Lab 1

Objective : The main purpose behind this lab is to get familiar with Scraping and NLP Pipeline.

Work to do :

1. Use libraries like Scrapy and Beautiful Soup to do scraping from server Arabic web sources concerning a specific domaine.

2. Store the raw data on a NoSql database "MongoDB".

3. Establishment of NLP Pipeline (Text Cleaning, Tokenization, Stop words, Discretization, Normalization).

4. Do the Stemming and Lemmatization, then compare the tow mechanisms.

5. Apply Parts of Speech technics based on both Rule based and Machine learning approaches.

5. Apply NER Methods.

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.