

# CSCI 572 HW5: Adding Spell Checking, AutoComplete and Snippets to Search Engine

-Chirayu Samarth

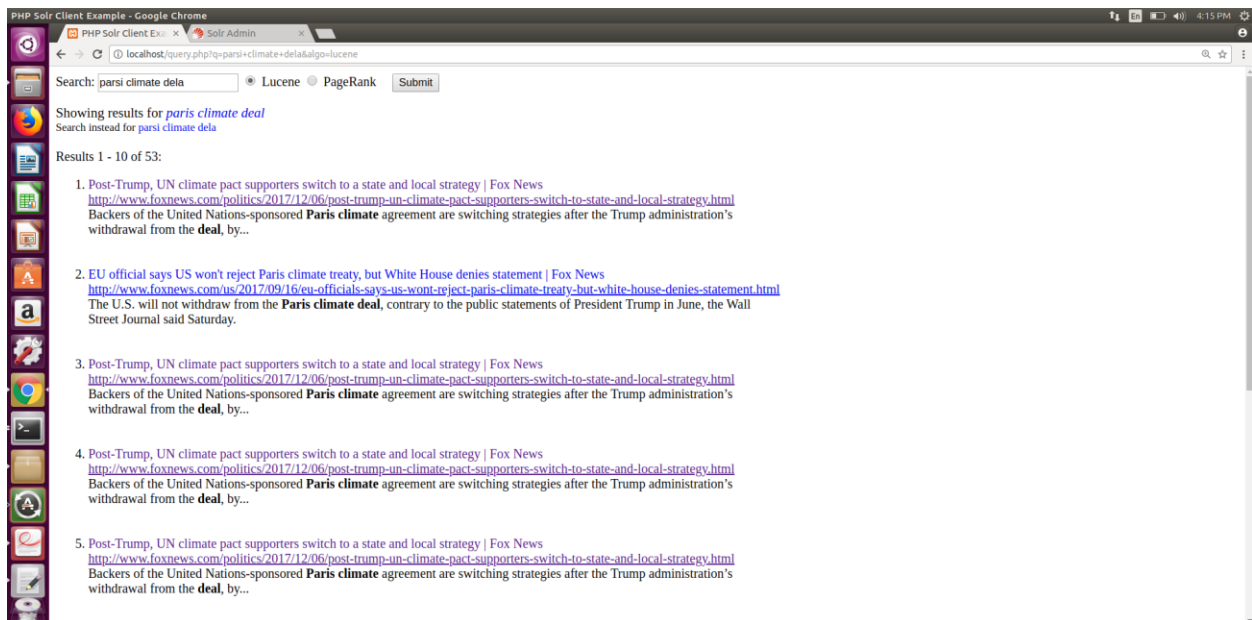
USC ID: 4144 1140 49

## News Website: FOX News

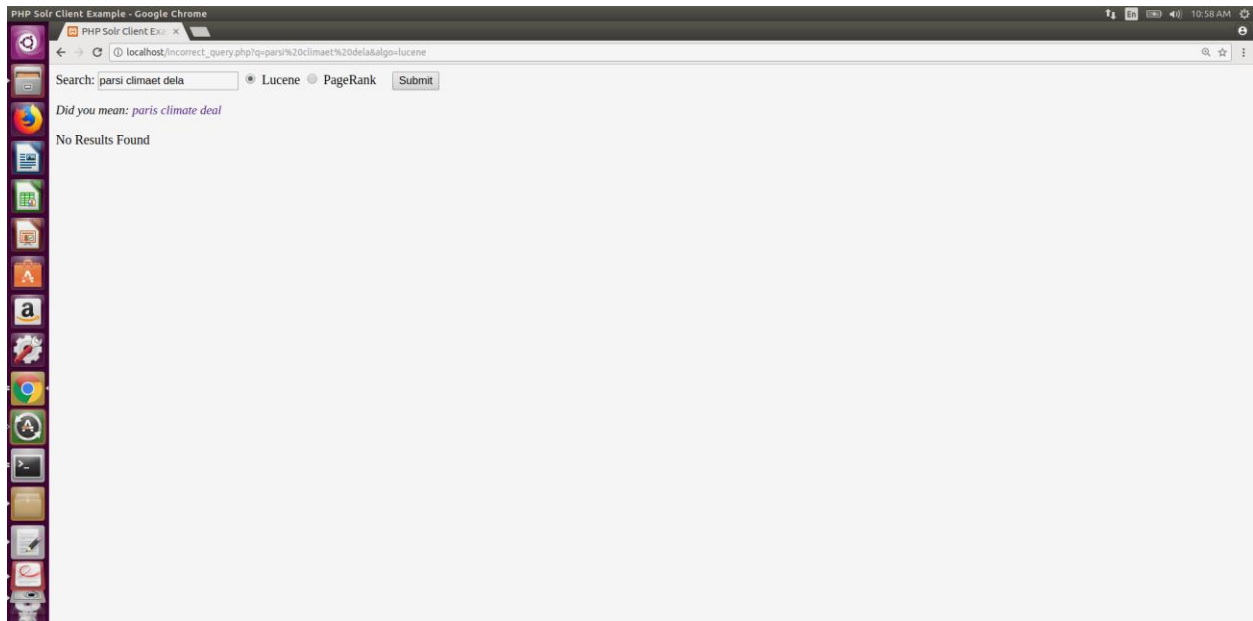
### STEPS:

#### I. Spell-Checking:

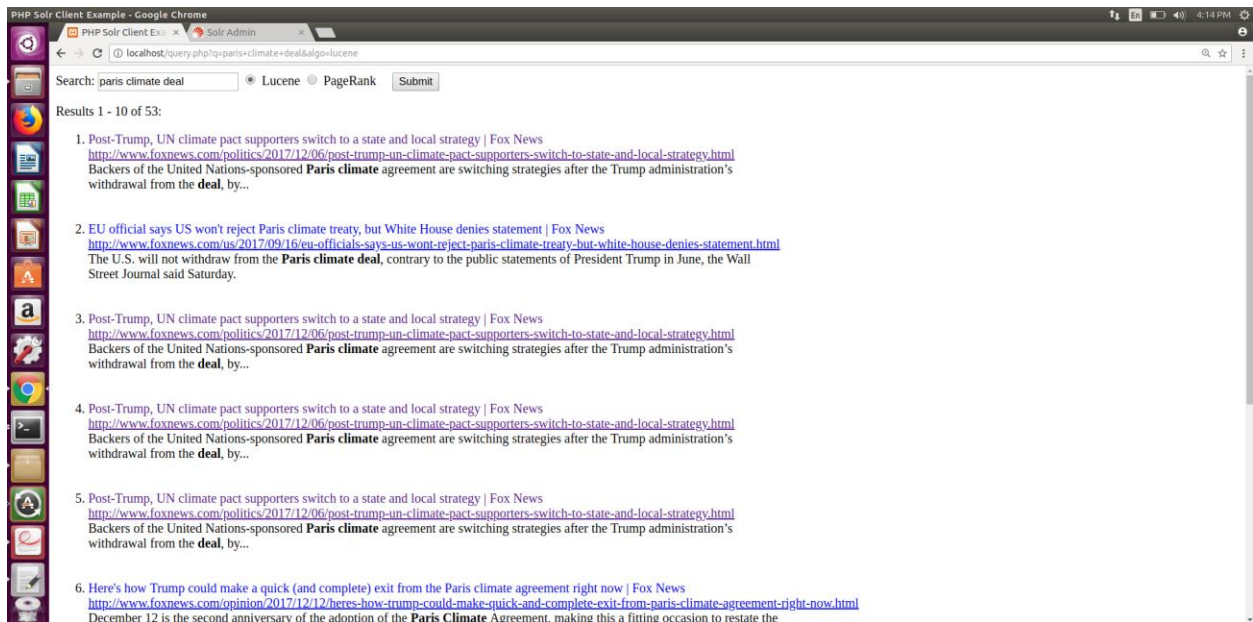
- For Norvig's spell correction program, I created a text file "big.txt" to get set of words to calculate edit distance using Apache Tika Parser.
- Adopted PHP version of Norvig's spelling program from and ran it on the server.
- When the user enters a query, program checks if it is a correct word or not using *SpellCorrector.php*. If it is a correct or a valid word then corresponding search results are displayed. If not, the program corrects the word and displays search results for the corrected word along with options to "Showing results for" *spell-corrected word* and "Search instead for" *incorrect word* as shown in figure. Furthermore, if the user selects to search instead for the incorrect word, the program gives a hint "Did you mean" *spell-corrected word* and displays search results for incorrect word, if any.



After clicking on “Search Instead for”:



After clicking on either “Showing results for” or “Did you mean” :



## II. Auto-Complete:

- a. The SuggestComponent in Solr provides users with automatic suggestions for query terms. Solr has a SuggestComponent designed for this functionality. The first step is to add a search component to solrconfig.xml and tell it to use the SuggestComponent.
- b. After adding the search component, a request handler must be added to solrconfig.xml. This request handler works the same as any other request handler, and allows us to configure default parameters for serving suggestion requests. The request handler definition must incorporate the "suggest" search component as defined in the previous step.
- c. Once these configurations are made in solrconfig.xml, I modified the client program and added an AJAX call to incorporate `"/suggest"` request. The url used to extract the suggestions:  
[http://localhost:8983/solr/fox\\_news/suggest?q=inputquery&wt=json](http://localhost:8983/solr/fox_news/suggest?q=inputquery&wt=json)
- d. The 5 suggestions are listed in the drop-down list.



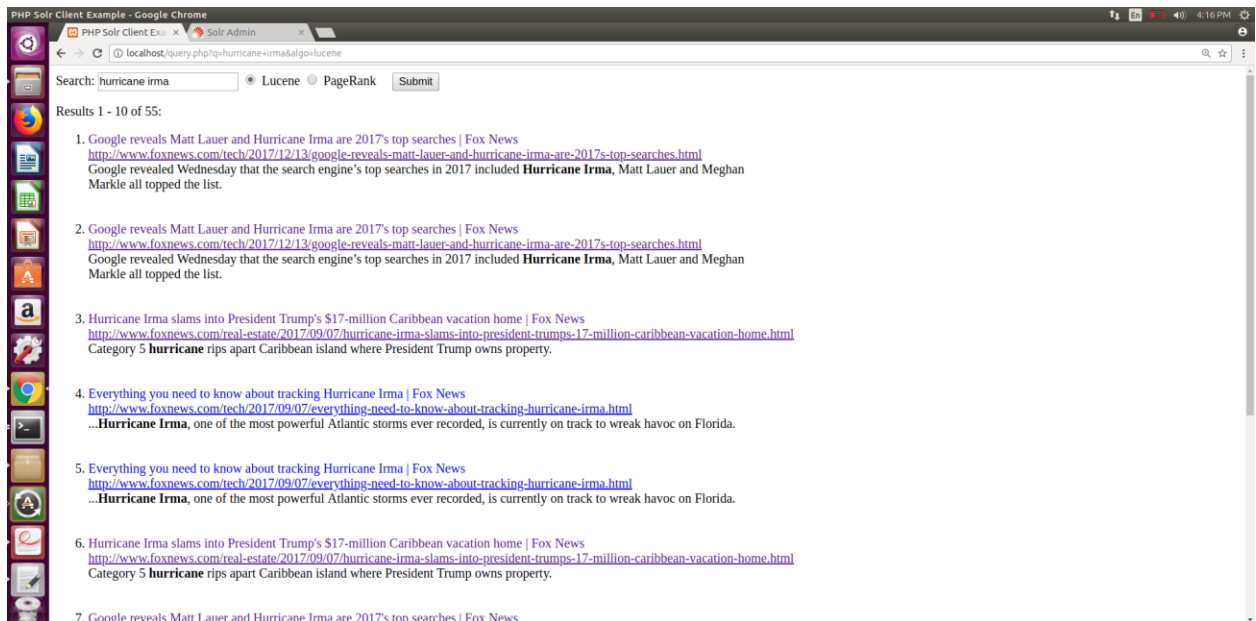
### III. Snippet:

- a. Used Solr Highlighting to incorporate snippets in search results. Highlighting in Solr allows fragments of documents that match the user's query to be included with the query response. Fragments are a portion of a document field that contains matches from the query and are sometimes also referred to as "snippets".
- b. To enable highlighting in solr, we need to set the hl and often hl.fl parameters to get results:

<!-- Highlighting defaults -->

```
<str name="hl">on</str>
<str name="hl.fl"> title og_url description </str>
<str name="f.description.hl.snippets">1</str>
<str name="f.description.hl.fragsize">160</str>
<str name="f.description.hl.simple.pre">&lt;b&gt;</str>
<str name="f.description.hl.simple.post">&lt;/b&gt;</str>
<str name="f.og_url.hl.simple.pre"></str>
<str name="f.og_url.hl.simple.post"></str>
<str name="f.og_url.hl.fragsize">1000</str>
<str name="f.title.hl.simple.pre"></str>
<str name="f.title.hl.simple.post"></str>
<str name="f.title.hl.fragsize">1000</str>
```

- c. Modified the client program to look for a string match of the query terms with the web page and return the first sentence that provides a match. If no match is found, then no snippet is returned.



**Spell-Check Examples:**

Query	Corrected query
faecbook	facebook
bitcni	bitcoin
jamse coemy	james comey
norh korae	north korea
syira wra	syria war

**Auto-Complete Examples:**

Query	Suggestions
face	face facebook fact faces faced
fo	fo fox foxnews.com foxnews for
re	re resourcename rect redistributed reserved
steph	steph step Stephen steps stephanie
hurric	hurricane hurricanes hurried hurricane_harvey hurriedly