**MT project Topics**

1. Arabic Dialects translation
2. Pre-trained model fine tuning
3. Arabic-Turkish Machine Translation System
4. Text Simplification using Machine Translation
5. Natural Language to SQL translation

If you want to work on a different topic, talk to me first

**Group Formation**

To work in groups on the project, fill the form <https://forms.office.com/r/0vJm6YC0fX>

**Project Phases**

1. Select Topic
2. Select Dataset and describe it with references, use the standard split from the source or split it by yourself.
3. Develop/ fine tune at least two models, describe them with citation, conduct experiments. Then compare the results of the two models tested on the same dataset with at least three standard MT evaluation metrics.
4. Submit the code and the notebooks and the report according to the template

Project Title: put here the project title

|  |  |
| --- | --- |
| Student1 Name (in Arabic) |  |
| Student1 ID |  |
| Student2 Name (in Arabic) |  |
| Student2 ID |  |

# Introduction

Describe the project briefly

# Dataset

list the selected here with references, then describe it (language pair, data format, # of sentences, average length of sentences, the training & the test parts, …….)

have a look on <https://mksaad.wordpress.com/2020/12/27/machine-translation-tools-data/>

# Models & Experiments

List the selected models / pre-trained models with references, describe them, and describe how you use these models (training from scratch, fine tuning, ensemble, ….)

# Results & Discussion

Present results in table like this, and Comments on the results.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **BLEU** | **METEOR** | **ROUGE** |
| Model1 |  |  |  |
| Model2 |  |  |  |

You should use these three: BLEU, METEOR, ROUGE

Have a look on <https://github.com/Maluuba/nlg-eval>

# 