



Verborgene Datenschätze heben Das FDMLab experimentiert mit KI im Archiv

Das Landesarchiv Baden-Württemberg engagiert sich bei NFDI4Memory gemeinsam mit anderen Partnern für den Ausbau der Forschungsinfrastruktur für historische Daten. <https://4memory.de/>

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

Das FDMLab berichtet regelmäßig auf seinem Blog über den Einsatz von KI-Werkzeugen im Archiv. <https://fdmlab.landesarchiv-bw.de>



1 Mit der Texterkennungssoftware Transkribus wird ein Erkennungsmodell trainiert, das Personennamen in Militärstammrollen finden kann.

Vorlage: LABW

Das Landesarchiv Baden-Württemberg digitalisiert seit Jahren Archivgut und macht die Ergebnisse über seine Online-Katalog zugänglich. Inzwischen können bereits 16,7 Millionen Digitalisate online zeit- und ortsunabhängig recherchiert, angesehen und heruntergeladen werden (Stand Dezember 2021). Um diesen wachsenden Schatz an digitalen Quellen noch besser zugänglich zu machen, hat das Landesarchiv das Projekt *Forschungsdatenmanagementlabor am Landesarchiv Baden-Württemberg (FDMLab@LABW)* ins Leben gerufen. Das Projekt steht im größeren Kontext des Aufbaus einer Nationalen Forschungsdateninfrastruktur (NFDI) für die historisch arbeitenden Geisteswissenschaften. Während der zweijährigen Projektlaufzeit erprobt das FDMLab Werkzeuge und Technologien auch aus dem Bereich der Künstlichen Intelligenz (KI) für den Einsatz im Archiv. Gerade KI hat das Potenzial, uns bei

der Digitalisierung, bei der Erstellung von Erschließungsdaten sowie beim Auffinden von Informationen im Archiv zu unterstützen.

Dazu ein Beispiel: In den militärischen Beständen des Landesarchivs gibt es umfangreiche Personalunterlagen aus der Zeit von 1871 bis ca. 1920, die als Digitalisate über das Online-Findmittelsystem (OLF) einsehbar sind. In diesen sind Angehörige der einzelnen Einheiten mit biografischen Angaben aufgelistet. Deswegen sind die Personalunterlagen für Genealogen eine gerne genutzte historische Quelle. Da die Namen der Personen in unseren Erschließungsinformationen nicht erfasst sind, werden zur Recherche bisher zusätzliche Angaben zur militärischen Formation benötigt. Um diese Situation zu verbessern, experimentiert das FDMLab damit, die Personennamen mittels automatischer Handschriftenerkennung (HTR) aus den Digitalisaten auszulesen und als Suchindex bereitzustellen. Hierfür wird mit der Software Transkribus ein spezifisches Erkennungsmodell sowohl für das Layout der Personalunterlagen als auch für die Handschriften trainiert. Wenn dies gelingt, steht für die Suche nach der berühmten Nadel im Heuhaufen ein Metalldetektor zur Verfügung.

Neben der Volltexterkennung von Digitalisaten gibt es weitere Einsatzgebiete für KI im Archiv. So kann KI für die Verschlagwortung von Texten und Bildern verwendet werden. Auch die Erstellung von Beschreibungen für Bilder und die Generierung von Untertiteln für Videos ist möglich.

Dabei ist es notwendig, vorhandene KI-Werkzeuge auf die archivischen Anforderungen anzupassen. Die Archive stehen hier noch am Anfang. Über aktuelle Entwicklungen berichtet das FDMLab in seinem Projektblog.

✱ Elisabeth Klindworth, Benjamin Rosemann