

Steps:

Spelling Correction:

1. I use tika library to implement a html parser file named ExtracContent.java. This file generates my big.txt file which contains all the terms in the inverted index of the Search Engine.
2. I downloaded the Peter Norvig's Spelling corrector code and use a test.php file to test the correct function and thus generating my serialized_dictionary.txt, which is the dictionary of words in my news file.
3. In my query.php file I call out the correct function and retrieve the correct word, completing the "Did you mean" correction part.

Autocomplete:

1. I first made the changes required in the solrconfig.xml as mentioned in the tutorials given under AutocompleteInSolr_V2.pdf
2. I use JQuery Autocomplete API to implement all this part, including ajax request, and other functions for the UI part.
3. My code will only take the last word as all the previous words are already corrected.

Snippet:

1. I have used simple_html_dom parser to generate the plaintext out of the html pages given to us.
2. Then I use a regex to globally match the query. And get the first result of matches.
3. I use regex to find the key words in query and highlight them in red color by preg_replace function.

Analysis of the results:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [stephen having](#)

Results 0 - 0 of 0:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [google doc](#)

Results 1 - 10 of 116:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [camera](#)

Results 0 - 0 of 0:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [retrieve document](#)

Results 1 - 10 of 2045:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [chines restaurant](#)

Results 1 - 1 of 1:

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean: [retrieval document](#)

Results

1.

Search:

☐ Lucene Algorithm ☐ pageRank Algorithm

Did you mean:

Results

1.

Search: china de

☐ Luce
submit

Did you
Results

1. I

- china de
- china device
- china description
- china death
- china default

Search: weeknd ne

☐ Luce
submit

Did you
Results

1. I

- weeknd ne
- weeknd news
- weeknd newscms
- weeknd new
- weeknd next

Search: apart

☐ Luce
submit

Did you
Results

1. I

- apart
- apartment
- apartments
- apartheid
- apathy