

Beitragsvorschlag zur 22. Tagung des Ak "AUdS" 2018 in Marburg

REFERENT: Stephan Lenartz, M.A. – Magisterstudium der Geschichte und Germanistik. Anschließendes Masterstudium der Konservierung Neuer Medien und digitaler Information.

BEITRAGSTITEL: Möglichkeiten der automatisierten Aufbereitung und Bewertung von Filablagen mit Python am Beispiel einer digitalen Fotosammlung.

KURZFASSUNG

Der Beitrag basiert auf einem Projekt, das ich anlässlich meiner Masterarbeit im Studiengang Konservierung Neuer Medien und Digitaler Information an der Kunstakademie Stuttgart in Kooperation mit dem Landesarchiv Baden-Württemberg – beteiligt waren die Staatsarchive Ludwigsburg und Sigmaringen – von Mai bis September 2017 durchgeführt habe. Ziel war die Entwicklung eines Workflows zur Aufbereitung und Bewertung des digitalen Nachlasses des Amateurfotografen Gunter Aipperspach. Die Datensammlung umfasste bei einem Datenvolumen von rund 450 GB über 200.000 Bilddateien im Format JPG. Einerseits sollte die von Herrn Aipperspach vorstrukturierte Sammlung für den Ingest aufbereitet werden. Dabei mussten Verzeichnisnamen korrigiert, die Verzeichnistiefe auf eine bestimmte Ebene reduziert und Dateien mitunter verschoben und umbenannt werden. Zudem sollte die Sammlung auf rund 10 % ihrer ursprünglichen Größe reduziert werden. Da eine intellektuelle Bewertung angesichts des Sammlungsumfangs nicht in Betracht kam, war ein geeigneter technischer Ansatz zu entwickeln. Letztlich wurden beinahe sämtliche Bearbeitungsschritte in der Skriptsprache Python realisiert, wozu insgesamt 16 Skripte geschrieben und angewendet wurden. Zwei Grundprinzipien wurden hierbei berücksichtigt: 1. sollten alle ausgeführten Änderungen in CSV-Files geloggt werden. 2. sollten sämtliche vollzogenen Schritte revidierbar sein. Zur Verdichtung der Sammlung kam einerseits ein randomisiertes Zufallssampling zum Einsatz, außerdem wurden unscharfe Fotografien kassiert. Darüber hinaus wurde mittels einer Ähnlichkeitssuche Bilder einer bestimmten Kategorie, die zuvor allgemein als archivwürdig bewertet wurde, über sämtliche Verzeichnisse hinweg gefunden. Der Beitrag zeigt exemplarisch, wie im praktischen Umgang mit Datensammlungen die Anwendung bekannter Tools sinnvoll durch den Einsatz einer vergleichsweise leicht zu erlernenden Skriptsprache ergänzt und damit die Effizienz der Bearbeitung maßgeblich gesteigert werden kann.