

Schopenhauer digital

Crowdsourcing-basierte Transkriptionsplattform zu Schopenhauers handschriftlichem Nachlass

Die Universitätsbibliothek JCS digitalisiert seit etwa 25 Jahren ihren Altbestand und präsentiert diesen in ihren *Digitalen Sammlungen*. Mittlerweile haben sich etwa 8 Millionen digitalisierte Seiten angesammelt. Um die Zugänglichkeit und Nutzbarkeit der Digitalisate zu optimieren, werden Drucke einer automatischen Texterkennung (Optical Character Recognition – OCR) unterzogen. Die damit hinterlegten Texte erlauben die Volltextsuche und damit zielgerichtete und intensive Recherchen. Das Verfahren, das bei Drucken

schon seit vielen Jahren zuverlässig funktioniert, lässt sich noch nicht adäquat auf handschriftliche Materialien übertragen. Zwar existieren mittlerweile äußerst leistungsfähige Tools, die sehr gute Ergebnisse bei der automatischen Texterkennung von Handschriften (Handwritten Text Recognition – HTR) leisten. Ohne intensive Vor- und Nacharbeiten sind die Resultate jedoch (noch) nicht zufriedenstellend. Um diese schiere Masse an Autographen im Bestand dennoch mit guten Volltexten versehen zu können, beabsichtigt die UB JCS, mittels Crowdsourcing eine engagierte Community zu aktivieren, die dabei hilft, Transkriptionen zu erzeugen, indem sie die Handschriften der Digitalisate in ein maschinenlesbares Format überträgt. Das Vorgehen, freiwillige Personen in die Bewältigung von Aufgaben einzubeziehen, zielt darauf ab, eine Gruppe von Individuen zu akquirieren, die sich aus intrinsischer Motivation als Crowdworker zur Ver-

fügung stellen und gemeinsam das Projekt des Crowdsourcers (in diesem Falle die UB) realisieren. Für die Pilotierungsphase wird der im Schopenhauer-Archiv bewahrte, vollständig digitalisierte Nachlass Arthur Schopenhauers herangezogen. Das Archivzentrum der UB beherbergt diesen sich über 25 Meter erstreckenden Bestand, der eine große Anzahl an heterogenen Materialien enthält. So lassen sich unter der Nachlass-Signatur Na 50 nicht nur Briefe von und an Arthur Schopenhauer, seine persönlichen Dokumente sowie die seiner Familienmitglieder oder Manuskripte, Tagebücher, Reiseberichte und Fotografien finden, sondern auch die mehrere Hundert Bände umfassende Privatbibliothek sowie persönliche Gebrauchsgegenstände und eine Porträtsammlung. Im Rahmen des Transkriptionsprojekts können interessierte Freiwillige aus über 1000 handschriftlichen Dokumenten wählen und in einem dafür vorgesehenen Editor Übertragungen der handschriftlichen Seiten vornehmen. Die Übertragung erfolgt in der vereinfachten Auszeichnungssprache Markdown. Dieses in vielen Bereichen etablierte Format hat den Vorteil, einerseits für menschliche Betrachter*innen (noch) gut lesbar zu sein, andererseits aber auch computergestützt verarbeitet werden zu können. In einem redaktionellen Workflow werden die eingereichten Transkriptionen geprüft und freigegeben. Die Redaktion, die neben



Mitarbeiter*innen der UB vor allem aus einer Gruppe von internationalen Schopenhauer-Forschern besteht, prüft, ob der erstellte Text den geforderten Qualitätsansprüchen entspricht, korrigiert ggf. und gibt die Transkription frei. Damit ist nicht nur eine Volltextsuche über die digitalisierten Dokumente möglich: Sukzessive werden die veröffentlichten Transkriptionen semi-automatisiert in das Format TEI/XML, einen Defacto-Standard für die digitale Aufbereitung und Archivierung von Textdaten in den Geisteswissenschaften, überführt und im gerade im Aufbau befindlichen Forschungsdatenrepositorium der GU gesichert. Auf diese Art und Weise werden der Forschung qualitativ hochwertige Daten für vielfältige Nutzungsszenarien zur Verfügung gestellt und langfristig vorgehalten.

Agnes Brauer, Michelle Kamolz



Transkriptionsumgebung der Digitalen Sammlungen der UB.

Weitere Informationen und Möglichkeit zur Registrierung und Partizipation:

<https://sammlungen.ub.uni-frankfurt.de/schopenhauer/transcript>