

# CSC 391/691 – Spring 2014

## Exercise Set #1

Due 3:30pm, January 21

Show work for partial credit. These exercises should be completed individually but feel free to discuss concepts and examples with other students.

1. Exercise 1.2.1 from the text.

(a)

(b)

(c)

2. Exercise 1.3.1 from the text.

(a)

(b)

3. Exercise 1.3.2 from the text

(a)

(b)

4. Here are three "documents," each consisting of a single sentence:

- a. cats and dogs like to fight
- b. take your cat to the dog store
- c. we all like cats and we all like dogs

First, stem the words by replacing plurals by their singular forms. (In general, stemming will require other transformations as well.) Compute the term frequency (TF) for each of the following words: "cat", "like", "store", and "to" in each of the three documents. Also, compute the inverse document frequency (IDF) for each of these words. Note: you can use 1.59 as the value of  $\log_2 3$ . Then report each of the following values:

- The TF for "cat" in document (a) is \_\_\_\_\_
- The IDF for "store" is \_\_\_\_\_
- The TF for "to" in document (c) is \_\_\_\_\_
- The TF for "like" in document (c) is \_\_\_\_\_
- The IDF for "cat" is \_\_\_\_\_
- The IDF for "store" is \_\_\_\_\_