

## Transkribus

*Galka, Selina; selina.galka@uni-graz.at*

Transkribus ist eine KI-gestützte Plattform, welche die Texterkennung (HTR, OCR), Layout-Analyse und Strukturerkennung von historischen Dokumenten erlaubt. Die Plattform wurde im Rahmen der Projekte “tranScriptorium” und “READ” (Recognition and Enrichment of Archival Documents) entwickelt und wird seit 2019 von der europäischen Genossenschaft READ-COOP betrieben. Aktuell gibt es einen Expertclient, der am eigenen Rechner verwendet werden kann und eine Onlineplattform, die in Zukunft die Clientsoftware sukzessive ablösen wird.

Transkribus ist kostenfrei nutzbar, allein für die automatische Texterkennung müssen sogenannte “Credits” erworben werden. Für diese können bereits vorhandene Schriftekennungsmodelle verwendet, aber auch eigene Modelle trainiert werden. Für das Training eines neuen Modells für die Erkennung von handgeschriebenem Text werden 25-75 Seiten manuelle Transkription vorgeschlagen, damit Transkribus lernt, welche Zeichen in der Handschrift digitalen Zeichen entsprechen. Die neu trainierten Modelle können zukünftigen Nutzerinnen und Nutzern zur Verfügung gestellt werden, was innerhalb der Community auch kontinuierlich gemacht wird.

Die transkribierten Dokumente können in Transkribus außerdem mit Markup versehen werden, so besteht z. B. die Möglichkeit, Personen, Orte oder Phänomene wie Streichungen und Hinzufügungen mit bereits vorgegebenen Annotationskategorien zu taggen. Es können aber auch eigene Tags definiert werden. Zudem unterstützt die Software das Durchsuchen der Dokumente nach Tags.

Transkribus stellt verschiedene Exportformate und -funktionen zur Verfügung – so können die Transkriptionen als PDF-, TEI- oder reines Textdokument über einen Client- oder Server-Export exportiert werden.

Die Bilder können unterschiedlich importiert werden – zum einen durch einfaches Hochladen, zum anderen beispielsweise auch via URL eines IIIF-Manifests. Mittlerweile wird auch ein ScanTent angeboten, das es in Kombination mit einer App ermöglicht, Dokumente zu scannen.

Die Dokumente können kollaborativ bearbeitet und in Sammlungen organisiert werden. Außerdem enthält die Software Funktionen zur Versionskontrolle und Änderungsverfolgung. Die Transkriptionssoftware hat sich mittlerweile als wesentliches Tool im Rahmen der Erstellung Digitaler Editionen etabliert; außerdem können die Editionen auch (kostenpflichtig) in einem Präsentationsmodus veröffentlicht werden. Im Ressourcenzentrum der READ-Webseite lassen sich zahlreiche Anleitungen zum Umgang mit Transkribus finden (vgl. READ: Ressourcenzentrum), außerdem bietet forText eine Anleitung zur Manuskript-

digitalisierung mit Transkribus (vgl. Horstmann 2019). Daneben existieren unterschiedliche Erfahrungsberichte (z. B. Eigner 2021).

### **Literatur:**

- Der erste Schritt zur digitalen Edition: Transkribieren mit Transkribus Hypothesis. URL: <https://hnb.hypotheses.org/115>
- Manuskriptdigitalisierung mit Transkribus. URL: <https://fortext.net/routinen/lerneinheiten/manuskriptdigitalisierung-mit-transkribus>
- Mühlberger, Günter; Seaward, Louise; Terras, Melissa: Transforming scholarship in the archives through handwritten text recognition Transkribus as a case study. In: Journal of Documentation: 2019.
- READ: Ressourcenzentrum Ressourcenzentrum. URL: <https://readcoop.eu/de/transkribus/ressourcen/>

### **Software:**

Transkribus, Dario Kampkaspar: XSLT Transkribus

### **Verweise:**

HTR, Transkription, Transkriptionswerkzeuge, TEI, Digitale Edition

### **Projekte:**

CoReMA - Cooking Recipes of the Middle Ages, Mittelalterlabor, Hyperdiplomatische Transkriptionsplattform, Peter Handke Notizbücher. Digitale Edition

### **Themen:**

Digitalisierung, Digitale Editionswissenschaft

### **Zitervorschlag:**

Galka, Selina. 2021. Transkribus. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.265>