
Foto: Valentin Marquardt, Universität Tübingen

Hands-On-Lab: Einführung in die Transkription von Handschriften und Drucken mit eScriptorium

BiblioCon 2023

Larissa Will, Stefan Weil (UB Mannheim)
Dorothee Huff (UB Tübingen)

Vorhaben

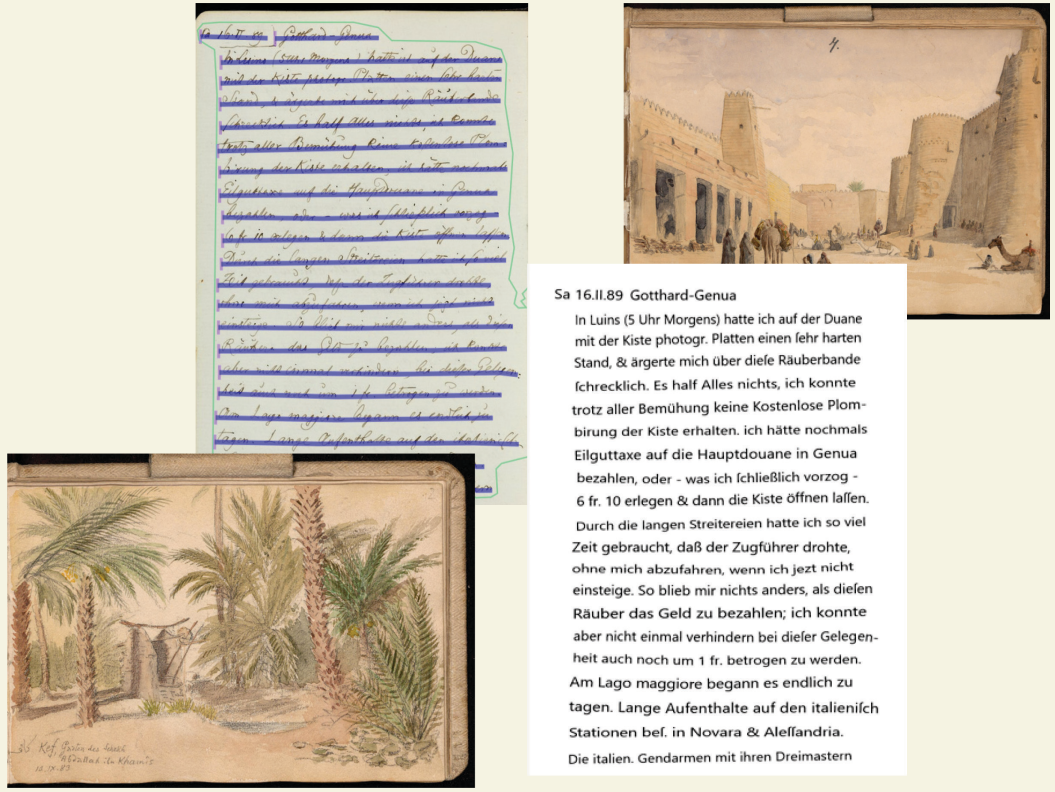
Vorstellung des Kompetenzzentrums OCR der UBs Mannheim und Tübingen
Kurze Vorstellungsrunde – bisherige Erfahrungen mit OCR
Wie funktioniert automatische Texterkennung?
Workflow moderner Texterkennung
Was ist eScriptorium?
Schritt-für-Schritt-Einführung
Bearbeiten eigener Dokumente
Erfahrungsaustausch

Kompetenzzentrum OCR

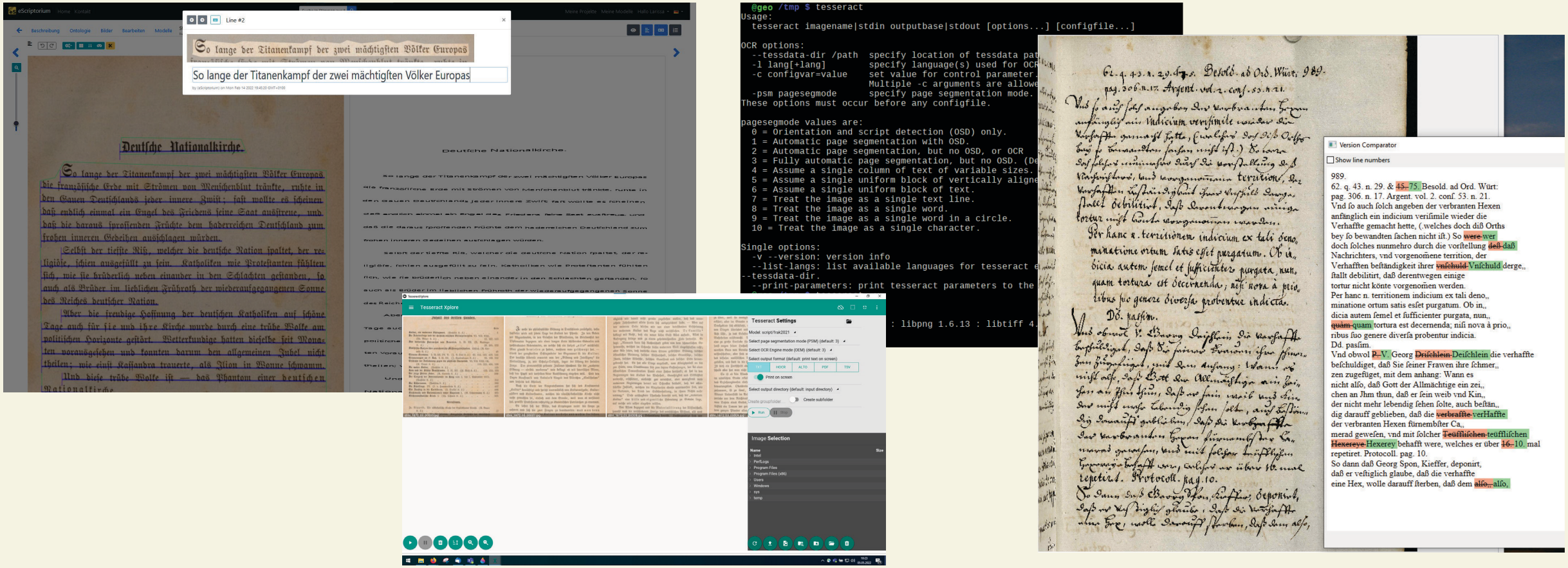
- Entstanden aus dem Projekt OCR-BW
  - Laufzeit: 2019–2022
  - Projektpartner: UB Tübingen und UB Mannheim
  - gefördert durch: Ministerium für Wissenschaft, Forschung und Kunst
  - Aufgabe: Unterstützung von Archiven, Bibliotheken und anderen kulturbewahrenden Einrichtungen bei Anwendung von OCR- und Transkriptionssoftware
  - Ziel: Aufbau eines Kompetenzzentrums für automatisierte Texterkennung von Drucken und Handschriften

Kompetenzzentrum OCR

- Kompetenzzentrum OCR mit Kooperationspartnern UB Tübingen und UB Mannheim bleibt weiterhin als Ansprechpartner bestehen
- Ausbau des Serviceangebots
- Fortführung bestehender Kooperationen, aber auch neue Projekte mit neuen Partnern
- Organisation eigener Veranstaltungen wie dem „Transcribathon durch den Orient“ in der Love Data Week 2023
- Offene Sprechstunde, jeden zweiten Donnerstag im Monat, online



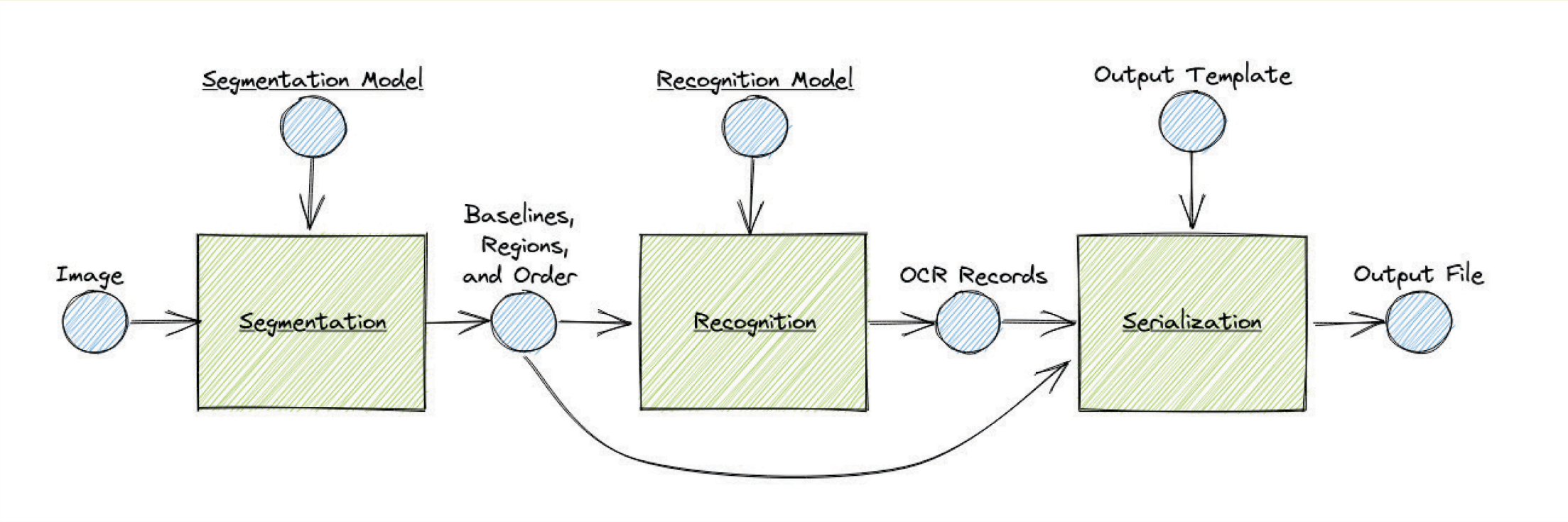
Bisherige Erfahrungen mit OCR?



Wie funktioniert automatische Texterkennung?

- Textliche Bildinhalte → digitale Textformate
- OCR: „Optical Character Recognition“ (optische Zeichenerkennung)
- Begriff mittlerweile veraltet
- Neuronale Netzen erkennen nicht mehr Zeichen für Zeichen, sondern ganze Zeilen
- Texterkennung und OCR wird im deutschen Sprachraum oft synonym verwendet

Workflow einer modernen Texterkennung



Was ist eScriptorium?

- entwickelt an der Université Paris PSL
- Kostenfrei und Open-Source
- Alternative zu Transkribus
- Handgeschriebene und gedruckte Texte manuell oder automatisiert segmentieren und transkribieren
- Einfache Weitergabe trainierter Modelle
- Jeder kann eigene Instanz installieren
- Für Windows, MacOS und Linux

Schritt-für-Schritt-Einführung

Gehen Sie auf:
https://ocr-bw.bib.uni-mannheim.de/escriptorium/

Melden Sie sich mit den folgenden Zugangsdaten an:

Benutzername: bibliocon1

Passwort: Hannover23!

Wir gehen zusammen Schritt für Schritt durch eScriptorium!

Weiterführende Ressourcen

Projekt OCR-BW/Kompetenzzentrum OCR
https://ocr-bw.bib.uni-mannheim.de

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

Modelle für eScriptorium/Kraken
https://ocr-bw.bib.uni-mannheim.de/faq/
https://zenodo.org/communities/ocr_models

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

Kontakt
Dorothee Huff
dorothee.huff@uni-tuebingen.de

Larissa Will
larissa.will@uni-mannheim.de

Stefan Weil
stefan.weil@uni-mannheim.de