**Quiz 9**

**DSCI 558 BUILDING KNOWLEDGE GRAPH**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Question 1 (2pts)

Given the cost of string operations as follows: insertion = 1, deletion = 2, substitution = 1. Compute the Levenstein distance between “Saturday” and “Sunday”

5

## Question 2 (3pts)

Compute the Jaccard similarity between “University of Southern California” and “University of California Los Angeles”. Strings are tokenized based on whitespace into words first.

|intersection| = 3

|union| = 6

Jaccard similarity = 0.5

## Question 3 (1pts)

Given the following set of strings:

[‘Actor Jackie Chan’, ‘Jackie Chan Kong-sang’, ‘Professor Alexander W. Brown’, ‘Alexander Williams Brown, Senior Professor’]

Assuming that you need to figure out which pairs of string that are referring to the same person, which of the following string similarities will you choose ?

1. Jaro-Winkler
2. Smith-Waterman
3. Jaccard Similarity
4. Needleman-Wunch

## Question 4 (1pts)

Given the following set of strings:

[“Apple Corporation”, “IBM Corporation”, “Intel Corp”, “Apple Corp”, “IBM Corp”]

Assuming that you need to figure out which pairs of string that are referring to the same corporation, which of the following string similarities will you choose ?

1. Jaro-Winkler
2. Levensteish
3. Jaccard Similarity
4. TF/IDF

## Question 5 (1pts)

Given the following set of strings:

[“Building mashups by example”, “Building Mashups By Example”, “Semi-automatically mapping structured sources into the semantic web”, “Semi-Automatically Mapping Structured Sources Into The Semantic Web”]

Assuming that you need to figure out which pairs of string that are referring to the same paper, which of the following string similarities will you choose ?

1. Jaro-Winkler
2. Levensteish
3. Jaccard Similarity
4. TF/IDF

## Question 6 (1pts)

Given the following set of strings:

[“Tom M. Mitchell”, “Michael I. Jordan”, “Jordan; Michael”, “Mitchell; Tom”]

Assuming that you need to figure out which pairs of string that are referring to the same person, which of the following string similarities will you choose ?

1. Jaro-Winkler
2. Soundex
3. Jaccard Similarity
4. TF/IDF

## Question 7 (1pts)

Given the following set of strings:

[“Gaddafi”, “Qaddafi”, “Gheddafi”, “Ashcraft”, “Ashcroft”]

Assuming that you need to figure out which pairs of string that are referring to the same person, which of the following string similarities will you choose ?

1. Levensteish
2. Soundex
3. Jaccard Similarity
4. TF/IDF