# Digital Tools for the Humanities

## Prof. Francesco Mambrini

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course explores the main research questions and methods in the field traditionally known as "Humanities computing", focusing on the computational analysis of the literary languages. The lectures will cover methods and techniques of corpus exploration that are used for tasks of literary analysis such as authorship attribution or character analysis. The Children’s Literature CCR Corpus (University of Birmingham) will be used as testbed for the different types of computational experiments and investigations.

By the end of the course, the students will have gained a comprehensive understanding of the benefits and limitations of computational methods to the analysis of texts for research in the Humanities. They will know how to perform simple and more advanced analyses on a corpus of novels using Python and how to include computational analyses in their research on textual sources and literary works.

***COURSE CONTENT***

The course is articulated in the following sections.

1. Introduction to Humanities Computing:
   1. Overview of the field and its history.
   2. The quantitative analysis of language for the Humanities: key concepts and problems.
   3. Humanities Computing in/and the Digital Humanities today.
2. Data and metadata for Humanities computing:
   1. Corpora and digital archives.
   2. The Children’s Literature CCR Corpus (University of Birmingham)
3. The computation of style: basic notions
   1. From words to vectors
   2. Token distributions and measures of lexical variety
4. The computation of style: more advanced explorations
   1. Burrows’ *Delta* measure
   2. Clustering
   3. Topic modeling

***READING LIST[[1]](#footnote-1)***

[All the readings reported below are not mandatory for the exam]

Karsdorp, F., M. Kestemont, and A. Riddell. 2021. *Humanities Data Analysis: Case Studies with Python*. Princeton University Press.

Jockers, M. 2014. *Text Analysis with R for Students of Literature*. Heidelberg: Springer.

Hoover, D. 2008. “Quantitative Analysis and Literary Studies.” A Companion to Digital Literary Studies, eds R. Siemens and S. Schreibman. Oxford: Blackwell. <http://www.digitalhumanities.org/companionDLS/>.

***TEACHING METHOD***

The course will consist of a combination of lectures and hands-on exercises that will be performed together by the lecturer and the students.

***ASSESSMENT METHOD AND CRITERIA***

The final assessment will consist in an oral exam. The aim of the examination is to assess the acquisition of the content presented in class and the ability to use the newly acquired concepts to solve some practical exercises related to the computational analysis of textual data for research in the Humanities.

In particular, the questions of the exam will be focused on: a) theoretical aspects related to the tools and the methods for the analysis of literary texts; b) practical scenarios, such as: 1. the choice of the best tool to extract a certain type of information or to solve a specific task in the Humanities, 2. the planning of a hypothetical scenario proposed by the lecturer (data collection, selection of tools, data processing) to carry out scientific activity related to research in the Humanities.

Both types of questions will have the same weight in view of the final grade.

***NOTES AND PREREQUISITES***

The course requires some basic familiarity with the problems and methods of Computational Linguistics, as well as with language resources. Students should also know how to install and execute open open-source applications and command-line tools.

*Place and time of consultation hours*

By appointment (to schedule, write to: [francesco.mambrini@unicatt.it](mailto:francesco.mambrini@unicatt.it)).

Place: CIRCSE Research Center, Franciscanum building, second floor.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)