

Digital tools for humanists - Summer school 2024

Program Week 1 – Working on Texts - June 3 to June 7 2024

Monday June 3

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.

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): Introduction

Giulia Pedonese (CNR-ILC & H2IOSC): CLARIN tools and services – hands-on session

Angelo Mario Del Grosso (CNR-ILC): Knowledge sharing and best practices

This 4-hour module is intended to be an introduction on research infrastructures (RIs) in the field of Social Sciences and Humanities. After a short recap on FAIR and Open Data, participants will learn how to incorporate and exploit RIs data, knowledge, and practices in their research. More specifically, we will concentrate on CLARIN, the European Common Language Resources and Technology infrastructure, with hands-on modules where participants will learn how to (F)ind resources, (A)ccess them, exploit them with (I)nteroperable tools and (R)euse them.

Central CLARIN tools such as the Virtual Language Observatory and the Switchboard will be presented, but also the wider network of data and knowledge centres (including those operated by the Italian CLARIN-IT consortium). We will focus in particular on the search and querying of historical records, both written and oral. We will also showcase highly interoperable resources, such as the ParlaMint corpus, as an example of best practices in the field, particularly well-suited for diachronic and comparative studies in contemporary history and social science.

We will conclude with an overview of the current panorama of RIs, extending our focus beyond CLARIN, to the Social Sciences and Humanities Open marketplace and to national initiatives such as the Humanities and Heritage Italian Open Science Cloud (H2IOSC), focussing on training materials as a special type of resources that RIs share and maintain.

Tuesday June 4

Methods and tools for digital philology

Roberto Rosselli Del Turco (University of Torino)

Elisabetta Magnanti (University of Wien)

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.

Wednesday June 5

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.

Thursday June 6

Historical GIS

Tiago Gil (University of Brasilia)

In recent years, cartography has made significant advances in the representation of narratives and historical processes, both in literary criticism and in studies on displacements in social history, such as migrations, pilgrimages, transhumance, and commercial flows, amongst other instances. The purpose of this course is to discuss the cartographic representation of the movement and to develop skills in the creation of cartography aimed at representing historical processes as well as (textual and oral) narratives. Journey diaries, literary sources, and general documentation on population displacement will be sources used to discuss cartographic methods and techniques. Purpose software for cartography and languages such as R and python will be used through specific “libraries” (software packages) for these approaches.

The session will consist of: 1) An introduction to cartographic language and preparation of simple maps; 2) An introduction to thematic cartography and databases focusing on narratives; 3) Training activities based on different sources; 4) A map production workshop.

Friday June 7

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.

Program Week 2 – Working with AI - June 10 to June 12 2024

Monday June 10

Introduction to Artificial Intelligence

Irene Sucameli (University of Pisa)

Large Language Models – a practical guide. 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.

Tuesday June 11

To be a historian in AI times

Salvatore Spina (University of Catania)

The lecture will begin with an introductory workshop on Transkribus, a tool for the Optical Character Recognition of Handwritten Text (HTR/OCR), providing an overview of the tool, the model configuration, and an introduction to its training. Following this, the lecture will delve into the potential and perspectives of utilizing ChatGPT for textual analysis of historical documents. The practical exercises that follow encompass downloading the software, cloud-based work, uploading document images, applying models, and performing all the tasks needed for the transcription of handwritten documents. We then transition to using ChatGPT for transcription correction, text analysis, and identification of entities of interest such as names of individuals and locations. These identified entities can be utilized for constructing archival records, building a historical network, or other relevant applications.

Wednesday June 12

Moral Imperatives of Technological Ethics in Historical Studies: Navigating the Ethical Maze in deploying AI and ML for Historical Analysis and Narrative Production

Seamus Ross (University of Toronto)

Artificial Intelligence (AI) and Machine Learning (ML) hold the potential to introduce novel paradigms in historical study by leveraging its capacity for pattern recognition, text mining, and quantitative analysis. Through the analysis of extensive datasets, ML algorithms can uncover hidden patterns and correlations, providing historians with new insights into historical events and relationships. Additionally, ML enables the quantitative examination of historical data on a scale that was previously impractical, allowing for perhaps a more perceptive exploration of historical trends and socioeconomic factors. For instance, network analysis contributes a sophisticated understanding of historical actors and influences. The capability of ML to process diverse data types, including images and multimedia, supports a multimodal analysis, enriching historical studies with a broader range of perspectives. But as technology advances, the ethical considerations surrounding the deployment of AI and ML for historical analysis and narrative production become increasingly central to the historical process.

Participants will navigate the intricate ethical maze, exploring the implications of leveraging these technologies in shaping historical narratives and analyses. The utilization of Machine Learning (ML) to autonomously process and scrutinize historical materials, and then construct historical

narratives from these extensive documentary resources introduces a spectrum of moral issues. Even using AI and ML to assist in the process raise issues. Foremost among these concerns is the inherent risk of perpetuating biases inherent in historical documents themselves, potentially reinforcing skewed representations and perpetuating historical injustices. The assignment of narrative construction to algorithms raises questions about ethical responsibility, as current generations of AI and ML services appear to lack the nuanced ethical values imagined to be inherent in human historians.

One must wonder whether this approach also risks diminishing the human perspective, empathy, and contextual interpretation in historical approach and narrative development. Moreover, there is a looming threat to historical authenticity, with the potential prioritization of sensationalism over accuracy. Privacy concerns emerge in particular in the context of contemporary history as historical data inevitably involves personal information, necessitating careful ethical considerations in data handling. The lack of transparency, or explainability, in complex ML algorithms complicates the assessment of ethical implications. Cultural sensitivity becomes paramount, as algorithms may struggle to navigate diverse cultural contexts appropriately. Addressing these moral quandaries requires interdisciplinary collaboration, involving historians, ethicists, technologists, and policymakers to ensure that the integration of ML in historical narrative construction aligns with ethical principles and respects the richness and complexity of human history.

Over a day consisting of seminar discussion and hands on group based activities participants will investigate the ethical and moral challenges that arise when AI and ML are deployed in the historical sciences.