

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

1. 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 library 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" class="solr.SearchHandler">` and Inside `<lst>` 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 are correctly handled by your spelling correction program. You should also provide FIVE examples of auto-completion.
- o spell check

Search Engine Using Solr

Search Keyword	<input type="text" value="california"/>
Ranking Type	<div>Solr Ranking</div>
<div>submit</div> <div>Clear</div>	

Did you mean: [california](#)

Search Engine Using Solr

Search Keyword	<input type="text" value="Donad Trup"/>
Ranking Type	<div>Solr Ranking</div>
<div>submit</div> <div>Clear</div>	

Did you mean: [donald trump](#)

Search Engine Using Solr

Search Keyword	<input type="text" value="nrth kora"/>
Ranking Type	<div>Solr Ranking</div>
<div>submit</div> <div>Clear</div>	

Did you mean: [north korea](#)

Search Engine Using Solr

Search Keyword	<input type="text" value="kapsack"/>
Ranking Type	<div>Solr Ranking</div>
<div>submit</div> <div>Clear</div>	

Did you mean: [knapsack](#)

Search Engine Using Solr

Search Keyword	<input type="text" value="computer secnce"/>
Ranking Type	<div>Solr Ranking</div>
<div>submit</div> <div>Clear</div>	

Did you mean: [computer science](#)

- o autocomplete

Search Engine Using Solr

Search Keyword	<input type="text" value="calif"/>
Ranking Type	<div><div>calif</div><div>california</div><div>caliber</div><div>calipers</div></div>

Search Engine Using Solr

Search Keyword	<input type="text" value="north ko"/>
Ranking Type	<div><div>north ko</div><div>north korea</div><div>north korean</div><div>north kourtney</div><div>north kostroun</div></div>

Search Engine Using Solr

Search Keyword	<input type="text" value="milk choco"/>
Ranking Type	<div><div>milk choose</div><div>milk choosing</div><div>milk chocolate</div><div>milk chooses</div><div>milk coco</div></div>

Search Engine Using Solr

Search Keyword	<input type="text" value="bitte"/>
Ranking Type	<div><div>better</div><div>battery</div><div>bite</div><div>butter</div><div>bites</div></div>

Search Engine Using Solr

Search Keyword

Ranking Type

ora

organization

orange

orlando

organizations