Due Date: No later than start of class on Mon Dec 5, 2011.

Problem 1. (25 points)
Use a merge-and-prune approach for Intersection (i.e Figure 12 of page 21 of Subject 7), to find the cost of an efficient evaluation of the query
\[ tA \text{ AND } tB \text{ AND } tC \text{ AND } tD \text{ AND } tE \]
where the doclists of \( tA, tB, tC, tD, tE \) are of length 1000, 10, 2000, 5, 10000 respectively. Explain.

Problem 2. (25 points)
For the terms and doclists of Problem 1, what is the cost of an efficient evaluation of the query
\[ (tA \text{ AND } tB) \text{ OR } (tC \text{ AND } tD) \]
? Explain.

Problem 3. (50 points)
Repeat the following study for Google and Bing. We execute two queries in Google and Bing and the collected results can be found in the last 4 pages of this homework as well as in the Handouts section of the course web-page. One query(Q1) is CS 485 NJIT and the other(Q2) is CS 485 Web-search. However, the CS 485 course number is a generic special topics one and thus NJIT CS 485 courses on Android development that ARE NOT RELEVANT TO OUR CS 485 course are output. Thus for Q1/Q2 relevant documents are only those of OUR CS 485 course. The first query is not very well thought-out. The second is more refined. You are asked to evaluate those two queries and also Google’s and Bing’s capabilities in generating relevant documents. We compared both engines in HW1; we conclude the course by evaluating these engines’ search capabilities. When you examine the output results tabulate by starting with Google Q1, then Q2, and then move to Bing and do not deviate from this order! Do not include ad links in considering relevant documents!

(a) Tabulate for each engine and query and engine, the number of hits reported in the result page attached (easy), the number of relevant documents (be cautious), and the precision for the single result page reported.

(b) Include 3-point effectiveness rates and 3-point averages of those rates separately for each one of the four queries/engines. (Note: Use the methods of Subject 8 as needed if incomplete information is available.)

(c) Fill the following table with data (columns 2 and 3). If Google is better give Google 1pt and Bing 0 in columns 4 and 5. If Bing is better, give Bing 1pt and Google 0. If it is a tie, give each one 1pt each. Add up the points of columns 4 and 5 and provide sum in the last row. Who wins? (Note that the interpolated precision figures asked might not be the ones computed from part (b). “Interpolate” appropriately as it was done in Subject 8 and in class.)
Problem 4. (25 points)
Find the page rank of the graph of Figure 1. Use the algorithm of page 14 of Subject 9. Initial values are \(1/N\). Also use \(a = (d =) = 0.85\) in the page rank formula. Iterate as many times as needed for the error to be less than \(10^{-4}\). Use the synchronous update version.

![PageRank Graph](image)

Figure 1: PageRank Graph.

Problem 5. (Bonus 35 points)
This is a bonus problem that can earn you 35 more points than the normally allotted ones (i.e. 160 instead of just 125 points).

In HW2 you were asked to create a web-page for HW2 that would point to the course web-page and include some form of a time-stamp in it. I hope that you had regularly updated the date/time-stamp information field, and searched for that page in Google or Bing regularly. What is the refresh rate of Google or Bing of your own page or alternatively of our course web-page?

Prove your claims by providing snapshots of Google’s or Bing’s storing your web-page or the CS 485 course web-page (otherwise) to determine refresh rates i.e you need two consecutive snapshots to have a convincing case for the 35 points.

(Take snapshots to prove your claims and compute a refresh rate. PrintScreen pages and submit them for the solution to be considered for the 25 points.)
Figure 2: Google for CS 485 NJIT
Figure 3: Google for CS 485 Web-search
Figure 4: Bing for CS 485 NJIT
Figure 5: Bing for CS 485 Web-search