Final project

Introduction

- NLP helps to solve many problems related to human-computer communication
- During the course we have reviewed some of the most relevant techniques and tasks in NLP
- In recent weeks we have discussed the topic of neural language models such as LLMs and explained their potential use today
- The aim of the final project is to integrate the techniques seen in class, including LLMs, to implement a solution to a specific task related to NLP technologies

Specifications

- Form a team of 3 to 4 people
- The team must propose a solution to the following task:
  - Information retrieval
  - Text classification
  - Sentiment analysis
  - Other NLP related task

Requirements

- For the solutions to be valid they must:
  - Use some of the NLP techniques reviewed in class
  - It must use LLMs
  - It is necessary to use metrics to evaluate the performance of the methods
  - The use of Spanish corpus is mandatory
  - A user interface to interact with the implemented solution is also recommended (but not mandatory)

Methodology

- The following methodology is mandatory for the implementation of the project:

    1) Project definition
    2) Corpus selection
    3) Selection of frameworks, libraries or technologies used
    4) Specification of normalization steps
    5) Specification of text representation
    6) Selection of machine learning/deep learning/LLMs methods
    7) Definition of evaluation metrics
    8) Definition of a baseline to compare model performance

Evidence

- The following evidence must be uploaded to TEAMS platform:
  - Source code
  - Resources like corpus, lexicons, datasets, etc.
  - Report of the project development

Report specifications

- The report must contain the following sections:
  - Introduction. Problem description, relevance of the task, possible applications and proposed solution
  - Corpus and datasets. Corpus and datasets used for the task, specifying:
    - Size, content description, tagged classes (when used)
  - Additional resources. Description of all the resources used as lexicons, pretrained LLMs, etc.
  - Solution. Description of the solution specifying:
    - Text preprocessing
    - Training phase
    - Test phase
  - Experiments and results
    - Description of the experiments performed
    - Evaluation of the training phase
    - Evaluation of test phase
  - Conclusions and future work