Homework 4: Adding Spell Checking and AutoComplete to Your Solr-based Search Engine

Objectives

- o Experience using a third party spell program
- o Developing efficient methods for accomplishing autocomplete

In the previous document (SpellAndAutocompleteInSolr.pdf) you saw how to enhance the Solr program with spelling correction and an autocomplete (suggest) function. In this exercise you are asked to replace the existing Solr functionality for spelling correction and to enhance Solr's autocomplete functionality. In the case of spelling correction you will use an existing third-party program adapted to your downloaded files. In the case of autocomplete you will need to enhance your client program that communicates with Solr to deliver autocomplete suggestions to the web interface you created in homework #3.

Description of the Exercise

Spelling Correction: in the class lecture you saw a complete spelling correction program developed by Peter Norvig. The program was written in Python. For this exercise you are welcome to use whatever third party spelling program you wish, or you may even write your own. Since most of you wrote your homework #3 client using PhP, you may want to adopt a version of Norvig's spelling program written in PhP. You can download the PhP version of Norvig's spelling corrector from here:

http://www.phpclasses.org/package/4859-PHP-Suggest-corrected-spelling-text-in-pure-PHP.html#download

(you will have to register at the site before being able to download the software, registration is free)

If you prefer to use Norvig's program in a different language, a wide variety of implementations can be found at the bottom of this page, http://norvig.com/spell-correct.html

Autocomplete: for the autocomplete portion of the exercise, you will first have to modify Solr to enable the suggest functionality. Next, you will have to modify your client program so it accepts single character insertions to the text box, and returns a list of completions/suggestions. You should also consider enhancing Solr's suggest feature with a set of terms that are likely to be entered for the particular USC school that you crawled.

Submission Instructions

There should be a report describing what you have done. This report should include:

1. Steps you followed to complete this assignment. Include the details of what tools and techniques you used to implement spelling correction and autocomplete.

2. Analysis of the results: In this you should provide examples of misspelled terms that are correctly handled by your spelling correction program. You should also provide some examples of autocompletion.

3. Using the submit command you should provide a single .zip (CSCI572_HW4.zip) file which contain the following files
● the external spelling correction program that was used
● all source code that you wrote, most especially the code implementing the autocomplete functionality.

You are required to submit your results electronically to the csci572 account on SCF so that it can be graded. To submit your file electronically, enter the following command from your Unix prompt:

Submit -user csci572 -tag hw4 CSCI572_HW4.zip