**Homework 5**

**Adding Spell Checking, AutoComplete and Snippets to Search Engine**

1. Steps followed to complete the assignment:

* Generating big.txt:

Generated big.txt by writing a java program that used Apache Tika library to parse the crawled web pages of Mercury News.

This file is used for the vocabulary of spell check and auto-completion.

* Spelling Correction:

Used the PHP version of Peter Norwig’s spelling correction program and provided big.txt as a dictionary for it.

Whenever a user inputs a query the spelling correction code checks if it is a valid query and if not, the code looks for spelling errors within one or two edit distance by performing inserts, deletes and substitutions on the input words/phrases and suggests the words as “Did you mean”.

When the user clicks on the “Did you mean version” of the query, new search results with the corrected query are displayed.

* Auto Complete:

Configured Solr to use Solr SuggestComponent in the solrconfig.xml

Used JQuery UI autocomplete component along with the Solr suggest to display a dropdown list consisting of query suggestions.

Auto complete starts suggesting words as soon as the user enters the first character of a query. Auto complete is done on a word by word basis for every newly typed word.

* Snippets:

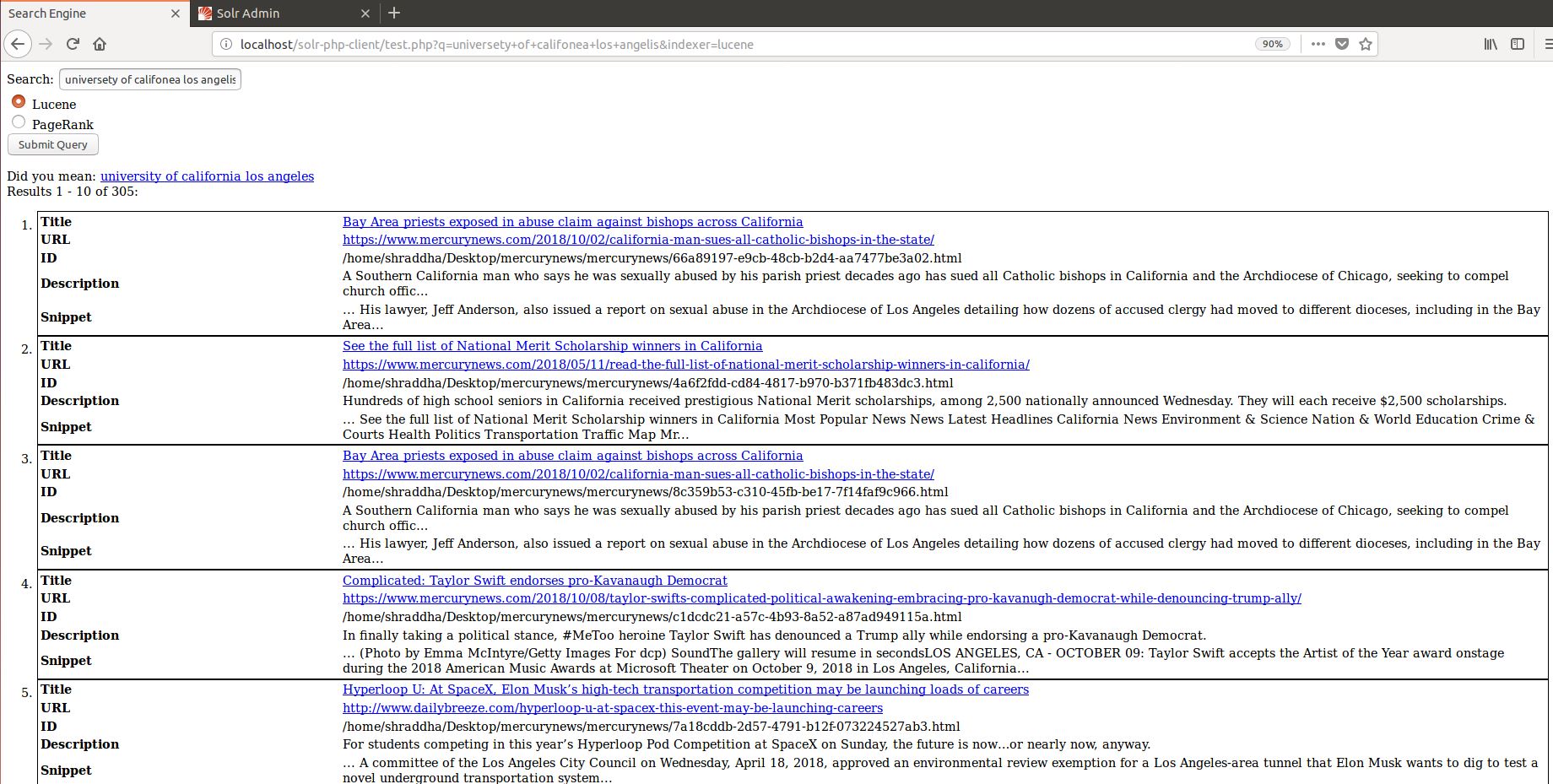
Used the simple\_dom\_html parser to get the content of the resulting webpages as plain text using str\_get\_html function.

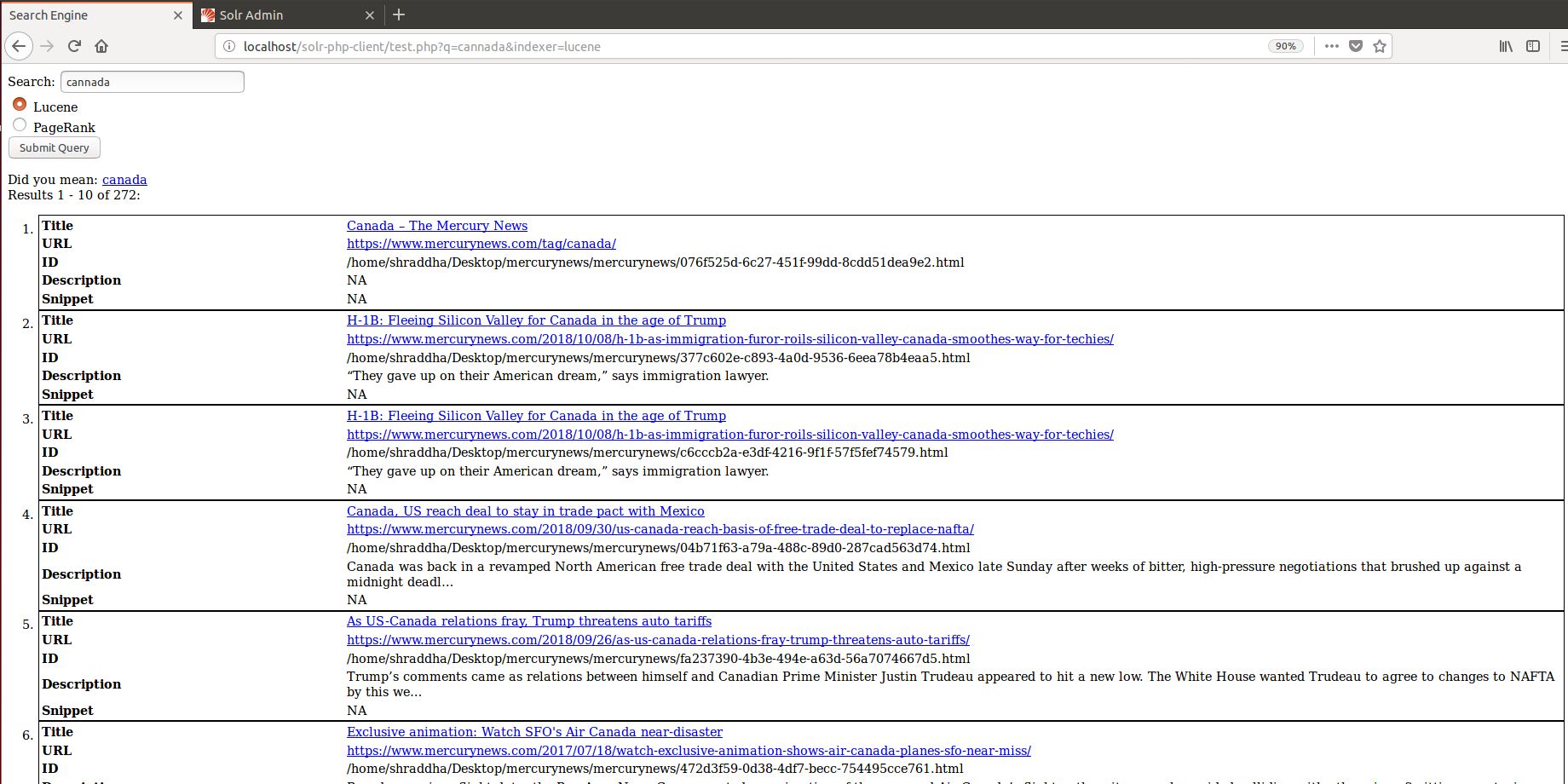
Used phrase match to find out the paragraph that matches most with the query and return 160 characters as snippet.

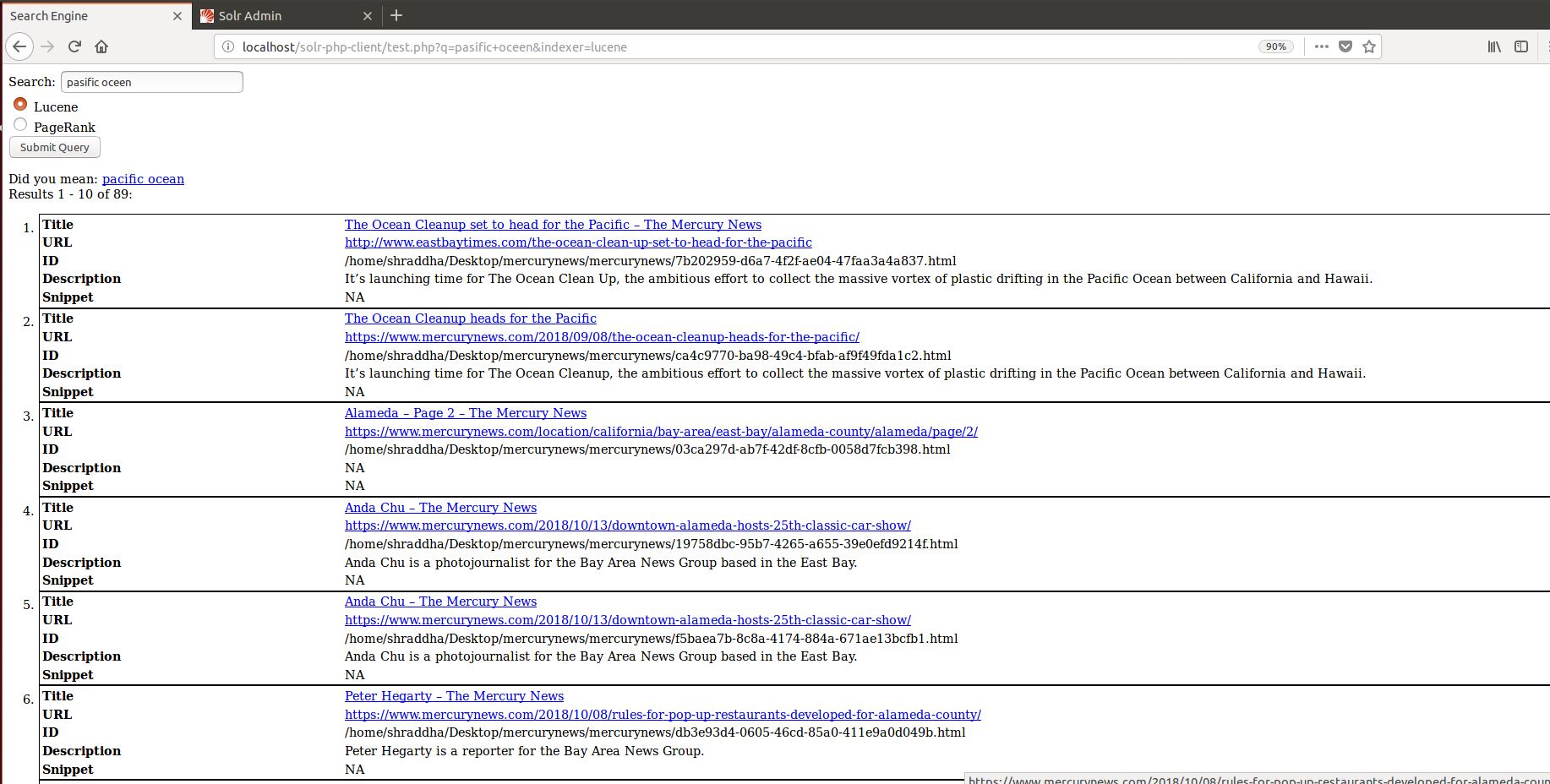
If no match is found then no snippet is returned.

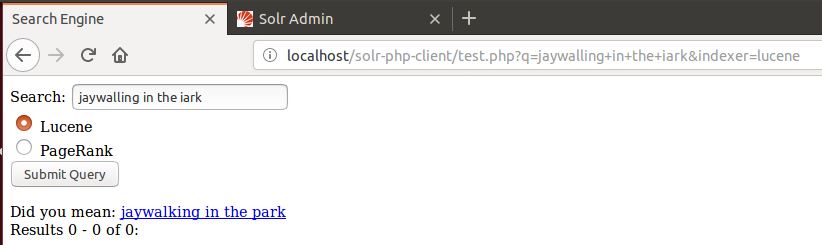
2. Analysis of the results:

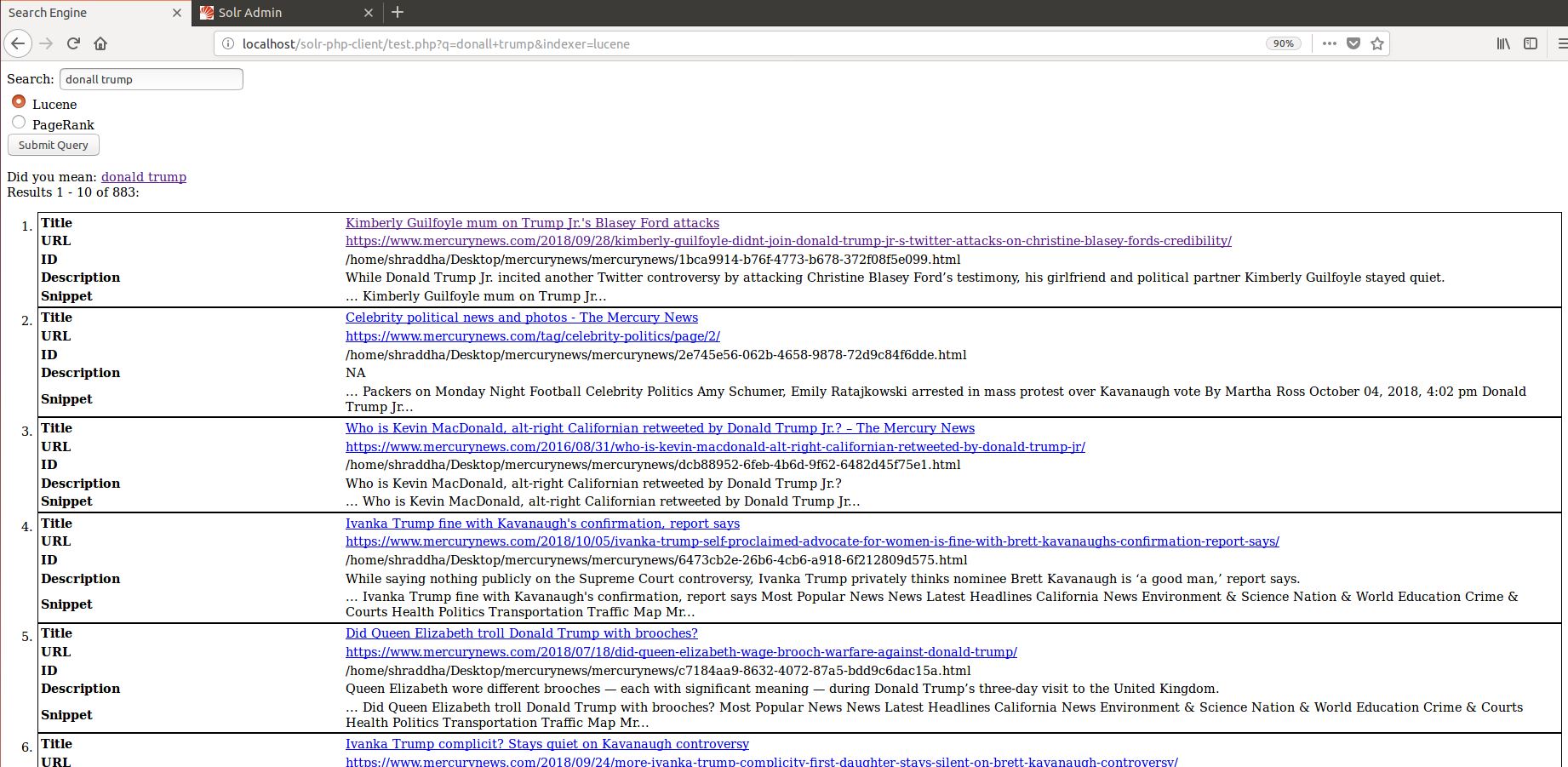
Five examples of misspelled terms that are correctly handled by spelling correction program.









’

FIVE examples of auto-completion:

