Part A: Steps followed to complete this assignment

1. Spelling Correction

1.1 Download the PHP version of Norvig's spelling corrector (SpellCorrector.php) and put it in the same folder with my index.php

1.2 Write a Java program (ExtractContent.java) to use Apache Tika to parse the downloaded news website pages for hw4 and generate big.txt which will be used by SpellCorrector.php. The Java program only extract English page content and replaces whitespaces with a single space. The size of big.txt is around 25.6 MB. Put the big.txt in the same folder with SpellCorrector.php.

1.3 Modify index.php to include SpellCorrector.php and write php codes to implement spelling correction function. The code will check the spelling of each query term and correct it based on Norvig's algorithm. It will send the corrected query to Solr server and notify user with the paragraph "Showing results for … " in the webpage. The program will also provide user the option to search using the original query by clicking the link of the "Search instead for … " paragraph. If the link is clicked, the program will send a query and will not perform spelling correction.

2. Autocomplete

2.1 Enable the SuggestComponent in Solr as described in hw5 tutorial. Modify solrconfig.xml as below:

```
 <searchComponent class="solr.SuggestComponent" name="suggest">
   <lst name="suggester">
     <str name="name">suggest</str>
     <str name="lookupImpl">FuzzyLookupFactory</str>
     <str name="field">_text_</str>
     <str name="suggestAnalyzerFieldType">string</str>
     <str name="nonFuzzyPrefix">4</str>
   </lst>
 </searchComponent>
```

```
<requestHandler class="solr.SearchHandler" name="/suggest">
  <lst name="defaults">
    <str name="suggest">true</str>
    <str name="suggest.count">5</str>
    <str name="suggest.dictionary">suggest</str>
  </lst>
  <arr name="components">
    <str>suggest</str>
  </arr>
</requestHandler>
```

2.2 Enable CORS in Solr to allow index.php directly making AJAX call to Solr for autocomplete function. The instruction can be found in this url: http://laurenthinoul.com/how-to-enable-cors-in-solr/

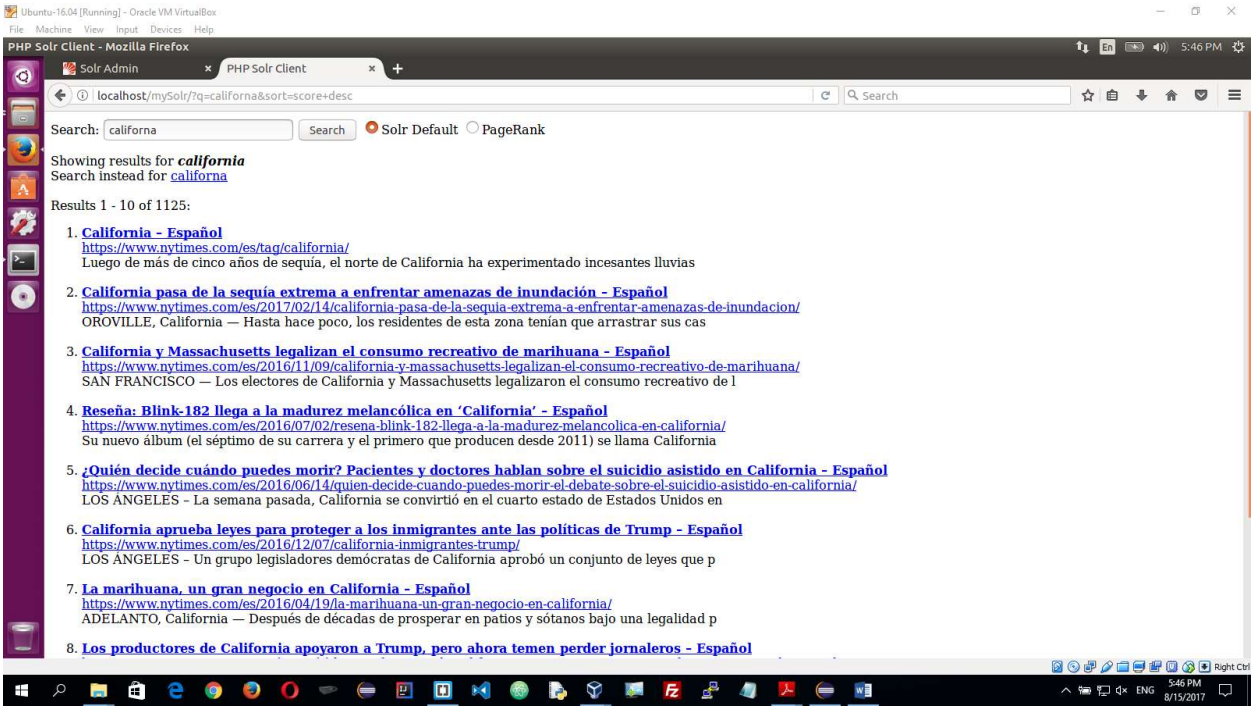
2.3 Modify index.php to implement autocomplete function. Whenever the user types a character in the search box, the page will make a AJAX call to Solr, get a list of suggestions based on the SuggestComponent in Solr, and display the suggestions as the options of the HTML5 <datalist> tag. The program also filter out suggestions containing special characters. For multiple term query, the autocomplete will be performed for the last word that the user is typing. The words which user already finish typing are fixed.

3. Snippets

3.1 Modified index.php to generate snippets for each page returned by Solr. The program will open local webpage files, find a sentence (might be metadata or body content) containing the query term for each query and return a snippet joining those sentences with " … ". The length of the snippet is also limited. If no match is found, then no snippet is returned.

Part B: Analysis of the results

1. Spelling Correction



2. Autocomplete