

Information Retrieval – Assignment-5 Report

STEPS FOR SPELLING CORRECTION:

- 1.The first step was to download the PHP version of Peter Norvig's Spelling correction program and to download of the apache tika jar(tika-app-1.14.jar)
- 2.The next step was to create the big.txt file necessary for the spell correction program to create the serialized dictionary. An intermediate file was created using the content extracted from all the html files using the tika command line utility and was fed to a java word count program which would count the number of occurrences of each valid word and redirect the output to big.txt.
- 3.The PHP program was modified to populate the features array using the above specified big.txt file and create the serialized_dictionary.txt file which would be used for spelling correction.
- 4.Once the user enters and submits a query, the spell correct program would do an isolated spell correction for every query term and the query results are fetched for the corrected query giving the user an option to search for the incorrect query as well.

STEPS FOR AUTOCOMplete:

- 1.For autocomplete, for every keystroke, there is a jquery event associated with it and the currently entered query in the text box is sent to an external php file called test.php.
- 2.The test.php file receives the input query and creates a suggest request to solr which returns an array of suggestions. The suggestions are done only for the last entered word.
- 3.The suggestions returned from test.php are populated as a dropdown text in mainPage.php.

STEPS FOR SNIPPET GENERATION:

- 1.The idea behind snippet generation was that the body of the specified html is extracted using the DOM apis available in php and the content is converted into a string.
- 2.The given query term is then preg_matched against the body content and the first occurrence found is returned.If there was no match, then no snippet is displayed.

EXAMPLES FOR SPELL CORRECTION:

INCORRECT SPELLING	CORRECTED SPELLING
Immigraton	Immigration
Favorte	Favorite
Donld Turmp	Donald Trump
Speling	Spelling
Brext	Brexit

EXAMPLES FOR AUTO-COMPLETION:

QUERY TERM	SUGGESTION-1	SUGGESTION-2	SUGGESTION-3
Fav	Favorite	Favor	Favorites
Ele	Election	Elect	Elected
Clu	Club	Clues	Clusters
Bre	Breadcrumbs	Breaking	Brewing
We	Week	Weather	wed