Homework05 Report

Adding Spell Checking, AutoComplete and Snippets to Your Search Engine

Name: Biru Lyu

USC ID: 1810570267

• 1. Steps you followed to complete this assignment. Include the details of what tools and techniques you used to implement spelling correction and autocomplete.

On the basis of Homework04, I added the functionality of autocomplete, spell check and generate snippet.

- For autocomplete Part
 - I added the SuggestComponent and request handler to solrconfig.xml under myexample core according the instruction _AutocompleteInSolr_V2.pdf _ and _ reload the core._
 - 2. I modified the service.php under Solr client PHP library by adding a HTTP request to the /suggest service on Solr's server.
 - 3. In my query.php, I use autocomplete AJAX function to request http://localhost:8983/solr/myexample/suggest?q= and the server will return well-formated JSON data which will be rendered as a drop-down list box.
- For spell check Part
 - 1. I make use of Apache Tika to parse all the html by using tika-app.jar and store the content in big.txt which is ready for Peter Norvig's spell corrector program.
 - 2. I downloaded PHP version of Peter Norvig's program and placed it into my PHP server for query.php to call out this component.
 - 3. When the query is not null, I use spellcorrector::correct to check the spell of
 each term in the query, once the corrected term is not match the input one, set the
 flag variable \$is_spell_error to false and print the suggest correct terms
 following Did you mean hint.
- For generate snippet Part
 - 1. In this part I use simple_html_dom parser libarary to read and extra the content of each webpage to generate the snippet. I will only focus on the word and try to find the postion of the first matching and then restrict it to around 50 words.
 - 2. I also implemented a function to decorate sinppet by using regular expression match to color the target query terms.
- Others
 - 1. I make 'AND' as default for multi-word queries in Solr by add <str

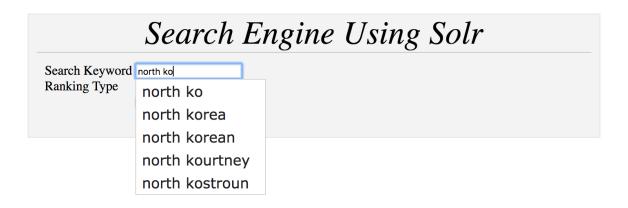
name="q.op">AND</str> within this tag: <requestHandler name="/select"</pre> class="solr.SearchHandler"> and Inside <1st> default tag within requestHandler tag: <lst name="defaults"> in solrconfig.xml and reload the Solr again.

- 2. Analysis of the results: In this you should provide FIVE examples of misspelled terms that ex
 - 0 SI

re correctly handled by your spelling correction program. You should also provide FIVE xamples of auto-completion.
pell check
Search Engine Using Solr
Search Keyword californa Ranking Type Solr Ranking submit Clear
Did you mean: california
Search Engine Using Solr
Search Keyword Donad Trup Ranking Type Soir Ranking submit Clear
Did you mean: donald trump
Search Keyword nroth korae Ranking Type Soir Ranking Submit Clear
Did you mean: north korea
Search Keyword kanpsack Ranking Type Solr Ranking submit Clear
Did you mean: knapsack
Search Engine Using Solr Search Keyword computer secnce Ranking Type Solr Ranking Submit Clear

Did you mean: computer science

Search Keyword Calif Ranking Type Calif california caliber calipers







Search Engine Using Solr

Search Keyword oran
Ranking Type ora
organization
orange
orlando
organizations