



# OCR-BW

*Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen*



## Workshop zu OCR und Handschriftenerkennung (19. Februar 2020)

Automatische Texterkennung mit Transkribus

19.02.2020, Dorothee Huff



---

## Hintergrund

- Projekt READ (Recognition and Enrichment of Archival Documents), 2016-2019
- ab Juli 2019 READ-COOP SCE
- UB Tübingen seit Januar 2020 Mitglied
- für Juni 2020 ist die Einführung von Nutzungsgebühren geplant



---

## Vorgesehene Quellen an der UB Tübingen

- Tagebücher der Tendaguru-Expedition des Tübinger Geologen und Paläontologen Edwin Hennig (1909-1911)
- Tagebücher des Tübinger Altphilologen Martin Crusius (1573-1604)
- Griechische Predignachschriften von Martin Crusius (1563-1604)
- ausgewählte Bände der juristische Konsilien (1602-1879)
- ausgewählte Bände der Senatsprotokolle (1524-1912)
- Handschriften und Drucke in Malayalam



## Tagebücher von Edwin Hennig (1897-1973)

Entwicklung der HTR (Handwritten Text Recognition)-Modelle

- M1: 100 Seiten Ground Truth (1910-1911) → 8.88% CER
- M2: 133 S. GT (1910-1911) → 5,58% CER
- M3: 165 S. GT (1910-1911) → 4,32% CER
  
- M4: 165 (1910-1911) + 30 (1944-1946) → 4,31% CER
- M5: 165 (1910-11) + 30 (1944-46) + 30 (1961-62) → 4,03% CER
- M6: 165 (1910-11) + 30 (1944-46) + 30 (1961-62) + 15 (1897-1901) → 4,05% CER (unter Verwendung eines Language Models 3,61% CER)



# Modellentwicklung

	UAT_M1	UAT_M2	UAT_M3	UAT_M4	UAT_M5	UAT_M6	UAT_M6/La anguageMod el	Comb_Deut sch_Latein_ XX_M2	German_Ku rrent_XIX_ M1
S. 1 (1910-11)	2,97	2,03	1,88	1,88	2,03	1,74	1,38	18,62	14,13
S. 2 (1910-11)	3,09	3,09	2,7	2,08	2,47	1,85	1,62	15,9	12,73
S. 3 (1910-11)	18,7	17,58	17,36	16,69	16,01	15,23	14,39	30,85	26,6
S. 4 (1910-11)	8,96	9,3	9,3	8,69	8,96	8,69	8,08	21,11	18,81
S. 5 (1910-11)	1,92	1,37	0,89	1,24	0,82	1,1	0,89	18,56	12,65
S. 6 (1910-11)	2,2	1,52	1,24	1,24	1,52	1,31	1,17	18,8	13,09
S. 7 (1910-11)	3,15	1,8	1,75	1,97	1,69	1,63	1,52	15,26	11,94
S. 8 (1910-11)	2,12	1,33	1,27	1,03	1,39	1,15	0,97	15,73	12,22
S. 9 (1911)	2,07	1,31	0,98	0,92	1,03	0,98	0,92	14,25	11,21
S. 10 (1944-46)	18,78	19,5	17,01	8,09	6,74	6,64	5,71	31,22	29,46
S. 11 (1944-46)	11,72	8,54	8,39	3,4	3,04	3,4	2,97	23,01	20,26
S. 12 (1961-62)	16,72	14,37	12,35	8,05	2,64	2,64	2,64	19,36	16,79
S. 13 (1961-62)	20,8	16,79	15,68	11,42	3,52	2,9	2,72	22,65	17,84
S. 16 (1897)	57,69	53,95	49,64	46,19	35,7	6,54	4,53	9,27	8,05
S. 15 (1899)	30,64	29,95	23,7	22,64	16,23	4,27	3,35	14,71	9,6
S. 16 (1901)	14,17	12,31	11,66	10,45	8,26	4,62	4,45	18,54	13,2
Durchschnitt	8,88 (12,94)	5,58 (11,6)	4,32 (10,52)	4,31 (8,86)	4,03 (6,81)	4,05	3,61	23,3	15,36
Seiten im Traini	95	126	156	184	212	224			
Testseiten	5	7	9	11	13	16			
GT insgesamt	100	133	165	195	225	240			







## M1: CER 2.07%, WER 8.53%

8

Büffel. Sandte Leute nach ihnen aus u. erhielt um 4 etwa Meldung. Darauf ging ich mit starker Begleitung bis zu ~~den-dem~~ überraschend entfernten Fluß (ca ½ Std. Die Leute schöpfen z. T. aus Wasserlöchern!), u. jenseits ~~im-in~~ d. Dickicht. ~~Plötz., Plötz.,~~ lich ging unmittelbar vor uns d. Gekrache los, wir ~~waren-ahn-waren~~, ehe ich ~~nach-noch~~ d. Gewehr selbst genommen hatte, der ~~starkein-starken~~ Herde auf vielleicht 25 m. ~~nate-nahe~~ gekommen, ohne daß auch nur das ~~Geriegste-Geringste~~ v. d. Tieren sichtbar wurde. Wir flüchteten ~~eilichst-eiligst~~, da sie ebensogut auf uns ~~zukommen-zukommen~~ konnten. Auch Brüllen der Kühe ertönte. Da hier nicht möglich war, zu Schuß zu kommen, wollte ich in eine angrenzende ~~gangwa-saffenen-yangwa~~ (offenen Wald) auf die sie zu geflüchtet waren. Meine Führer gingen furchtlos, wenn auch vorsichtig in gleicher Richtung ~~weiter-weiter~~, wobei plötzlich auch auf d. andern Seite d. Krachen losging. Das war mir doch etwas unheimlich, zumal jede Kugel nach spätestens 5 m. abge-, lenkt werden mußte u. so nahe wohl nur ein bereits an-, ~~stärmender-stürmender~~ Büffel kommt. Wir gingen dann am Rande rings um d. Gehölz, ohne etwas zu ~~sehen-sehen~~ (als die Wald-bestandene Fläche, die hier ~~anscheinend-anscheinend~~ beim Zusammenstoß der Täler d. ~~Maruai-Mavudyi~~ u. ~~Kikende-Kihendye~~ in den Busch u. d. Hügelland unvermittelt eingestreut ist.) Auf d. Weg zurückgekehrt trafen wir den ~~Jumben-Jumben~~, einen alten freundlichen hübschen ~~Grnis-Greis~~ mit 2 kleinen ~~Kindern-Kindere~~ ruhig auf d. Nachhausewege, er berichtete daß die Spuren nahe am Fluß ü. d. Weg führten. In d. Tat ~~wart~~ war die ganze Herde im Bogen ~~rückwärts durchgebrochen-rückwärts durchgebrochen~~ u. zwar im übrigen völlig lautlos. Es war überflüssig in dem Dickicht zu folgen, ich kehrte wieder unverrichteter Sache ~~heim-heim~~, aber doch um ~~einige-seltsam-einige-seltsame~~ Erfahrungen reicher. Die ~~Leuthen-Leuthen~~ hier in diesem verlorenen Talkessel kamen mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.

4. Aug. Rückkehr nach ~~Makangaga-Makangaga~~. Morgens schöne Wanderung

## M6: CER 0.92%, WER 3,47%

8

Büffel. Sandte Leute nach ihnen aus u. erhielt um 4 etwa Meldung. Darauf ging ich mit starker Begleitung bis zu dem überraschend entfernten Fluß (ca ½ Std. Die Leute schöpfen z. T. aus Wasserlöchern!), u. jenseits ~~im-in~~ d. Dickicht. Plötz., lich ging unmittelbar vor uns d. Gekrache los, wir ~~waren-waren~~, ehe ich noch d. Gewehr selbst genommen hatte, der starken Herde auf vielleicht 25 m. nahe gekommen, ohne daß auch nur das Geringste v. d. Tieren sichtbar wurde. Wir flüchteten ~~nlichst-eiligst~~, da sie ebensogut auf uns zukommen konnten. Auch ~~Brüllen-Brüllen~~ der Kühe ertönte. Da hier nicht möglich war, zu Schuß zu kommen, wollte ich in eine angrenzende yangwa (offenen Wald) auf die sie zu geflüchtet waren. Meine Führer gingen furchtlos, wenn auch vorsichtig in gleicher Richtung ~~weiter-weiter~~, wobei plötzlich auch auf d. andern Seite d. Krachen losging. Das war mir doch etwas unheimlich, zumal jede Kugel nach spätestens 5 m. abge-, lenkt werden mußte u. so nahe wohl nur ein bereits an-, stürmender Büffel kommt. Wir gingen dann am Rande rings um d. Gehölz, ohne etwas zu sehen (als die Wald-bestandene Fläche, die hier ~~anscheinend-anscheinend~~ beim Zusammenstoß der Täler d. Mavudyi u. ~~Kihendze-Kihendye~~ in den Busch u. d. Hügelland unvermittelt eingestreut ist.) Auf d. Weg zurückgekehrt trafen wir den ~~Jumben-Jumben~~, einen alten freundlichen hübschen ~~Grnis-Greis~~ mit 2 kleinen ~~Kindern-Kindere~~ ruhig auf d. Nachhausewege, er berichtete daß die Spuren nahe am Fluß ü. d. Weg führten. In d. Tat war die ganze Herde im Bogen rückwärts durchgebrochen u. zwar im übrigen völlig lautlos. Es war überflüssig in dem Dickicht zu folgen, ich kehrte wieder unverrichteter Sache ~~heim-heim~~, aber doch um einige seltsame Erfahrungen reicher. Die Leuthen hier in diesem verlorenen Talkessel ~~kannen-kamen~~ mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.

4. Aug. Rückkehr nach Makangaga. Morgens schöne Wanderung





x von hier be-  
 stiegen wir  
 noch den Hülm,  
 einen Berg,  
 von dem mag  
 eine sehr schö-  
 ne Aussicht  
 auf die See sein.
 
  
 beitet. Um  $\frac{1}{2}$  1 gingen wir nach Kurth's  
 Gasthaus und aßen dort sehr gut zu  
 Mittag. Von dort gingen wir wieder nach  
 Hause und waren unschlüssig, ob wir ge-  
 hen sollten, oder nicht. Ich arbeitete noch  
 etwas und, da es nur ganz wenig regnete,  
 wollten wir zum Förster gehen. Als wir  
 ein Stück die Chaussee gegangen waren,  
 fing es an stärker zu regnen, doch gingen  
 wir weiter und es hörte bald wieder auf.  
 Wir kehrten in einem großen Restaurant  
 mit einer Kapelle ein, das mir gar nicht  
 sehr förstermäßig vor kam, da auch Kellner  
 da waren. Nachher gingen wir weiter und  
 merkten auch, dass wir gar nicht beim För-  
 ster gewesen waren, sondern kamen erst  
 jetzt daran vorbei und stiegen auf einen  
 Berg mit dem einem Turm von Dingen





## M1: CER 57.69%, WER 93.73%

(17.)

belete.aerimachdr-beitete. Um ½1 gingen wir nach Kurth's  
Barlharvne afindrrsvhrhgabas-Gasthaus \* und aßen dort sehr gut zu  
Nitbag. Ba dors gmmrarmibeerrach. Mittag. Von dort gingen wir wieder nach  
Nlaave nad raren ihlarhirbwir Gle-Hause und waren unschlüßig, ob wir ge-  
kenollbenoerast. Sharbaibobenreh-hen sollten, oder nicht. Ich arbeitete noch  
Froaecears vr Mir wer rembl- etwas und, da es nur ganz wenig regnete,  
KolbbienrurwBndlrgeser. Lbwir- wollten wir zum Förster gehen. Als wir  
aa Mik ein Stück die Shaaaa Geerepraarer-Chaussée gegangen waren,  
EineDa-Bleeskerarlaen dochaga-fing es an stärker zu regnen, doch gingen  
Moraeebonne eibiolebalear- wir weiter und es hörte bald wieder auf.  
Pirkehrlern MaemroenNalauruk. Wir kehrten in einem großen Restaurant  
inenPapellee Nagrrarer mit einer Kapelle ein, das mir gar nicht  
kbermirrbara, Laaachiellari-sehr förstermäßig vorkam, da auch Kellner  
NawaremNalkaoroaaamaraerna-dawaren. Nachher gingen wir weiter und  
kerkiendaei daibeiaarrabbeiForu-merkten auch, daß wir garnicht beim För-  
ler. Gererer worvo, Wadera Kareners-ster gewesen waren, sondern kamen erst  
d. MiabDaraa orbemababiegeauberer-jetzt daran vorbei und stiegen auf einen  
Berye mirrdura-Berg, auf dem einen Turm war. Bieseraer-Dieser war  
Aaber rerseklosseruadrar Neeenaden- aber verschlossen und wir stiegen an den  
Dbrond hiab. Dordsmaarein omekya-Strand hinab. Dort war ein ziemlich gro-  
ser Ber, aber BrnjerPandberabesaus-junger Hund, der alles aus dem  
Maverkeraushalae rraswankrar-Wasser herausholte, was man hineinwarf,  
brih. Naerihmrrkorile doch da er nicht schwimmen konnte,  
Donderrrdowers, aberBaraahake-ging er nur so weit, als er Grund hatte,  
Nac, abenkeererbruaererberWaro- und, als ein kleinerer Hund weiter hinaus-  
awarra Nararo Nedisch Nacer Aer-schwamm, war er so neidisch, daß er den  
Mdr rbweairberawrrarerlebe- andern nicht wieder herauslassen wollte  
balabaraikbarrerjaelrba-wellen und ihn nachher verjagte. Als wir weiter  
denilaiberakelBorne ngenaben!-der Kaiser Wilhelm Brücke zuginen, sahen  
esenraberollen Berenbogerra-wir einen wundervollen Regenbogen von  
Neodrog a Neerraensderndee-Misdroy aus ins Meer ragen, denn die  
Grmbwar Mederkemrracbooojerwaa-Sonne war wieder hervorgekommen und  
Kooneker-ben-\* vorher be-  
ziaBessurir-stiegen wir  
roek-Neaala, noch den Kulm,  
iela Beref-einen Berg,  
on-Neraz-von dem man

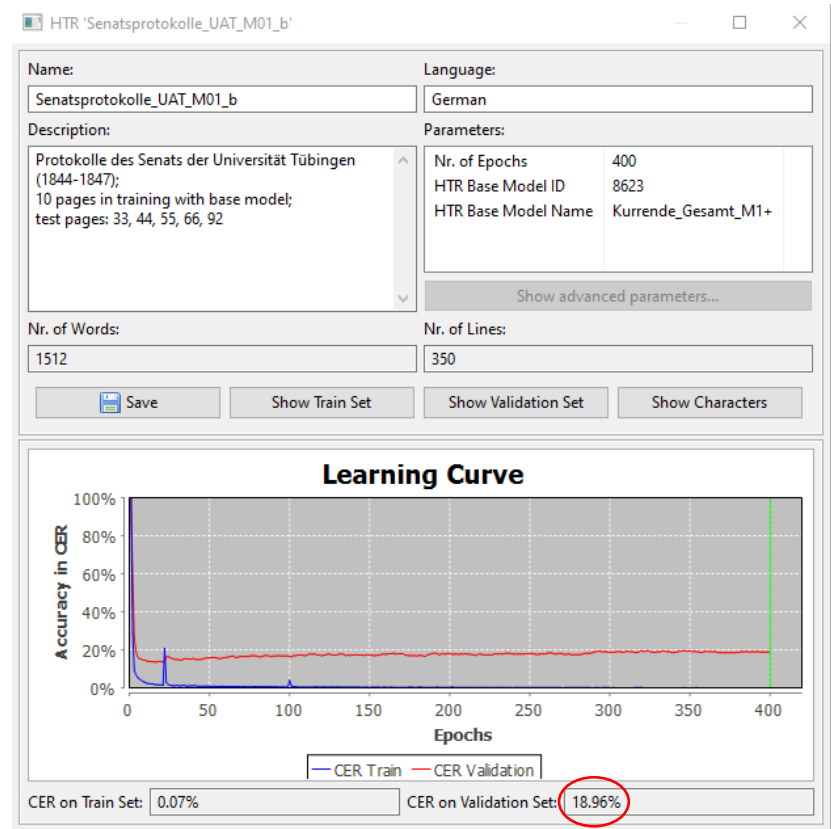
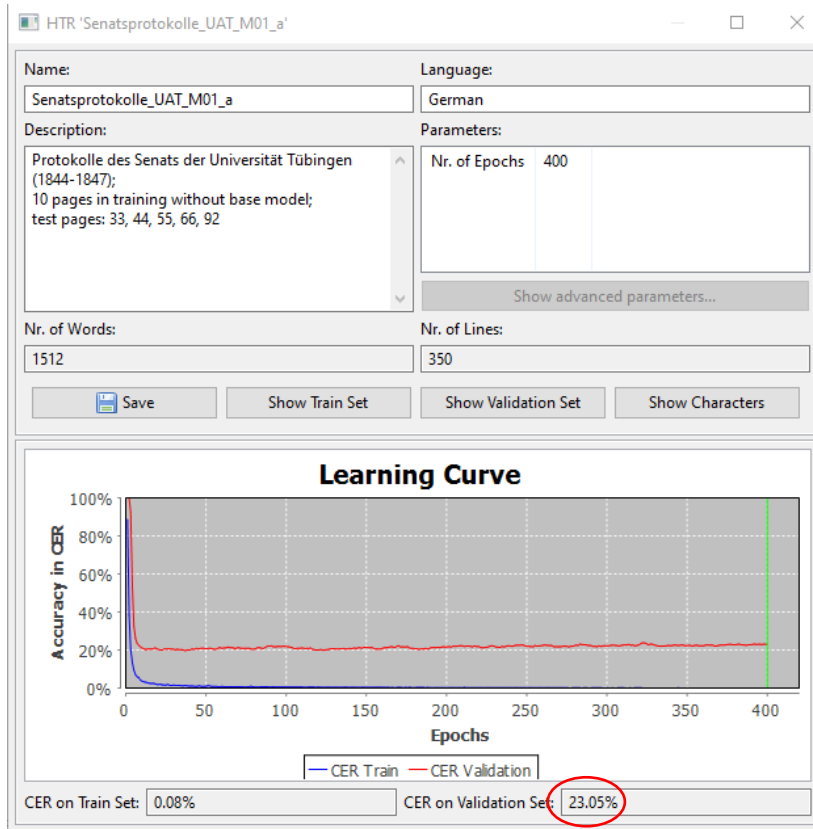
## M6: CER 4.53%, WER 18.82%

(9-7.)

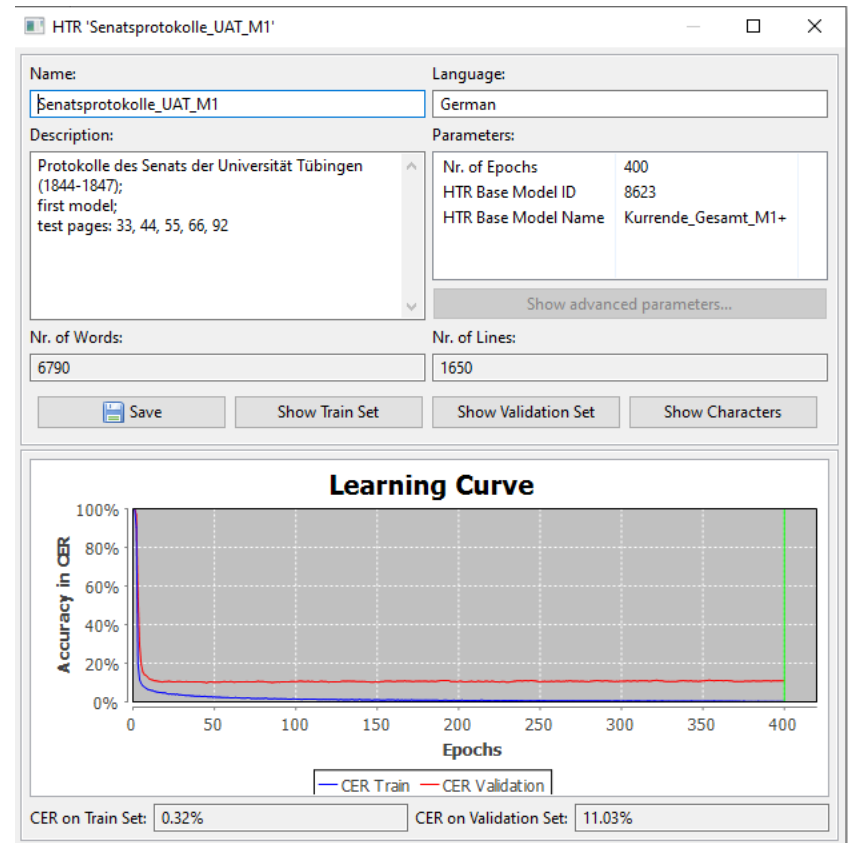
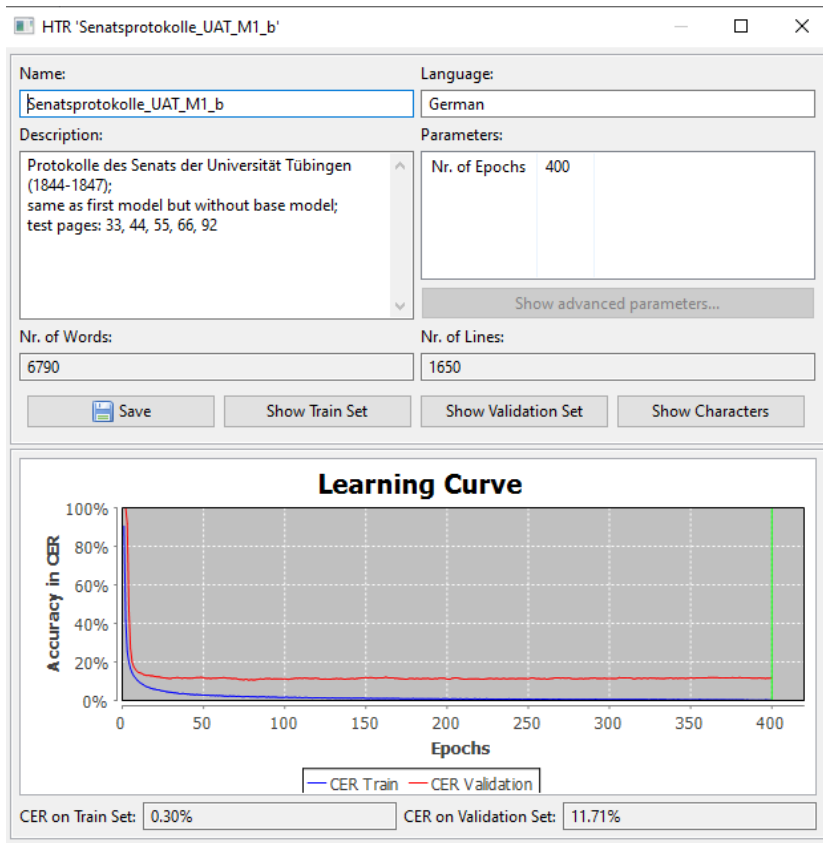
beitete. Um ½1 ginden-gingen wir nach Kurtes-Kurth's  
Gasthaus \* und aßen dort sehr gut zu  
Mittag. Von dort gingen wir wieder nach  
Hause und waren unschlüßig, ob wir ge-  
hen sollten, oder nicht. Ich arbeitete noch  
etwas und, da es nur gant-ganz wenig regnete,  
wollten wir zum Förster gesen-gehen. Als wir  
ein Stück die Chaussee-Chaussée gegangen waren,  
fing es an stärker zu rechnen-regnen, doch gingen  
wir weiten-weiter und es hörte bald wieder aus-auf.  
Wir kehrten in einem großen Restauran-Restaurant  
mit einer Kapelle ein-ein, das mir gar nicht  
sehr förstermäßig vor kam, förstermäßig vorkam, da auch Kellzer-Kellner  
dawaren. Nachher gingen wir weiter und  
merkten auch, daß wir garnicht beim Hor-För-  
ster gewesen waren, sondern kamen erst  
jetzt daran vorbei undistiegen-und stiegen auf einen  
Berg, auf dem einen Turm war. Sieser-Dieser war  
aber verschlossen und wir stiegen an den  
Strand hinab. Dort war ein ziemlich gro-  
ßer Ber, aber junger Hund, der alles aus dem  
Wasser herausholte, was man hinein wart, hineinwarf,  
doch da er nicht schwimmen konnte-konnte,  
gind-ging er nur so weit, es-als er Grund hatte,  
und, als ein kleinerer Hund weiter hinaus-hinaus-  
schwamm, war er so neidisch, daß er den  
andern nicht wieder herausassen-herauslassen wollte  
und ihn nachher vergagterts-verjagte. Als wir weiten-weiter  
der Haiser-Kaiser Wilhelm Brücke zu-gingen, zuginen, sahen  
wir einen wundervollen Gegenbogen-Regenbogen von  
Misdrog-Misdroy aus ins Meer ragen, denn die  
Sound-Sonne war wieder hervorgekommen-hervorgekommen und  
Kvenher bei-\* vorher be-  
stiegen wir  
noch den Duhm-Kulm,  
einen Bergt-Berg,  
von dem Tag-man



# Senatsprotokolle, 10 Seiten GT



# Senatsprotokolle, 50 Seiten GT





---

## Vorteile

- Transkriptionsoberfläche, die verschiedene Funktionen vereint, ohne dass spezielles technisches Wissen vonnöten ist
- aufgrund der großen Nutzergemeinschaft und Menge der verarbeiteten Daten werden zunehmend leistungsfähigere Modelle entwickelt und veröffentlicht
- Verwaltung von Dokumenten
- kollaboratives Arbeiten
- Nutzung der Daten in anderen Kontexten, z.B. Erstellung von Editionen
- permanente Weiterentwicklung durch die Betreiber
- mit Keyword Spotting ist eine nutzbringende Durchsuchbarkeit auch bei schlechtem HTR-Ergebnis möglich





---

## Nachteile

- die Entwicklung eines eigenen HTR-Modells ist relativ aufwändig (wobei mit zunehmender Zahl veröffentlichter Modelle oftmals ein passendes gefunden werden kann, mit welchem schon recht gute Ergebnisse erzielt werden können)
- Nutzeroberfläche ist nicht intuitiv
- In der Regel sind manuelle Nacharbeiten notwendig, wenn das Ergebnis perfekt sein soll



# Registrierung & Installation

<https://transkribus.eu/Transkribus/>

Transkribus [Register](#) [Login](#)

## Transcribe. Collaborate. Share...

...and benefit from cutting edge research in Handwritten Text Recognition!

[Download version 1.10](#) [Download version 1.10 for Mac](#) [Wiki > How-to guide \(pdf\) >](#)

### Scholars

Are you transcribing historical documents? Handwritten or printed, from the middle ages or from the 20th century? Would you like to do this in a highly standardized, flexible and reliable way? And do you appreciate to get support from automated tools such as Handwritten Text Recognition and Layout Analysis?

[View details >](#)

### Archives

Are you responsible for large collections of handwritten and printed documents? Do you believe that digitisation paves the way to realise new opportunities to access, enrich and explore archival material? And are you open to involve humanities scholars and volunteers so that they can work with these documents in an effective way – producing data which can also be integrated in your repository?

[View details >](#)

### Volunteers

Are historical letters, postcards, manuscripts or medieval documents fascinating for you? Do you enjoy deciphering handwriting – this wonderful feeling when you can read something which may be hidden to most other people? And do you believe that everyone can make a valuable contribution to scholarship and science?

[View details >](#)

### Scientists

Are you a computer scientist and in the fields of computer vision, analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, and publishing? And you would enjoy to get some reference data in a well-acknowledged format, such as PAGE?

[View details >](#)

### Create account

**Email\***

**Password\***

**Repeat password\***


**Given name\***

**Family name\***

**ORCID**

**Gender**

**I accept the terms of use\***

**Ich bin kein Roboter.**   
reCAPTCHA  
Datenschutzerklärung - Nutzungsbedingungen

\* Required

[Sign up](#)

Hinweis: Probleme bei der Installation sind oft darauf zurückzuführen, dass nicht die neue Java Version installiert ist.



Transkribus v1.10.0 (23\_01\_2020\_15:03)

Server: Overview | Layout | Metadata | Tools

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: No collection selected Col-ID

Documents

ID	Title	Pages	Uploader	Uploaded	Collections
----	-------	-------	----------	----------	-------------

100 Filter



Transkribus v1.10.0 (23\_01\_2020\_15:03)

Search current document... 0 / 0

Server Overview Layout Metadata Tools

Logout handschriften@ub.uni-tuebingen.de

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: nicole.domka@ub.uni-tuebingen.de Collection (36743, Owner) **Col ID**

Documents HTR Model Data

1-4 / 4

ID	Title	Pages	Uploader	Uploaded	Collections
139200	German Handwriting 0.1	1	nicole.domka...	Fri Mar 22 10:0...	(nicole.domka...
139199	Wiener Diarium 4.0	1	nicole.domka...	Fri Mar 22 10:0...	(nicole.domka...
139198	Dutch Handwriting 0.1	1	nicole.domka...	Fri Mar 22 10:0...	(nicole.domka...
139197	English Handwriting 0.1	1	nicole.domka...	Fri Mar 22 10:0...	(nicole.domka...

**Collections**

**Dokumente**

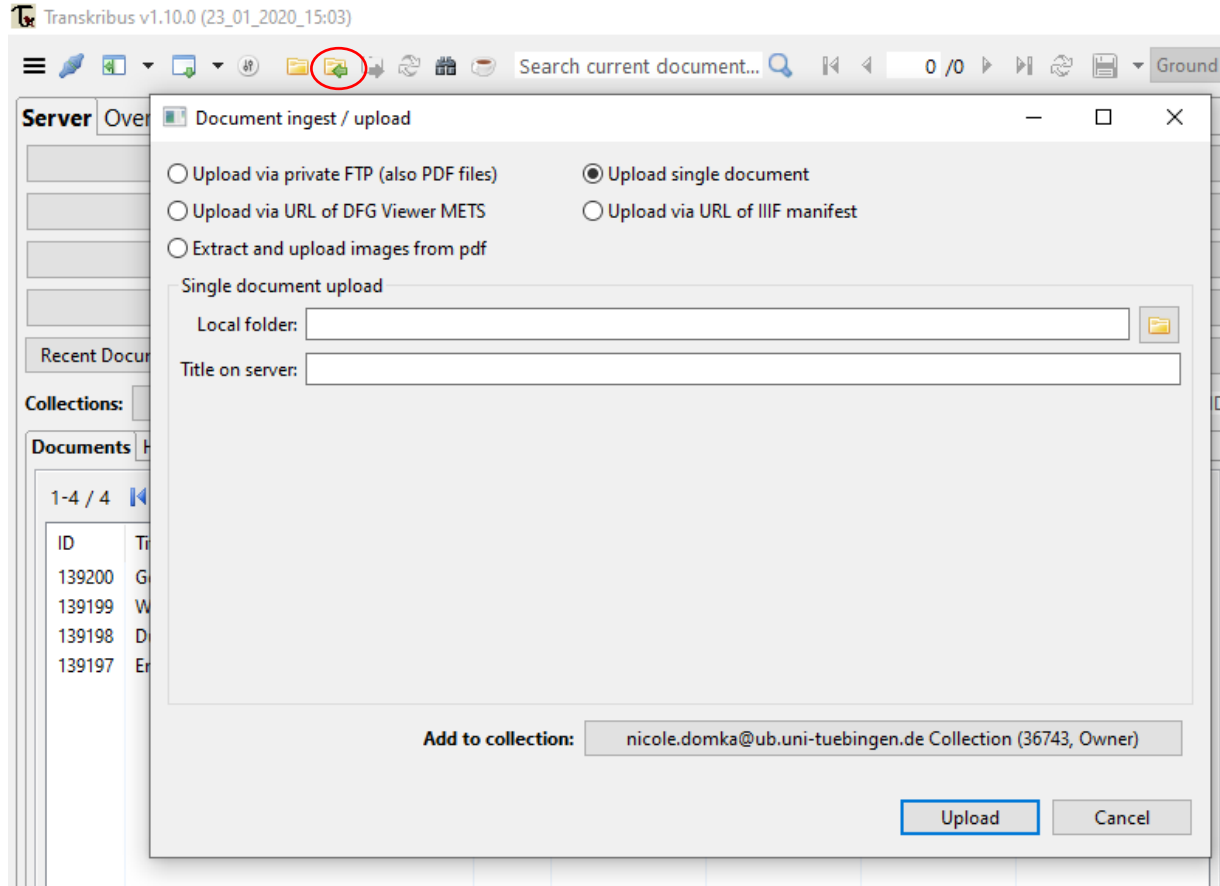




---

## Upload und Management von Dokumenten

- die hochgeladenen Dokumente werden auf Servern in Innsbruck gespeichert (abgesehen vom Programm selbst werden keine Kapazitäten auf der eigenen Festplatte und des Arbeitsspeichers benötigt)
- die Dokumente sind grundsätzlich privat und von anderen Nutzern nicht einsehbar
- über den User Manager können andere Personen der eigenen Collection hinzugefügt werden und ihnen über die Zuweisung von Rollen bestimmte Rechte eingeräumt werden
- Verwaltung der Dokumente in Collections



- mindestens 300 dpi
- Upload des gesamten Ordners, nicht eines einzelnen Bildes
- es können auch PDFs hochgeladen werden



## Grundlegende Funktionen



- Versions: es ist möglich zu jeder abgespeicherten vorherigen Version des Dokuments zurückzukehren
- User Manager: es ist möglich, andere Personen zu einer Collection hinzuzufügen und ihnen über Rollen bestimmte Rechte einzuräumen
- Jobs: hier lassen sich laufende und vergangene Prozesse überwachen
- User activity: welcher User wann was in welchem Dokument gemacht hat



## Segmentierung: Layout Analyse

- HTR ist zeilen- und nicht zeichenbasiert wie klassische OCR
- Bild und Text müssen miteinander verbunden werden, damit die Handschriftenerkennung funktioniert. Dies geschieht dadurch, dass das Dokument segmentiert wird. Die Hauptelemente sind hier Textregions und Baselines. Zunächst werden Textregionen definiert, die den Text beinhalten, und diese anschließend wiederum in Zeilen gegliedert. Diese Zeilen korrespondieren mit der Transkription.
- manuelle Segmentierung oder automatische Layout Analyse
- bei kompliziertem Layout mit sich wiederholenden Elementen ist das Training eines eigenen Strukturmodells über das Tool P2PaLA möglich





---

## Transkription/Ground Truth Daten

- damit eine automatische Texterkennung möglich ist, muss Transkribus trainiert werden
- Voraussetzung für die Erstellung eines eigenen HTR-Modells sind Ground Truth Daten, also korrekte Transkriptionen, anhand derer die Software lernen kann
- Ground Truth Daten sollten repräsentativ für das Dokument sein
- 2 Optionen
  - Erstellung einer manuellen Transkription in Transkribus (der segmentierte Text wird Zeile für Zeile transkribiert)
  - Verwendung von vorhandenen Transkriptionen über das Tool Text2Image (dazu muss die Transkription so aufbereitet werden, dass Text und Bild übereinstimmen)



- 
- Als Grundlage für wissenschaftliche Editionen kann auch eine anspruchsvollere Transkription angefertigt werden, z.B. über
    - Festlegung der Lesereihenfolge von Textregionen und Zeilen
    - Hinzufügung von Tags und Metadaten
  - Transkriptionsregeln
    - eine einheitliche Transkription ist sinnvoll, damit die Software dementsprechend lernen kann
    - Beispiele:
      - Transkription von „f“ als „f“ oder „s“ (ähnlich i/j, u/v)
      - Auflösung/Beibehaltung von Abkürzungen (am Wortende funktioniert dies gut, innerhalb des Wortes ist es schwieriger)
      - siehe Editorial Declaration
      - die aktuell fehlende Standardisierung sollte bei der Nutzung von HTR-Modellen beachtet werden



---

## Modell-Training

- Voraussetzung: Ground Truth Daten im Umfang von ca. 25-75 Seiten (5000 bis 15000 Wörter), bei gedrucktem Material weniger
- das Modelltraining ist nicht automatisch freigeschaltet (Email an Transkribus, die entweder den entsprechenden Button freischalten oder das Training selbst übernehmen)
- Aufteilung der Ground Truth Daten in ein Trainings Set und in ein Validation Set
- es kann ein Basismodell hinzugeschaltet werden
- Festlegung der Anzahl der Epochen



---

## Texterkennung

- Auswahl eines HTR-Modells (eigenes oder fremdes)
- Hinzuschaltung eines Dictionarys oder Language Models
- Erstellung eines Sample Sets zum Vergleich
- es gibt verschiedene Optionen, (automatische) Transkriptionen zu vergleichen und so die Performance eines Modells zu bewerten



---

## Keyword Spotting

- während die vom Programm am wahrscheinlichsten gewertete Transkription angezeigt wird, sind im Hintergrund weitere Varianten abgespeichert, die über Keyword Spotting abgerufen werden können
- Keyword Spotting ermöglicht also auch bei einer höheren Fehlerrate die Nutzbarkeit der automatischen Transkription für die Durchsuchung des Volltextes



---

## Export von Dokumenten

- unterschiedliche Formate → Nachnutzbarkeit in verschiedenen Kontexten
- Export per Downloadlink oder direkte Speicherung auf den eigenen Computer
- Export von Tags
- Export von Tabellen in Excel
- Export bestimmter Versionen
- Schwärzung vertraulicher Informationen





---

# Literatur

Registrierung/Download Transkribus: <https://transkribus.eu/Transkribus/>

Allgemeine Informationen: <https://read.transkribus.eu/>

Anleitungen und Tutorials (Englisch): <https://read.transkribus.eu/transkribus/>

Wiki: [https://transkribus.eu/wiki/index.php/Main\\_Page](https://transkribus.eu/wiki/index.php/Main_Page) (hier auch Anleitungen in anderen Sprachen)

Blog vom Universitätsarchiv Greifswald, der eine gute Ergänzung zu den von Transkribus zur Verfügung gestellten Anleitungen darstellt, und einzelne Aspekte näher beleuchtet (z.B. Erste Schritte): <https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/de/>

- <https://mittelalter.hypotheses.org/21828>
- <https://latex-ninja.com/2019/11/10/how-to-historical-text-recognition-a-transkribus-quickstart-guide/>
- <https://www.emerald.com/insight/content/doi/10.1108/JD-07-2018-0114/full/html>
- <https://fortext.net/tools/tools/transkribus>
- <https://fortext.net/routinen/lerneinheiten/manuskriptdigitalisierung-mit-transkribus>

<https://ocr-bw.bib.uni-mannheim.de/projektuebersicht/>



# Danke.

Kontakt: Dorothee Huff

**Universitätsbibliothek Tübingen**  
Wilhelmstraße 32, 72074 Tübingen  
Telefon: +49 7071 29-72852  
[Dorothee.huff@uni-tuebingen.de](mailto:Dorothee.huff@uni-tuebingen.de)