

Absichtserklärung

Im Rahmen des Förderprogramms „e-Research-Technologien“ beabsichtigt die Thüringer Universitäts- und Landesbibliothek (ThULB)/Friedrich-Schiller-Universität Jena¹ gemeinsam mit der Verbundzentrale des Gemeinsamen Bibliotheksverbundes GBV (VZG)² die „Implementierung der OCR-D-Software zur Volltextdigitalisierung historischer Drucke“ in das Open-Source-Framework „My Content Repository“ (MyCoRe)³.

Projektziel

Ziel des Kooperationsvorhabens ist es, die Software OCR-D in einer **Pilotphase** in Anwendungen auf Basis von MyCoRe (z.B. Dokumenten- und Publikationsserver, Archivanwendungen, Sammlungen von Digitalisaten oder vergleichbare Repositorien) zu implementieren. Dazu soll das OCR-D-Modul der MyCoRe-Community insgesamt zur Systemintegration zur Verfügung gestellt werden. Die MyCoRe-Infrastruktur erlaubt es, das Modul sofort in alle auf MyCoRe basierenden Systeme zu integrieren und anzuwenden.

ThULB/FSU, VZG und die MyCoRe-Community bündeln dabei das entsprechende technische und bibliothekarische Know-How, um das Implementierungsprojekt im beabsichtigten Zeitraum in einer Pilotphase umzusetzen und entsprechende Anwendungskontexte zu generieren. Dazu können bereits jetzt digitalisierte Bestände aus MyCoRe-Repositorien (neben Dokumenten- und Publikationsservern auch Forschungs-, Digitalisierungs- und Archivanwendungen, Zeitschriftenserver, biografische Informationssysteme und Bilddatenbanken) zur Verfügung gestellt werden.⁴

Kooperationspartner und Vorarbeiten

Die Thüringer Universitäts- und Landesbibliothek Jena (ThULB) ist Hochschulbibliothek der Friedrich-Schiller-Universität Jena und als solche auch Landesbibliothek. Mit einem umfassenden Portfolio an Medien und Informationsdienstleistungen ist die ThULB die größte wissenschaftliche Bibliothek im Freistaat Thüringen. Die Aufbereitung und die Online-Bereitstellung der kulturellen Überlieferung Thüringens zählt zu den zentralen Aufgaben der ThULB.

Die Verbundzentrale des GBV (VZG) ist Katalogisierungs- und Dienstleistungszentrum für wissenschaftliche und öffentliche Bibliotheken sowie für Forschungseinrichtungen im Bereich des Gemeinsamen Bibliotheksverbunds der Bundesländer Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und der Stiftung Preußischer Kulturbesitz (SPK). Sitz der Verbundzentrale des GBV (VZG) ist Göttingen. Sie hat die Aufgabe einen Rahmen für eine abgestimmte Bibliotheksautomation zu schaffen, neuartige Bibliotheks- und Informationsdienstleistungen zu entwickeln und deren Anwendung zu fördern.

¹ www.uni-jena.de bzw. www.thulb.uni-jena.de

² www.gbv.de/Verbundzentrale/

³ mycore.de

⁴ mycore.de/site/applications/list

Das Open-Source-Projekt MyCoRe, hervorgegangen aus dem MILESS-Projekt der Universität Duisburg-Essen, ist ein Framework zur Entwicklung von Dokumenten- und Publikationsservern, Archivanwendungen, Sammlungen von Digitalisaten oder vergleichbaren Repositorien. Auf MyCoRe basierende Anwendungen nutzen einen gemeinsamen Kern, der die typischerweise für solche Anwendungen erforderliche Funktionalität, wie Metadatenverwaltung, Suchfunktionen, OAI-Schnittstelle sowie ein Bildbetrachtungsmodul, bereitstellt.⁵ In der Community organisieren sich Entwickler/-innen verschiedener deutscher Universitäten und bibliothekarischer Einrichtungen mit dem Ziel der gemeinsamen Weiterentwicklung des Frameworks. ThULB und VZG stellen auf Basis von MyCoRe eine Vielzahl zentraler, zum Teil gemeinschaftlich entwickelter und betriebener Dienste für die Bibliothekscommunity insgesamt zur Verfügung.

Anwendungsszenario

In der **Projektphase** soll, in Abstimmung mit der Niedersächsischen Staats- und Universitätsbibliothek (SUB) Göttingen als Koordinierungsstelle des von der DFG geförderten Vorhabens zur kooperativen Digitalisierung und Erschließung der im deutschen Sprachraum erschienenen Drucke des 18. Jahrhunderts (**VD18**), speziell für diese (digitalisiert vorliegenden) Bestände ein OCR-Dienst zur Verfügung gestellt werden. Die VZG ist der nationale Host für die Koordinierung der VD18-Digitalisierung, auf dieser Grundlage soll das geplante VD18-Portal gemeinsam von VZG und SUB Göttingen entwickelt werden.⁶ Sämtliche VD18-Bibliotheken könnten Digitalisate in einer MyCoRe-Umgebung verwalten, normieren, anreichern und präsentieren. Entsprechende Geschäftsgänge und Datenmanagementpläne liegen vor und ermöglichen eine rasche Umsetzung und Produktivphase für ein Digital Asset Management auf MyCoRe-Basis.

Die durch Anwendung der OCR-D-Software zur Verfügung gestellten Daten sollen im Rahmen der Projektphase zusätzlich mit **Normdaten** angereichert werden. Dazu kann seitens des GBV als zentraler Dienst die „DATendrehscheibe für Normdaten und Terminologien – DANTE“ (dante.gbv.de) zur Verfügung gestellt werden.⁷ Die ThULB ist an der Entwicklung der „Toolbox“ der Gemeinsamen Normdatei (GND) beteiligt und wird in Fortführung des DFG-geförderten Projektes „GND für Kulturdaten – GND4C“, ein Teilprojekt des DFG-geförderten „GND-Entwicklungsprogramms 2017–2021“, Mittragsteller in der zweiten Projektphase.⁸ Zudem kann auf weitere kontrollierte Vokabulare und wissenschaftliche Thesauri nationaler und internationaler kulturtragender Einrichtungen zugegriffen werden.

Neben der Bereitstellung eines verlässlichen Dienstes zur OCR-Erkennung sollen die OCR-Bibliotheken von OCR-D durch das Projekt deutlich erweitert und durch Anbindung von Erschließungsnormdatenquellen qualitativ ergänzt werden.

⁵ mycore.de/site/features/

⁶ <https://gso.gbv.de/DB=1.65/>

⁷ dante.gbv.de

⁸ wiki.dnb.de/display/GND/GND+Wiki

In Anlehnung an die Personalmittelsätze der DFG für das Jahr 2020 sollen zur Implementierung eines OCR-D-Workflows für die Text- und Strukturerkennung zur Volltextdigitalisierung historischer Drucke in Anwendungen und Infrastrukturen der MyCoRe-Community zwei Stellen (100%, E13, St. 2) für 24 Monate beantragt werden.

Ansprechpartner

Michael Lörzer
Direktor der Thüringer Universitäts-
und Landesbibliothek (ThULB) Jena
michael.loerzer@uni-jena.de

Reiner Diedrichs
Direktor der Verbundzentrale des GBV
(VZG)
reiner.diedrichs@gbv.de