








# KI-basierte Analyse von E-Portfolios

Alexander Gantikow <sup>1</sup>, Sara Durski <sup>1</sup>, Andreas Isking <sup>1</sup>, Paul Libbrecht <sup>1,2</sup>,  
Wolfgang Müller <sup>1</sup>, Simon Ostermann <sup>3</sup> und Sandra Rebholz <sup>1,4</sup>






**Abstract:** E-Portfolios stellen eine effektive Form des Assessments im Kontext des kompetenzbasierten Lernens dar. Die Sichtung und Bewertung solcher Lerntagebücher bringt allerdings bislang einen hohen Zeitaufwand mit sich. Die natürliche Sprachverarbeitung (NLP) liefert vielversprechende Ergebnisse in der Analyse von Texten und eröffnet auch in den Bereichen des formativen und summativen Assessments sowie Feedback von E-Portfolios neue Möglichkeiten, Lehrende bei der Bewertung, aber auch Studierende bei der Komposition von E-Portfolios zu unterstützen. Im Rahmen des hier vorgestellten Projekts wurde eine webbasierte Infrastruktur entwickelt, die E-Portfolios von Studierenden analysiert und für das Assessment relevante Indikatoren und Aspekte in Form von Dashboards visualisiert, um so neue Zugänge zu deren Inhalten zu schaffen. Verfolgte Ansätze und Methoden werden beschrieben und erste Evaluationsergebnisse präsentiert, die die Machbarkeit des Ansatzes demonstrieren, sowie erste Erkenntnisse über den potenziellen Mehrwert des Systems und dafür notwendige Weiterentwicklungen abgeleitet.

**Keywords:** Learning Analytics, Natural Language Processing, Artificial Intelligence, E-Portfolios, Dashboard, Visualisierung

## 1 E-Portfolios als Instrument zur Kompetenzbewertung

E-Portfolios sind eine bewährte Form, um individualisiertes Lernen zu unterstützen [Ra05]. Die Erstellung von E-Portfolios fördert Studierende in ihrer Kompetenzentwicklung, indem sie dieser durch selbst verfasste Texte und ausgewählte multimediale Lernartefakte Ausdruck verleihen. Auf Seiten der Lehrenden ermöglichen E-Portfolios einen Einblick in individuelle Lernprozesse und versetzen sie in die Lage, den Studierenden gezieltes Feedback zu geben. Werden E-Portfolios als Instrument zur Leistungsbewertung eingesetzt, müssen die Lehrkräfte die verschiedenen in einem E-Portfolio enthaltenen Artefakte anhand geeigneter Kompetenzraster bewerten [BC15].

E-Portfolios sind digitale Sammelmappen, in denen Lernende ihr Wissen und ihre Lernprozesse dokumentieren. Durch die individuelle Zusammenstellung digitaler Artefakte (Texte,

- 1 Pädagogische Hochschule Weingarten, Mediendidaktik und Visualisierung, Kirchplatz 2, 88250 Weingarten, Deutschland, gantikow@ph-weingarten.de,  <https://orcid.org/0000-0002-6571-4683>;  
sara.durski@ph-weingarten.de,  <https://orcid.org/0000-0003-0744-7983>;  
andreas.isking@ph-weingarten.de,  <https://orcid.org/0000-0003-2531-376X>;  
muellerw@ph-weingarten.de,  <https://orcid.org/0000-0001-6474-3733>
- 2 IU Internationale Hochschule, IT & Technologie, Juri-Gagarin-Ring 152, 99084 Erfurt, Deutschland,
- 3 Deutsches Forschungszentrum für Künstliche Intelligenz, Multilingualität und Sprachtechnologie, Stuhlsatzenhaus 3, 66123 Saarbrücken, Deutschland,  
simon.ostermann@dfki.de,  <https://orcid.org/0000-0002-0899-0657>
- 4 Ostbayerische Technische Hochschule Amberg-Weiden, Elektrotechnik, Medien und Informatik, Kaiser-Wilhelm-Ring 23, 92224 Amberg, Deutschland,

Grafiken, Videos, ...) können sich Lernende mit einem Vokabular ausdrücken, das sowohl für sie selbst als auch für ihr Zielpublikum zugänglich ist. Gängige E-Portfolio-Systeme sind webbasiert und erlauben sowohl die Präsentation eigens erstellter Artefakte als auch die Einbettung und Verlinkung existierender Artefakte (z.B. Videos oder relevante Lernressourcen). Die dadurch entstehende (Re-)Kombination und Vernetzung von Wissen stellt den Lernprozess und seine Ergebnisse auf sehr individuelle Weise dar. Die Umsetzung kann weit über eine bloße Reproduktion des Lehrstoffs hinausgehen: Das Wissen wird in eigenen Worten wiedergegeben, grafisch aufgearbeitet (z.B. in Form von Concept Maps) oder im Rahmen von Projekten angewendet. Individuelle Schwerpunktsetzungen und Vertiefungen erlauben den Studierenden, den Lernstoff gemäß ihren eigenen Interessen und Kompetenzen zu erarbeiten. Reflexionen des Lernprozesses fordern von den Studierenden zudem eine bewusste Auseinandersetzung mit den eigenen Stärken und Schwächen und helfen dabei, den Lernprozess gezielt zu steuern und anzupassen. Im Gegensatz zu vielen anderen Lernmethoden nutzen E-Portfolios offene, digitale Dokumentenformate, die eine Verknüpfung mit frei verfügbaren Online-Ressourcen erlauben. Diese als Teil der Informationskompetenz geltende Fähigkeit, eigenes Wissen mit bestehenden Wissensquellen zu vernetzen, auf einen bestimmten Kontext anzuwenden und das Ergebnis darzustellen, ist eine Schlüsselkompetenz, die im modernen Berufsleben unabdingbar ist [Eh20].

Für das Assessment der E-Portfolios wurden detaillierte Bewertungsraster (Rubrics) entwickelt, erprobt und optimiert [MRL17]. Mit Hilfe dieser Rubrics bewerten die Lehrenden die erworbenen Kompetenzen und das Wissen der Studierenden anhand von Kriterien wie thematische Vollständigkeit, fachspezifische Tiefe, Verwendung relevanter Quellenmaterialien, Originalität der präsentierten Artefakte, Qualität der Erklärungen, formalen Aspekten und der Reflexion über die eigenen Lernprozesse sowie selbst hergestellte, fächerübergreifende Verknüpfungen der Kursthemen. Beim Lesen der E-Portfolios beurteilen die Lehrenden, wie gut diese Kriterien erfüllt sind, und erhalten einen Einblick, welche Themen (mehr oder weniger ausführlich) behandelt wurden. Diese Bewertung erfolgte bisher größtenteils manuell und ist in bestimmten Kohorten aufgrund der hohen Anzahl von E-Portfolios und des Zeitaufwands für die Bewertung der einzelnen Kriterien sehr aufwändig und zeitraubend. Um eine Vorstellung des insgesamt notwendigen Zeitaufwands zu vermitteln, sei hier außerdem auf ein im Grundstudium zu absolvierendes Projektmodul zur Förderung der digitalen Kompetenzen von Lehramtsstudierenden verwiesen: Rund 600 Teilnehmende dokumentieren jedes Jahr ihre Lernprozesse mit E-Portfolios, wobei die Bewertung eines E-Portfolios in der Regel ein bis zwei Stunden in Anspruch nimmt. An dieser Stelle setzt das Projekt AISOP an und untersucht, wie KI-basierte Methoden und Werkzeuge eingesetzt werden können, um Lehrende beim Assessment (formativ und summativ) und Studierende bei der Erstellung von E-Portfolios zu unterstützen. Speziell im Bereich der automatisierten Textanalyse kann mit einer Themenerkennung schnell und effizient überprüft werden, ob alle geforderten Themenbereiche abgedeckt sind und in welcher Tiefe diese behandelt wurden. Dies entlastet die Lehrenden und ermöglicht es ihnen, sich stärker auf die qualitative Bewertung der Inhalte zu konzentrieren.

## 2 Relevanz der Textanalyse

Die Analyse geschriebener Sprache ist ein breites Feld der künstlichen Intelligenz und befasst sich mit der Verarbeitung von Informationen in Textform und der Ableitung nützlicher Dienste. Die Textanalyse hat in den letzten Jahren mehrere große Entwicklungsschritte durchlaufen: Nach den semantischen Bereicherungen von Wortähnlichkeiten auf Basis von Ontologien oder distributionellen, zählbasierten Methoden, wurden vektorbasierte Modelle entwickelt, die Wörter als kontinuierliche Vektoren darstellen, die mithilfe von neuronalen Netzen aus dem Satzkontext berechnet werden und den Sinn von geschriebener Sprache deutlich besser erfassen können. Aktuelle neuronale Sprachmodelle sind der Höhepunkt dieser Entwicklung. Sie ermöglichen eine Modellierung von Sprache durch unüberwachtes Training auf allgemeinen Sprachkorpora (etwa Texte aus Enzyklopädien oder Nachrichten). Durch dieses sog. "Pre-training" lernt das Sprachmodell Grundlagen, die für spezielle Anwendungszwecke optimiert ("fine-tuning") werden können. BERT [De19b] oder RoBERTa [Li19] sind Beispiele für solche Modelle. Seit der Etablierung der Technologie sind Sprachmodelle in Größe und Datenhunger immens gewachsen, was mit einer spürbaren Performanzverbesserung einherging.

Mit Hilfe von automatischer Textanalyse lassen sich verschiedene Aufgaben durchführen, darunter die Themenerkennung (engl. Topic Recognition). Sie erlaubt es, Wörter und benachbarte Wörter in Texten zu analysieren, um das zugrundeliegende Thema bzw. darin behandelte Konzept zu erkennen und zu extrahieren. Im Supervised Topic Modeling werden neuronale Netze oder maschinelle Lernmodelle, wie zum Beispiel die oben erwähnten Sprachmodelle, dahingehend trainiert (im Falle von Sprachmodellen optimiert), solche Themen automatisch erkennen zu können. Dafür müssen händisch Trainingsdaten erzeugt werden, indem Textfragmente mit Hilfe einer vorgegebenen Sammlung an Konzepten annotiert und als Korpus gebündelt werden.

Eine solche Themenerkennung wird beispielsweise in der Software "PapagAI" [So23] eingesetzt. Dort werden anhand mehrerer Erkennungssysteme verschiedene Dimensionen studentischer Reflexionen analysiert und als Feedback in Form eines Netzdiagramms den Studierenden präsentiert. [Wu23] wenden ebenfalls NLP-Methoden basierend auf BERT an, um schriftliche Reflexionen zu analysieren und qualitative Indikatoren zu deren Bewertung zu extrahieren. Erste Ergebnisse zeigen, dass Indikatoren wie die Anzahl an behandelten fachspezifischen Konzepten, die Wortanzahl und die Textkohärenz Hinweise auf die Textqualität geben und dies als Ausgangspunkt zur Entwicklung von automatisierten Bewertungsmethoden genutzt werden kann. In [Mu23] werden mit Themenerkennung und anderen Analysen wissenschaftliche Texte im Hinblick auf deren Qualität analysiert. Die Rückmeldung wird als Formulierungsvorschlag im Text ausgegeben, was den weiteren Schreibprozess erleichtern kann. [SR19] entwickelten mit Hilfe von Textanalyse eine Methode, um anhand von Lernmaterialien Ähnlichkeiten zwischen Lehrveranstaltungen festzustellen, um es Studierenden wie Lehrenden zu erleichtern, Relationen zwischen Lehrveranstaltungen zu entdecken. [MRL17] haben ein automatisiertes Werkzeug zur Analyse von E-Portfolios entwickelt, das auf semantischen und deskriptiven Analysen

beruht. Um diesen Ansatz zu erweitern, wird im Folgenden eine ganzheitliche, webbasierte Infrastruktur zur KI-basierten Themenerkennung in E-Portfolios vorgeschlagen.

### **3 KI-gestützte Analyse von E-Portfolios**

Das Projekt AISOP hat das Ziel, Lehrende beim Assessment und Studierende bei der Erstellung von E-Portfolios zu unterstützen. Eine Themenerkennung in E-Portfolios soll speziell auf die Fachkonzepte der in der Lehrveranstaltung behandelten Themengebiete ausgerichtet werden, um so die in einem E-Portfolio enthaltenen Themen in einer interaktiven Concept Map [SAH07] hervorzuheben und die Themenabdeckung in einem Learning-Dashboard zu visualisieren. Neben der inhaltlichen Breite hat sich auch die inhaltliche Tiefe als wichtige Bewertungsgrundlage erwiesen. Hierfür gibt es verschiedene Indikatoren, für die Daten erhoben und auf einem Dashboard visualisiert werden können. Beispielsweise die Häufigkeiten von Konzepten in einem Portfolio, durchschnittliche oder absolute Wortanzahl pro Konzept, die lexikalische Dichte oder die Verwendung von Fachbegriffen und Quellenmaterialien. Hinzu kommen formelle Kriterien wie die Gesamtwortanzahl des Portfolios oder die Anzahl und Kategorisierung eingesetzter Medien. Außerdem soll das Dashboard klassische Elemente aus dem Bereich Learning Analytics beinhalten, wie z.B. eine Zeitleiste aus der das Nutzungsverhalten von Studierenden ersichtlich wird. Ein Vergleich verschiedener Messgrößen zwischen Studierenden ist für die Bewertung von großem Interesse und soll ebenfalls Teil des Dashboards sein. Im Gegensatz zu den gängigen Essay-Scoring-Algorithmen, wie etwa in [Sc23] oder [KB20] bietet AISOP eine visuelle und interaktive Unterstützung, damit Dozierende ein vertrauenswürdigeres und fundierteres Assessment durchführen können.

Die automatisierte Unterstützung beim Assessment von E-Portfolios basiert auf den folgenden Schritten:

- Transformation des bewährten Kompetenzrasters: Damit sich die Kriterien auch beim automatisierten Assessment von E-Portfolios quantifizieren lassen, wurden diese als Indikatoren in objektive Messgrößen überführt. So sollen z.B. die thematische Vollständigkeit über die An- oder Abwesenheit behandelter Themen und die Qualität der Erklärungen mit Textmetriken gemessen werden.
- Erfassen und Darstellen erwarteter Inhaltskonzepte: Um die erwarteten Konzepte zu ermitteln, wurde eine SKOS-Ontologie [MPA07] erstellt, die das zugrundeliegende Fachwissen der Kurse konzeptualisiert. Das Ergebnis dieses Schrittes ist eine Concept Map, die als Expertenmodell dient.
- Erstellung eines domänenspezifischen Korpus zum Training von Machine-Learning-Modellen (ML-Modelle) zur automatischen Textanalyse: Verschiedene Arten von Lernressourcen (z.B. Folien, Webquellen, Videotranskriptionen) sowie Portfolios aus vergangenen Kursen bildeten die Grundlage für ein domänenspezifisches Korpus, das mit Annotationen angereichert wurde, die auf die entsprechenden Konzepte des

Expertenmodells verweisen. Das Korpus wurde für das Training diverser ML-Modelle zur Analyse der Inhalte von E-Portfolios genutzt.

- Wissensextraktion aus E-Portfolios: Der Inhalt eines E-Portfolios wird als Input für das trainierte ML-Modell verwendet. Darüber hinaus werden Textmetriken sowie der Aufbau des HTML-Dokuments erhoben. Das identifizierte Wissen wird im E-Portfolio über HTML-Attribute kenntlich gemacht.
- Vergleich des extrahierten Wissens mit den erwarteten Lernergebnissen: Die semantische Ähnlichkeit zwischen dem extrahierten Wissen und dem Expertenmodell wird berechnet und in einer kombinierten Concept Map visuell dargestellt: Vorhandene, überlappende und fehlende Konzepte werden hervorgehoben und geben Aufschluss über Wissenslücken, spezifische Expertise und individuelle Interessensgebiete.
- Vergleich des extrahierten Wissens mit dem aggregierten Lernerwissen: Alle E-Portfolios, die im Rahmen einer Gruppenanalyse ausgewertet wurden, bilden das aggregierte Wissen aller Lernenden. Auf diese Weise lässt sich das in einem E-Portfolio enthaltene Wissen mit der Grundgesamtheit vergleichen.
- Visuelle Navigation innerhalb eines E-Portfolios: Die kombinierte Concept Map aus Studierenden- und Expertenwissen eignet sich auch als Werkzeug für die visuelle Navigation [SAH07] innerhalb eines E-Portfolios: Zum einen können die Knotenpunkte der Concept Map als interaktive Links zu den entsprechenden Abschnitten im E-Portfolio genutzt werden, zum anderen werden die Konzepte im aktuell sichtbaren Abschnitt des Portfolios erkannt und in der Concept Map hervorgehoben.
- Kontinuierliche Erweiterung des Korpus: Das zugrundeliegende Korpus für das Training des ML-Modells wird kontinuierlich durch neu erstellte und annotierte E-Portfolios erweitert.

### 3.1 Training eines Modells zur Textklassifizierung

Von besonderem Interesse für AISOP ist es herauszufinden, welche Konzepte die Studierenden in ihren E-Portfolios vollständig, teilweise oder gar nicht abgedeckt haben. Für das Kriterium der thematischen Vollständigkeit wurde dementsprechend die quantitative Abdeckung der Seminarthemen als einer der Indikatoren definiert. Zu diesem Zweck wurde zunächst eine Concept Map des Seminars “Fundamentale Ideen der Informatik” erstellt und deren Konzepte als Label-Set (“geschlossenes Label-Vokabular”) für das Training eines seminarspezifischen ML-Modells verwendet. Damit das ML-Modell auch auf reflexive Inhalte sowie fächerübergreifende Verknüpfungen trainiert werden konnte, wurde das Label-Set um die übergeordneten Konzepte “Reflexion” und “Link” ergänzt.

**Datensatz** Für die Erstellung eines Trainingsdatensatzes wurden neben den Foliensätzen die E-Portfolios von 18 Studierenden mit deren Einverständnis zunächst anonymisiert und in

insgesamt 4787 Paragraphen zerlegt. Die Zerlegung wurde von geschulten Mitarbeitenden durchgeführt, die mit einem Browserwerkzeug ausschließlich Portfolioabschnitte mit einem einzelnen Thema selektierten. Die extrahierten Textabschnitte sind im Schnitt 400 Zeichen lang. Alle identifizierten Abschnitte wurden anschließend mit Hilfe der Software “Prodigy” [Ex23a] und dem Label-Set in einer Mehrfachauswahl annotiert. Drei Annotatoren konnten so insgesamt 4396 Datensätze annotieren. Fälle, die von den Annotatoren als unklar markiert wurden, wurden von einer Expertin manuell verbessert. Eine Säuberung der Daten (Löschen von Duplikaten und nicht auflösbaren Fällen) resultierte in insgesamt 3618 Einzelinstanzen, die in 3 Teil-Datensätze aufgesplittet wurden: Trainingsset (2894 Instanzen), Validierungsset (362 Instanzen) und Testset (362 Instanzen).

Abbildung 1 zeigt die Verteilung annotierter Klassen auf den finalen Daten. Die Kategorie “Rest” beinhaltet alle Klassen mit weniger als 50 Instanzen (*Virtualisation*: 44, *Computer Science*: 35, *IT Security*: 30, *Input-Process-Output Model*: 3, *Von Neumann Architecture*: 2). Abschnitte, die als *Algorithmization* gelabelt wurden, stellen mit über einem Viertel der Gesamtdaten die größte Klasse dar. Über 50% der Daten können einer von 4 Klassen zugeordnet werden: *Algorithmization*, *Representation*, *Encryption*, und *Operation System*.

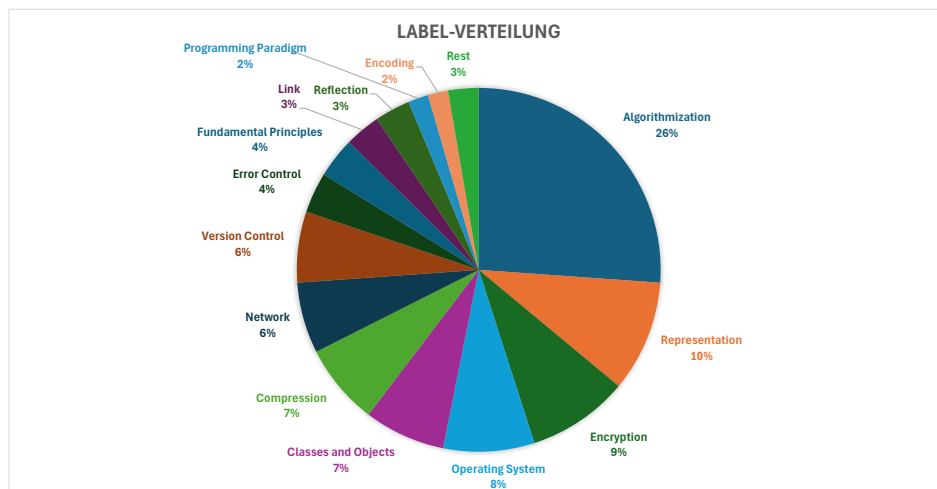


Abb. 1: Verteilung der annotierten Klassen auf den bereinigten Daten im Kurs “Fundamentale Ideen der Informatik”

**Klassifikationsmodell** Auf Grundlage des annotierten Korpus wurden verschiedene Machine-Learning-Modelle (ML-Modelle) zur Textklassifikation trainiert:

- Wir benutzen ein “spaCy”-basiertes Pipeline-Modell zur Textklassifizierung. Wir experimentieren mit einer Standard-Preprocessing-Pipeline und einem tok2vec-Modell [Ex23b], welches mit einem Multiclass-Head auf den annotierten Trainingsdaten

trainiert wird. Die Vorhersage der Labels erfolgt im Ensemble mit einem 1-Gramm-basierten linearen Bag-of-Words-Modell.

- Wir benutzen mono- und multilinguale ML-Modelle basierend auf vortrainierten mittelgroßen Sprachmodellen (ein deutsches BERT-Modell [Ch23] und ein multilinguales XLM-R-Modell [Co19], was in einem ersten Schritt auf deutschen ConLL-Daten [TKSDM03] optimiert wurde). Für beide Modelle erfolgte eine weitere Optimierung.

Die Genauigkeit der ML-Modelle beträgt 90,1% (spaCy), 92% (BERT) und 91,7% (XLM-R). Die Performanz-Unterschiede der benutzten Modelle fallen nur minimal aus. Die generell hohe Performanz aller Modelle zeigt, dass die Textklassifikation der Abschnitte ein wenig komplexes Problem darstellt. Aufgrund der geringen Performanz-Unterschiede und der einfachen Integrierbarkeit fand das spaCy-basierte ML-Modell in der durchgeführten Evaluation (s.u.) Verwendung.

Im Rahmen einer genaueren Untersuchung der Ergebnisse des besten Modells mit einer Konfusions-Matrix konnten wir keine problematischen Paarungen zwischen Klassen finden, die besonders häufig verwechselt würden; die Anzahl der Fehlklassifizierungen ist nahezu uniform über alle Klassen verteilt. Das absolut am häufigsten falsch klassifizierte Label ist *Algorithmization*, was aber zugleich auch die größte Klasse darstellt, mit lediglich 7 falsch klassifizierten Instanzen.

### 3.2 Die AISOP-Web-Application

Die AISOP-Web-Application wird als eigenständige Node.js-Anwendung neben dem E-Portfolio-System Mahara [Ca22] betrieben und über dessen Schnittstellen angebunden. Über eine Autorisierung mittels OAuth 1.0a können Dozierende und Studierende die Dienste der Webanwendung auf die Portfolios von Mahara anwenden. Nach einer Synchronisierung der Portfolio-Metadaten können beide Zielgruppen zunächst ihre eigenen E-Portfolios analysieren. Da auch Mahara-Gruppen in der Web-Application abgebildet werden, können deren Administratoren (i.d.R. Lehrende) darüber hinaus die E-Portfolios ihrer Gruppenmitglieder zunächst thematisch kategorisieren und diese gebündelt analysieren.

Im Rahmen der automatisierten Analyse werden die von Mahara bereitgestellten Webservice-Schnittstellen genutzt: Die zu analysierenden E-Portfolios werden für die weitere Bearbeitung zunächst auf Mahara gesperrt, deren "HTML Lite Export" angestoßen sowie heruntergeladen und anschließend wieder entsperrt. Das empfangene ZIP-Archiv enthält eine HTML-Datei, die mit geeigneten Python-Analyse-Skripten ausgewertet wird. Mit der Analyse gängiger Textmetriken sowie einer HTML-Analyse liegen zwei generische Skripte vor, die sich auch anwendungsunabhängig nutzen lassen. Ergänzt werden diese durch eine Textanalyse, die auf das zuvor mit spaCy trainierte ML-Modell zur Textkategorisierung zurückgreift. Es erkennt die vom Studierenden beschriebenen Konzepte und reichert die entsprechenden

Paragrafen um das ermittelte Konzept sowie den zugehörigen Wahrscheinlichkeitswert (Score) an. Auf diese Weise werden nicht nur die Treffer gezählt, sondern auch deren Fundstellen im E-Portfolio markiert. Wurde die Analyse auf Gruppenebene angestoßen, so werden die erhobenen Daten entlang der zuvor gebündelten E-Portfolios aggregiert.

Die Auswertungen werden für Lehrende und Studierende in zielgruppenspezifischen Learning-Dashboards dargestellt. Während für Studierende nur die Ergebnisse der eigenen E-Portfolios einsehbar sind, können Dozierende die Analyseergebnisse aller Studierenden einer Kohorte aufrufen. Für jedes analysierte E-Portfolio steht eine Übersicht bereit, die die Ergebnisse zusammenfassend visualisiert. Einen tieferen Einblick erhalten die Nutzenden mit der "Portfolioansicht". Diese präsentiert das zuvor angereicherte E-Portfolio und erweitert dieses um zusätzliche Interaktions- und Navigationsmöglichkeiten. Eine Visualisierung der kursspezifischen Concept Map bildet hierbei das Herzstück in Abb. 2: Durch Anwählen eines Konzepts können die Nutzenden zu den entsprechenden Fundstellen im E-Portfolio springen. Gleichzeitig wird die Scrollbar um eine Visualisierung ergänzt, die die Verteilung des Konzepts über das E-Portfolio hinweg abbildet. Die Concept Map macht außerdem nicht erwähnte Konzepte sowie abgeleitete Informationen sichtbar, wie beispielsweise die vom Studierenden in Relation verwendeten Konzepte. Darüber hinaus erlauben auch die Ergebnisse der HTML-Analyse (z.B. Abbildungen, Tabellen, Hyperlinks, Tags) sowie ein automatisch erzeugtes Inhaltsverzeichnis einen Sprung zu den jeweiligen Fundstellen. Die Concept-Map-basierte Navigation schafft so neue Zugänge zu den Portfolioinhalten und verknüpft diese mit den für das Assessment erhobenen Daten.

The screenshot shows a web interface with two main sections. On the left is a 'Concept Map' (Concept Map) with a central node 'Symmetrische Verschlüsselung' (Symmetric Encryption) and several related nodes: 'Nichtlo-Funktionsstelle', 'Systemisches Konzept', 'Public-Key-Systeme', 'IT-Security-Risiko', 'Sitz-Information', 'Private-Key', 'Public-Key', and 'Man-in-the-Middle-Angriff'. Below the map is an 'Auswertung' (Evaluation) table. On the right is the 'E-Portfolio' content, which includes a section on 'Symmetrische Verschlüsselung' with three tasks (Aufgabe 1, 2, 3) and a 'Kryptanalyse' section with sub-tasks 'a) Häufigkeitsanalyse' and 'b) Brute-Force-Angriff'.

Auswertung	Wert	Link
Anzahl	29	
Durchschnittlicher Wert	0.72	
Höchster Wert	1.0	<a href="#">↗</a>
Niedrigster Wert	0.26	<a href="#">↘</a>

Abb. 2: Die Portfolioansicht besteht aus einer Visualisierung der kursspezifischen Concept Map (links) und dem E-Portfolio (rechts), das angereichert wurde. Durch Anwählen und Weiterblättern eines Konzepts kann ein Nutzer zu den entsprechenden Fundstellen im E-Portfolio springen.



## 4 Erprobung

Um eine erste Erprobung der webbasierten Infrastruktur zu ermöglichen, wurde die Portfolioprüfung im Wintersemester 2023/2024 so gestaltet, dass sie wenig von der traditionellen Prüfungsform abweicht: Die Studierenden haben ihr E-Portfolio weiterhin im Mahara der Hochschule erstellt, zum Assessment freigegeben und anschließend eine mündliche Prüfung absolviert. Darüber hinaus sollten sie einen Export ihres E-Portfolios einreichen, der in das projekteigene Mahara importiert wurde. Diese Mahara-Instanz wurde eingerichtet, um die technische Machbarkeit des vorgestellten Assessment- und Analyseworkflows zu erproben. Sie entspricht der Version an der Hochschule und dient als Testumgebung, bis das Produktiv-Mahara der Hochschule entsprechend konfiguriert ist. Der hier beschriebene Evaluationsdurchgang beschränkt sich auf den Einsatz des Dashboards zur Analyse und Bewertung von E-Portfolios durch Lehrende. Eine umfassende Evaluation mit Studierenden kann erst durchgeführt werden, wenn die Web-Application an das Hochschul-Mahara-System angebunden ist. Diese Evaluation ist für das Sommersemester 2024 geplant. Die Prüfenden hatten die Möglichkeit, die E-Portfolios sowohl im (Entwicklungs-)Mahara als auch in der angebundenen Web-Application zu betrachten. Mit letzterer wurden die E-Portfolios analysiert und deren interaktive Dashboards sowie Portfolionavigation zur Bewertung herangezogen. Auf diese Weise konnten zehn E-Portfolios importiert und vollständig in der Mahara-Testumgebung sowie Web-Application betrachtet werden. Die E-Portfolios wurden von insgesamt vier Gutachterinnen und Gutachtern mit Hilfe des Bewertungsrasters überprüft. Zwei der Teilnehmenden leiteten die zu prüfenden Seminare im laufenden Semester, während die anderen beiden in den vorhergehenden Semestern zuständig waren. Eine der beiden Gutachterinnen nahm zudem in der Rolle als UX-Expertin an der Evaluation teil. Alle Teilnehmenden teilen dasselbe informationstechnische Wissen.

Folgende Fragestellungen sollten mit der Evaluation beantwortet werden:

- Ist der mit der webbasierten Infrastruktur umgesetzte Workflow funktionsfähig? Ist insbesondere eine reibungslose Integration des E-Portfoliosystems, der NLP-Pipeline und des Lehrenden-Dashboards über die verfügbaren Schnittstellen in der Praxis umsetzbar und das System für das vorgesehene Bewertungsszenario nutzbar?
- Ist die Concept-Map-basierte Navigation und die um die Ergebnisse der Textanalyse angereicherte Darstellung der E-Portfolios intuitiv und einfach zu verwenden?
- Wie gut nehmen die Lehrenden das interaktive Dashboard an?
- Welchen Mehrwert bieten die im Lehrenden-Dashboard dargestellten Ergebnisse der Textanalyse bei der Bewertung von E-Portfolios?

### 4.1 Methodik und Durchführung

Um die obenstehenden Fragen zu beantworten, wurde ein halbstrukturiertes Interview mit den an der Evaluation teilnehmenden Lehrenden und einer Expertin im Bereich User-

Experience-Design (UX-Design) durchgeführt. Das Interview wurde auf der Grundlage eines Interviewleitfadens durchgeführt, der basierend auf Konstrukten der Usability [De18], User Experience [De19a] und des Technologieakzeptanzmodells (TAM) [Da85] konzipiert wurde. Somit konnten die Gebrauchstauglichkeit und das Benutzererlebnis der Befragten Aufschluss darüber geben, inwieweit sie ihre Ziele bei der Verwendung der Web-Application effektiv, effizient und zufriedenstellend erreichen konnten und welche Vorstellungen, Wahrnehmung und Vorgehensweisen sie hatten. Um Aufschluss über die Akzeptanz und den potenziellen Mehrwert des Lehrenden-Dashboards zu erhalten, wurden außerdem Fragen zur wahrgenommenen Nützlichkeit (Perceived Usefulness, PU) und zu möglichen Erweiterungen des Dashboards in den Leitfaden mit aufgenommen. Das halbstrukturierte Interview wurde als Evaluationsmethodik gewählt, um möglichst tiefgreifende Einblicke in die Erfahrungen der Lehrenden sowie die Beurteilung der UX-Expertin zu erhalten und bei Bedarf Rückfragen stellen zu können. Die Interviews wurden in einem zeitlichen Abstand von einer Woche nach der Nutzung der Web-Application in einer Online-Session geführt. Die qualitativ erhobenen Daten wurden durch eine Inhaltsanalyse, in welcher Kategorien gebildet wurden, ausgewertet.

Die Durchführung der Bewertung wurde von jedem Prüfenden unabhängig vorgenommen, d.h. es fanden keine Absprachen zwischen den Prüfenden statt. Zur Einführung in die Web-Application erhielten die Lehrenden vorab eine videobasierte Anleitung (Anmeldung, grundlegenden Nutzung). Informationen zur Concept-Map-basierten Navigation und der interaktiven Visualisierung der Analyseergebnisse wurden nicht gegeben (Gegenstand der Evaluation). Um den datenschutzrechtlichen Anforderungen gerecht zu werden, wurden nur solche E-Portfolios gesichtet, für die eine schriftliche Einwilligung der Studierenden vorlag.

## **4.2 Ergebnisse**

Für die Nutzung der Web-Application mussten sich die Teilnehmenden zunächst bei Mahara einloggen. Dieser Weg war manchen nicht bewusst und hat den Einstieg in die Evaluation erschwert. In der Web-Application angekommen, konnten alle Teilnehmenden zügig zur Prüfungsgruppe wechseln und darüber die Dashboards der Studierenden aufrufen. Alle konnten die Analyse eines E-Portfolios anstoßen, wurden durch eine durchschnittliche Wartezeit von zwei Minuten allerdings verunsichert und hätten sich Feedback über den Analysefortschritt gewünscht. Alle Teilnehmenden konnten die zur Analyse erzeugte Übersicht öffnen. Die enthaltenen Visualisierungen wurden positiv bewertet, deren Datenvielfalt allerdings bemängelt. So hat ein Teilnehmer beispielsweise kritisiert, dass lediglich die lexikalische Dichte berechnet wurde und weitere Textmetriken nicht angezeigt wurden.

Die Benutzeroberfläche der Portfolioansicht wurde positiv aufgenommen. In Bezug auf die Produktivität gaben die Teilnehmenden an, dass sie schnell zwischen den bereitgestellten Auswertungen (z.B. Inhaltsverzeichnis, Konzepte, Hyperlinks, Bilder) wechseln und mit Hilfe der internen Verlinkungen zu den entsprechenden Trefferstellen im E-Portfolio springen konnten. Die Teilnehmenden haben die Möglichkeit der Navigation über die interaktive

Concept Map rege in Anspruch genommen, bemängelten aber eine fehlende Systematik z.B. durch eine Abhebung von bereits inspizierten Konzepten von den verbleibenden Konzepten, um den Assessmentprozess effektiver zu gestalten. Die Effektivität wurde laut den Teilnehmenden ebenfalls durch eine unpassend gewählte Farbgebung eingeschränkt, sodass die Teilnehmenden nicht erkennen konnten, welche Konzepte im E-Portfolio nicht abgedeckt waren. Zudem fühlten sich die Teilnehmenden unsicher, welchen Mehrwert die für jedes Konzept berechneten Kennzahlen für ihr Assessment bieten. Um die Nützlichkeit dieser Auswertungen zu steigern, wurde eine strengere Strukturierung der Analyseergebnisse entlang des Kompetenzrasters gefordert. In Bezug auf die wahrgenommene Gebrauchstauglichkeit der Web-Application wurde die Concept Map als zu klein bewertet und die Möglichkeit gefordert, die statische durch eine benutzerdefinierte Aufteilung des Viewports abzulösen. Fachfremde Teilnehmende haben die Concept Map zudem als sehr komplex beschrieben und gaben an, mehr Zeit zur Orientierung zu benötigen. Außerdem bedürfen die für jedes Konzept berechneten Kennzahlen einer Erklärung - so einer der Teilnehmenden. Beispielsweise erlaubt die Portfolionavigation das Springen zu dem Paragraphen, der von der Textanalyse mit dem höchsten Wahrscheinlichkeitswert bewertet wurde - eine Funktionalität, die nicht alle Teilnehmenden verstanden haben. Durch eine abweichende Sortierung der Analyseergebnisse von den Indikatoren im Bewertungsraster konnten die Teilnehmenden nur mit Verzögerung die Kriterien des Rasters mit den dargestellten Werten in der Benutzeroberfläche der Web-Application in Verbindung bringen.

Darauf angesprochen, stellten die Teilnehmenden folgende Vorteile der Web-Application heraus: In der Möglichkeit, die Ergebnisse unterschiedlicher Analysen direkt im E-Portfolio zu verorten, sehen sie den Vorteil, sich schnell einen Überblick über das Dokument zu verschaffen. Durch das Hervorheben der entsprechenden Fundstellen wirft die Portfolionavigation den Fokus auf bestimmte Aspekte des E-Portfolios. Laut den Prüfern fällt es ihnen so leichter, dem E-Portfolio zusätzliche Informationen zu entnehmen, die ihrem Assessment zugutekommen. Sie bewerten die Web-Application als hilfreich, sehen in ihr ein sinnvolles Tool für das Assessment von E-Portfolios und trauen ihr bei zukünftiger Verwendung eine Zeitersparnis zu. Alle Teilnehmenden können sich eine Nutzung der Web-Application in kommenden Portfolioprüfungen vorstellen.

Die durchgeführte Evaluation liefert bereits einen Einblick, welchen Mehrwert die Web-Application in der Praxis bietet. Auch wenn der Autorisierungsprozess noch Verbesserungspotenzial aufweist, konnten alle Teilnehmenden den gesamten Workflow aus Login, Portfolioanalyse und Betrachtung der Ergebnisse durchlaufen. Die Teilnehmenden haben die Web-Application als Werkzeug verstanden, das sie während des Bewertungsszenarios unterstützend begleitet hat. Obwohl die Concept Map noch nicht alle gewünschten Informationen abbildet, zeigt sie bereits jetzt ihr Potenzial der vereinfachenden Navigation innerhalb eines E-Portfolios. Die derzeit im Lehrenden-Dashboard integrierten Visualisierungen können hingegen in verschiedenen Use Cases (z.B. vergleichende Benotung) das Potential noch nicht vollständig entfalten.

## 5 Zusammenfassung und Ausblick

Dieser Beitrag stellt ein KI-gestütztes Portfolio-Analysesystem vor, das Lehrende im Assessment von E-Portfolios unterstützen kann. Mit der Evaluation legt das Projekt sowohl erste Erkenntnisse zum Mehrwert des Systems als auch eine Machbarkeitsstudie vor: Mit der Kopplung an eine Mahara-Instanz kann das vorgestellte Softwaresystem auf deren Portfolioinhalte autorisiert zurückgreifen. Die empfangenen E-Portfolios lassen sich mit Hilfe spezieller Analyse-Skripte auswerten und die Ergebnisse werden in Dashboards und der Portfolionavigation zugänglich gemacht. Mit der durchgeführten Evaluation konnte der ganzheitliche Workflow aus Login, Portfolioexport, Analyse, Aggregation und Visualisierung erprobt werden. Dennoch hat die Evaluation weiteres Verbesserungspotenzial aufgedeckt, das im Folgenden überblicksartig dargelegt wird.

Aufgrund seines begrenzten Label-Sets entdeckt das vom Projekt trainierte ML-Modell nur die Konzepte des Expertenmodells. Mit dieser Vorgehensweise lassen sich zwar die vorhandenen Konzepte in einem E-Portfolio, nicht aber über das Themenspektrum hinausragende Inhalte identifizieren. Als Ergänzung zur geschlossenen kommt deshalb auch eine offene Themenerkennung in Betracht. Außerdem ist der Einsatz künstlicher Intelligenz in AISOP noch auf die Analyse der Portfoliotexte begrenzt. In Zukunft könnten auch Abbildungen und Videos sowie Code- und PDF-Einreichungen analysiert werden. Darüber hinaus soll überprüft werden, ob sich weitere Textmetriken zur Erhebung qualitativer Indikatoren in die Analysen integrieren lassen (z.B. [Mo23], [HOE23]). Eine über das Dashboard nachvollziehbare Zuordnung dieser Indikatoren zu den im Bewertungsraster hinterlegten Kriterien soll ein effizientes Assessment der Dozierenden ermöglichen. Für eine Verbesserung des trainierten ML-Modells sollen die E-Portfolios der kommenden Kohorten ebenfalls anonymisiert, annotiert und in das Korpus übernommen werden. Sowohl das Korpus als auch die trainierten Modelle werden mit Projektende veröffentlicht.

Da die Testumgebung auf derselben Version gehalten wird, wie das Mahara der Hochschule, kann mit deren finaler Zustimmung die AISOP Web-Application für die gesamte Institution angeboten werden. Mit der Kopplung beider Softwaresysteme kann eine umfassendere Evaluation geplant werden, die auch Nutzertests mit Studierenden umfasst.

Die Web-Application, Analyse-Skripte sowie die trainierten ML-Modelle sind auf GitLab<sup>5</sup> öffentlich zugänglich und werden bis zum Projektende laufend aktualisiert. Mit Hilfe der ausführlichen Dokumentation kann die Web-Application an Mahara-Systeme anderer Hochschulen angeschlossen und für den dort gegebenen Anwendungsfall erweitert werden. Lediglich ein ML-Modell muss für den eigenen Anwendungsfall selbst trainiert werden, alle anderen Komponenten der entwickelten Infrastruktur können direkt eingesetzt werden. Das im Experiment verwendete Modell wurde mit Hilfe annotierter E-Portfolios und Foliensätze trainiert; das Korpus und das Modell werden in einem RADAR-Repository<sup>6</sup> veröffentlicht, sobald Klarheit über das anzuwendende Anonymisierungsverfahren herrscht.

---

<sup>5</sup> <https://gitlab.com/aisop>

<sup>6</sup> <https://www.radar-service.eu/radar/de/workspace/OcssWywfBIOotMSi>

## Literaturverzeichnis

- [BC15] Brookhart, Susan M.; Chen, Fei: The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3):343–368, Juli 2015. <https://doi.org/10.1080/00131911.2014.929565>.
- [Ca22] Catalyst.Net Limited: , Mahara ePortfolio System, Oktober 2022.
- [Ch23] Chan, Branden; Möller, Timo; Pietsch, Malte; Soni, Tanay: , German BERT, April 2023. <https://huggingface.co/google-bert/bert-base-german-cased>.
- [Co19] Conneau, Alexis; Khandelwal, Kartikay; Goyal, Naman; Chaudhary, Vishrav; Wenzek, Guillaume; Guzmán, Francisco; Grave, Edouard; Ott, Myle; Zettlemoyer, Luke; Stoyanov, Veselin: Unsupervised Cross-lingual Representation Learning at Scale. *arXiv preprint arXiv:1911.02116*, 2019.
- [Da85] Davis, Fred D.: , A technology acceptance model for empirically testing new enduser information systems: Theory and results, 1985.
- [De18] Deutsches Institut für Normung e. V.: , Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte, 2018.
- [De19a] Deutsches Institut für Normung e. V.: , Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme, 2019.
- [De19b] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In (Burstein, Jill; Doran, Christy; Solorio, Thamar, Hrsg.): *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, S. 4171–4186, Juni 2019.
- [Eh20] Ehlers, Ulf-Daniel: *Future Skills: Lernen der Zukunft - Hochschule der Zukunft*. Springer VS, Wiesbaden, Januar 2020.
- [Ex23a] Explosion AI GmbH: , Prodigy. An annotation tool for AI, Machine Learning & NLP, November 2023.
- [Ex23b] Explosion AI GmbH: , SpaCy Model de\_core\_news\_md. German pipeline optimized for CPU., Oktober 2023. [https://spacy.io/models/de#de\\_core\\_news\\_md](https://spacy.io/models/de#de_core_news_md).
- [HOE23] Hansen, Lasse; Olsen, Ludvig Renbo; Enevoldsen, Kenneth: TextDescriptives: A Python package for calculating a large variety of metrics from text. *Journal of Open Source Software*, 8(84):5153, April 2023.
- [KB20] Kumar, Vivekanandan; Boulanger, David: Explainable Automated Essay Scoring: Deep Learning Really Has Pedagogical Value. *Frontiers in Education*, 5:572367, Oktober 2020.
- [Li19] Liu, Yinhan; Ott, Myle; Goyal, Naman; Du, Jingfei; Joshi, Mandar; Chen, Danqi; Levy, Omer; Lewis, Mike; Zettlemoyer, Luke; Stoyanov, Veselin: , RoBERTa: A Robustly Optimized BERT Pretraining Approach, Juli 2019. <https://doi.org/10.48550/arXiv.1907.11692>.
- [Mo23] Morse, Colin: , ReadabilityAnalyserDE, Juni 2023.

- [MPA07] Miles, Alistair; Pérez-Agüera, José R.: SKOS: Simple Knowledge Organisation for the Web. *Cataloging & Classification Quarterly*, 43(3-4):69–83, April 2007. [https://doi.org/10.1300/J104v43n03\\_04](https://doi.org/10.1300/J104v43n03_04).
- [MRL17] Müller, Wolfgang; Rebholz, Sandra; Libbrecht, Paul: Automatic Inspection of E-Portfolios for Improving Formative and Summative Assessment. In (Wu, Ting-Ting; Gennari, Rosella; Huang, Yueh-Min; Xie, Haoran; Cao, Yiwei, Hrsg.): *Emerging Technologies for Education*, Jgg. 10108, S. 480–489. Springer International Publishing, Cham, 2017. Series Title: *Lecture Notes in Computer Science*.
- [Mu23] Muecke, Justin; Waldow, Daria; Metzger, Luise; Schauz, Philipp; Hoffman, Marcel; Lell, Nicolas; Scherp, Ansgar: , Fine-Tuning Language Models for Scientific Writing Support, Juni 2023. <https://doi.org/10.48550/arXiv.2306.10974>.
- [Ra05] Ravet, Serge: ePortfolio for a learning society. *E Learning Conference*, Brussels. Retrieved February, Januar 2005.
- [SAH07] Steiner, C.; Albert, Dietrich; Heller, Juergen: Concept mapping as a means to build e-learning. In (Buzetto-More, Nicole A., Hrsg.): *Advanced Principles of Effective e-Learning*, S. 59–111. Informing Science Press, Santa Rosa, Januar 2007.
- [Sc23] Schlippe, Tim; Stierstorfer, Quintus; Koppel, Maurice ten; Libbrecht, Paul: Explainability in Automatic Short Answer Grading. In (Cheng, Eric C. K.; Wang, Tianchong; Schlippe, Tim; Beligiannis, Grigorios N., Hrsg.): *Artificial Intelligence in Education Technologies: New Development and Innovative Practices*. Springer Nature Singapore, Singapore, S. 69–87, 2023.
- [So23] Solopova, Veronika; Gruszczynski, Adrian; Rostom, Eiad; Cremer, Fritz; Witte, Sascha; Zhang, Chengming; Plöbl, Fernando Ramos López Lea; Hofmann, Florian; Romeike, Ralf; Gläser-Zikuda, Michaela; Benzmüller, Christoph; Landgraf, Tim: , PapagAI:Automated Feedback for Reflective Essays, Juli 2023. <https://doi.org/10.48550/arXiv.2307.07523>.
- [SR19] Seidel, Niels; Rieger, Christian Moritz: Semantic Textual Similarity von textuellen Lernmaterialien und ihre Anwendungen. In: *DELFI 2019 - Die 17. Fachtagung Bildungstechnologien*. Gesellschaft für Informatik e.V., Berlin, 2019. [https://doi.org/10.18420/delfi2019\\_360](https://doi.org/10.18420/delfi2019_360).
- [TKSDM03] Tjong Kim Sang, Erik F.; De Meulder, Fien: Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. S. 142–147, 2003.
- [Wu23] Wulff, Peter; Westphal, Andrea; Mientus, Lukas; Nowak, Anna; Borowski, Andreas: Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers’ written reflections. *Frontiers in Education*, 7, Januar 2023. Publisher: Frontiers.