**Master BeNeFri in Computer Science**

Course: Digital Humanities
Fall 2022

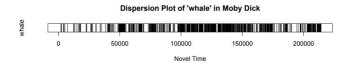# Exercise #3:  Literature and Moby Dick

**Instructions**

Download from the ILIAS website the Moby Dick text (filename: "MobyDick.txt").
This file contains the ebook of the famous novel Moby Dick, or The Whale, written
by Herman Melville.
This exercise series consists of 8 practical questions. Upload your answers and the
source code used for your computations to the ILIAS website. You can submit either
a .pdf file or comment the source .py file. IPython/Jupyter notebook files (.ipynb) are
allowed as well.

**Practical Questions (consider the Moby Dick data)**

1) How many word-tokens do you count in this novel?

2) What is the vocabulary size of this novel?

3) How many occurrences of "Swiss" do you count in this novel?

4) Can you count the number of plays per author?

Consider the following picture. It describes with which intensity the world "whale"
and its close variations ("whales", "Whale", …) appear in the novel.



Dispersion Plot of 'whale' in Moby Dick

Hint: **don't** try to reproduce **exactly** the same picture (it was generated in R and not in
Python).

5) Can you generate a similar view for the word "whale"?

6) Repeat the previous exercise with the word "Ahab" (the captain's name).

7) Repeat the previous exercise with the word "upon".

8) Do you notice any similarities/differences between these three views?