**CPSC5340– Text Processing and Search  
Assignment 1**

Details of the project are in the Jupyter Notebook in the course repository.

You will hand in a Zip file containing three elements:

1. A Jupyter notebook containing your solution. Please remove everything from the notebook except code that is required to test your solution.
2. A folder config containing the configuration for your SOLR collection
3. A file Retrospective.pdf which contains the following material
   1. Your name
   2. Is your assignment fully working? If not, what pieces are working and what pieces are not
   3. How much time did you spend on the assignment?
   4. Overall how did you find the homework in terms of understanding and applying concepts covered in class

To evaluate your solution I will do the following:

1. Start with a clean SOLR service (no collections loaded)
2. Put your notebook and config file in the same folder
3. Put a set of test documents in the same folder
4. Fix filenames in your notebook so they align with the test environment
5. Put code at the bottom of your notebook that
   1. creates the SOLR collection
   2. scrapes the test documents
   3. loads the test documents
   4. uses your functions to run test queries
6. Run your notebook from the beginning

So you don't need anything in your notebook apart from your functions. You can put comments in the notebook explaining your design and implementation decisions, but please delete the narrative cells (i.e. those that define the problem) and the intermediate tests I put in the notebook for clarification.