

# Birzeit University

Department of Electrical & Computer Engineering

Second Semester, 2016/2017

ENCS311, Assembly LAB

## Lab Project – Measuring Similarity between Two Sentences

---

### **Objective:**

Write an 8086 assembly program to find the similarity between two sentences.

### **Procedure:**

1. Read sentence S1 from S1.txt file and read sentence S2 from S2.txt file.
2. Remove punctuation marks from both sentences.
3. Make all characters as a small letters.
4. Remove Stop words from both sentences. Stop Words are words, which do not contain important information. Use the following list of these words:  
[I, a, an, as, at, the, by, in, for, of, on, that]
5. Remove the duplication of words from both sentences. In other words, each word will appear once per sentence.
6. Calculate the similarity as the size of the intersection of words between the two processed sentences divided by the size of the union of the two processed sentences:

$$Similarity = \frac{(S1 \cap S2)}{(S1 \cup S2)}$$

A value “0” means the two sentences are completely dissimilar, “1” that they are identical, and values between 0 and 1 representing a degree of similarity.

### **Notes:**

1. Work in a groups of at most two students
2. Deadline: Thursday, 18 may, 2017 at 11:59pm. Please submit your project through Ritaj.
3. This project is per group effort: instances of cheating will result in you failing the lab.