

ENCODE Workshop ‘AI and ancient writing cultures’

Bologna 23rd-27th January 2023

The workshop has seen the participation of several experts in the field of Artificial Intelligence applied to disciplines studying ancient writings and scripts, who alternated between presentations of existing projects and hands-on training sessions. Participants had the opportunity to experiment with tools related to different branches of AI, such as pattern recognition, NLP, data visualisation, automatic annotation, tagging and linguistic parsing, and to discuss possible applications and future developments with experts.

CONTENT OF THE WORKSHOP

- An overview of AI-based methods and tools for DH researchers
- Signal and Noise: Epigraphic Ventures in Machine Learning
- The Artificial Papyrologist at Work – Digital Papyrology and the AI
- Analyze Tabular Data: some Tools and Workflows
- Annotating papyrus images for the palaeographer and for AI: What, how and why?
- AI and manuscript studies
- Using Transkribus in a wider architecture, some case studies
- Challenges and issues of using Transkribus in large late medieval manuscript collections: The Memoriali Project (MemoBo)
- Automatic tagging and parsing of Latin texts: methods, tools and challenges
- Basics of Python and Jupyter Notebooks: Q&A
- A practical introduction to machine learning and natural language processing on papyrus data
- The Pattern Analysis Software Tools

LEARNING OUTCOMES

At the end of the workshop trainees:

- Understand the different methods and instruments provided by the broader concept of Artificial Intelligence for the study of ancient written artefacts
- Being able to prepare datasets and models in order to manipulate data to answer research questions
- Being able to build simple machine learning models for classifying different characteristics of ancient texts
- Being able to train specific text recognition models in order to transcribe handwritten documents and reuse the output formats within a research or publication workflow
- Understand how linguistic annotation and automatic tagging work and critically evaluate results and discuss the reliability of these methods



- Being able to exploit automatic analysis of visual and tabular patterns in the research data from the study of ancient written artefacts.

COMPETENCES

The workshop has been organised taking into account the Digital Competence Framework for Citizens (DigComp.2.2: <http://dx.doi.org/10.2760/115376>) and, more specifically, aimed at providing training in the following areas and levels:

Competence area 1: Information and data literacy

- 1.2 EVALUATING DATA, INFORMATION AND DIGITAL CONTENT
Can evaluate instruments and methods pertaining to the field of Artificial Intelligence applied to ancient documents and critically discuss their reliability; are aware that the data, on which AI depends, may include biases.
(DigComp2.2: 1.2 level 5)
- 1.3 MANAGING DATA, INFORMATION AND DIGITAL CONTENT:
Are aware that AI systems use statistics and algorithms to process data and generate outcomes; can prepare and manipulate research data for automatic analysis of visual and tabular patterns (Pattern Analysis Software Tools) and prepare and train machine learning models for classifying different characteristics of texts (Orange Data Mining, Jupyter Notebook, Streamlit).
(DigComp2.2: 1.3 level 5).

Competence area 2: Communication and collaboration

- 2.1 INTERACTING THROUGH DIGITAL TECHNOLOGIES: are able to interact and give feedback to AI systems and to influence what it next recommends. (DigComp 2.2: 2.1 level 4)
- 2.2 COLLABORATING THROUGH DIGITAL TECHNOLOGIES: can share data, information and digital content with others through appropriate digital technologies, i.e. know how to export the available output formats from Transkribus in TEI XML.
(DigComp2.2: 2.2 level 4)

Competence area 3: Digital content creation

- 3.1 DEVELOPING DIGITAL CONTENT: are aware that AI systems can be used to automatically create digital content using existing digital content as its source; can prepare a dataset and train classification models in order to achieve desired outcomes using different tools for specific tasks (Transkribus, Jupyter Notebook, PAST).
(DigComp2.2: 3.1 level 4)
- 3.2 INTEGRATING AND RE-ELABORATING DIGITAL CONTENT: are aware that AI systems can help the users to edit and process digital content; can modify, refine, improve and integrate information and content into an existing body of



knowledge to create new, original and relevant content and knowledge; know how to incorporate AI edited/manipulated digital content in one's own work.
(DigComp2.2: 3.2 level 4)

- 3.4 PROGRAMMING: can plan and develop a sequence of understandable instructions for a computer to solve a given problem or perform a specific task (i.e. code in Python and manipulate Jupyter notebooks).
(DigiComp2.2/3.4 level 4/5).

Competence area 4: Safety

- Can evaluate the results and critically discuss the reliability of AI systems for the study of ancient documents. Can evaluate the ethical consequences of AI systems through their life-cycle, such as their environmental and societal impacts. (DigiComp2.2: 4 level 3)

Competence area 5: Problem solving

- 5.1 SOLVING TECHNICAL PROBLEMS: are aware that AI is a product of human intelligence and decision making (humans choose and clean data, design algorithms, train the models and apply human values to the outputs); can identify technical problems when operating devices and using digital environments, and solve them (from trouble-shooting to solving more complex problems), e.g. are aware and can use forums, tutorials and handbooks to find out why a query does not work. (DigComp2.2: 5.1 level 3/4)
- 5.2 IDENTIFYING NEEDS AND TECHNOLOGICAL RESPONSES: can recognise different examples of AI systems and select digital tools and possible technological responses to answer specific research questions. Can select ways to adjust and customise digital environments to personal needs.
(DigComp2.2: 5.2 level 3/4)
- 5.3 CREATIVELY USING DIGITAL TECHNOLOGIES: can use digital tools and technologies to create knowledge and to innovate processes and products, to engage individually and collectively in cognitive processing to understand and resolve conceptual problems and problem situations in digital environments; are aware that humans can use AI to enhance their creativity and are willing to engage in collaborative processes to co-design new products and services based on AI systems to support and enhance citizen's participation in society.
(DigComp2.2: 5.3 level 3/4)
- 5.4 IDENTIFYING DIGITAL COMPETENCE GAPS: are aware that AI is a constantly evolving field; can understand where one's own digital competence needs to be improved or updated, being able to support others with their digital competence development and seek opportunities for self-development (e.g. being aware of existing tools and projects using machine learning, NLP, pattern analysis, can use forums, tutorials and handbooks to learn Python...), are open and curious towards today's



emerging technologies and applications and intentionally discuss about their use with others.
(DigComp2.2: 5.4 level 4)

ENTRY REQUIREMENTS

According to the Calohee competence framework (<https://www.calohee.eu/>) trainees were required to show at entrance a level 7 of the Humanistic competences with special reference to the following sub-dimensions:

- **DIM 2 - TEXT AND CONTEXT**
 - Show knowledge of their own research data and of its nature and structure. Can define and describe the different types of data in relation to typology, source, area and time period.
 - Can retrieve metadata from different relevant sources for their own field of research.
 - Can analyse different aspects of their material and specific research questions which can guide the modelling of their data.
- **DIM. 3: THEORIES AND CONCEPTS**
 - Apply appropriate theoretical and clear conceptual approaches to humanities research.
- **DIM. 4: INTERDISCIPLINARITY**
 - Have a good understanding of the role of Digital Humanities within the broader context of the humanities.
 - Are aware of methods of different areas of research within the humanities.
- **DIM. 5: COMMUNICATION**
 - Demonstrate an active knowledge of English as a second language.
 - Demonstrate a good knowledge of digital and communication technologies and their uses.
- **DIM. 6 INITIATIVE AND CREATIVITY**
 - Understand the dialogic nature of the humanities within scientific and public debate: approach issues with critical awareness; think in scientific terms; pose problems.

