible, showing suggestions: "applauded", "applause", "apple", "applewood", "appliances", and "applicable". The "Submit" button is highlighted in red. Below the search bar, the results are displayed as a list of three items, each with a link, title, ID, and snippet.

Results 1 - 10 of 25:

- [Link](#)

Title : Donald Trump's noteworthy tweets as presid

ID : /home/radhika/Desktop/crawl\_data/data/c1878

Link : <https://www.newsday.com/news/nation/donald-trump-s-noteworthy-tweets-as-president-1.12632966>

Snippet : 'Hurricane Irma is of epic proportion, perhaps bigger than we have ever seen' Photo Credit: @realDonaldTrump via Twitter Trump showed his preference for grand descriptions while relaying a public safety message in this tweet on Friday, Sept...
- [Link](#)

Title : Poll: Sexual misconduct allegations voted top news story of 2017 | Newsday

ID : /home/radhika/Desktop/crawl\_data/data/6dfa71abaa10408ec9f6645ff76018f0.html

Link : <https://www.newsday.com/news/nation/ap-poll-sexual-misconduct-allegations-voted-top-news-story-1.15619939>

Snippet : No snippet for given query
- [Link](#)

Title : Traveling to the Virgin Islands, Puerto Rico and more destinations post-hurricane season | Newsday

ID : /home/radhika/Desktop/crawl\_data/data/507ba32c27e335ae557715f9a58fe94f.html

Link : <https://www.newsday.com/travel/caribbean-vacation-after-hurricane-1.16912094>

Snippet : Beaches Turks and Caicos resort, pictured, which sustained damage during Hurricane Irma, opened Dec...

CSCI572-Linux [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Solr Search Solr Admin

localhost/IR\_HW5.php?q=hurricane&hid=no&algo=PageRank+Algorithm

Search: Donald Tru

Submit

Page

Donald truck

Donald trudeau

Donald true

Donald truffie

Donald truly

Donald trumaine

Donald trump

Results 1 - 10 of 339:

- [Link](#)

Title : Editorials, Columns, Op-Eds, Letters & Opini

ID : /home/radhika/Desktop/crawl\_data/data/261db

Link : https://www.newsday.com/opinion

Snippet : No snippet for given query
- [Link](#)

Title : Long Island Travel Ideas, Sightseeing, Articles & Advice | Newsday

ID : /home/radhika/Desktop/crawl\_data/data/21404c91c7a0cee0aed08de46f8c7c2c.html

Link : https://www.newsday.com/travel

Snippet : CONNECT ACCOUNT SUBSCRIBE Special projects How LI political connections helped Oheka's owner thrive One man's rise shines light on LI's corrosive system Manage the Mets: Create your lineup card What hospitals offer expecting parents Travel 25 weekend getaways every New Yorker should take One-tank beach getaways Cruises, tours, more travel deals this week A glimpse inside Resorts World Catskills Tips to save money on European vacations 7-Elevens in Hawaii have things you've probably never heard of In Bhutan, gross national happiness trumps GDP 5 girlfriend getaways close to LI, NYC Planning a **post-hurricane** Caribbean trip 10 real-world locations from popular films to visit 4 spring trips to take with your family Tips for making road trips less stressful, expensive Top U...
- [Link](#)

Type here to search

95%

3:24 PM

4/25/2018

## Spelling correction and Snippet Generation

CSCI572-Linux [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Solr Search Solr Admin

localhost/IR\_HW5.php?q=hurricane&hid=no&algo=PageRank+Algorithm

Search: hurricane

Submit Query

PageRank Lucene

Results 1 - 10 of 339:

- [Link](#)

Title : Editorials, Columns, Op-Eds, Letters & Opinion Articles | Newsday

ID : /home/radhika/Desktop/crawl\_data/data/261dbb56ccff1929ee407e304b34585a.html

Link : https://www.newsday.com/opinion

Snippet : No snippet for given query
- [Link](#)

Title : Long Island Travel Ideas, Sightseeing, Articles & Advice | Newsday

ID : /home/radhika/Desktop/crawl\_data/data/21404c91c7a0cee0aed08de46f8c7c2c.html

Link : https://www.newsday.com/travel

Snippet : CONNECT ACCOUNT SUBSCRIBE Special projects How LI political connections helped Oheka's owner thrive One man's rise shines light on LI's corrosive system Manage the Mets: Create your lineup card What hospitals offer expecting parents Travel 25 weekend getaways every New Yorker should take One-tank beach getaways Cruises, tours, more travel deals this week A glimpse inside Resorts World Catskills Tips to save money on European vacations 7-Elevens in Hawaii have things you've probably never heard of In Bhutan, gross national happiness trumps GDP 5 girlfriend getaways close to LI, NYC Planning a **post-hurricane** Caribbean trip 10 real-world locations from popular films to visit 4 spring trips to take with your family Tips for making road trips less stressful, expensive Top U...
- [Link](#)

Type here to search

93%

3:20 PM

4/25/2018

CSCI572-Linux [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Solr Search Solr Admin

localhost/IR\_HW5.php?q=happiness&hid=no&algo=PageRank+Algorithm

Search: happiness

Submit Query

☒ PageRank ☐ Lucene

Showing results for happiness

Show results for [hapiness](#)

Results 1 - 10 of 478:

- [Link](#)  
Title : National & World News Stories | Newsday  
ID : /home/radhika/Desktop/crawl\_data/data/114bf516e02601968da6b2d48b7a9f2a.html  
Link : <https://www.newsday.com/news/nation>  
Snippet : , Canada rank in world **happiness** 2:17 The life & legacy of Stephen Hawking See stunning images from around the world Scenes from the National School Walkout 0:26 Sled dog team spotted without musher on highway Tillerson firing just the latest Trump notable departure 2:15 Students stage walkout to White House, Capitol Scenes from the deadly school shooting in Florida 13 moments that tell the story of Trump's first year 1:39 Lamb declares victory, Saccone not giving up See photos of the super blue blood moon Snowy scenes from across the U...
- [Link](#)  
Title : Long Island Travel Ideas, Sightseeing, Articles & Advice | Newsday  
ID : /home/radhika/Desktop/crawl\_data/data/21404c91c7a0cee0aed08de46f8c7c2c.html  
Link : <https://www.newsday.com/travel>  
Snippet : CONNECT ACCOUNT SUBSCRIBE Special projects How LI political connections helped Oheka's owner thrive One man's rise shines light on I.I.'s corrosive system Manage the Mets: Create your lineup card What hospitals offer expecting parents Travel 25 weekend getaways every New Yorker

Type here to search

96% 3:26 PM 4/25/2018

CSCI572-Linux [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Solr Search Solr Admin

localhost/IR\_HW5.php?q=happiness&hid=yes&algo=PageRank+Algorithm

Search: hapiness

Submit Query

☒ PageRank ☐ Lucene

Results 0 - 0 of 0:

Type here to search

96% 3:29 PM 4/25/2018