

# Project Description

Nova IMS, 2019

## Bag-of-Authors

Identifying the author who wrote a text excerpt.

The goal of this project is to develop a model capable of understanding the who wrote a text excerpt with 500 or 1000 words. The core of this project is a non-balanced corpora containing several books from different Portuguese authors and text excerpt that contain no information about the author.

With the NLP techniques you have learned during class you must create an algorithm that receives text excerpts with 500 or 1000 words and identifies the author. Note that the provided test corpora is not entirely reliable (the labels are missing).

**In your final report you must tell me who was the writer of those excerpts!** I will not provide the labels for the test set.

I recommend the usage of python 3 and libraries such as [NLTK](#) and [Scikit Learn](#). Also, the project is simple and can be solved in various different ways which means that there is no correct solution! Be creative and keep it simple!

### Important Notes:

- The delivery deadline for the project and report is March 27th, at 17:00.
- The report must contain 3 pages explaining your approach to the problem and your results for the test excerpts.
- Students should not use code from each other!
- The project is to be developed in groups of up to three.