

# Assignment 5 Report

Yuchuan Chen 4096061963

## 1. Steps to complete the assignment

### 1) Preparation

Write a Java program **parser.java** to generate **big.txt** from **mercurynews** html files using Tika parser package. Save it and put it into server folders as vocabulary dictionary for spelling correction and auto-completion.

### 2) Edit the **solrconfig.xml** file for spelling correction and auto-completion:

Add this line: **<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">**

Add a searchComponent named "suggest" and its corresponding requestHandler for auto-completion.

### 3) Spelling Correction

Spelling correction is implemented by Norvig's spelling correction program (php version).

Download file **SpellCorrector.php** from given link in homework description and included it into my **index.php** file. When user input a query, the php program will check it with generated dictionary **serialized\_dictionary.txt** to see if the query is valid. If not, the program will return correction word and my frontend page will show message "Did you mean" + correct word to user.

### 4) Auto-complete

First, add prefix and suffix to user input to generate new query for auto-completion. Then, use the previously generated query to call solr back-end

and get its response with 5 auto-completion suggestion words. Finally, display the suggestion words in a dropdown list under the input box.

## 5) Snippets

First, download and include **simple\_html\_dom.php** in the frontend php program to ensure only plain text will be extracted from html files.

Then, Split the content in html files by dot (extract each sentence) and query by whitespace (extract each query word).

And then, write php codes to check every sentence in html file to see if it contains one or more query words. If so, stress the query words in the sentences and display in each searching result.

## 2. Examples

### 1) Five examples of spelling correction

- a) Input: conection; Correction: connection
- b) Input: peformance; Correction: performance
- c) Input: tecnolog; Correction: technology
- d) Input: nagation; Correction: navigation
- e) Input: entertinment; Correction: entertainment

### 2) Five examples of auto-completion

- a) Input: week; auto-completion: week, weekly, weekend, weeks, weekday
- b) Input: americ; auto-completion: american, america, americans, america's, americas
- c) Input: applic; auto-completion: application, applied, applications, applies, applicants
- d) Input: univers; auto-completion: university, university's, universities, universal, universe
- e) Input: goog; auto-completion: google, googletag, googletag.cmd.push, googletag.display, good