

CSCI 572 ASSIGNMENT 5 – REPORT

Name: Rahul Shreerang Lele

USC ID:2483165273

Steps followed to implement Spelling Correction:

1. Generated big.txt by parsing html files using tika parser in Java program.
2. Download Norvig's program in php.
3. Update the php file to check for spelling correction and show results as Google does.

Steps followed to implement AutoComplete:

1. Modified the solrconfig.xml for suggestions as given in the assignment guidelines.
2. Used jquery for ajax call to solr for suggestions on every keypress.
3. Used HTML datalist to display the suggestions to autocomplete.

Steps followed to implement Snippets:

1. Used php's simple-html-dom parser to parse html pages.
2. Search for the query in html's body and div tags.
3. Used regex to filter special characters and bolded the query words.
4. Displayed the snippet on the page.

Examples of misspelled terms:

1. brek – spelling corrected to break
2. californa- spelling corrected to california
3. chna - spelling corrected to china
4. donld tump- spelling corrected to donald trump

5. rusisa - spelling corrected to russia

Examples of autocomplete:

1. de Suggestions: de, device, description, death, default

2. pr Suggestions: pr, privacy, press, president, progress

3. un Suggestions: un, undefined, under, universal, university

4. al Suggestions: al, all, alternate, also, alert

5. le Suggestions: le, left, learn, legal, leftfield