

HW5 Report

#Steps

Step1: Autocomplete

I use the FuzzyLookupFactory feature in Solr to fulfill this function. In order to use it, I configure solrconfig.xml file in my own core, mySolarCore which I created in the HW4 to enable Solr to use SuggestComponent. There are two parts should be added, details as below:

```
<!-- add for autocomplete -->
<searchComponent name="suggest" class="solr.SuggestComponent">
  <lst name="suggester">
    <str name="name">suggest</str>
    <str name="lookupImpl">FuzzyLookupFactory</str>
    <str name="field">_text_</str>
    <str name="suggestAnalyzerFieldType">text_general</str>
  </lst>
</searchComponent>

<!-- add for autocomplete -->
<requestHandler name="/suggest" class="solr.SearchHandler">
  <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>
```

And then, I modify the front-end API to activate the autocomplete function. The main functions are as following. First, I make it possible by the JavaScript function- `autoComplete(strURL)` which I made, that when typing some characters in the search box, it could send the URL and parameters to my Solr core to communicate with Solr and get a list of suggestions. The important point is that the URL format should be formed like this “<http://localhost:8983/solr/mySolarCore/suggest?wt=json&suggest=true&suggest.dictionary=suggest&suggest.q=star>”. Second, I make the function- `listCorrection(str)` to receive the list from Solr and list it below the query box. Also, I make it possible that clicking one suggestion, it could fix the query in the query box and search the results automatically.

Step2: Spelling Correction

In this part, due to the JavaScript language I use in HW4, I download the codes of Norvig's complete spelling correction program in JavaScript language developed by Panagiotis Astithas from the website <https://github.com/past/speller/tree/master>. Because the training function in the Norvig's program should use a big.txt which contains a large set of words to calculate edit distance, I use the tika-app-1.16.jar and build a Java project to parse the webpages

files of New York Daily to make the big.txt file. Finally, I import the spelling correction program file in JavaScript into my own front-end and modify my code to load the big.txt file at the initial loading of page and use it as the input of training function and use the output of training function as the input of correction function. When I do the search, I trigger the correction function simultaneously and display the output of correction function on the above of results table.

Step3: Snippets

I simply follow the instruction on this site https://lucene.apache.org/solr/guide/6_6/highlighting.html and use the Highlighting in Solr to generate snippets. To be more specific, I modify the URL which sent to Solr to get the results containing highlighting part. Like set hl=true, hl.fl=*, hl.simple.pre=<i>, hl.simple.post=</i>, which is to enable highlighting function and search all the possible fields which contain the query term and generate snippet, and make the query terms occur in the snippets bolded and italicized.

(** Regarding detail implementation, please check files zipped together)

#Analysis of the results

#1 five examples of misspelled terms handled by spelling correction program

(1)

Query: ☒ Default ☐ PageRank

Did you mean: los *angeles*

(2)

Query: ☒ Default ☐ PageRank

Did you mean: *star* wars

(3)

Query: ☒ Default ☐ PageRank

Did you mean: *north*

(4)

Query: ☒ Default ☐ PageRank

Did you mean: *story*

(5)

Query: ☒ Default ☐ PageRank

Did you mean: *love*

#2 five examples of auto-completion.

(1)

Query: ☒ Default ☐ PageRank

north

nothing

northern

northeast

northwestern

(2)

Query: south kor| ☒ Default ☐ PageRank

korea
kardashian
korean
kristaps
kris

(3)

Query: hurri| ☒ Default ☐ PageRank

hurricane
hurricanes
harris
harrison
horrific

(4)

Query: love stor ☒ Default ☐ PageRank

stream_content_type
stream_size
sports
start
stories

(5)

Query: big apple fa| ☒ Default ☐ PageRank

fa
facebook
family
far
face