

Sachthematische Zugänge im Archivportal-D am Beispiel Weimarer Republik

Meyer, Nils

nils.meyer@la-bw.de

Landesarchiv Baden-Württemberg, Deutschland

Ein neuer Researchweg

Das Archivportal-D (www.archivportal-d.de) ist ein Subportal der Deutschen Digitalen Bibliothek (DDB, <https://www.deutsche-digitale-bibliothek.de/>), das im Rahmen eines DFG-Projekts speziell für die Recherche nach Archiven und Archivgut entwickelt wurde. Derzeit sind im Archivportal zwei Nutzungsszenarien möglich: Nutzende können entweder nach einzelnen Archiven suchen und davon ausgehend durch die Bestände des jeweils ausgewählten Archivs browsen oder sie können mit der Freitextsuche nach einzelnen Objekten suchen. In beiden Fällen stehen verschiedene Facetten zur Einschränkung der Ergebnisse bereit.

In dem aktuellen, ebenfalls von der DFG geförderten Projekt sollen diese Nutzungsmöglichkeiten um einen dritten Zugang erweitert werden. Am Beispiel zweier Digitalisierungsprojekte des Landesarchivs Baden-Württemberg und des Bundesarchivs zum Thema „Weimarer Republik“ soll ein neuer, themenbasierter Zugang zu den Objekten im Archivportal entwickelt werden. Im Zentrum stehen dabei die Anreicherung der Erschließungsdaten des Archivguts mit sachthematischen Begriffen und die Nutzbarmachung der angereicherten Informationen für ein intuitives, forschungsgerechtes Angebot. Das Projekt läuft über zwei Jahre seit dem 1. Juni 2018. Projektpartner sind neben Bundesarchiv und Landesarchiv Baden-Württemberg die Deutsche Nationalbibliothek als Betreiberin der DDB sowie FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur.

Entwicklung von Systematik und Vokabularen

In einem ersten Schritt soll gemeinsam mit Vertreterinnen und Vertretern der Archivcommunity, der Geschichtswissenschaft und der Digital Humanities eine an den Bedürfnissen der Forschung ausgerichtete sachthematische Systematik zur Weimarer Republik erstellt werden, die sowohl als Sucheinstieg als auch als hierarchische Facette bei der Filterung der Suchergebnisse dienen kann. Den einzelnen Objekten werden dabei kontrollierte Indexbegriffe aus dem Themenbereich

„Weimarer Republik“ zugewiesen, welche wiederum mit den Klassen der Systematik verknüpft sind. Für den Aufbau einer nutzerorientierten Systematik werden verschiedene wissenschaftliche und nicht-wissenschaftliche Systematiken als Grundlage genommen. Das kontrollierte Vokabular der Indexbegriffe wird aus einer umfangreichen Wortgutsammlung ausgehend von den vorhandenen Erschließungsdaten sowie externen Quellen (z.B. Normdateien, Wikidata) gespeist. Daneben soll ein kontrolliertes Vokabular zu den einzelnen Staaten der Weimarer Republik eine grundlegende geografische Suchmöglichkeit schaffen. Indexbegriffe und Systematik sollen dabei mit Wikidata-Datenobjekten und Normdateien verknüpft werden. Gleichzeitig soll auch die normdatenbasierte Indexierung bekannter Personen bei den einzelnen Archivobjekten verbessert werden.

Tools und algorithmische Zuordnung

Ein Schwerpunkt des Projekts und zentrale Forschungsaufgabe neben der Entwicklung von Systematik und Vokabularen ist die Entwicklung von Tools zur Zuordnung der einzelnen Archivobjekte zu den verschiedenen Vokabularen. Ziel der Arbeit soll ein komfortables Zuordnungswerkzeug zur einfachen Verknüpfung von ganzen Beständen sowie auch einzelnen Archivalien mit den verschiedenen sachthematischen, geografischen und Personen-Indexbegriffen sein. Dieses Tool soll zunächst anhand der Weimar-Bestände des Bundesarchivs und des Landesarchivs Baden-Württemberg entwickelt und getestet werden. In einem späteren Schritt soll das Tool auch dem Personal in anderen Archiven sowie den Nutzenden zur Verfügung gestellt werden. Auf diese Weise können die Nutzenden kollaborativ an der Anreicherung der Erschließungsdaten und der Erweiterung des Themenzugangs mitwirken. Zugleich wird die bisher nur wenig genutzte Möglichkeit der Sachindexierung in den Archiven gestärkt.

Dem Zuordnungstool zur Seite gestellt werden soll ein Werkzeug zur algorithmischen Zuordnung von Archivobjekten zu den verschiedenen Indexbegriffen. Über Maschinenlernprozesse sollen die einzelnen Objekte beziehungsweise die Objektmetadaten und ihr jeweiliger Kontext analysiert werden. Dabei werden verschiedene externe Linked-Data-Ressourcen abgefragt und in die Analyse mit einbezogen. Daraus werden im Ergebnis Vorschläge für die Zuordnung der einzelnen Archivobjekte zu den verschiedenen Vokabularen generiert. Die Vorschläge dienen der Unterstützung des Fachpersonals und der Nutzer bei der Auswahl der passenden Indexbegriffe und können von diesen mit dem Zuordnungstool umgesetzt werden. Der hohe Anteil unstrukturierter Daten aufgrund der Erschließung in verschiedenen Freitextfeldern in den Archiven stellt die Arbeit in diesem Arbeitspaket vor besondere

Herausforderungen. Unterschiedliche Ansätze (Improved Fisher Vector, Convolutional Neural Networks) sollen daher auf ihre Eignung für archivische Erschließungsdaten erprobt werden.

Grundlagenentwicklung für Archivportal und DDB

Um die sachthematische Systematik adäquat abbilden zu können, sollen die bisherigen Facetten des Archivportals und der DDB um eine hierarchische Dimension erweitert werden. Daneben werden die neuen sachthematischen, geografischen und personenbezogenen Metadaten der einzelnen Archivobjekte in einem eigenen, neu zu entwickelnden Repositorium unabhängig von den Ursprungsdaten des jeweiligen Archivs gesichert und sollen über eine Exportfunktion auch an das datengebende Archiv zurückgespielt werden können. Hierfür sollen umfangreichere Entwicklungsarbeiten im gemeinsamen Backend von DDB und Archivportal angestoßen werden.

Mit der Grundlagenarbeit aus dem Projekt sollen später mit geringem Aufwand auch Themenzugänge zu anderen Bereichen (Nationalsozialismus, Erster Weltkrieg, verschiedene Archivaliengattungen) möglich sein. Über das Frontend des Archivportals und der DDB sind diese unmittelbar nutzbar, über das API der DDB können die angereicherten Metadaten auch in anderen Kontexten weitergenutzt werden.

Bibliographie

Becker, Irmgard Christa / Maier, Gerald / Uhde, Karsten / Wolf, Christina (eds.) (2016): *Netzwerken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung.* Beiträge zum 19. Archivwissenschaftlichen Kolloquium (= Veröffentlichungen der Archivschule Marburg 61), Marburg: Archivschule Marburg.

Fähle, Daniel / Maier, Gerald / Schröter-Karin, Tobias / Wolf, Christina (2015): *Archivportal-D. Funktionalität, Entwicklungsperspektiven und Beteiligungsmöglichkeiten*, in: *Archivar* 68: 10-19.

Hentschel, Christian / Wiradarma, Timur Pratama / Sack, Harald (2015): *If we did not have imagenet.* Comparison of fisher encodings and convolutional neural networks on limited training data, in: *Advances in Visual Computing* (= Lecture Notes in Computer Science 9475) 400-409.

Herrmann, Tobias / Zahnhausen, Vera (2018): *Auf dem Weg zum Digitalen Lesesaal.* Das Projekt "Weimar – Die erste deutsche Demokratie", in: *Kompetent! Archive in der Wissensgesellschaft*. 86. Deutscher Archivtag 2016 in Koblenz (= Tagungsdokumentationen zum Deutschen Archivtag 21), Fulda: Verband deutscher Archivarinnen und Archivare [Veröffentlichung Herbst 2018].

Hollmann, Michael (2016): *Deutschland in zwei Nachkriegszeiten.* Der Einstieg in das Online-Archiv des Bundesarchivs, in: *Archivar* 69, 6-9.

Laux, Susanne / Wolf, Christina (2016): *Forschungsprojekt "Von der Monarchie zur Republik" gestartet.* Digitalisierung von Quellen zur Demokratiegeschichte im deutschen Südwesten 1918-1923, in: *Archivnachrichten* 52, 28-29.

Lindenthal, Jutta (2016): *Datenqualität und Retrieval.* Vorschläge zur Verbesserung der Suche in der Deutschen Digitalen Bibliothek http://jl.balilabs.de/DDB/DQ/DDB_Datenqualit%C3%A4t_Retrieval_1.0.pdf [letzter Zugriff 24. September 2018].

Reisacher, Martin / Krauth, Wolfgang (2015): *Vernetzen als Herausforderung.* Die Deutsche Digitale Bibliothek, in: *Archivnachrichten* 51, 36 f.

Rühle, Stefanie / Schulze, Francesca / Büchner, Michael (2014): *Applying a Linked Data Compliant Model.* The Usage of the Europeana Data Model by the Deutsche Digitale Bibliothek <http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/231/225> [letzter Zugriff 24. September 2018].

Sack, Harald (2014): *Linked Data Technologien.* Ein Überblick, in: **Pellegrini, Tassilo / Sack, Harald / Auer, Sören (eds.):** *Linked Enterprise Data.* Management und Bewirtschaftung vernetzter Unternehmensdaten mit Semantic Web Technologien. Berlin: Springer 21-62.

Wolf, Christina (2016): *Eines für alle: Das Archivportal-D.* Neue Zugangswege zu Archivgut, in: *Neue Wege ins Archiv. Nutzer, Nutzen, Nutzung.* 84. Deutscher Archivtag 2014 in Magdeburg (= Tagungsdokumentationen zum Deutschen Archivtag 19). Fulda: Verband deutscher Archivarinnen und Archivare 47#63.