
Foto: Valentin Marquardt, Universität Tübingen

Kompetenzzentrum OCR – Automatische Texterkennung als Serviceangebot

BiblioCon 2023
Dorothee Huff, Kristina Stöbener (UB Tübingen)
Larissa Will (UB Mannheim)

Vorhaben

1. Projektvorstellung: OCR-BW
2. Kompetenzzentrum OCR
    2.1. Serviceangebot
    2.2. Institutionen
    2.3. Wissenschaft und Forschung
3. Fazit
4. Kontakt

1. Projektvorstellung: OCR-BW

Projektpartner:	UB Tübingen (Handschriften) & UB Mannheim (Drucke)
Laufzeit:	2019 – 2022
Finanzielle Förderung:	Ministerium für Wissenschaft, Forschung und Kunst
Ziel:	Aufbau eines Kompetenzzentrums für Volltexterschließung von handschriftlichen und gedruckten Werken
Aufgabe:	Unterstützung von Archiven, Bibliotheken und anderen kulturbewahrenden Einrichtungen sowie Wissenschaft und Forschung bei Anwendung von OCR- und Transkriptionssoftware

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

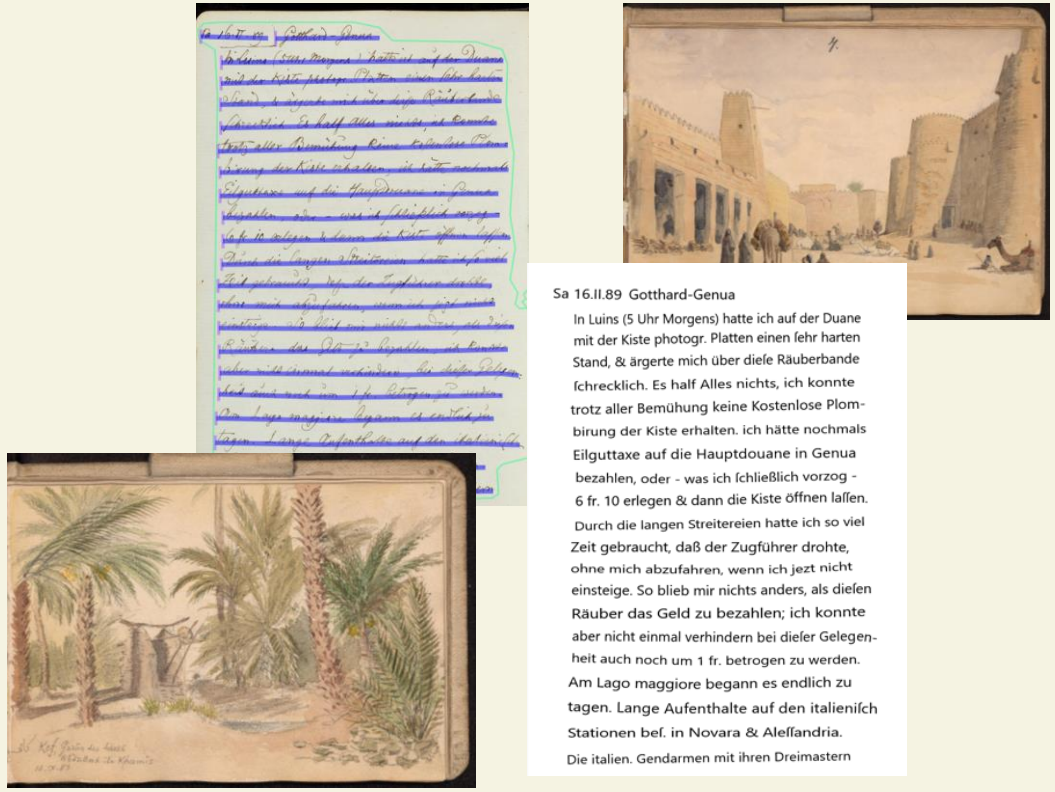
1. Projektvorstellung: OCR-BW

Meilensteine

- Softwareentwicklung und -verbesserungen
- Erstellung von Ground-Truth-Daten
- Modelltraining für Tesseract, Kraken und Transkribus
- umfangreiche Schulungsmaterialien und Einführungsveranstaltungen zu gängiger OCR-Software
- Bereitstellung einer eScriptorium-Instanz
- Beratung und Unterstützung kulturbewahrender Institutionen sowie Wissenschaft und Forschung bei der Anwendung von Texterkennungsprogrammen

2. Kompetenzzentrum OCR

- Kompetenzzentrum OCR mit Kooperationspartnern UB Tübingen und UB Mannheim bleibt weiterhin als Ansprechpartner bestehen
- Ausbau des Serviceangebots
- Fortführung bestehender Kooperationen sowie neue Projekte
- Organisation eigener Veranstaltungen wie dem „Transcribathon durch den Orient“ im Rahmen der Love Data Week 2023



2.1. Serviceangebot

Beratung

- Beratung und Unterstützung bei Fragen zur automatisierten Texterkennung per Mail, Videokonferenz oder in persönlichem Austausch
- offene Sprechstunde via Zoom, jeden zweiten Donnerstag im Monat

Schulungen

- Durchführung von Einführungskursen in eScriptorium und Transkribus
- Schulungsmaterialien über ZOERR

2.1. Serviceangebot

Unterstützung bei der Anwendung von OCR

- Hinweise zur Digitalisierung
- Bearbeitung von Testseiten
- Empfehlung von Modellen zur Layout- und Texterkennung
- Texterkennung im kleineren Umfang
- Bereitstellung einer eScriptorium-Instanz durch die UB Mannheim
- Durchführung von Einführungskursen in eScriptorium und Transkribus
- für kleine Projekte Bereitstellung von Ressourcen bzw. Übernahme des Modelltrainings

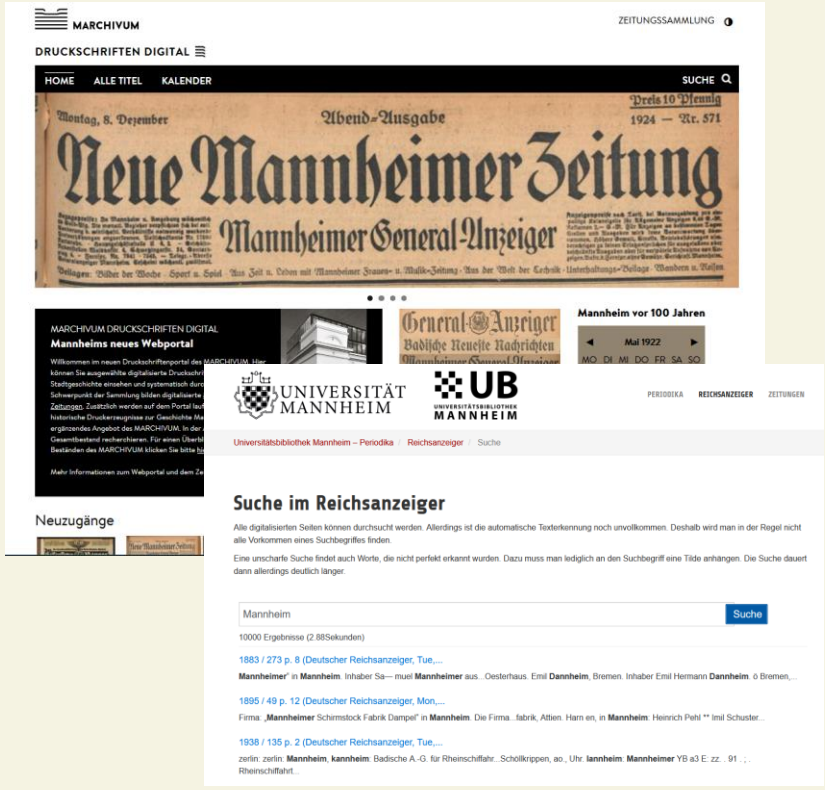
2.2. Institutionen

- Badische Landesbibliothek Karlsruhe
- Bibliothèque Nationale et Universitaire de Strasbourg
- Friedrich-Ebert-Stiftung
- Internet Archive
- MARCHIVUM (Mannheimer Stadtarchiv)
- Stadtarchiv Freiburg
- Stadtarchiv Ladenburg
- Stadtarchiv Nagold
- Universitätsbibliothek Mannheim
- Universitätsbibliothek Tübingen
- …

2.2. Institutionen

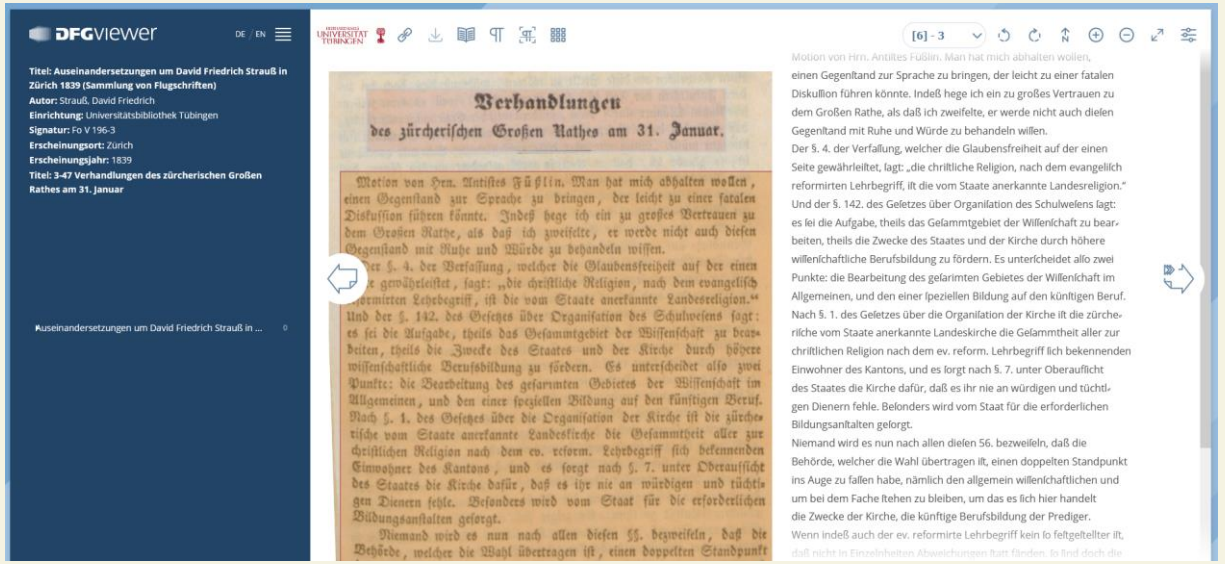
- Mannheimer Zeitungen bei MARCHIVUM:
- https://druckschriften-digital.marchivum.de/

- Reichsanzeiger bei UB Mannheim:
- https://digi.bib.uni-mannheim.de/periodika/reichsanzeiger/



2.2. Institutionen

- Teile des Druckschriftenbestands der UB Tübingen:
- https://github.com/UB-Mannheim/digitue-gt/wiki#results-examples



2.3 Wissenschaft und Forschung

Serviceleistungen

- Probelauf mit konkreten Anwendungsbeispielen
- Einführungsveranstaltungen speziell auf Projektbedürfnisse zugeschnitten
- Unterstützung bei der Projektplanung und -durchführung

Ziel: Erzeugung von möglichst guten automatischen Transkripten mit möglichst geringem Aufwand zur Weiterverarbeitung der Daten in anderen Kontexten

Wer: Anglistik, Altorientalische Philologie, Biologie, Germanistik, Geschichte, Judaistik, Kulturanthropologie, Orient- und Islamwissenschaft, Paläontologie, Rechtswissenschaft, Romanistik, Skandinavistik, Theologie etc.

Wozu: Lesehilfe, maschinelle Auswertung und Weiterverarbeitung, Vorbereitung von Editionen, Aufbereitung von Druckeditionen, Durchsuchung automatischer Transkriptionen

Anwendungsfall: Edition

- **Ziel:** Erstellung einer Hybridedition für ein Textkorpus, das auf ca. 600 mittelalterliche Handschriften verteilt ist
- **Prämisse:** Transkriptionsgrundlage
- **Anforderungen an OCR:**
  - Unterstützung der Editionsrichtlinien
  - generisches Modell
  - vielfältige Diakritika

- **Vorgehen:**
  - Nachnutzung bereits erstellter Transkriptionen für Modelltraining → seitdem sukzessives Nachtraining bzw. werksspezifisches Training mit korrigierten Daten
  - UB: Layouterkennung, Einfügung der Transkriptionen, Modelltraining
  - Projekt: manuelle Textkorrektur

- 1. Modell: 130 S. GT aus 18 Handschriften → 5,76% CER
- 11. Modell: 348 S. GT aus 43 Handschriften → 5,40% CER

9,00% CER (Modell RKE_7)

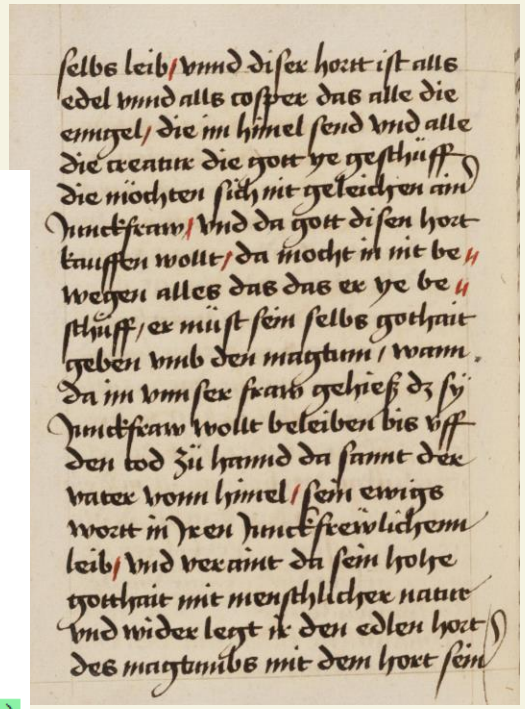
selbs leib ⁄ vnnd diser hortt ist ~~als~~ alls
edel vnnd alls cosper das alle die
~~einigel~~ enngel ⁄ die im himel send vnd ~~als~~ alle
die creatur die gott ~~ge geschuff~~ ye geschüff
die ~~mochten~~ möchten sich nit geleichen ~~an~~ ain
~~juckfram~~ junckfraw ⁄ vnd da gott disen ~~hört~~ hort
kauffen ~~wont~~ wollt ⁄ da mocht in nit be
wegen alles das das er ye be ~~i~~
~~schuff~~ schüff ⁄ er ~~must~~ müst sein selbs gothait
geben vmb den ~~magtun~~ magtum ⁄ wann
da im ~~vnn ser fratb~~ vnnser fraw gehieß daz sy
~~muetfraw~~ junckfraw wollt beleiben bis ~~vß~~ vff
den tod zü hannd ~~du sanmit~~ da sannt der
~~nater nonn jnirel~~ vater vonn himel ⁄ sein ewigs
wortt in ~~iren junctkfrewlichem~~ jren junckfrewlichenn
leib ⁄ vnd ~~verainit da~~ veraint du sein ~~hoge~~ hohe
gotthait mit ~~menschlichen~~ menschlicher natur
vnd ~~vmider~~ wider legt ir ~~denn~~ den edlen hort
des magtumbs mit dem hort ~~sein~~ sein(er)

3,79% CER (Modell RKE_8)

selbs leib ⁄ vnnd diser hortt ist alls
edel vnnd alls ~~tosper~~ cosper das alle die
enngel ⁄ die im himel send vnd ~~allt~~ alle
die creatur die gott ye ~~geschuff~~ geschüff
die möchten sich nit geleichen ain
junckfraw ⁄ vnd da gott disen hort
kauffen wollt ⁄ da mocht in nit be
wegen alles das das er ye be
~~schuff~~ schüff ⁄ er müst sein selbs gothait
geben vmb den magtum ⁄ wann
da im vnnser ~~fraib~~ fraw gehieß daz sy
junckfraw wollt beleiben bis ~~vf~~ vff
den tod ~~zu~~ zü hannd da sannt der
vater vonn himel ⁄ sein ewigs
wortt in jren junckfrewlichenn
leib ⁄ vnd veraint ~~da~~ du sein hohe
gotthait mit menschlicher natur
vnd wider legt ir den edlen hort
des ~~magtubs~~ magtumbs mit dem hort ~~sein~~ sein(er)

2,84% CER (Modell RKE_11)

selbs leib ⁄ vnnd diser hortt ist alls
edel vnnd alls cosper das alle die
enngel ⁄ die im himel send vnd alle
die creatur die gott ye ~~geschuff~~ geschüff
die möchten sich nit geleichen ain
junckfraw ⁄ vnd da gott disen ~~hört~~ hort
kauffen wollt ⁄ da mocht in nit be
wegen alles das das er ye be
~~schuff~~ schüff ⁄ er müst sein selbs gothait
geben vmb den magtum ⁄ wann
da im vnnser ~~fraib~~ fraw gehieß daz sy
junckfraw wollt beleiben bis vff
den tod zü hannd da sannt der
vater vonn himel ⁄ sein ewigs
wortt in jren junckfrewlichenn
leib ⁄ vnd veraint ~~da~~ du sein hohe
gotthait mit menschlicher natur
vnd wider legt ir den edlen hort
des ~~magtmubs~~ magtumbs mit dem hort ~~sein~~ sein(er)

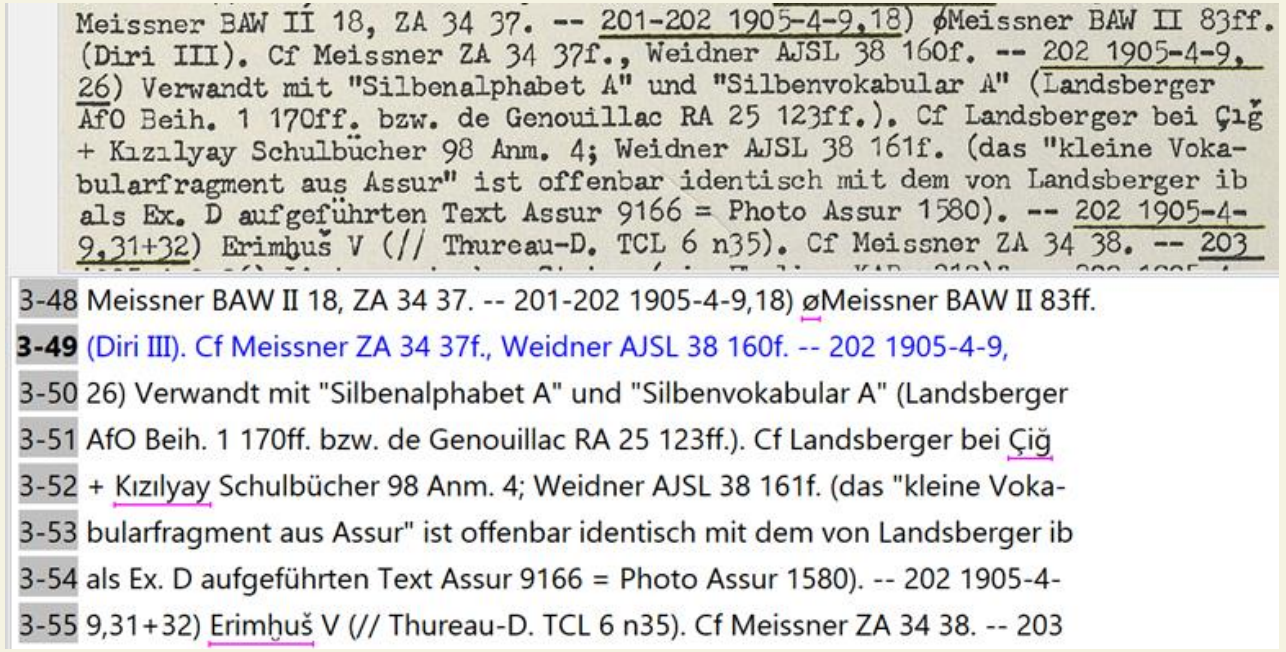

selbs leib/ vnnd diser hortt ist alls
edel vnnd alls cosper das alle die
enngel/ die im himel send vnd alle
die creatur die gott ye geschüff
die möchten sich mit geleichen ain
junckfraw/ vnd da gott diesen hort
kauffen wollt/ da mocht in mit be
wegen alles das das er ye be
schüff/ er müst sein selbs gothait
geben vmb den magtum/ wann
da im vnnser fraw gehieß daz sy
junckfraw wollt beleiben bis vff
den tod zü hannd da sannt der
vater vonn himel/ sein ewigs
wortt in jren junckfrewlichem
leib/ vnd veraint da sein holtze
gotthait mit menschlicher natur
vnd wider legt ir den edlen hort
des magtumbs mit dem hort sein

http://idb.ub.uni-tuebingen.de/opendigi/Md123#p=182

Anwendungsfall: Datenbank

- **Ziel:** Überführung von gedruckten und maschinenschriftlichen Bibliographien in eine Datenbank
- **Prämisse:** strukturierter Volltext
- **Anforderungen an OCR:**
  - Erkennung sprachspezifischer Sonderzeichen

- **Vorgehen:**
  - werksspezifisches Training nach manueller Korrektur
  - UB: Erstellung von GT-Daten für einen ersten Testlauf, Layouterkennung, automatische Texterkennung, Modelltraining, Weiterverarbeitung der Daten
  - Projekt: manuelle Textkorrektur


Meissner BAW II 18, ZA 34 37. -- 201-202 1905-4-9,18) øMeissner BAW II 83ff.
(Diri III). Cf Meissner ZA 34 37f., Weidner AJSL 38 160f. -- 202 1905-4-9,
26) Verwandt mit "Silbenalphabet A" und "Silbenvokabular A" (Landsberger
AfO Beih. 1 170ff. bzw. de Genouillac RA 25 123ff.). Cf Landsberger bei Çığ
+ Kızılyay Schulbücher 98 Anm. 4; Weidner AJSL 38 161f. (das "kleine Voka-
bularfragment aus Assur" ist offenbar identisch mit dem von Landsberger ib
als Ex. D aufgeführten Text Assur 9166 = Photo Assur 1580). -- 202 1905-4-
9,31+32) Erimḫuš V (// Thureau-D. TCL 6 n35). Cf Meissner ZA 34 38. -- 203

3-48 Meissner BAW II 18, ZA 34 37. -- 201-202 1905-4-9,18) øMeissner BAW II 83ff.
3-49 (Diri III). Cf Meissner ZA 34 37f., Weidner AJSL 38 160f. -- 202 1905-4-9,
3-50 26) Verwandt mit "Silbenalphabet A" und "Silbenvokabular A" (Landsberger
3-51 AfO Beih. 1 170ff. bzw. de Genouillac RA 25 123ff.). Cf Landsberger bei Çığ
3-52 + Kızılyay Schulbücher 98 Anm. 4; Weidner AJSL 38 161f. (das "kleine Voka-
3-53 bularfragment aus Assur" ist offenbar identisch mit dem von Landsberger ib
3-54 als Ex. D aufgeführten Text Assur 9166 = Photo Assur 1580). -- 202 1905-4-
3-55 9,31+32) Erimḫuš V (// Thureau-D. TCL 6 n35). Cf Meissner ZA 34 38. -- 203

Borger, Rykle: Handbuch der Keilschriftliteratur. Bd. 1: Repertorium der sumerischen und akkadischen Texte. Berlin, 1967, S. 337.

Ç	LATIN CAPITAL LETTER C WITH CEDILLA
É	LATIN CAPITAL LETTER E WITH ACUTE
Ú	LATIN CAPITAL LETTER U WITH ACUTE
à	LATIN SMALL LETTER A WITH GRAVE
á	LATIN SMALL LETTER A WITH ACUTE
ä	LATIN SMALL LETTER A WITH DIAERESIS
ç	LATIN SMALL LETTER C WITH CEDILLA
è	LATIN SMALL LETTER E WITH GRAVE
é	LATIN SMALL LETTER E WITH ACUTE
ê	LATIN SMALL LETTER E WITH CIRCUMFLEX
î	LATIN SMALL LETTER I WITH CIRCUMFLEX
ó	LATIN SMALL LETTER O WITH ACUTE
ö	LATIN SMALL LETTER O WITH DIAERESIS
ø	LATIN SMALL LETTER O WITH STROKE
ú	LATIN SMALL LETTER U WITH ACUTE
û	LATIN SMALL LETTER U WITH CIRCUMFLEX
ü	LATIN SMALL LETTER U WITH DIAERESIS
ý	LATIN SMALL LETTER Y WITH ACUTE
ā	LATIN SMALL LETTER A WITH MACRON
č	LATIN SMALL LETTER C WITH CARON
ē	LATIN SMALL LETTER E WITH MACRON
ğ	LATIN SMALL LETTER G WITH BREVE
ī	LATIN SMALL LETTER I WITH MACRON
ı	LATIN SMALL LETTER DOTLESS I
ś	LATIN SMALL LETTER S WITH ACUTE
Š	LATIN CAPITAL LETTER S WITH CARON
š	LATIN SMALL LETTER S WITH CARON
ū	LATIN SMALL LETTER U WITH MACRON
ᵈ	MODIFIER LETTER SMALL D
Ḫ	LATIN CAPITAL LETTER H WITH BREVE BELOW
ḫ	LATIN SMALL LETTER H WITH BREVE BELOW
Ṣ	LATIN CAPITAL LETTER S WITH DOT BELOW
ṣ	LATIN SMALL LETTER S WITH DOT BELOW
₆	SUBSCRIPT SIX
←	LEFTWARDS ARROW
◦	SQUARE LOZENGE

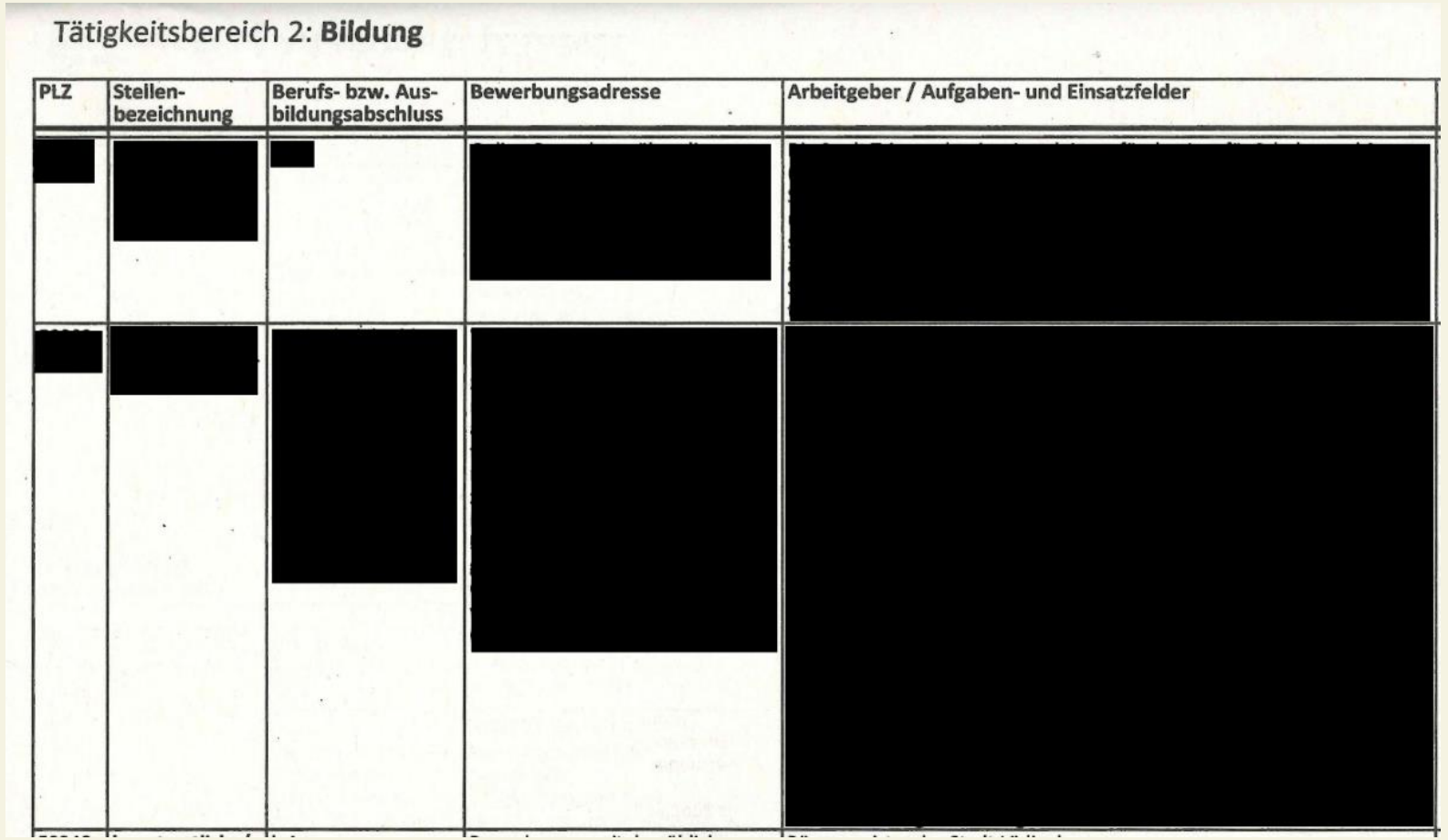
Anwendungsfall: linguistische Textanalyse

- **Ziel:** Sprachanalyse eines großen Textkorpus (moderne Druckwerke)
- **Prämisse:** durchsuchbarer Volltext
- **Anforderungen an OCR:**
  - gesuchte Entitäten müssen zuverlässig gefunden werden
    - → 0% CER
    - → keine Worttrennungen
- **Vorgehen:**
  - Testlauf mit 150 Werken
  - Modelltraining für Sonderzeichen
  - UB: Layouterkennung, automatische Texterkennung, Modelltraining
  - Projekt: manuelle Textkorrektur

Anwendungsfall: Tabelle

- **Ziel:** Auswertung von Tabelleninhalten
- **Prämisse:** durchsuchbarer Volltext
- **Anforderungen an OCR:**
  - Layouterkennung von Tabellen

- **Vorgehen:**
  - UB: Prüfung verschiedener Methoden der Layouterkennung, Einführung in Transkribus
  - Projekt: selbstständige Bearbeitung



Tätigkeitsbereich 2: **Bildung**

PLZ	Stellen-bezeichnung	Berufs- bzw. Aus-bildungsabschluss	Bewerbungsadresse	Arbeitgeber / Aufgaben- und Einsatzfelder
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Anwendungsfall: Biologie

- **Ziel:** Lesehilfe, Einbindung der Daten in verschiedene Applikationen
- **Prämisse:** durchsuchbarer Volltext, exportierbare Tags
- **Anforderungen an OCR:**
  - Layouterkennung und Export von Tabellen
  - Listen mit Fachtermini
  - Tagging

- **Vorgehen:**
  - Modelltraining auf Grundlage vorhandener Transkriptionen
  - UB: Layouterkennung, Modelltest und -training
  - Projekt: Tagging

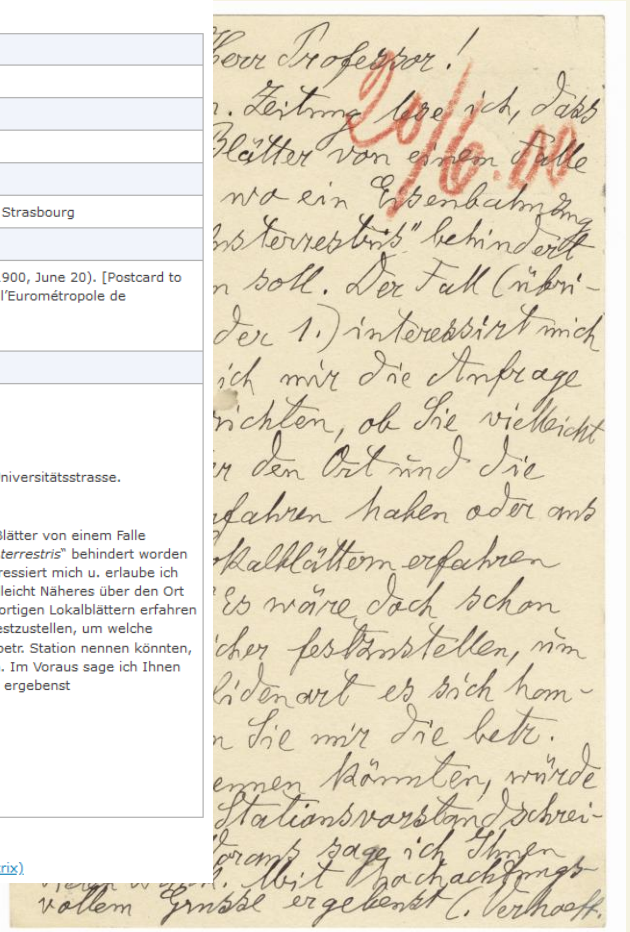
Einbindung in die virtuelle Forschungsumgebung Myriatrix

[Postcard to Ludwig Döderlein]

Publication Type:	Manuscript
Year of Publication:	1900
Authors:	C. Wilhelm Verhoeff
Collection Title:	Correspondence with Ludwig Döderlein
Date Published:	20/06/1900
Library/Archive:	Archives de la Ville et de l'Eurométropole de Strasbourg
City:	Strasbourg, Grand Est, France
Abstract:	**Recommended citation:** Verhoeff, C. W. (1900, June 20). [Postcard to Ludwig Döderlein]. Archives de la Ville et de l'Eurométropole de Strasbourg, Strasbourg, Grand Est, France.
Citation Key:	1429
Full Text	BONN 20.6.00 **An** Herrn Prof. Dr. L. Döderlein. **in** Strassburg i/E. **Wohnung** Zoolog. Anstalt der Hochschule. Universitätsstrasse. Verehrter Herr Professor! In der Köln. Zeitung lese ich, dass Elsässer Blätter von einem Falle berichten, wo ein Eisenbahnzug durch „*Julus terrestris*" behindert worden sein soll. Der Fall (übrigens nicht der 1.) interessiert mich u. erlaube ich mir die Anfrage an Sie zu richten, ob Sie vielleicht Näheres über den Ort und die Sachlage erfahren haben oder aus dortigen Lokalblättern erfahren können. Es wäre doch schon wichtig sicher festzustellen, um welche Julidenart es sich handelt. Wenn Sie mir die betr. Station nennen könnten, würde ich an den Stationsvorstand schreiben. Im Voraus sage ich Ihnen vielen Dank. Mit hochachtungsvollem Grusse ergebenst C. Verhoeff. C. Verhoeff Dr. phil. Zoologe BONN

Taxonomic name:
Julida (Myriatrix), Julus terrestris (Myriatrix), Schizophyllum sabulosum (Myriatrix)

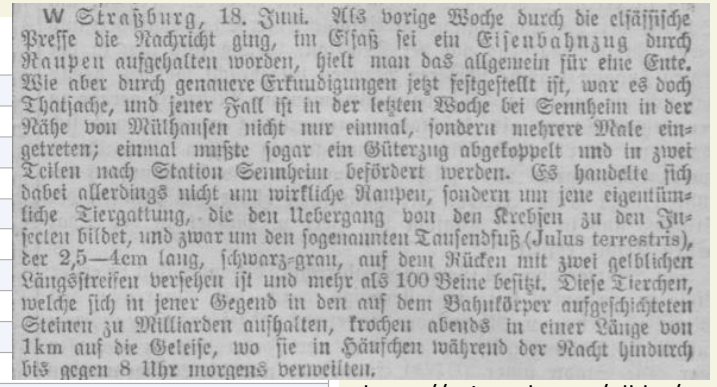
https://myriatrix.myspecies.info/content/postcard-ludwig-d%C3%B6derlein



[Tausendfuß (*Julus terrestris*)]

Publication Type:	Newspaper Article
Year of Publication:	1900
Authors:	Anonymous
Newspaper:	Kölnische Zeitung
Volume:	473
Section:	Vermischte Nachrichten
Edition:	Erste Morgen-Ausgabe
Pagination:	2
Issue Date:	20/06/1900
City:	Köln
Abstract:	**Recommended citation:** Anonymous (1900): [Tausendfuß (*Julus terrestris*)]. *Kölnische Zeitung*, Mittwoch, 20. Juni. 1900 - Nr. 473, Erste Morgen-Ausgabe, Vermischte Nachrichten, page 2. https://zeitpunkt.nrw/ulbbn/periodical/zoom/7594565
URL:	https://zeitpunkt.nrw/ulbbn/periodical/zoom/7594565
Citation Key:	1489
Full Text	Straßburg, 18. Juni. Als vorige Woche durch die elsässische Presse die Nachricht ging, im Elsaß sei ein Eisenbahnzug durch Raupen aufgehalten worden, hielt man das allgemein für eine Ente. Wie aber durch genauere Erkundigungen jetzt festgestellt ist, war es doch Thatsache, und jener Fall ist in der letzten Woche bei Sennheim in der Nähe von Mülhausen nicht nur einmal, sondern mehrere Male eingetreten; einmal mußte sogar ein Güterzug abgekoppelt und in zwei Teilen nach Station Sennheim befördert werden. Es handelte sich dabei allerdings nicht um wirkliche Raupen, sondern um jene eigentümliche Tiergattung, die den Uebergang von den Krebsen zu den Insecten bildet, und zwar um den sogenannten Tausendfuß (Julus terrestris), der 2,5—4cm lang, schwarz-grau, auf dem Rücken mit zwei gelblichen Längsstreifen versehen ist und mehr als 100 Beine besitzt. Diese Tierchen, welche sich in jener Gegend in den auf dem Bahnkörper aufgeschichteten Steinen zu Milliarden aufhalten, krochen abends in einer Länge von 1km auf die Geleise, wo sie in Häufchen während der Nacht hindurch bis gegen 8 Uhr morgens verweilten.

Taxonomic name:
Julus terrestris (Myriatrix), Schizophyllum sabulosum (Myriatrix)


W Straßburg, 18. Juni. Als vorige Woche durch die elsässische Presse die Nachricht ging, im Elsaß sei ein Eisenbahnzug durch Raupen aufgehalten worden, hielt man das allgemein für eine Ente. Wie aber durch genauere Erkundigungen jetzt festgestellt ist, war es doch Thatsache, und jener Fall ist in der letzten Woche bei Sennheim in der Nähe von Mülhausen nicht nur einmal, sondern mehrere Male eingetreten; einmal mußte sogar ein Güterzug abgekoppelt und in zwei Teilen nach Station Sennheim befördert werden. Es handelte sich dabei allerdings nicht um wirkliche Raupen, sondern um jene eigentümliche Tiergattung, die den Uebergang von den Krebsen zu den Insecten bildet, und zwar um den sogenannten Tausendfuß (Julus terrestris), der 2,5—4cm lang, schwarz-grau, auf dem Rücken mit zwei gelblichen Längsstreifen versehen ist und mehr als 100 Beine besitzt. Diese Tierchen, welche sich in jener Gegend in den auf dem Bahnkörper aufgeschichteten Steinen zu Milliarden aufhalten, krochen abends in einer Länge von 1km auf die Geleise, wo sie in Häufchen während der Nacht hindurch bis gegen 8 Uhr morgens verweilten.

https://zeitpunkt.nrw/ulbbn/periodical/zoom/7594565

https://myriatrix.myspecies.info/content/tausendfu%C3%9F-julus-terrestris

Anwendungsfall: Vaybertaytsh

- **Ziel:** Kollationierung und Textanalyse
- **Prämisse:** durchsuchbarer Volltext
- **Anforderungen an OCR:**
  - linksläufige Schrift

- **Vorgehen:**
  - Erstellung einer Transkriptionsgrundlage mit einem öffentlichen Modell, anschließend eigenes Modelltraining
  - UB: Prüfung verschiedener Optionen für die Erkennung von hebräischer Schrift, Layouterkennung, Modelltraining
  - Projekt: manuelle Textkorrektur

1-1 בשׂורת

1-2 ווייז ·ווערטס ·אויך ·גין ·אויף ·דען ·טג ·וועז ·דעס ·מענשן ·זון ·זול ·אופֿנבארט ·ווערדן :אן

1-3 דעם ·זעלביגין ·טג ·ווער ·אויף ·דעם ·דך ·איז ·אונ׳ ·זיין ·הויס ·ראט ·אין ·דעם ·הויס ·דער ·שטייג ·ניט

1-4 הרנידר ·דש ·זעלב ·צו ·הולן :.דש ·זעלבן ·גלייכֿן ·ווער ·אויף ·דעם ·ועלד ·איז ·דער ·וועגד ·ניט ·אוים ·נוך

1-5 דעם ·דש ·הינטר ·אים ·איז :.גידענק ·אן ·דש ·ווייב ·לוט :.ווער ·דא ·זוכֿט ·זיין ·זיל ·צו ·ערהלטן ·דער

1-6 וועדט ·זי ·ור ·לירן :.אונ׳ ·ווער ·זי ·ור ·לירן ·וועדט ·דער ·וועדט ·זי ·צום ·לעבן ·גיבערן :איך ·זג

1-7 אויך ·אן ·דעם ·טג ·וועדן ·צווין ·אויף ·איינם ·בעט ·ליגן ·איינר ·וועדט ·אן ·גינומן ·דער ·אנדר ·וועדט

1-8 ור ·לוסן ·וועדן :.צוויי ·וועדן ·מאלן ·מיט ·אנדר :.אייני ·וועדט ·אן ·גינומן ·די ·אנדר ·וועדט ·ור ·לוסן

3. Fazit

- über die Projektlaufzeit hat sich eine gute Zusammenarbeit entwickelt, die weiter fortgeführt wird
- Kompetenzen der Projektpartner ergänzen sich
- OCR-Angebot war von Anfang an in die Bibliotheksstruktur eingebunden
- das Thema OCR wird für Bibliotheken sowie Wissenschaft und Forschung immer relevanter
  - Anreicherung von Digitalisaten mit Volltexten
  - neue Forschungsfragen z.B. aus den Digital Humanities
- OCR kann flexibel auf unterschiedliche Bedürfnisse angepasst werden
  - Spannungsverhältnis: Benutzerfreundlichkeit/Nachnutzbarkeit der Daten
- abgesehen von Einzelwerken kann der komplette OCR-Prozess in der Regel nicht durch das Kompetenzzentrum geleistet werden
  - Hilfe zur Selbsthilfe
  - begleitende Unterstützung

4. Kontakt

Projekt OCR-BW/Kompetenzzentrum OCR
https://ocr-bw.bib.uni-mannheim.de
https://uni-tuebingen.de/de/179298
https://www.bib.uni-mannheim.de/ihre-ub/projekte-der-ub/

eScriptorium
https://gitlab.com/scripta/escriptorium
https://ocr-bw.bib.uni-mannheim.de/escriptorium

Mailingliste
https://listserv.uni-tuebingen.de/mailman/listinfo/ocr_htr_ub

Kontakt
Dorothee Huff
dorothee.huff@uni-tuebingen.de
+49 7071 29–72852

Larissa Will
larissa.will@uni-mannheim.de
+49 621 181–2754

Kristina Stöbener
kristina.stoebener@uni-tuebingen.de
+49 7071 29–72834