CS 599(02) - Information Retrieval

Programming Assignment #4 (optional)

Due date: 11/15/2014 11:59 pm

Statement

You have become familiar with different retrieval models, scoring and ranking schemes, so far. In this assignment, you will build your toy search engine using Apache Solr. For this, you need to perform the following steps:
- crawl
  - you can use any crawler of your own to crawl webpages; i would prefer you to use nutch
  - you need to crawl atleast 1000 webpages from the domain csupomona.edu
- index
  - index these pages using Solr
- search
  - use relevant query of your own to search documents you have indexed
  - you need to provide me with a list of search queries which your system can return

Submission

You need to zip the following into one file pa4.zip:
- solr directory (which is present under ../example/ dir)
- query file containing a list of at least 50 queries which will work in your search system

Evaluation

For this assignment, you need to individually show me your results. Besides you need to submit the .zip file to the Blackboard.

Honor Code

Students should not share any source code or crawled data.