Assignment: Natural Language Processing

Due date : May 16, 2021 11:59pm

Deliverables:

1. A report (with introduction, body and conclusion) describing the task, your experimentation protocol and your results in a pdf format (10 to 20 pages).

AND

2. The Python code.

Dataset:

The dataset contains thousands of abstracts of scientific papers dealing with agriculture and machine learning.

Tasks:

Students must choose one of the following tasks:

- **Relation extraction:** this task consists of two sub-tasks. First, find a set of keywords related to agriculture (crop, smart farming, etc.) and another set of keywords related to machine learning (CNN, RNN, logistic regression, etc.) Secondly, find relationships between the two sets (e.g. Crop - CNN) and represent them graphically.

- **Abstract summarization:** this task consists of summarizing a given abstract into two or three sentences. Students are highly encouraged to consider both "abstractive summarization" and "extractive summarization".

- **Question Answering**: this task consists of responding a question (e.g. What are CNNs used for?) given an abstract that contains the answer. The model must be validated on a relevant and representative sample of abstracts and questions.

- **Knowledge graph**: this task consists if building a graphical representation of a chosen subtopic (smart farming, precision agriculture, milk, etc.) The graph should highlight the subtopic keywords as well as the relationships between the keywords.