Master BeNeFri in Computer Science

Course: Digital Humanities
Fall 2022

Exercise #2: Literature and the French Theater

Instructions

Download from the ILIAS website the French Theater data (path: "/french-theater/"). This folder contains french theater plays (stored in XML format), without any additional information. The author name is provided according to the pattern "LastName, FirstName" (e.g., "Smith, Adam"). Lowercase and uppercase make no differences (e.g., "Rousseau" and "ROUSSEAU" are the same author).
This exercise series consists of 2 theoretical questions and 4 practical questions. Upload your answers and the source code used for your computations to the ILIAS website. You should submit your answers to the theoretical questions in .pdf format. For the practical ones, you can submit either a .pdf file or comment the source .py file. IPython/Jupyter notebook files (.ipynb) are allowed as well.

Theoretical Questions (motivate your answers!)

1) Cite four differences between XML and HTML standards.

2) Are both XML and HTML fully declarative languages?

Practical Questions (consider the French Theater data)

1) How many unique author names can you find?

2) Are you sure that all these unique names refer to distinct authors?

3) Can you reduce the variability around the author names?

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