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 tasks in NLP

- The aim of the final project is to integrate the techniques seen in class 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 task and a solution

- Some of the tasks considered but not limited to the project are:

  - Information retrieval

  - Sentiment analysis

  - Text summarization

  - Machine translation

  - Conversational agents

Requirements

- For the proposed project to be considered valid, it must cover at least the following aspects:

  - It must use the NLP techniques reviewed in class

  - It must integrate Machine Learning methods

  - It is necessary to use metrics to evaluate the performance of the methods

  - The use of Spanish corpus is strongly recommended

  - A user interface to interact with the implemented solution is expected

Project examples

1) Information retrieval system for searching and retrieving books from an abstract

2) System for detecting polarity/emotions in opinions/reviews

3) System for summarizing news articles

4) Conversational agent to solve questions related to characters from science fiction novels or series

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 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 mus 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:
    - Normalization steps
    - Text representation
    - 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