PYTHON. FOR THE DIGITAL HUMANITIES. Course Syllabus

Instructor: Elena Fernández Fernández

Course content

- **Data Acquisition**
  - Webscrapping
  - APIs
  - PDF text Extraction
  - Databases
- **Cleaning and Pre-processing**
- **Data Analysis: Information Extraction**
- **Data Visualization**
  - Geospatial Analysis
  - Network Analysis

Course content may vary depending on the development of the course

Github repository url: ZI IT Training / APPH / Python For The Digital Humanities January 2025 · GitLab

- ❖ IMPORTANT: if we end up doing the Databases exercise (it will depend on how much time we have left) you will need to install the UZH VPN to access content from the Gale Databases (we will be learning how to use Digital Scholar Lab: for Digital Humanities & Scholarship). Please find in here the information to do so: External UZH Network Access (VPN) | Central IT | UZH

If you experience any technical problems installing Jupyter Notebooks or the UZH VPN feel free to contact the **UZH IT Help desk**:

Telefon +41 44 634 33 33
support@zi.uzh.ch

DAY 1

17.00-17.50

- Introduction
- What is Digital Humanities?
- Steps to follow in a Digital Humanities Project
- **Data Acquisition**: **Webscraping**
- Introduction to Webscrapping I: Project Guttenberg

18.00-18.50

- Webscrapping II: HTML
- Webscrapping Wikipedia

- Webscrapping Spanish Government Website (BOE)

19.00-19.50

- **Data Acquisition**: **APIs 1**
- Arxiv

DAY 2

17.00-17.50

- **Data Acquisition: PDF text extraction**

18-18.50

- **Data Acquisition: APIs 2**. Using the API of The Guardian

19.00-19.50

- **Cleaning and Pre-Processing Data**

Extra exercise (if there is time): Databases. Learning how to use Gale Digital Scholar Lab (Digital Scholar Lab: for Digital Humanities & Scholarship).

DAY 3

17:00-17:50

- Information Extraction

- NLTK

18.00-18.50

- Exercise 1

- Spacy

- Exercise 2

19.00-20.00

- Mapping Jules Verne. NER with Spacy

- Exercise 3

DAY 4

17:00-17:50

- **Geospatial Analysis**

- Exercise 1

18:00-18:50

- **Network Analysis**: Sociocentric Networks (weighted and unweighted).

19:00-20.00

- Exercise 1

- Exercise 2