



## Lab 2

**Objective :** The main purpose behind this lab is to get familiar with NLP Rule based, Regex and NLP Word embedding.

### Work to do:

#### Part 1:Rule Based NLP and Regex:

Using Regex write a python code that can generate bill from a text given by user.

Use case:

“I bought three Samsung smartphones 150 \$ each, four kilos of fresh banana for 1,2 dollar a kilogram and one Hamburger with 4,5 dollar”.

#### **Generated Bill :**

Product	Quantity	Unit Price	Total Price
Samsung smart phone	3	150	450
Banana	4	1,2	3,2
Hamburger	1	4,5	4,5

#### Part 2: word Embedding :

1. Apply one hot encoding, bag of words, TF-IDF technics on the Data vectors collected during the lab 1.
1. Apply one Word2Vec Approach (Skip Gram, CBOW) on the same DataSet.
3. Apply Glove and FastText approaches on the same DataSet.
4. Plot all the encoded / vectorized vectors by using Tsne Algorithm, evaluate those approaches and give a general conclusion.



**Université Abdelmalek Essaadi**  
**Faculté des Sciences et techniques de Tanger**  
**Département Génie Informatique**  
Master : AISD  
NLP  
Pr . ELAACHAK LOTFI



**Notes :**

- **At the end each student must give a brief synthesis about what he has learn during the proposed lab.**
- **Push the work in the Github repository and write a brief report in Github readme file.**

**Tools:**

Google colab or Kaggle, gitlab/github, spacy , NLTK, Sklearn.