

BEDEUTUNG VON TRAININGSDATEN UND
MODELLEN FÜR DEN DIGITALISIERUNGSPROZESS
IM BEREICH DIGITAL HUMANITIES.

LEISTUNGSNACHWEIS 1

VON

ANDRÉS BAUMELER

TELEFON: 076 443 04 71, E-MAIL: ANDRES@BAUMELER.DEV

ALTE RIEDIKERSTRASSE 5C, 8610 USTER

BETREUER

CLEMENS NEUDECKER



Universität
Zürich UZH



UNIVERSITÄT ZÜRICH, PHILOSOPHISCHE FAKULTÄT /
ZENTRALBIBLIOTHEK ZÜRICH

CAS DATENMANAGEMENT UND INFORMATIONSTECHNOLOGIEN

Inhaltsverzeichnis

1	Einleitung	2
2	Hauptteil	3
3	Schluss	4
	Abkürzungen	5
	Literaturverzeichnis	6

Kapitel 1

Einleitung

Bei der Digitalisierung von historischen Schriften in den Digital Humanities (DH) nimmt die Optical Character Recognition (OCR) eine wichtige Rolle ein. OCR wird im Digitalisierungsprozess dazu verwendet Text von digitalisierten Dokumenten maschinenlesbar zu machen. Romanello, Sven, and Robertson nennen in [3] die Qualität der verfügbaren Trainingsdaten sowie Komplexität des Layouts der Dokumente als Herausforderungen bei der Digitalisierung von historischen Dokumenten. In den letzten Jahren konnte die Qualität der OCR Resultate durch den Einsatz von Maschinelernen verbessert werden. So wird etwa das Erkennen von Handschrift möglich. Der Einsatz von Verfahren aus dem Maschinelernen setzt voraus, dass entsprechende Daten in genügender Qualität und Quantität verfügbar sind um die Algorithmen zu trainieren.

Kapitel 2

Hauptteil

Kapitel 3

Schluss

Das ist der Text. Quelle zitieren mit [1] oder Author Zitieren mit Niranjan

Abkürzungen

OCR Optical Character Recognition

DH Digital Humanities

Literaturverzeichnis

- [1] Adrian Paschke, Georg Rehm, Clemens Neudecker, Lydia Pintscher, editor. *Proceedings of the Third Conference on Digital Curation Technologies (Qurator 2022)*, volume 3234 of *CEUR Workshop Proceedings*, 2022. URL <https://ceur-ws.org/Vol-3234/paper2.pdf>.
- [2] Akella Niranjana. `niranjanstudy06/Multilingual-PDF-OCR-on-Google-Colab: Multilingual-PDF-OCR-on-Google-Colab`, January 2021. URL <https://doi.org/10.5281/zenodo.4415832>.
- [3] Matteo Romanello, Najem-Meyer Sven, and Bruce Robertson. Optical Character Recognition of 19th Century Classical Commentaries: the Current State of Affairs, September 2021. URL <https://doi.org/10.5281/zenodo.5524352>.