Digital Tools for Humanists Summer School 2025
Program Week 1 – June 16 to June 20 2025

Monday June 16

AM: *Refresher on computers and networking*

Vittore Casarosa (ISTI-CNR and University of Pisa)

One (simple) way to think of Digital Humanities is to think that it is just the use of "digital tools" in the study and research activities carried on by scholars in the Humanities. To better understand how digital tools work, and for the benefit of all those who were exposed to Computer Science a long time ago, or have been only marginally touched by it, we will briefly review the basics of computer architecture and the representation of information within a computer.
We will also see how the evolution of computer technology and of communication networks has led, in the early '90, to the explosive growth of the Internet and the Web, and how the actual Web is (slowly) evolving towards the Semantic Web.

PM: *Research Infrastructures supporting FAIR and Open data, tools, practices in the humanities and social sciences. The case of CLARIN*

Monica Monachini (CNR-ILC, CLARIN-IT & H2IOSC), , Francesca Frontini (CNR-ILC, CLARIN ERIC & H2IOSC), Giulia Pedonese (CNR-ILC & H2IOSC), Michele Mallia (CNR-ILC, CLARIN-IT & H2IOSC)

The lecture is designed to provide students with a practical and theoretical introduction to FAIR principles and Open Science, with a focus on the CLARIN research infrastructure. In the first part, through a presentation enriched with interactive quizzes, the fundamentals of proper data management in line with the FAIR principles and the strategic role of CLARIN in supporting the discovery and repository of linguistic resources and data will be explored. The importance of responsible data management will be emphasized, highlighting how a structured approach adhering to the FAIR principles fosters transparency, reproducibility and reuse of data in research.
The second part will be dedicated to an interactive laboratory session, during which students will participate in a role-playing game. Randomly choosing a simulated research project, they will work in small groups to develop a draft Data Management Plan. During this activity, students will address practical issues, such as the management of personal data, and identify the necessary releases for the processing of sensitive data. The lesson aims to develop practical skills and raise awareness among students on the importance of ethical and sustainable data management in scientific research.

Tuesday June 17

Local festivity to honor San Ranieri, the Patron Saint of Pisa

Wednesday June 18

GIS – Geographical Information Systems

Augusto Ciuffoletti (University of Pisa)

This tutorial explores the diverse facets of Geographic Information System (GIS) technology from the perspective of a digital humanist seeking to leverage it for research and documentation purposes. It begins with an introduction to fundamental concepts and terminology, highlighting the impact of GIS resources on humanistic research.

The tutorial then showcases four representative tools that exemplify different approaches to GIS technology. OpenStreetMap (OSM) is a collaborative project aimed at providing free, high-quality maps. It serves as an example of a cloud-based GIS service with an inclusive focus, offering basic editing capabilities but no storage for user-generated content. In contrast, QGIS is a standalone application dedicated to map processing and production. While the tutorial only covers a small portion of its functionalities, it provides a glimpse into its vast potential. UMap bridges the gap between these two tools, allowing users to access OSM maps, add custom features, and store maps for sharing or embedding in web resources. Lastly, Gaia GPS, an Android app, represents the category of mobile applications used for fieldwork. It utilizes a smartphone's GPS to track movements, record notes, and capture photos, with the collected data being processed or published using the previously introduced tools.
In the second part of the tutorial, the focus shifts to implementing a web service with a GIS interface. Participants are introduced to the Leaflet library, which extends JavaScript with GIS capabilities. This section requires minimal programming skills and guides learners through the step-by-step creation of a basic application with GIS functionalities. The final product exemplifies a GIS web application that integrates seamlessly with QGIS.
The tutorial adopts a "learn-by-doing" methodology, equipping participants with practical tools and actionable insights to practice the concepts and gain hands-on experience. All tools presented in the first part are free to use, while the programming activities in the second part are facilitated through the StackBlitz platform, eliminating the need for software configuration on participants' computers.

Thursday June 19

Designing a project in Digital Public History

Enrica Salvatori (University of Pisa)

The main characteristics of a Digital Public History project involving private and public realities of the territory will be illustrated, with the description of the main phases of its organization, implementation, maintenance and conservation. In the practical part we will try to create a work team on a concrete project and to design a possible work plan. Some existing projects will be analyzed by evaluating their characteristics from the point of view of the structure, sustainability, transparency of information and relations with different audiences.

Friday June 20

Methods and tools for digital philology

Roberto Rosselli Del Turco (University of Torino)

Digital philology is a fairly recent discipline aiming at applying ICT methods and tools to textual criticism. Quite a number of new digital editions have been published during the last twenty years or so. Many of them, however, are achieved by programming and configuring complex frameworks only accessible to medium-large research groups. Although the encoding of text in TEI-XML format allows the individual scholar to prepare a digital edition, the online publication and navigation still remain a complicated, potentially expensive, operation. EVT (Edition Visualization Technology) is an open-source tool the purpose of which is to allow the publication of scholarly TEI-based editions in an easy way, through a user-friendly interface and making available several research tools. This course will introduce the subject of digital philology and text encoding using the TEI-XML standard. It will be followed by a hands-on final session in which students will be able to experiment with EVT.

Program Week 2 – June 23 to June 27 2025

Monday June 23

Introduction to Natural Language Processing

Rachele Sprugnoli (University of Parma)

Natural Language Processing (NLP) is an interdisciplinary field whose goal is to create machines that understand (and even generate) natural languages. NLP applied to Humanities disciplines helps in dealing with large amount of data, extracting information and finding relationships and patterns between words.

The lesson will feature: (i) an introduction to the main concepts and approaches related to the NLP field; (ii) hands-on activities on some NLP tasks, such as lemmatization, part-of-speech tagging, named entity recognition, topic modelling.

Tuesday June 24

Large Language Models – a practical guide

Irene Sucameli (University of Pisa)

These days, the new language models – on which technologies like Chat-GPT are based – have been matter of debate, even for the perspectives and the social implications that derive from them. But what a Large Language Model (LLM) really is? And how could we get the best from them? To discover that, during the morning we will talk about what LLMs are, the idea behind their creation, and how they are implemented and trained to be adapted to different tasks. In the afternoon, a hands-on workshop will take place, during which we'll test some practical applications of these models to build a starter digital toolkit useful for researchers and digital humanists.

Wednesday June 25

Multimodal Artificial Intelligence

Fabrizio Falchi, Fabio Carrara, Nicola Messina (ISTI-CNR)

This course explores the foundations and applications of deep learning, focusing on multimodal artificial intelligence — the integration of data from multiple sources, such as images and text. Deep learning methods are particularly effective at creating data representations that support tasks like classification and retrieval. The course emphasizes how these techniques can be applied to analyze and connect information across different types of data, focusing on images. The theoretical part of the course covers the core principles of deep learning and its applications in computer vision and multimedia retrieval. The practical sessions will provide hands-on experience with tools and methods for building and using representations suitable for multimodal tasks, including image-based classification and retrieval.

Thursday June 26

AI Meets the Archive: Refining Generative Tools for Historical Research

Seamus Ross (University of Toronto)

How can Generative AI help us unlock the past—and where does it fall short? This class invites participants to explore the transformative potential of AI tools like ChatGPT, DeepSeek, and Claude.ai for historical research. Through a combination of lectures and hands-on experimentation, we will examine how these tools can be applied to three distinct historical document sets: 19th century plantation business records, a mid-18th-century cookbook, and an archive of early 19th century personal correspondence.

Participants will refine prompting strategies to maximize accuracy, compare outputs across different AI models, and identify validation techniques to assess reliability and bias in AI-generated interpretations. Key topics include text normalization, named entity recognition, thematic analysis, and historical validation, all framed within a critical approach to AI that emphasizes active engagement over passive acceptance.
Special attention will be given to examining strategies for working with large document collections that require analysis in manageable parts due to model limitations (e.g., token constraints). By the end of the session, participants will have a deeper understanding of how to shape AI outputs for research, recognize its limitations, and apply best practices for AI-assisted humanities inquiry.

Friday June 27

To be a historian in AI times

Michela Galli (Università di Roma - La Sapienza)

This lecture will focus on automatic text recognition (ATR) software, providing participants with a theoretical basis, key terminology, contexts of use and an overview of current functionalities. The session will include a detailed introduction to the two main ATR platforms: eScriptorium and Transkribus. In the practical part of the workshop, participants will work in groups on both platforms using real handwritten materials. Practical exercises will cover software use in the cloud, uploading images, applying models and the entire workflow required for transcribing manuscript documents.