

Name: Anchal Kapoor

Student ID: 2650399913

News Website: LA Times

Steps followed for Spell Correction, Autocomplete, and adding snippets:

- Configuration settings for Autocomplete:
  - Updated the solrconfig.xml file based on the changes suggested in the homework specifications.
  - Reloaded the core after the update in the Solr Admin panel.
- Configuration and implementation of Spell Correction:
  - Updated the solrconfig.xml file based on the changes suggested in the homework specification.
  - Reloaded the core after the update in the Solr Admin panel.
  - Downloaded the SpellCorrector.php file based on Peter Norvig's code.
  - Generated the big.txt file that contains all the content of all the html files of the news site.[Jsoup]
  - Added code for spell correction in the ui.php file.
- Implementation of Autocomplete:
  - Included jquery and jquery ui to the code.
  - Used jquery autocomplete function to add query suggestions as per user's input.
  - Removed stopwords to enhance query suggestions
- Implementation of Snippets:
  - Included simple\_html\_dom.php to get html
  - Generate snippets relevant to user query
  - Added regex to remove special characters

Some queries with spelling error and the corrected queries by the Spell Checker:

1. Showing results for: donald trump (corrected query)  
Search instead for: donald trum (spelling error in query)
2. Showing results for: california (corrected query)  
Search instead for: california (spelling error in query)
3. Showing results for: november (corrected query)  
Search instead for: Novembr (spelling error in query)
4. Showing results for: illegal immigration (corrected query)  
Search instead for: illegal immgration (spelling error in query)

5. Showing results for: terrorism (corrected query)  
Search instead for: terrorsms (spelling error in query)  
Autocomplete results for some queries:

1. Incomplete query: Donal

- donal
- Donald
- donations
- donated
- donation

2. Incomplete query: calif

- calif
- california
- Californians

3. Incomplete query: novem

- november
- novel
- novels
- novelist
- novelty

4. Incomplete query: ill

- ill
- illegal
- island
- illegally
- Islamic

5. Incomplete query: ter

- ter
- text
- turbine
- trb
- tertiarynavbar