

Assignment 5 Report

Steps Followed

Spelling Correction

In order to implement spelling correction, setup for Assignment 4 was improvised. To start with, “**norvig**” algorithm in php was downloaded. In it, “**SpellCorrector.php**” was moved to html root folder and included in “**user.php**” file(frontend file). This algorithm is called in following cases:

- Case 1: When user submits a query
- Case 2: When user is typing query in textbox

For both of the cases, first the input string is converted to lower case and array is formed by splitting the query by space. For each element of the array, “**SpellCorrector**” is called separated and result is again joined by inserting space between each element.

In Case 1, when a user submits a query, as first step “**SpellCorrector**” is called to validate if query has any spelling errors or not. If query has error, then results are shown for correct spelling along with message “Showing result of ***black panther*** instead of ***Black Panthr***”.

In Case 2, when user is typing the query, again as first step “**SpellCorrector**” is called to validate if query has any spelling errors or not. If query has error, then the corrected query is displayed as first item of autocomplete list. Corrected query is shown in bold and italics to differentiate from autocomplete results.

In order to improve the dictionary, “big.txt” referred by “**SpellCorrector.php**” by “big.txt” file generated by compressing html content of all Fox_News pages by writing Java program. For a page’s html content, content was compressed by replacing tabs and newline with spaces and then multiple spaces with single space.

Autocomplete

As part of improvising Assignment 4 setup, firstly search component “**solr.SuggestComponent**” was added along with request handler “**solr.SearchHandler**” in “**solrconfig.xml**”. With this, whenever “**/suggest**” is called, call would be redirected to the new request handler. Now, whenever user type character in text field, **onKeyUp** invokes **showResult** function. This function makes XMLHttpRequest to **user.php?a=\$query**. When this request is received, then it is recognized as autocomplete request. Thus, result is send to solr with additional parameter as '**qt**' => '**suggest**'. In order to enable solr to process this request, changes are made in **solr-php-client/Apache/Solr/Service.php** . In this file, a new url for “**/suggest**” is added. In function

search(\$query, \$offset = 0, \$limit = 10, \$params = array(), \$method = self::METHOD_GET), code is modified to call “/suggest” endpoint if \$params[“qt”] == “suggest”.

For single word query, input is send as it is and top five results are displayed to user as link. When user clicks on the particular item of result list, then result for that particular item is displayed.

For multiple word query, it is assumed that only last word needs to be autocompleted. That is why, query is send only for last word and result are shown similar to single word query.

In autocomplete list, if there is spelling error in query then first element in list spell correction output.

Snippet

Snippet is generated for top 10 results for a query. For each result html content is fetched via file_get_html function. In all the <p> tags inside “<div class=’article-body’>” are extracted. For each paragraph tags, lines are extracted by splitting paragraph by full stop (“.”). Then for each line, number of words matching the query words are calculated. Line with maximum matched words is selected. If no such line exist then, metadata description is used. If it is also not presented then, “N/A” is displayed. Now if snippet was more than 160 characters, then part of sentence with most number of query word is taken and “...” is added as prefix and/or as suffix based on position of extracted sentence with respect to original file.

Snippet is shown always as lower case along with url, title and id in search result. In snippet, word corresponding to query words are highlighted.

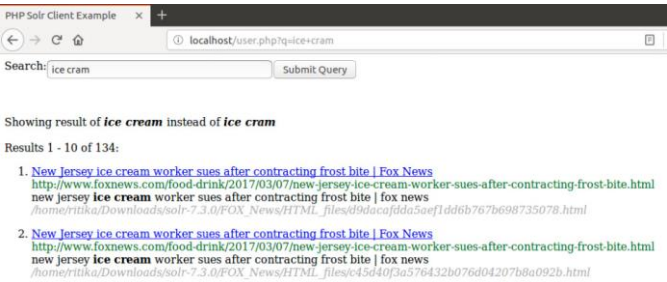
In the request handler of “/select”, <str name=“q.op”>AND</str> is added so that query result would have web page containing all the query terms.

Analysis of Result

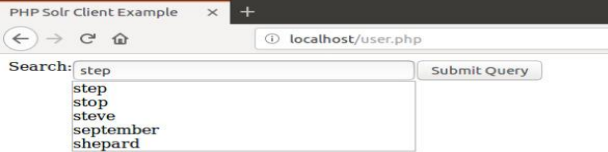
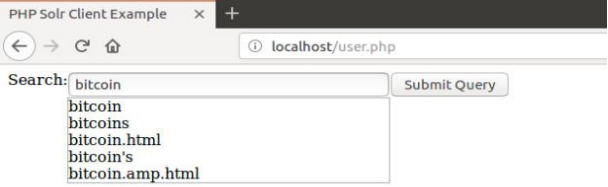
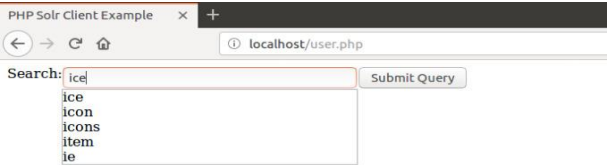
Spelling Correction Examples

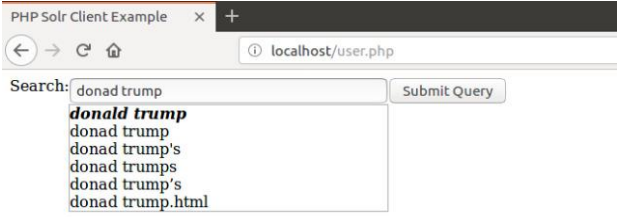
<u>Query Number</u>	<u>Input</u>	<u>Output</u>	<u>Screenshot</u>
1	black panthr	black panther	

			 <p>PHP Solr Client Example x +</p> <p>localhost/user.php?q=black+panthr</p> <p>Search: black panthr Submit Query</p> <p>Showing result of black panthr instead of black panther</p> <p>Results 1 - 10 of 10:</p> <ol style="list-style-type: none"> 1. Stan Lee, 95, 'feeling great' after return from LA hospital Fox News http://www.foxnews.com/entertainment/2018/02/02/stan-lee-95-feeling-great-after-return-from-la-hospital.html ...black panthr movie, but has faced reports of poor health in the past /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/69f808413af7aa57d89b6378bd860cfd.html 2. Stan Lee, 95, 'feeling great' after return from LA hospital Fox News http://www.foxnews.com/entertainment/2018/02/02/stan-lee-95-feeling-great-after-return-from-la-hospital.html ...black panthr movie, but has faced reports of poor health in the past /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/69f808413af7aa57d89b6378bd860cfd.html
2	bitcin	bitcoin	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php?q=bitcin</p> <p>Search: bitcin Submit Query</p> <p>Showing result of bitcoin instead of bitcin</p> <p>Results 1 - 10 of 110:</p> <ol style="list-style-type: none"> 1. Lawsuit alleges bitcoin pioneer became a thief Fox News http://www.foxnews.com/tech/2018/03/01/lawsuit-alleges-bitcoin-pioneer-became-thief.html ...bitcoin mining and research firm that controlled between 550,000 and 1 /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/015ceb45f7726e2a99287ada6eea 2. Lawsuit alleges bitcoin pioneer became a thief Fox News http://www.foxnews.com/tech/2018/03/01/lawsuit-alleges-bitcoin-pioneer-became-thief.html ...bitcoin mining and research firm that controlled between 550,000 and 1 /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/69ad9b81695a19c8cfbb7062cde8f
3	stephen hawng	stephen hawking	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php?q=stephen+hawng</p> <p>Search: stephen hawng Submit Query</p> <p>Showing result of stephen hawking instead of stephen hawng</p> <p>Results 1 - 10 of 408:</p> <ol style="list-style-type: none"> 1. A look back on the life of renowned physicist Stephen Hawking Fox News http://www.foxnews.com/science/slideshow/2018/03/14/look-back-on-life-renowned-physicist-stephen-hawking.htn renowned physicist stephen hawking died wednesday at the age of 76. /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/96039c86b6584cf266dfc5e892cd00e2.html 2. A look back on the life of renowned physicist Stephen Hawking Fox News http://www.foxnews.com/science/slideshow/2018/03/14/look-back-on-life-renowned-physicist-stephen-hawking.htn renowned physicist stephen hawking died wednesday at the age of 76. /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/c0e5ba375b7fd8f331ca09e7a2237de8.html
4	donad	donald	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php?q=donad</p> <p>Search: donad Submit Query</p> <p>Showing result of donald instead of donad</p> <p>Results 1 - 10 of 2685:</p> <ol style="list-style-type: none"> 1. 2016 Fox News Exit Polls Politics Fox News http://www.foxnews.com/politics/elections/2016/exit-polls n/a /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/b33c9b40cf6175afe78e3ea9e973722c.html 2. 2016 State Primaries and Caucuses Politics - SOUTH DAKOTA Fox News http://www.foxnews.com/politics/elections/2016/primary-caucus/south-dakota n/a /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/69aad3c19ae2468044d4ad07b63ed641.htm

5	ice cram	ice cream	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php?q=ice+cram</p> <p>Search: ice cram Submit Query</p> <p>Showing result of ice cream instead of ice cram</p> <p>Results 1 - 10 of 134:</p> <ol style="list-style-type: none"> 1. New Jersey ice cream worker sues after contracting frost bite Fox News http://www.foxnews.com/food-drink/2017/03/07/new-jersey-ice-cream-worker-sues-after-contracting-frost-bite.html new jersey ice cream worker sues after contracting frost bite fox news /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/d9daca5aef1dd6b767b698735078.html 2. New Jersey ice cream worker sues after contracting frost bite Fox News http://www.foxnews.com/food-drink/2017/03/07/new-jersey-ice-cream-worker-sues-after-contracting-frost-bite.html new jersey ice cream worker sues after contracting frost bite fox news /home/ritika/Downloads/solr-7.3.0/FOX_News/HTML_files/c45d40f3a576432b076d04207b8a092b.html
---	----------	-----------	---

Autocomplete Examples

Query Number	Input	Output	Screenshot
1	step	step stop steve september shepard	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php</p> <p>Search: step Submit Query</p> <p>step stop steve september shepard</p>
2	bitcoin	bitcoin bitcoins bitcoin.html bitcoin's bitcoin.amp.html	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php</p> <p>Search: bitcoin Submit Query</p> <p>bitcoin bitcoins bitcoin.html bitcoin's bitcoin.amp.html</p>
3	ice	ice icon icons item ie	 <p>PHP Solr Client Example x +</p> <p>localhost/user.php</p> <p>Search: ice Submit Query</p> <p>ice icon icons item ie</p>

4	donad trump	donald trump donad trump donad trump's donad trumps donad trump`s donad trump.html	
5	ru	ru russia run russian running	