

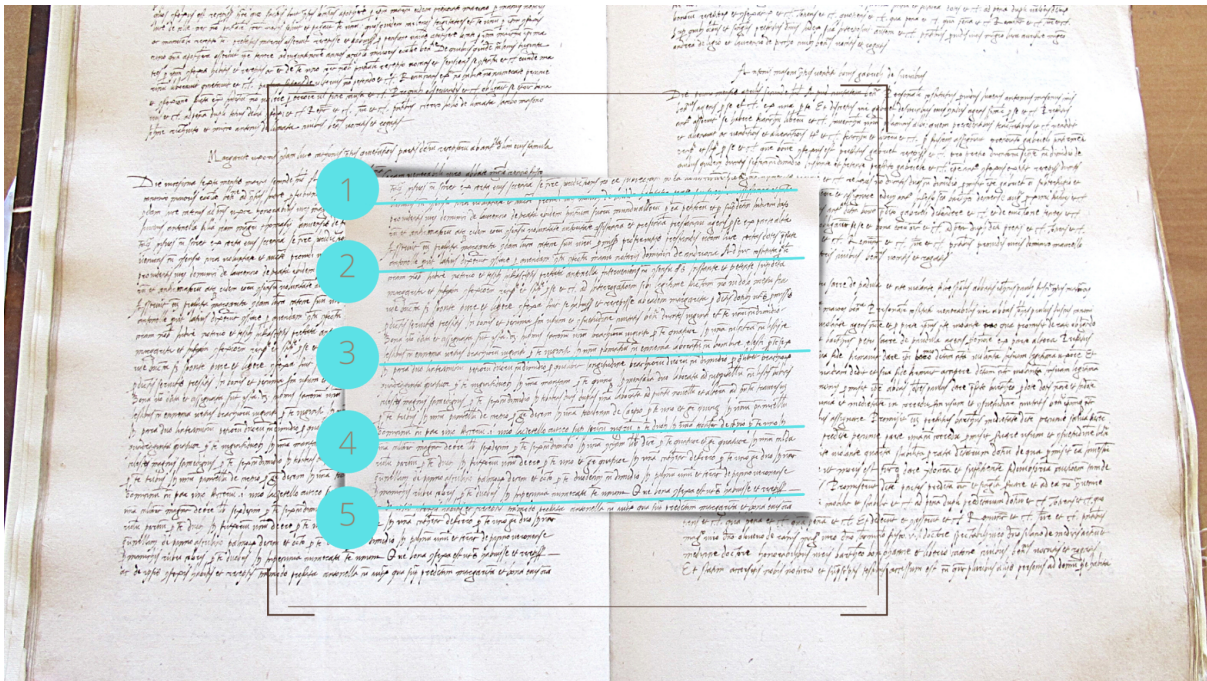
LE SCIENZE DEL PATRIMONIO

CULTURALE *on Air*

"Handwritten Text Recognition for Manuscripts and Documents. Comparing Experiences"

January, 20th 2022
9.00-13.00 CET

Organisers *Gemma Colesanti, Vera Isabell Schwarz-Ricci*



Handwritten text recognition is a promising and relatively new field of research, of particular importance due to the increasing amount of digitized cultural heritage objects, especially historical sources. In recent years, **several neural networks** have been developed for this purpose, and **user-friendly interfaces** are now available that allow these deep learning techniques to be used also by scholars without specific computer skills.

This webinar brings together researchers from different fields involved in ongoing or recently concluded projects dealing with the application or development of neural networks for the automatic recognition of handwriting on different media, but with a specific focus on **medieval documents and manuscripts**. The main goal is to offer an overview of the tested technologies or under development in this field and to discuss the different experiences.

The initiative takes place under the auspices of the **State Archive of Benevento**, Campania Region, and the **Notmed** project of the University of Barcelona, "The public notary in the western mediterranean: writing, institutions, society and economy (13th-15th centuries)".

The webinar will be held in English and Italian on Zoom platform.

Thursday, 20th of January 2022



9.00-9.10 CET
GREETINGS & WELCOME

Costanza Miliani
DIRECTOR CNR ISPC



Coordination

Alfonsina Pagano
CNR ISPC

9.15-09.25 CET

OPENING SPEECH



Gemma Colesanti

CNR ISPC



Daniel Piñol Alabart

UNIVERSITAT DE BARCELONA

A <https://cnr-it.academia.edu/GemmaColesanti>

 <https://webgrec.ub.edu/webpages/000012/cat/danielpinol.ub.edu.html>



existing initiatives around it.

Peter Stokes

ÉCOLE PRATIQUE DES HAUTES ÉTUDES
UNIVERSITÉ PSL (PARIS)

 <http://www.peterstokes.org/cv/>

 <https://prosopo.ephe.psl.eu/peter-stokes>

9.30-9.55 CET

OPEN HTR WITH ESCRIPTORIUM AND KRAKEN

eScriptorium and kraken are freely available open-access tools that work together to provide a user-friendly platform for manual and automatic transcription (HTR), line and region detection. The platform is designed to accommodate a very wide range of different scripts and languages, and both data and models can be freely imported, exported and shared with other users and other instances. This talk will present the platform and discuss its strengths, limitations, and some of the



10.00-10.25 CET

This speech will be held in Italian

LANGUAGE POST-PROCESSING TO INCREASE THE ACCURACY OF THE HTR

The use in the post-processing phase of linguistic analysis tools based on historical dictionaries can significantly increase the accuracy of the HTR, both coming from open and non-centralized recognition solutions such as Kraken (<https://github.com/mittagessen/kraken>), and from centralized solutions such as Traskribus (<https://readcoop.eu/transkribus>). The exemplification is

based on texts in Middle French and Old Italian.

IL POST-PROCESSING LINGUISTICO PER AUMENTARE L'ACCURATEZZA DELL'HTR

Il ricorso in fase di post-processing a strumenti di analisi linguistica basati su dizionari storici può aumentare significativamente l'accuratezza dell'HTR, sia proveniente da soluzioni di riconoscimento aperte e non centralizzate come Kraken (<https://github.com/mittagessen/kraken>), sia proveniente da soluzioni centralizzate come Traskribus (<https://readcoop.eu/transkribus>). L'esemplificazione è basata su testi in medio francese e italiano antico.

Federico Boschetti
CNR ILC



<http://www.ilc.cnr.it/en/content/federico-boschetti>

Benedetta Salvati
UNIVERSITÉ DE LAUSANNE



<https://medialitt.hypotheses.org/benedetta-salvati>



10.30-10.55 CET

IN CODICE RATIO: MACHINE LEARNING MEETS HISTORICAL RESEARCH

Codice Ratio, a research project born in 2016 at Roma Tre University, sees the collaboration of software engineers and paleographers for creating a tool for text transcription, extraction and use of data from handwritten historical sources, using machine learning and crowdsourcing techniques. It benefits from the collaboration of the Vatican Apostolic Archive and the State Archive of Rome. We will present our research experience, the results achieved so far and the future prospects in the field of humanities research and beyond.



Serena Ammirati

UNIVERSITA' DEGLI STUDI DI ROMA 3



url.it/3qv20



<https://romatre.academia.edu/SerenaAmmirati>



<https://www.linkedin.com/in/serena-ammirati-0a5b58b/>

Paolo Merialdo

UNIVERSITA' DEGLI STUDI DI ROMA 3



<http://merialdo.inf.uniroma3.it>



<https://uniromatre.academia.edu/paolomerialdo>



<https://www.linkedin.com/in/paolomerialdo/>



11.00 - 11.25 CET

This speech will be held in Italian

HANDWRITTEN TEXT RECOGNITION FOR MEDIÉVAL NOTARY REGISTERS: EXPERIMENTATION IN PROGRESS

Notarial protocols are one of the most important and numerous sources for the economic and social history of southern Italy in the Late Middle Ages and early modern times, but have rarely been edited or digitized. Due to the very cursive writing rich in abbreviations, access to their contents requires a long paleographic training and

knowledge of notarial forms. Automatic handwriting recognition could therefore be the key to making these sources searchable and accessible even to a wider audience. The contribution proposes a first experimentation of the automatic recognition of the writing of the registers of the Benevento notary and chronicler Marino Mauriello, active between the end of the fifteenth century and the first half of the sixteenth century, whose volumes are well preserved and suitable for experimentation. In fact, the protocols used contain the notarial needs in expanded form in relatively legible cursive writing. The platform adopted is Transkribus which provides an all-in-one solution for layout and handwriting recognition.

HANDWRITTEN TEXT RECOGNITION PER REGISTRI NOTARILI MEDIÉVALI: UNA SPERIMENTAZIONE IN CORSO

I protocolli notarili sono una delle fonti più importanti e più numerose per la storia economica e sociale dell'Italia meridionale nel Basso Medioevo e nella prima età moderna, ma raramente sono stati editi o digitalizzati. A causa della scrittura molto corsiva ricca di abbreviazioni l'accesso ai loro contenuti richiede un lungo training paleografico e una conoscenza dei formulari notarili. Il riconoscimento automatico della scrittura potrebbe quindi essere la chiave d'accesso per rendere queste fonti consultabili, ricercabili e accessibili anche a un pubblico più largo. L'intervento propone una prima sperimentazione del riconoscimento automatico della scrittura dei registri del notaio e cronista beneventano Marino Mauriello, attivo tra la fine del Quattrocento e la prima metà del Cinquecento ed i cui volumi sono ben conservati e adatti alla sperimentazione. Difatti i protocolli utilizzati contengono gli atti scritti in forma estesa con una scrittura corsiva relativamente leggibile. La piattaforma adottata è Transkribus che fornisce una soluzione all-in-one per il riconoscimento del layout e della scrittura.

Vera Schwarz-Ricci



<https://orcid.org/0000-0003-3790-4241>

CNR ISPC

11.30-12.00 CET

QUESTIONS & CONCLUSIONS



Chair session

Gemma Colesanti

CNR ISPC

The session is open to speakers and online participants. It will be held both in Italian and English.