CS160 - Homework 1
Due: Wednesday Sept. 9, Before class

1. (15 points) Find 5 different IR systems

   For each system, find one query where the system performs well and one query where the system does not perform well. In both cases, describe what the system is doing well/poor. You don't need to write much, just a sentence or two. At least one of your IR systems must be a non-web page IR system.

2. (20 points) Find 2 articles in popular media (e.g. nytimes, cnn, digg, slashdot, ...) discussing IR or issues related to IR (e.g. online advertising).

   The article should be discussing technological or social issues (i.e. I don't want articles talking about Google's quarterly earnings). If there's any question about the appropriateness of an article, contact the professor. For each article:

   (a) Write a one paragraphs abstract/summary of the article.
   (b) Write 3 "discussion" questions that we can use in class to talk about the article.

   Time permitting, throughout the semester, I'll assign one or two of these a week for reading and we'll take 10 min. in class to discuss the articles.

3. (5 points) Describe a set of term postings lists and a boolean query where the heuristic for merging the postings lists of the query terms with the shortest lists first is sub-optimal. State what the optimal query order should be.

4. (5 points) 2.5 pg. 35 - 1 to 2 sentences is sufficient

5. (6 points) 2.9 pg. 41 - For part (a), also state the position(s) where the query matches.

6. (5 points) What is the potential problem of using stop words in combination with positional indexing? How could you solve this problem? A sentence or two for each question sufficient.

7. (5 points) Read all of the "Administrative" handout. What is the URL of that handout?