

Transkribus



Alte Handschriften mit KI entziffern

Transkribus gehört zu den bekanntesten Tools zur Texterkennung und Transkription historischer Dokumente. Grund genug, es für das Kursprogramm der WLB zu berücksichtigen, zumal die Kurse mit KI-Technologie im Sommersemester 2025 einen Schwerpunkt bilden sollten.

Die vielfältigen Möglichkeiten der automatischen Texterkennung, die schnelle Umwandlung handschriftlicher und gedruckter Texte in digitale Formate sowie die realisierbaren speziellen Schriftartenanpassungen wurden im Kurs niederschwellig vermittelt – insgesamt also ein lohnenswerter Einstieg. Grundkenntnisse im

Lesen alter Schriften sind aber dennoch wichtig, da die Nutzungsmöglichkeiten von Transkribus gerade auch von der Erfahrung mit der Transkription historischer Texte abhängig sind. Die Intention des Kurses war zunächst, für eigene Projekte und Abschlussarbeiten ein geeignetes und nützliches Hilfsmittel im Umgang mit handschriftlichen Quellen kennenzulernen. Für größere Projekte ist wie bei vielen anderen KI-Anwendungen jedoch eine kritische Kontrolle und Steuerung unerlässlich, was in diesem Fall Erfahrungen mit Handschriften voraussetzt.

Das Programm wird von der Genossenschaft Read-Coop betrieben, zu deren Mitgliedern zahlreiche große europäische Universitäten und Institute zählen. Seit 2024 ist auch die Würt-

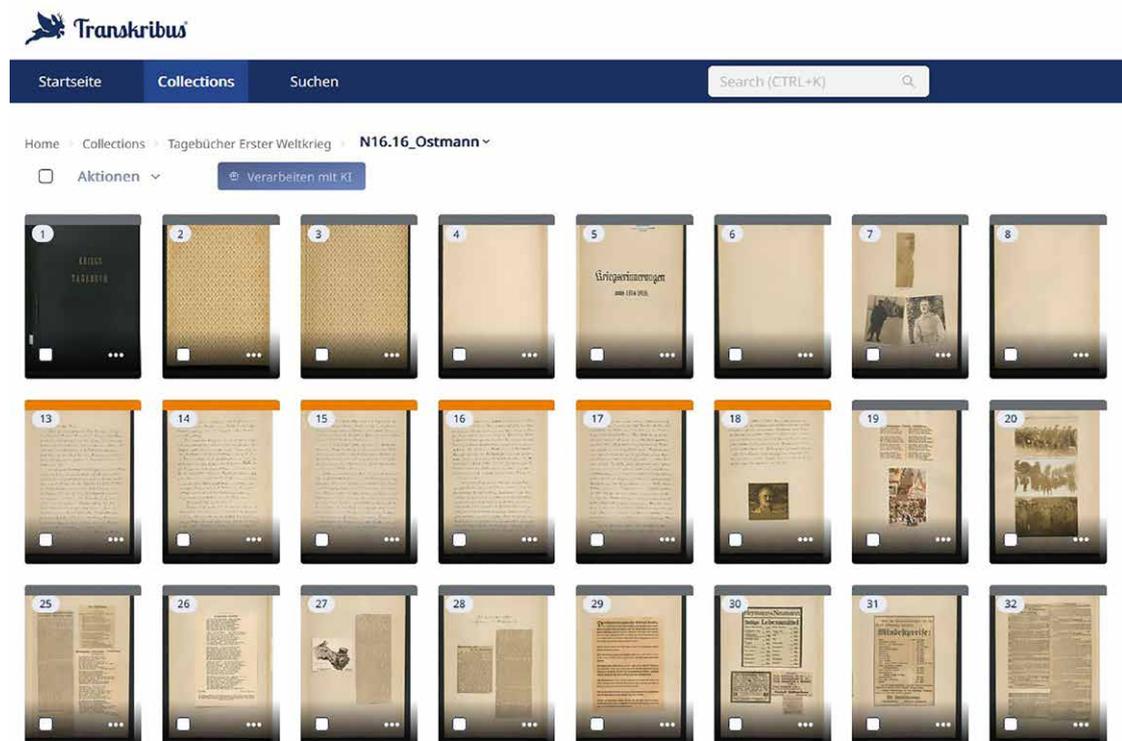


Abb. 1: Das Programm ist mittlerweile komplett browserbasiert. Im hier gezeigten Beispiel handelt es sich um das Tagebuch von Carl Ostmann, der als Eisenbahner im Ersten Weltkrieg eingesetzt wurde.

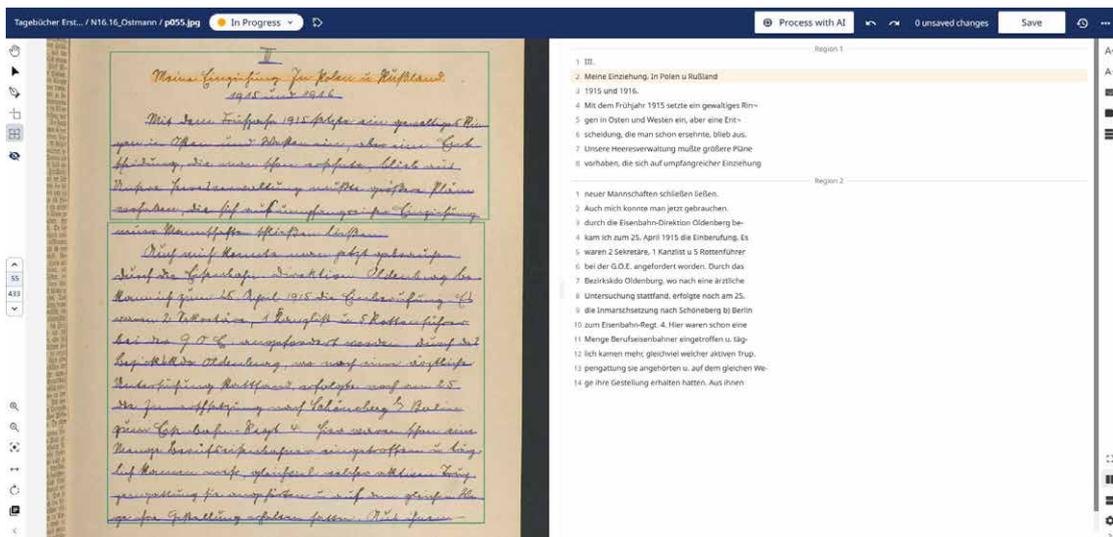


Abb. 2: Die Transkription wird als XML-Datei gespeichert und passt somit für den Import in verschiedene Präsentationsanwendungen zur Darstellung der Transkription.

tembergische Landesbibliothek Mitglied. Aus gutem Grund, denn im Haus befinden sich in den verschiedenen Sondersammlungen große Bestände mit handschriftlichem Material.

Die ersten Versuche mit der Anwendung von Handschriftenerkennung finden in der Bibliothek für Zeitgeschichte statt. Die Lebensdokumentensammlung enthält unter anderem zahlreiche Tagebücher aus dem Ersten Weltkrieg, die sich für die Bearbeitung mit Transkribus anbieten. Für das Themenportal *Erster Weltkrieg* wurden bereits einige dieser Quellen digitalisiert, womit die erste Hürde zur Transkription genommen ist, denn die zu bearbeitenden Texte müssen als Bilddateien vorliegen. Diese Dateien werden dann in einem ersten Schritt in die Transkribus-Plattform importiert (siehe Abb. 1). Die zu transkribierenden Seiten werden zur weiteren Bearbeitung markiert und mit der Funktion „Verarbeiten mit KI“ gelangt man zur Auswahl der Handschriftenerkennungsmodelle. Hat man ein möglichst passendes Modell für die entsprechende Schrift gewählt, kann man die Erkennung starten. Das KI-Modell segmentiert die vorliegende Bilddatei und unterteilt den darin abgebildeten Text in einzelne Zeilen und innerhalb der Zeilen in Wörter. Je regelmäßiger das Schriftbild, desto besser ist die Segmentierung. Einschübe, Errata, Bildunterschriften und ähnliche Unregelmäßigkeiten können jedoch in der Nachbearbeitung in die richtige Reihenfolge geschoben werden. Nach dem Abschluss der Segmentierung wird der Text transkribiert. Die verfügbaren Modelle funktio-

nieren mittlerweile sehr gut und liefern eine Transkription mit einer sehr geringen Fehlerquote. Diese ist aber, wie bereits erwähnt, auch von der Regelmäßigkeit des Schriftbilds abhängig. Die hier gezeigte Tagebuchseite (Abb. 2) ist eine Reinschrift und eignet sich deshalb besonders gut.

Als Fazit lässt sich festhalten: Transkribus ist für die Arbeit mit historischen Handschriften ein leistungsfähiges Werkzeug. Es hat großes Potenzial für die Nutzbarmachung der sonst eher schwierig zugänglichen Bestände der Württembergischen Landesbibliothek, vor allem in den Sondersammlungen. Durch die große Anzahl sofort verfügbarer, leistungsfähiger Erkennungsmodelle ist das Programm aber auch für die Transkription von kleineren Beständen wie Familiennachlässen im privaten Kontext geeignet, da hier in der Regel sofort gute Ergebnisse erzielt werden.

Tobias Thelen