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
Aprecition	Appreciation
Ohnesty	Honesty
Muntain	mountain

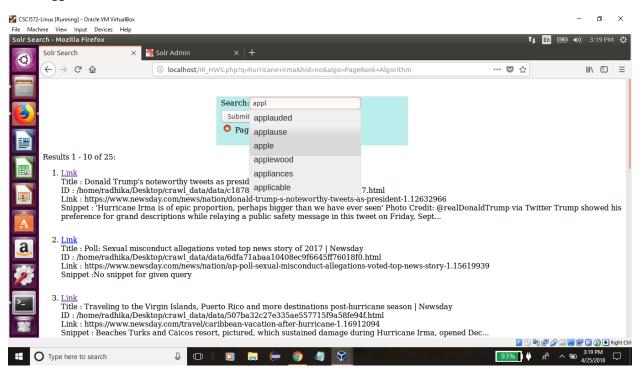
Examples of Autosuggestion

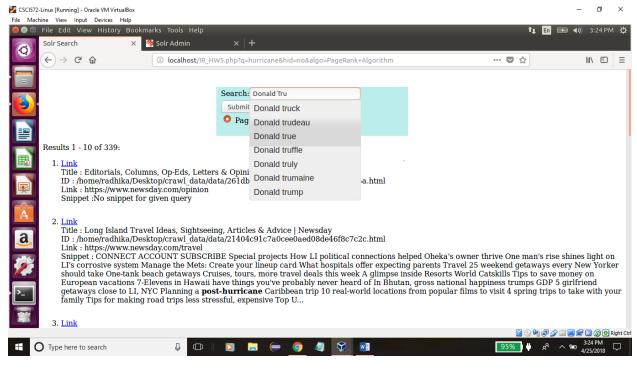
Query	Autocompletion
Congr	Congratulate
	Congress
	Congressional
	Congressman

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

SCREENSHOTS

Autosuggest





Spelling correction and Snippet Generation

