

## Wie wir Sie unterstützen möchten

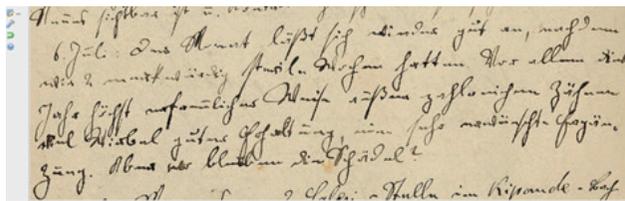
Mit Volltexten angereicherte Digitalisate bieten einen zeitgemäßen, umfassenden Zugang zum Kulturgut und können als Grundlage für Anwendungen im Bereich der „Data Science“ dienen.

Die UB Tübingen unterstützt Wissenschaftler/innen und Studierende, diese Zugänge zu schaffen und steht als Ansprechpartner bei Fragen bereit.

Wenn Sie große Mengen an Scans von Handschriften, Archivalien oder alten Drucken analysieren wollen und nach einer Möglichkeit suchen, diese leichter und effizienter les- und durchsuchbar machen zu können, beraten wir Sie gerne in Hinblick auf die Einsatzmöglichkeiten von Transkribus, eScriptorium und Tesseract für Ihre eigenen Materialien und Projekte.

Weitere Informationen erhalten Sie auf unserer Homepage, bei Informationsveranstaltungen und Schulungsangeboten sowie über unseren Newsletter.

Natürlich können Sie uns auch jederzeit gerne persönlich ansprechen, damit wir gemeinsam überlegen können, wie sich Ihre Bedürfnisse mithilfe der aktuellen technischen Hilfsmittel am besten realisieren lassen.



- 1-8 Juli: Der Monat läßt sich wieder gut an, nachdem
- 1-9 wir 2 merkwürdig sterile Wochen hatten. Vor allem dies
- 1-10 Jahr höchst erfreulicher Weise außer zahlreichen Zähnen
- 1-11 viel Wirbel guter Erhaltung, eine sehr erwünschte Ergän,
- 1-12 zung. Aber wo bleiben die Schädel?

## Kontakt

Dorothee Huff  
Tel.: 07071/29-72852  
E-Mail: dorothee.huff@uni-tuebingen.de

Mehr zu unserem Service:  
<https://uni-tuebingen.de/de/230831>

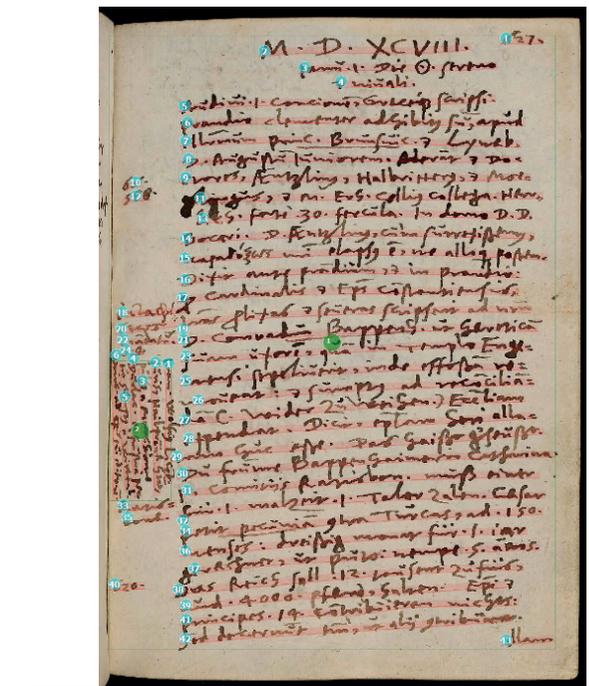
Projekt OCR-BW an der UB Tübingen:  
<https://uni-tuebingen.de/de/179298>

Homepage mit aktuellen Aktivitäten der UB Tübingen und Mannheim:  
<https://ocr-bw.bib.uni-mannheim.de>

Eberhard Karls Universität Tübingen  
Universitätsbibliothek  
Wilhelmstr. 32 · D-72074 Tübingen  
Tel.: 07071/29-72852  
E-Mail: ocr@ub.uni-tuebingen.de  
<https://www.ub.uni-tuebingen.de>  
aktuelle Meldungen auch auf Facebook / Instagram

## OCR-BW Kompetenzzentrum für automatische Texterkennung

# 31



UNIVERSITÄTSBIBLIOTHEK



## Das Projekt OCR-BW

Das 2019 als Service der Universitätsbibliothek (UB) Tübingen und Mannheim eingerichtete Kompetenzzentrum „Volltexterkennung von handschriftlichen und gedruckten Werken“ unterstützt Wissenschaftler/innen, Bibliotheken und Archive in Baden-Württemberg bei der Anwendung von automatischer Texterkennungs- und Transkriptionssoftware.

Bei einer Volltexterkennung werden textliche Bildinhalte in digitale Textformate übersetzt. Erkannte Texte können durchsucht, kopiert, bearbeitet und für eine Extraktion von Forschungsdaten verwendet werden.

Die UB Mannheim stellt mehrere Open-Source Softwareprodukte wie z.B. Tesseract aus dem Bereich OCR als nutzerfreundliche und systemunabhängige Anwendung für historische Drucke zur Verfügung. Die UB Tübingen testet die Einsatzmöglichkeiten der Transkriptionsplattform Transkribus für die automatische Handschriften- und Druckererkennung. Außerdem betreibt die UB Mannheim seit Oktober 2021 eine Instanz der Transkriptionsplattform eScriptorium.

Zusammen kann ein breites Spektrum an Materialien und Technik abgedeckt werden. Das Kompetenzzentrum erarbeitet sich so ein umfassendes Knowhow im Bereich der automatischen Volltexterkennung und vermittelt dieses weiter.

## Die Transkriptionsplattform Transkribus



Rainer Zenz (<https://commons.wikimedia.org/wiki/File:Wolpertinger.jpg>), „Wolpertinger“, <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

Die Universitätsbibliothek Tübingen evaluiert Transkribus mit seinen Tools für Layoutanalyse und HTR (Handwritten Text Recognition). Es hat sich gezeigt, dass für alle bisher getesteten Sprachen und Schriftarten das nach Vorgaben der DFG angestrebte Ergebnis einer CER (Character Error Rate) von unter 5% erreicht werden kann. Mittlerweile gibt es schon viele öffentliche Texterkennungsmodelle für Drucke und Handschriften, die bei geringem Eigenaufwand mit gutem Ergebnis auf eigene Texte angewandt werden können.

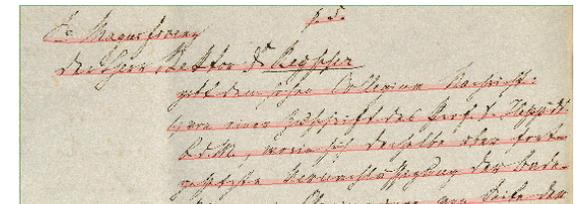
Die UB Tübingen erzeugt Ground Truth Daten (korrekte Transkriptionen anhand derer die Software trainiert wird) und Texterkennungsmodelle für verschiedene forschungsrelevante Korpora der Handschriftenabteilung und des Universitätsarchivs. Die mit Volltext angereicherten Digitalisate werden sukzessiv online zur Verfügung gestellt.

Auf Basis der Ergebnisse können Handlungsempfehlungen für unterschiedliche Materialgruppen gegeben werden, um den Zugang zu diesen Texten zu erleichtern und neue wissenschaftliche Fragestellungen und Auswertungsmöglichkeiten zu schaffen.

Weitere Informationen:  
<https://readcoop.eu/>

## Bisher bearbeitete Quellen aus der Handschriftenabteilung und dem Universitätsarchiv

- Tagebücher der Tendaguru-Expedition des Tübinger Geologen und Paläontologen Edwin Hennig (1909-1911)
- Tagebücher des Tübinger Altphilologen Martin Crusius (1573-1605)
- Griechische Predignachschriften von Martin Crusius (1563-1605)
- ausgewählte Bände der juristische Konsilien (1602-1879)
- ausgewählte Bände der Senatsprotokolle (1524-1912)
- Handschriften und Drucke in Malayalam
- Inkunabeln
- vermisches, loses Schriftgut (auch Maschinenschrift) und wissenschaftliche Nachlässe



Se Magnificenz  
der Herr Rektor D. Reyfcher  
gibt dem hohen Collegium Nachricht:  
1.) von einer Zuschrift des Prof. D. Hepp dt-dd.  
6. d. M. M., worin sich derfelbe über fort= gefetzte Vernachlässigung der Bade= fläge-Plätze u. Stagierrängen-Spaziergänge von Seite der städtischen Behörden beklagt; u.  
2.) von einem Schreiben der K. Stadt-Direc= tion dd. 13. Mai d. J., womit die Bitte des Stadtraths dahier um Ueber= lassung des bisherigen Univerfitäten Univerfitäts= Gebäudes Saarla) (Aula) zu Einrichtung einer Madchenfehule Mädchenfehule empfohlen wird.  
Antr. u. Befchluß:  
zu 1.) an die K. Stad-Direction StadtDirection dahier unter Bezugnahme auf den Minift. Er-Er- laß v. 7. Juni 1843. des das Erfuchen er= gehen zu laffen, daß für die Erhalt= ung der Bedeglätze Badeplätze geeignete Sorge getragen werden möchte, möchte;