



## Assignment 2: Implementing TF-IDF

### Description:

Implement TF-IDF from **scratch** using the equations presented in the lecture then apply it to a 4 different pdf documents from your choice. P.S. you can generate the documents.

### Requirements:

1. Preprocess texts (Tokenization & Case normalization “Lowering case” & Punctuation removal)
2. Compute Term Frequency (TF) and Inverse Document Frequency (IDF).
3. Display the words with the highest TF-IDF value.
4. Use TF-IDF to encode/embed your documents.

### Delivery:

- The assignment will be done in **teams of Four**.
- The assignment discussion will be held next Sunday 16/3 at lab time (the grades will be on discussion only there is no submission).