

Automatisierte RVK-Sacherschließung mit Annif in der Ausbildung des wissenschaftlichen Bibliotheksdienstes

Ein Praxisprojekt mit der Hessischen Bibliographie

1. Einleitung

Die Ausbildung von Referendar*innen und/oder Volontär*innen in Wissenschaftlichen Bibliotheken zeichnet sich durch die enge Verzahnung von Theorie und Praxis aus. Durch den vielschichtigen Einblick in alltägliche Arbeiten bzw. Workflows sowie die weitergegebenen Erfahrungen werden zukünftige Kolleg*innen auf Fach- und Führungspositionen vorbereitet. Doch wie bilden Wissenschaftliche Bibliotheken in Bereichen aus, zu denen wenig oder noch gar kein Know-how und Kompetenzen in den Häusern vorhanden sind und die zudem auch in der begleitenden theoretischen Ausbildung – z. B. am Institut für Bibliotheks- und Informationswissenschaft in Berlin (IBI) oder der Bibliotheksakademie Bayern – nicht vertiefend oder noch nicht behandelt werden?

Hinsichtlich der automatisierten Sacherschließung (AutoSE) – ein bisher wenig bearbeitetes Feld an der Universitätsbibliothek Johann Christian Senckenberg (UB JCS) – liegt in der theoretischen Ausbildung (am IBI) der Fokus primär auf den intellektuellen Aspekten der Sacherschließung und den theoretischen Grundlagen des automatischen Indexierens und Data Mining.

Mit Blick auf die steigende Bedeutung von digitalen Kompetenzen¹ für wissenschaftsnahe Dienste sowie die am Puls der Zeit liegende Entwicklung maschinellen Lernens und KI-Anwendungen haben wir deshalb im Sommer 2025 im Sachgebiet Sacherschließung & Bibliographien ein Hands-on-Projekt zum Einsatz von Annif² zur RVK-Sacherschließung am Beispiel der Hessischen Bibliographie aufgesetzt. Bei Annif handelt es sich um ein Open-Source-Werkzeug zur automatisierten Schlagwortvergabe (Subject Indexing), das von der Finnischen Nationalbibliothek entwickelt wurde. An der Projektplanung waren Ausbilder*innen, Bibliotheks-IT (IT) und eine Referendarin beteiligt, wobei mit Projektstart die Federführung und operative Verantwortung auf die Referendarin Anna-Maria Hünnes überging. Neben dem Ausloten der Grenzen und Herausforderungen in der praktischen Umsetzung von Tools (auf der Kommandozeile) war es explizit Teil der Zielsetzung, einen gemeinsamen Wissens- und Kompetenzstand zwischen fachlichen Ausbilder*innen und Referendarin zu erreichen.

- 1 Siehe dazu auch Berichte von ähnlich gelagerten Annif-Projekten im Kontext der (teil-)automatisierten Sacherschließung: Seifert, Wolfgang; Schneider, Sophie: Anforderungsanalyse für die semi-automatisierte Sacherschließung an der Staatsbibliothek zu Berlin, in: *O-bib*, 11(4), 2024, S. 1–7. <https://doi.org/10.5282/o-bib/6075> sowie Engel, Alexandra; Hafner, Ralph: KI-gestützte Automatisierung der Sacherschließung mit Annif. ein Pilotprojekt, in: *KIM kompakt* 109, 2025, S. 53–56. <https://ojs.ub.uni-konstanz.de/kimkompakt/article/view/8609>. Ebenso startete 11/2025 ein Verbundprojekt „KI gestütztes Mappen und Generieren von Sacherschließung und Formalerschließung“ (KISEFEE), <https://www.kim.uni-konstanz.de/das-kim/ueber-das-kim/projekte/aktuelle-projekte/kisefee/>, Stand 08.05.2026.
- 2 Annif, <https://annif.org/>, Stand 08.05.2026.

2. Sacherschließung & Bibliographien an der Universitätsbibliothek Johann Christian Senckenberg

Die UB JCS³ gliedert sich in fünf Abteilungen⁴, vier Stabsstellen sowie die Verbundzentrale des Hessischen Bibliotheksinformationssystems hebis (hebis VZ). Das Sachgebiet *Sacherschließung & Bibliographien* ist als Teil der Abteilung *Kuratieren, Fachinformation & Vermittlung* (KFV) an der Zentralbibliothek angesiedelt. Das Sachgebiet betreut die verbale und klassifikatorische Inhalterschließung der sechs Fachinformationsdienste⁵ sowie drei von fünf Bibliographien der UB JCS.⁶ In der Abteilung *Lernorte & Wissenschaftsunterstützung* (LWU) wird ebenfalls klassifikatorische Sacherschließung geleistet. In der Abteilung LWU sind die Bereichsbibliotheken⁷ der UB JCS subsumiert, in denen nach der Regensburger Verbundklassifikation (RVK) erschlossen bzw. aufgestellt wird.

Bis 1995 erfolgte die klassifikatorische Inhalterschließung an der UB JCS – zu diesem Zeitpunkt noch die Stadt- und Universitätsbibliothek Frankfurt am Main – nach der Eppelsheimer Systematik⁸. Diese wird bis heute für die zwei regionalen Fachinformationsdienste (Afrikastudien und Jüdische Studien) der UB JCS zur Facettenbildung der jeweiligen Portale angewendet. Nach der Beendigung der klassifikatorischen Inhalterschließung wird an der UB JCS seit 1996 die inhaltliche Erschließung verbal durch die Schlagwortkatalogisierung (RSWK) vorgenommen. In den drei vom Sachgebiet *Sacherschließung & Bibliographien* betreuten Bibliographien – die Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL), die Bibliography of Linguistic Literature (BLL) sowie die Hessische Bibliographie – erfolgt die Inhalterschließung sowohl verbal als auch klassifikatorisch. Im Fachinformationsdienst Biodiversitätsforschung (biofid) erfolgt die klassifikatorische Erschließung mit einer eigens angepassten Variante (bioDDC) der DDC.⁹

Die folgende Auflistung gibt an, auf Basis welcher Systematiken bzw. Systeme die verbale und klassifikatorische Sacherschließung an der UB JCS erfolgt:

- 3 Organigramm Universitätsbibliothek J. C. Senckenberg, https://www.ub.uni-frankfurt.de/ueber/dateien/Organigramm_UB_Frankfurt.pdf, Stand 08.05.2026.
- 4 Information, Bereitstellung & Originalerhalt (IBO); Lernorte & Wissenschaftsunterstützung (LWU); IT-Services (ITS); Kuratieren, Fachinformation & Vermittlung (KFV); Erwerbung, Lizenzen & Metadaten (ELM) und hebis-Verbundzentrale (hebis VZ).
- 5 FID Afrikastudien; FID Allgemeine und Vergleichende Literaturwissenschaft; FID Biodiversitätsforschung; FID Darstellende Kunst; FID Jüdische Studien und FID Linguistik.
- 6 Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL); Bibliography of Linguistic Literature (BLL) und Hessische Bibliographie. Ergänzend kommen die Bibliographie Kinder- und Jugendliteraturforschung sowie die Hochschulbibliographie hinzu, die beide in der Abteilung LWU betreut werden.
- 7 Standorte der Universitätsbibliothek J. C. Senckenberg, <https://www.ub.uni-frankfurt.de/standorte.html>, Stand 08.05.2026.
- 8 Riplinger, Thomas: Die Bedeutung der Methode Eppelsheimer für Theorie und Praxis der bibliothekarischen und der dokumentarischen Sacherschließung, in: BIBLIOTHEK – Forschung und Praxis, 28 (2), 2004, S. 252–262.
- 9 bioDDC, https://www.dnb.de/DE/Professionell/DDC-Deutsch/DDCausserhalbDNB/_content/ddcBioDDC_akk.html, Stand 08.05.2026.

Tab. 1: Aufstellung der verbalen und klassifikatorischen Sacherschließung an der UB JCS

Verbale Inhaltsschließung	Klassifikatorische Inhaltsschließung
Fachinformationsdienste <ul style="list-style-type: none"> • Afrikastudien • Allgemeine und Vergleichende Literaturwissenschaft (AVL) • Darstellende Kunst • Jüdische Studien • Linguistik 	Fachinformationsdienste <ul style="list-style-type: none"> • Biodiversitätsforschung: bioDDC • Afrikastudien: Eppelsheimer Systematik • Jüdische Studien: Eppelsheimer Systematik
Bibliographien <ul style="list-style-type: none"> • Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL) • Bibliography of Linguistic Literature (BLL) • Hessische Bibliographie 	Bibliographien <ul style="list-style-type: none"> • Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL): eigene Klassifikation • Bibliography of Linguistic Literature (BLL): eigener Thesaurus • Hessische Bibliographie: RVK

3. Die Hessische Bibliographie

Grundlage für das Training von Annif mit der RVK waren die Daten der Hessischen Bibliographie. Für die Erstellung der Hessischen Bibliographie kooperieren die vier hessischen Landesbibliotheken, die Universitätsbibliothek Kassel sowie das Hessische Institut für Landesgeschichte in Marburg. Die hebis VZ übernimmt sowohl die Zentralredaktion als auch die technische Betreuung der Hessischen Bibliographie.

Für die Hessische Bibliographie werden systematisch Zeitschriften, E-Books, Sammelbände, graue Literatur und Publikationen mit Hessenbezug ausgewertet. 2004 wurde die Hessische Bibliographie in das Hessische Bibliotheksinformationssystem (hebis) integriert. Die Daten werden direkt im Verbund erfasst und in einem eigenen Portal für die Recherche bereitgestellt.¹⁰ Seitdem stehen Nutzenden der Bibliographie sowohl die aktuellen Bestandsinformationen als auch die Bestellmöglichkeiten im hebis-Verbund zur Verfügung.

Für die Sacherschließung der Hessischen Bibliographie wird sowohl die verbale Sacherschließung mit der Vergabe von GND-Normdaten als auch die klassifikatorische Sacherschließung nach RVK¹¹ praktiziert.¹² Für die GND-Normdaten gelten folgende Regelungen: Es werden nur Einzelschlagworte vergeben, keine Schlagwortketten. Zudem muss mindestens ein Geographikum vergeben werden.

10 Hessische Bibliographie, <http://cbsopac.rz.uni-frankfurt.de/LNG=DU/DB=2.4/>, Stand 08.05.2026.

11 Im April 2013 wurde in der AG Sacherschließung des hebis-Verbundes beschlossen, dass parallel zu den HessBib-Notationen auch die nach RVK-Notationen vergeben werden sollen. (Siehe dazu Protokoll der AG Sacherschließung, https://www.hebis.de/uploads/2020/06/FAG_Sacherschlie%C3%9Fung_und_Recherche_2012_11_08-1.pdf, Stand 08.05.2026.) Diese Übergangslösung wurde im Jahr 2015 mit der Beendigung der Klassifikation nach der originären Systematik (HessBib-Notation) eingestellt und danach ausschließlich mit der Vergabe von RVK-Notationen gearbeitet (Protokoll der Jahressitzung der Arbeitsgemeinschaft Hessische Bibliographie am 04.09.2015).

12 Ab 1. März 2016 wurden in allen Katalogisaten der Hessischen Bibliographie neben den RVK-Notationen nur noch GND-Schlagwörter vergeben bzw. verknüpft. Weiterhin wurden keine Notationen aus der HessBib-Systematik mehr vergeben. Verpflichtend für die Sacherschließung der Hessischen Bibliographie sind seitdem mindestens eine passende RVK-Notation sowie mindestens ein passendes Orts- oder Regionenschlagwort.

Im vorgestellten Annif-Projekt wurden ausschließlich Titeldaten aus der Hessischen Bibliographie als Grundlage der automatischen Sacherschließung verwendet. Das Training mit anderen bibliographischen Daten oder die Erweiterung auf andere inhaltliche Metadatenfelder (z. B. Abstracts) wurde aufgrund der zur Verfügung stehenden Rechenleistung sowie der zeitlichen Dimension des Projekts nicht systematisch erprobt.

4. Projektplanung

Die Ausbildung in der Abteilung KFV und die Arbeit an dem Hands-on-Projekt erfolgten jeweils halbtags. Vormittags wurden Einblicke in die verschiedenen Sachgebiete (SGs)¹³ der Abteilung ermöglicht. Dabei wurden auch die Arbeitsprozesse der Sacherschließung an der UB JCS demonstriert und erläutert. Die Arbeit mit den unterschiedlichen Klassifikationen, die für die Sammlungsschwerpunkte Verwendung finden, ist dabei anspruchsvoll.

Der zweite Teil des Tages war in der Regel dem Annif-Projekt gewidmet. Die Arbeitspakete des Projektplans (Tabelle 2, im Anhang) wurden auf Basis mehrerer Vorgespräche zwischen Ausbilder*innen, der Referendarin sowie der IT – mit einem Gesamtarbeitsaufwand von ca. 3 Arbeitstagen – zusammengestellt. Die genaue zeitliche Einteilung (insbesondere der ersten Arbeitspakete zur Skript- und Programmiersprache Python, der Arbeit mit der Linux-Kommandozeile oder der Versionsverwaltung mit Git) wurden auf die individuellen Vorkenntnisse der Referendarin abgestimmt.

In der Planung des Projekts mussten vorab gewisse Rahmenbedingungen zwischen den verantwortlichen Stellen bzw. Abteilungen abgestimmt werden. Zwischen IT und den Ausbilder*innen waren insbesondere Fragen der IT-Sicherheit und Verantwortlichkeiten sowie Fragen zum Support und Eskalationswege zu klären. Hier haben die Kolleg*innen aus der IT geholfen, die Virtuelle Maschine (VM) (vgl. 4.1. Technik) in die Infrastruktur der UB einzugliedern und zeitgemäß gegen Angriffe und Sicherheitslücken zu sichern.

4.1 Technik

Für das Projekt wurde ein Linux-Server (Ubuntu 24.04.3 LTS) in der KVM/QEMU-basierten Virtualisierungsumgebung des Rechenzentrums der Goethe-Universität verwendet. Es standen vier virtuelle CPUs mit einer Taktfrequenz von 2.7 GHz, 16 GB Arbeitsspeicher und 340 GiB Festplattenspeicher zur Verfügung. Mit Hinblick auf den Charakter des Projekts wurde aus Kostengründen auf eine dezidierte GPU verzichtet. Für das Training der Modelle wurden Python 3.12.3 und Annif 1.4 genutzt.

4.2 Vorbereitung und Datenaufbereitung

Zu Projektbeginn wurde von den Ausbilder*innen ein Austausch mit Expert*innen in kollegialen Netzwerken zum Thema automatisierte Erschließung organisiert. Thematisiert wurden sowohl die Eckpunkte dieses Ausbildungsprojekts als auch die jeweiligen Arbeitsschwerpunkte der Kolleg*innen.¹⁴

13 SG 1 – Sammlungskuratierung; SG 2 – Erschließung Spezialbestände und Antiquaria; SG 3 – Sacherschließung & Bibliographien; SG 4 – Forschungslesesaal und Digitale Kuratierungsservices; SG 5 – Sammlungskoordination der Goethe-Universität; SG 6 – Langzeitarchivierung und -verfügbarkeit.

Die Arbeit am Projekt begann daraufhin mit dem Durcharbeiten der Library Carpentry Lesson "The Unix Shell", da das Ausbildungsprojekt auf einem Linux Server durchgeführt wurde. Parallel dazu fand Literaturrecherche zur automatischen Sacherschließung statt. Für den Einstieg in die Arbeit mit Annif wurden, zusätzlich zur Bearbeitung des Annif-tutorials¹⁵, die Arbeitsschritte und Ergebnisse einer Masterarbeit von 2023 zur Automatisierten Vergabe von RVK-Notationen mit Annif nachvollzogen und reproduziert.¹⁶ In der Arbeit untersuchte Elisabeth Mecking die Vergabe von RVK-Notationen der Hauptgruppe S und verglich die Qualität der Ergebnisse nach Trainingsdurchläufen mit Titeln. Dabei hat sie festgestellt, dass das Training mit Volltexten im Vergleich zu Trainings mit Abstracts und/oder Inhaltsverzeichnissen wesentlich größeren Aufwand bedeutet, während die Ergebnisse dadurch nicht bedeutend verbessert werden. Mecking erläutert die einzelnen Schritte der Arbeit mit Annif sowie die notwendigen Vorarbeiten der Erstellung des Vokabulars (RVK Notationen der Hauptgruppe S) und der Zusammenstellung von Test- und Trainingsdaten (Abfragen aus dem K10plus).

Für das Projekt zur AutoSE für die Hessische Bibliographie wurde zunächst überlegt, wie ein passendes Vokabular für Annif aus der RVK erstellt werden könnte. Weil im Rahmen des Projekts nicht mit allen RVK-Notationen trainiert werden konnte und auch die Anzahl aller bislang in der Hessischen Bibliographie verwendeten Notationen (Stand 2025: 15.613 unterschiedliche Notationen) zu umfangreich für das Projekt war, fand eine andere Beschränkung statt: Es wurden lediglich Stellen gewählt, die explizit mit Hessen verbunden sind, schließlich sollen sie vorrangig bei Einträgen in der Hessischen Bibliographie vergeben werden; Grundlage für das Vokabular war der XML-Abzug der RVK.¹⁷ Die XML-Datei wurde auf ihre Struktur untersucht und schrittweise ein Python-Skript geschrieben, um aus dem Abzug eine TSV-Datei mit ausgewählten Notationen zu generieren. Mittels einer trunkierten Suche in dem XML-Abzug wurden alle Elemente ausgewählt, in deren Benennungs-Attribut die Werte „Hessen“, „hessisch“ oder entsprechende Namensvarianten vorkamen. Ebenfalls eingeschlossen wurden alle Stellen, bei denen diese Benennungen in übergeordneten Ebenen enthalten waren.¹⁸ Diese Einschränkung führte zu einer Grundgesamtheit von etwa 1.100 RVK Stellen, die für das Training von Interesse wären. Die ausgegebene TSV-Datei umfasste dabei drei Spalten. Zunächst die URL zur jeweiligen Stelle in der RVK online, dann die Benennung und schließlich die Notation. Im Laufe des Projekts zeigte sich, dass es sich anbot, wenn auch die Benennung der übergeordneten Knoten in die zweite Spalte übernommen wurde. Die aufgelöste Schreibweise der Hierarchie hatte dabei keinen Einfluss auf die Ergebnisse von Annif, sondern half später, die Vorschläge von Annif intellektuell zu disambiguieren.

Mittels Anfragen über die SRU-Schnittstelle der Hessischen Bibliographie¹⁹ wurde anschließend ermittelt, welche der identifizierten RVK-Notationen bisher verwendet wurden. Das Ergebnis zeigte, dass nur 464 der RVK-Notationen, die in ihrer Bezeichnung einen expliziten Bezug zu Hessen aufwie-

14 Hier gilt unser Dank Clara Wan Ching Ho (FID Linguistik – UB Frankfurt) sowie Christoph Poley und Sandro Uhlmann (DNB) für ihre Zeit und Expertise.

15 Suominen, Osmo; Lehtinen, Mona; Inkinen, Juho; Kasprzik, Argie; Fürneisen, Moritz: Annif-tutorial, <https://github.com/Nat-LibFi/Annif-tutorial?tab=readme-ov-file>, Stand 20.02.2026.

16 Mecking, Elisabeth: Automatisierte Vergabe von Notationen der Regensburger Verbundklassifikation RVK mithilfe des Toolkits Annif, 2023 (Masterarbeit, Technische Hochschule Wildau), https://doi.org/10.15771/MA_2023_1.

17 Es wurde der XML-Abzug 2025_2 genutzt, der bis September 2025 unter <https://rvk.uni-regensburg.de/regensburger-verbundklassifikation-online/rvk-download> zur Verfügung stand.

18 So wurde etwa „NZ 60020 - Mittelalter“ auch ausgewählt, da in dem übergeordneten Bereich „NZ 60000 - NZ 69999 Landesgeschichte Hessen“ Hessen als Bezeichnung vorkommt.

sen, bislang auch vergeben wurden. Ein weiterer einschränkender Faktor war, dass man eine gewisse Anzahl an Trainingsdaten für jede RVK-Notation benötigt, auf die man die Annif-Backends trainieren möchte.²⁰ Daher erfolgte über die SRU-Schnittstelle mittels des Python-Skripts auch eine Zählung der Vorkommen für jede der verwendeten RVK-Notationen mit Hessenbezug. Es blieben 132 RVK-Stellen übrig, die in der Hessischen Bibliographie bislang mindestens 50-mal verwendet wurden.²¹ Die so ausgewählten und formatierten Notationen bilden das Vokabular, das in Annif-Projekten hinterlegt werden kann. Die starke Reduzierung der RVK-Stellen kann grundsätzlich durch die Verwendung weiterer Quellen für Trainingsdaten vermieden werden. Dies könnte unter Umständen auch durch eine Konkordanz erreicht werden, die beispielsweise RVK-Notationen und GND-Schlagwörter in Beziehung setzt.

Anschließend wurden die Trainings- und Testdaten für das Projekt zusammengestellt. Dafür wurde die Hessische Bibliographie über die SRU-Schnittstelle nach den 132 RVK-Stellen abgefragt. Wenn eine der Stellen in einem Datensatz vorkam, wurden bestimmte Felder ausgelesen und in einer TSV-Ausgabedatei gespeichert. Ausgelesen wurden dabei lediglich der Titel, Titelzusatz bzw. Paralleltitel und die RVK-Notationen. Das Ergebnis dieser Abfragen war eine TSV-Datei mit etwa 62.000 Einträgen, welche zufällig in 80 % Trainingsdatensätze und 20 % Testdatensätze aufgeteilt wurde.

4.3 Automatische Sacherschließung mit Annif

Nach der Installation des Programms und der gewünschten optionalen Features können einzelne Projekte eingerichtet werden, in denen bestimmte Parameter – wie das zu verwendende Vokabular, das genutzte Backend und die Anzahl der von Annif vorzuschlagenden Notationen – individuell eingestellt werden. In mehreren Projekten wurden hier unterschiedliche Parameter ausgetestet. Die Einrichtung von Projekten, die Durchführung von Training und die Evaluation des Trainings mit Testdaten dauerten dabei insgesamt fünf bis zehn Minuten. Der Ansatz des Projekts war es, Annif als *Machine in the Loop* (im Gegensatz zum *Human in the Loop*) zu testen. Es ging also nicht darum, die Arbeit der Sacherschließ*innen durch Annif zu ersetzen. Vielmehr war die Frage, ob Annif genutzt werden kann, um möglichst passende Vorschläge für RVK-Stellen zu geben.

Nach einigen Trainings- und Testdurchläufen mit unterschiedlichen Annif-Backends zeigte sich, dass die Ergebnisse mit dieser begrenzten Datengrundlage bereits bemerkenswert sind. In den intellektuell erstellten Datensätzen der Hessischen Bibliographie sind die meisten Datensätze lediglich mit einer RVK-Stelle erschlossen. Daher wurden die Projekte besonders auf einen hohen *Recall* und weniger auf *Precision* optimiert.²² Die Evaluation der Annif-Projekte ergab, dass die Wahrscheinlichkeit hoch ist, mindestens eine zutreffende RVK-Stelle zu erhalten, wenn man drei RVK-Stellen pro Datensatz vorschlagen lässt. Dass dadurch neben tatsächlich relevanten RVK-Stellen vom Programm auch nicht-

19 SRU-Schnittstelle der Hessischen Bibliographie, <http://sru.hebis.de/sru/DB=2.4>, Stand 08.05.2026.

20 Die Annif-Entwickler setzen den unteren Grenzwert bei etwa 50, <https://github.com/NatLibFi/Annif/wiki/Achieving-good-results>, Stand 08.05.2026.

21 Von den 464 verwendeten Stellen mit Hessenbezug sind also 28 % häufig verwendet und wurden für den Test ausgewählt. Im Vergleich dazu werden von den 15.613 insgesamt in der Bibliographie genutzten RVK-Stellen nur 995 (ca. 6 %) mindestens 50-mal verwendet.

22 Zu den Begriffen siehe: Lepsky, Klaus: B 3 Automatisches Indexieren, in: Kuhlen, Rainer; Lewandowski, Dirk; Semar, Wolfgang und Womser-Hacker, Christa (Hg.): Grundlagen der Informationswissenschaft, Berlin 2023, S. 171–182, <https://doi.org/10.1515/9783110769043-015> und Zeya, LT: Precision and Recall Made Simple. Towards Data Science, in: Towards Data Science, Nov 9, 2021, <https://towardsdatascience.com/precision-and-recall-made-simple-afb5e098970f/>, Stand 08.05.2026.

zutreffende vorgeschlagen werden, würde bei einem Einsatz als *Machine in the Loop* nicht unbedingt ins Gewicht fallen, da den Erschließer*innen die Bewertung und eventuelle Übernahme der Vorschläge überlassen bliebe. Vor einem Praxiseinsatz müsste aber noch eine qualitative Auswertung der Ansätze erfolgen und auch die Auswahl der Trainingsdaten evaluiert werden. Zentraler für das Praxis-Projekt waren jedoch die Erkenntnisse zu den benötigten Datenpaketen, technischen Strukturen und Vorkenntnissen für eine solche Projektplanung. Sie wurden durch Projektvorstellungen im Haus und die Bereitstellung der Daten samt Dokumentation der Prozesse gesichert und weitergegeben.²³

Fazit

Für die klassifikatorische Sacherschließung werden im Sachgebiet unterschiedliche Systematiken bzw. Systeme für einzelne Bereiche genutzt. Eine Arbeitshilfe in Form von automatisch generierten Vorschlägen könnte dabei die Arbeit unterstützen. Für die Entwicklung und dauerhafte Nutzung automatisierter Verfahren zur Sacherschließung sind kürzere Projekte wie im beschriebenen Beispiel jedoch nicht ausreichend.

Für das Referendariatsprojekt selbst waren Grundkenntnisse im Skripten – in diesem Fall mit Python – zentral. Sie waren durch die Ausbildung in Berlin vorhanden, können aber auch durch verschiedene kostenfrei zugängliche Selbstlernkurse erworben und erweitert werden. Es gibt grundsätzlich für alle Bereiche des Projekts ausführliche und leicht zugängliche Erklärungen und Dokumentationen. Beim Erstellen und Debugging von Skripten sowie bei Fragen zur Unix Shell waren Large Language Models sehr hilfreich.²⁴ Die Arbeit mit dem Tool Annif ist gut zu erlernen und braucht nur wenig Zeit. Die zuvor notwendigen Schritte der Datenbeschaffung und der Datenaufbereitung dauerten dagegen länger als erwartet.

Dieser Praxisbericht ist ein Plädoyer für einen Hands-on-Ansatz in der Ausbildung für Referendar*innen und Volontär*innen. Dieser nutzt die Freiheiten in den Vorgaben zur Ausbildung, um früh Verantwortung zu übertragen und Projekte anzugehen, die Raum für Fehler lassen und von denen – unabhängig von ihrer Berufserfahrung – alle Beteiligten etwas lernen können. Insgesamt war das Projekt erfolgreich, auch unter Berücksichtigung der aufwändigen Planung. Mit dem Projekt begann ein anhaltender Austausch in der Institution zu diesem Thema.

Anhang

Tab. 2: Projektplan

Arbeitspakete (AP)	Zeit*	Inhalt
(1) Einführung: Arbeiten mit SSH und CLI	4	<ul style="list-style-type: none"> • Erstellung und Verwaltung von SSH-Schlüsselpaaren • Windows-Terminal mit PowerShell sowie Linux-Shell mit Bash • SSH-Verbindungen und Navigation auf der Kommandozeile (Linux)

23 Die erstellten Skripte zur Zusammenstellung des Vokabulars und der Trainingsdaten wurden auf <https://gitlab.rz.uni-frankfurt.de/huennes/projekt-automatische-verschlagwortung/-/tree/5f80da382b6789e7aaa33c57b451dff9f2ef3d35/> zur Verfügung gestellt, Stand 08.05.2026.

24 Es wurde gpt-4o-mini von OpenAI verwendet.

Arbeitspakete (AP)	Zeit*	Inhalt
		<p>Ressourcen:</p> <p>https://librarycarpentry.org/lessons/ https://librarycarpentry.github.io/lc-shell/</p>
(2) Einführung: Python und Git	7	<ul style="list-style-type: none"> • Installation und Verwendung von Python und virtuellen (Python-)Umgebungen • Kernkonzepte der Versionierung (von Programmiersprachen und (Text-)Dateien) • Grundlegende Git-Befehle und -workflows • Kennenlernen von GitLab (bzw. der selbstbetriebenen Instanz der Goethe-Universität) und GitHub • Einführung in Visual Studio Code <p>Ressourcen:</p> <p>https://realpython.com/ https://learnpythonthehardway.org/book/appendixa.html https://rogerdudler.github.io/git-guide/index.de.html https://code.visualstudio.com/docs/getstarted/getting-started</p>
(3) Literaturrecherche: Automatisierte Sacherschließung (in Deutschland)	2	<ul style="list-style-type: none"> • Welche Werkzeuge nutzen Wissenschaftlichen Bibliotheken oder andere Universitätseinrichtungen zur automatisierten Sacherschließung? • Gibt es Einrichtungen, die automatisierte Sacherschließung als Service für Forschung, Studium und Lehre anbieten?
(4) Literaturrecherche: Fachbegriffe automatisches Indexieren	1	<ul style="list-style-type: none"> • Einführung ins automatische Indexieren • Die Evaluationsmetriken in Annif (Precision, Recall, F1 Score, ...) <p>Ressourcen</p> <p>https://github.com/NatLibFi/Annif/wiki/Achieving-good-results</p>
(5) Annif: Einrichtung und Tutorial	10	<ul style="list-style-type: none"> • Installation und Einrichtung der benötigten Komponenten auf einer virtuellen Linux-Maschine • Nachvollziehen und ggf. Reproduktion der Ergebnisse der Masterarbeit von Elisabeth Mecking <p>Ressourcen</p> <p>https://github.com/NatLibFi/Annif-tutorial?tab=readme-ov-file</p>
(6) Kollegiale Beratung	2	<ul style="list-style-type: none"> • Austausch mit Expert*innen in kollegialen Netzwerken
(7) Experimente: Anforderungskatalog für die Sacherschließung	5	<ul style="list-style-type: none"> • Überlegungen zu den Voraussetzungen für die automatische Sacherschließung • Formulierung von Kriterien für die Zusammenstellung von repräsentativen Trainingsdaten
(8) Dokumentation und Präsentation des Projekts	4	<ul style="list-style-type: none"> • Dokumentation der zentralen Arbeitsschritte im internen Wiki der UB. Mit einem besonderen Schwerpunkt auf die UB- spezifischen Arbeitsschritte, die nicht aus den Arbeitshilfen zu Annif hervorgehen. • Präsentation des Projekts auf Abteilungsebene (inkl. kurzer Einführung zum aktuellen Stand der automatisierten Sacherschließung).
Summe	35	*Die Bilanzierung der Tabelle bezieht sich auf jeweils halbe Arbeitstage!

Anna Maria Hünnes, Universitätsbibliothek Johann Christian Senckenberg, Frankfurt am Main,
<https://orcid.org/0000-0002-2843-3022>

Christoph Marutschke, Universitätsbibliothek Johann Christian Senckenberg, Frankfurt am Main,
<https://orcid.org/0000-0002-4810-6918>

Franziska Voß, Universitätsbibliothek Johann Christian Senckenberg, Frankfurt am Main, <https://orcid.org/0000-0002-6472-4177>

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/6263>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).