

**DS 320: Homework 2**  
**Programming Assignment**

Due: Tuesday, October 2, 2018, 11:59pm (EST)

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

<b>Question</b>	<b>Points</b>	<b>Score</b>
1	10	
2	30	
3	10	
Total	50	

1. (10 points) Implement the following function for determining the Jaccard similarity between two strings:

```
jaccard_sim (x, y, k)
"""
x: the first string
y: the second string
k : integer value represents the length of k-grams

Return: the Jaccard similarity between two sets of k-grams
       derived from x and y strings, respectively
"""
```

Let  $s_1 = \text{"Apple Corporation, CA"}$  and  $s_2 = \text{"Apple Corp, CA"}$

Print the output of:

`jaccard_sim(s1, s2, k)` for  $k = 1, 2, 3$ , and  $4$

2. (30 points) Implement the following function for computing the edit distance between two strings:

```
edit_distance(x, y)
"""
x: the first string
y: the second string

Return: the minimal cost of transforming x into y
"""
```

Let  $s_1 = \text{"Apple"}$  and  $s_2 = \text{"Apel"}$ . Print `edit_distance(s1, s2)`. Validate your output by manually applying the recursive dynamic programming equations.

3. (10 points) Implement the following function for computing the similarity between two strings based on their edit distance score:

```
ed_similarity(x, y)
"""
x: the first string
y: the second string

Return:  $1 - \text{edit\_distance}(x, y) / \max(\text{length}(x), \text{length}(y))$ 
"""
```

Let  $s_1 = \text{"Apple Corporation, CA"}$  and  $s_2 = \text{"Apple Corp, CA"}$

Print the output of `ed_similarity(s1, s2)`.