CS458 - Homework 3
Due: Thursday Oct. 11 at the beginning of class

1. (2.5 points) 6.22 (1-2 sentences)
2. (2.5 points) 7.1 (1-2 sentences)
3. (20 points) Calculating tf-idf

   Feel free to use matlab, python or excel to calculate these answers.

   (a) 6.10

   (b) 6.15

   (c) Compute the document similarities with the document vectors above for the query "insurance car insurance for new autos" using boolean term frequncies for the query.

   (d) Compute the document similarities with the document vectors above for the query "insurance car insurance for new autos" using "**lnc**" query weighting (see the fifth from last slide on 9/25). Use your proposal from above to handle out of vocabulary terms in the query. Did your ranking change from the previous question? What type of query would you expect to see a change in answer?

4. (10 points) Let $K = 2$ and $r = 3$. Give an index and a query such that the list of candidate documents generated using the champion list approach does NOT contain the best K documents using a nnn.nnn weighting model (i.e. only tf weighted).