

# CSCI-572-hw5

## Report

### Steps:

Based on hw4, I added spelling checking , autocomplete and snippets in hw5. I first took the configuration suggestion in Appendix – added `<str name="q.op">AND</str>` in solrconfig.xml for both Lucene and PageRank. Then I mapped the results' id to local mapping file to get URLs as required. I achieved the 3 objectives as follow.

#### *Spelling Correction*

1. Downloaded the PHP version implementation of Perter Norvig's spelling program.
2. Generated big.txt specified for news website BG with Apach Tika, and located it under Sites directory for SpellCorrector.php's calling.
3. Included "SpellCorrector.php in my PHP program for spelling checking of submitted queries.
4. When a query is submitted, first check whether the term(s) is(are) in the big.txt. If not, give spelling correction by "showing results for [corrected term\(s\)](#)", and return results for the corrected term(s). Also keep "Search instead for [submitted words](#)" for users to choose. If user click the later link, the spelling correction will be showed by "Did you mean [corrected term\(s\)](#)".

#### *Autocomplete*

1. Configured the solrconfig.xml for both Lucene and PageRank by inserting a searchComponent and a suggest requestHandler.
2. In PHP program, fetch the results from Solr suggest every time a character is inserted to the textbox with Ajax.

#### *Snippets*

1. Implemented getSnippets() method in PHP program by open corresponding local html files and get the text contents.
2. Processed fetched contents with regular expression, divided the contents into sentences.
3. Processed query in order, single term query, multiple terms query (all terms together), multiple terms query (all terms included), multiple terms query (at least one term included).
4. Returned the first sentence mentioning the query.

### Result Analysis:

#### *Spelling Correction*

Query	Spelling Correction
instrauction	instruction
huicane	hurricane
ihpone	iphone
univesity	university
feotball	football

#### Auto-completion

Input characters	First Suggestion
c	charset
ca	canonical
cal	challenge
cali	California

Input characters	First Suggestion
e	encoding
el	else
ele	else
elec	elected
elect	election
electr	electronic

Input characters	First Suggestion
u	user
un	undefined
uni	united
univ	university

Input characters	First Suggestion
p	policy
pa	partners
par	partners
pari	paris

Input characters	First Suggestion
l	loid
le	left
leg	log
lega	least
legac	legacy