

Approcci basati sull'intelligenza artificiale per l'analisi della stampa storica

Workshop del gruppo R-NHEP

26 giugno 2026 (online)

Organizzazione
Giulia Mantovani e Franz Meier

09.45-10.00	Introduzione Giulia Mantovani (Dresda) e Franz Meier (Dresda)
10.00-10.45	AI-based data processing in the historical disciplines: use cases from Vienna Thomas Wallnig (Vienna)
10.45-11.30	Dai nomi al sapere: introduzione alla <i>Named Entity Recognition</i> per la ricerca dei periodici storici Iris Ferrazzo (Bonn)
11.30-12.00	<i>Pausa</i>
12.00-12.45	Dalla pagina allo schermo: un caso di studio di digitalizzazione di un repertorio cartaceo con ChatGPT Stefano Miani (Augusta/Firenze)
12.45-13.00	Conclusioni Giulia Mantovani (Dresda) e Franz Meier (Dresda)

AI-based data processing in the historical disciplines: use cases from Vienna

Thomas Wallnig (Vienna)

The presentation provides insight into some use cases from Vienna-based DH projects that are currently testing or adopting AI pipelines for data processing. This includes ATR/HTR (traditional layout recognition and VLMs), the use of LLMs for text processing (including NER) and the automated creation of user interfaces. Examples range from card catalogues to state handbooks, and from projects on papyri to the medieval history of Vienna and modern newspaper collections. (The selection may still change due to the highly dynamic developments in the field.) The talk takes the angle of university infrastructure, and thus addresses the questions of technical implementation along with those of sustainability, ethics as well as synergy and competition.

Dai nomi al sapere: introduzione alla *Named Entity Recognition* per la ricerca dei periodici storici

Iris Ferrazzo (Bonn)

I testi giornalistici sono ricchi di nomi di persone, luoghi, istituzioni ed eventi. L'identificazione manuale di tali entità è estremamente onerosa in termini di tempo; i metodi automatizzati possono accelerare notevolmente il processo. Il Named Entity Recognition (NER) è una tecnica di Natural Language Processing (NLP) che permette di identificare e classificare automaticamente queste entità all'interno dei testi.

La presente relazione introduce i concetti fondamentali del NER, illustrando le modalità di addestramento e applicazione dei modelli. Particolare attenzione è rivolta a strumenti e framework consolidati, accessibili anche a chi non possiede competenze avanzate di programmazione. Attingendo ad esempi tratti da periodici storici, il contributo dimostra come il NER possa essere utilizzato per analizzare reti di persone e luoghi, indicizzare vasti corpora di testo o preparare i testi a successive fasi di analisi.

La presentazione è esplicitamente rivolta a ricercatori privi di esperienza pregressa nell'ambito del NLP, offrendo sia una base concettuale, sia un punto di accesso pratico all'utilizzo degli strumenti NER.

Dalla pagina allo schermo: un caso di studio di digitalizzazione di un repertorio cartaceo con ChatGPT

Stefano Miani (Augusta/Firenze)

L'intervento illustra un caso di studio sulla co-progettazione con modelli linguistici di grandi dimensioni (LLM), in particolare ChatGPT, di un glossario elettronico, ottenuto a partire da materiale non nativo digitale e reso, quindi, fruibile "dalla pagina allo schermo". L'occasione nasce da una collaborazione esterna al progetto *Archivio della lessicografia dell'Ottocento-Novecento* (ALON), per il quale sono state redatte schede del *Vocabolario Nomenclatore illustrato* (Milano, Aldo Manuzio, 1909-1912) di Palmiro Premoli (1856–1917). In particolare presenteremo la digitalizzazione di un glossario di parole nuove posto in calce a un'edizione del *Vocabolario Nomenclatorie* del 1932.

L'intervento intende focalizzarsi non tanto sul glossario in sé – di cui comunque verranno illustrate struttura e consistenza –, quanto sulle potenzialità e punti critici del lavoro che un umanista può svolgere con ChatGPT, utilizzato come un informatico *prêt-à-porte*. Ci focalizzeremo sulle pratiche di interazione con il modello e sui controlli necessari per evitare che lo strumento produca delle "allucinazioni", soprattutto nella fase di scrittura e revisione del codice. Quali competenze è necessario avere? Come rapportarsi allo strumento affinché resti uno strumento — utile, ma sotto il controllo dello studioso?

L'obiettivo è offrire un modello operativo replicabile per la digitalizzazione di repertori non nativi, spendibile in altri progetti lessicografici, mantenendo l'equilibrio tra tempo, usabilità del prodotto digitale finale e rigore filologico.