

CALL FOR PARTICIPATION WINTER SCHOOL | OCTOBER - DECEMBER 2022

INTRODUCTION INTO HTR HANDWRITTEN TEXT RECOGNITION TECHNOLOGIES OF MEDIEVAL MANUSCRIPTS LATIN|GERMAN|CZECH



Organisation:

Institute for Medieval Research of
the Austrian Academy of Sciences
Manuscript, Rare Books and
Archival Studies Initiative
(MARBAS), Princeton University

In cooperation with:

CRC 1288 Practices of Comparing,
University Bielefeld

Department of German Studies,
University of Vienna

Digital Humanities,
Walter Benjamin Kolleg,
Universität Bern

History Department, University
of Tübingen

Contact:

Dr. Jan Odstrčilík
Austrian Academy of Sciences

Jan.Odstrcilik@oeaw.ac.at

4 Zoom online sessions | Oct 21, Nov 4 and 18, Dec 4

3-day-workshop in person at Vienna | December 19-21

DEADLINE FOR APPLICATION: 11TH SEPTEMBER 2022

CALL FOR PARTICIPATION
WINTER SCHOOL | OCTOBER - DECEMBER 2022
INTRODUCTION INTO HTR | HANDWRITTEN TEXT RECOGNITION
TECHNOLOGIES OF MEDIEVAL MANUSCRIPTS | LATIN | GERMAN | CZECH

A revolution has slowly begun in the study of historical documents: Machine Learning tools have been developed to allow for the automatic transcription of documents. Over the last decade, these tools can now help assist in the production of texts from medieval manuscripts at previously unobtainable levels of accuracy. Today, libraries have used these tools to make their collections searchable, while researchers have sped up the process of creating editions of texts and adopted them for the study of medieval documents.

The course will offer an introduction into some of these ongoing projects, but more importantly provide an introduction into the practice of studying medieval documents with Handwritten Text Recognition (HTR) technologies. The course will have two main parts: 4 online sessions and a three-day in person workshop in Vienna. During the first phase, participants will be introduced to both the theory of handwritten text recognition and its practical application using the Transkribus (transkribus.eu) tool. We will then work in four groups, focusing on four different periods and languages: Carolingian Latin, late medieval Latin, late medieval German, and late medieval Czech. Each group will have its own supervisor and its goal will be to train an HTR model for each type of writing.

During the in-person workshop in Vienna, we will finalize the four projects and publish our results online: both the transcriptions and Handwritten Text Recognition models. Additionally, we will also visit libraries in Vienna to see selected manuscripts in person. Finally, we will test other machine learning tools for their automatic transcription outcomes and use other digital tools. The course will be taught by a team of experts in HTR, medieval manuscript studies and Latin, German and Czech philology. At the end of the course, you will receive a certificate.

REQUIREMENTS

The course is primarily designed for Masters or PhD students, however, we will consider other applications as well. You are expected to be familiar with the language of the group you want to join: Medieval Latin, Medieval German, and Medieval Czech. We expect you to have at least basic knowledge of medieval palaeography and manuscript studies. However, we would also be able and delighted to offer resources for self-training in these languages and manuscript studies to prepare for the course.

COSTS

There is no participation fee, but you will be expected to cover the costs of your trip to Vienna, including accommodation (c. 300 Euros). We do hope, however, to be able to offer bursaries for students who do not have support from their institutions. If you would like to apply for a bursary, please let us know as early as possible.

APPLICATION

Please, send a **short CV** and **cover letter** to
Jan Odstrčilík | jan.odstrcilik@oeaw.ac.at
before 11th September 2022.

As a subject, use „**HTR School 2022**“.

Please don't forget to indicate the team/module you would like to take (i.e. Carolingian Latin, late medieval Latin, medieval German, medieval Czech).
If you have any questions, don't hesitate to ask.

Michael Berger, University of Vienna | michael_berger@univie.ac.at

Tim Geelhaar, Bielefeld University | tim.geelhaar@uni-bielefeld.de

Tobias Hodel, University of Bern | tobias.hodel@unibe.ch

Anna Michalcová, Czech Academy of Sciences | a.michalcova@ujc.cas.cz

Stephan Müller, University of Vienna | stephan.mueller@univie.ac.at

Jan Odstrčilík, Austrian Academy of Sciences | jan.odstrcilik@oeaw.ac.at

Steffen Patzold, University of Tübingen | steffen.patzold@uni-tuebingen.de

Leon Pürstinger, Austrian Academy of Sciences | leon.puerstinger@oeaw.ac.at

Helmut Reimitz, Princeton University | hreimitz@princeton.edu

Dennis Wegener, University of Vienna | dennis.wegener@univie.ac.at

SCHEDULE

4 ZOOM ONLINE SESSIONS

OCT 21	What is HTR?
	Transkribus 1 (<i>first hands-on</i>)
	Introduction into manuscripts and working in groups
NOV 4	Transkribus 2 (<i>uploading documents, layout recognition, simple transcription</i>)
	Working in groups (first transcriptions)
NOV 18	Transkribus 3 (<i>training of custommodels, learningcurves, tagging</i>)
DEC 4	Presentation of resulting models
	Transkribus 4 (<i>exporting documents and further processing</i>)

3-DAY-WORKSHOP IN PERSON AT VIENNA

DEC 19	Publication of the Ground truth
	Creating a simple website 1
	Alternatives to Transkribus
DEC 20	Creating a simple website 2
	VCEditor and other tools
	Library visit
DEC 21	Time to finish the work