



KI-generierte Metadaten

Wie Künstliche Intelligenz die Auffindbarkeit
offener Lernressourcen verbessern kann

Axel Klinger, Technische Informationsbibliothek (TIB)

Online-Fachtagung: OER im Zeitalter von KI
6. November 2025

OER-Suche für die Hochschullehre



- **OERSI als zentrale Suchmaschine**

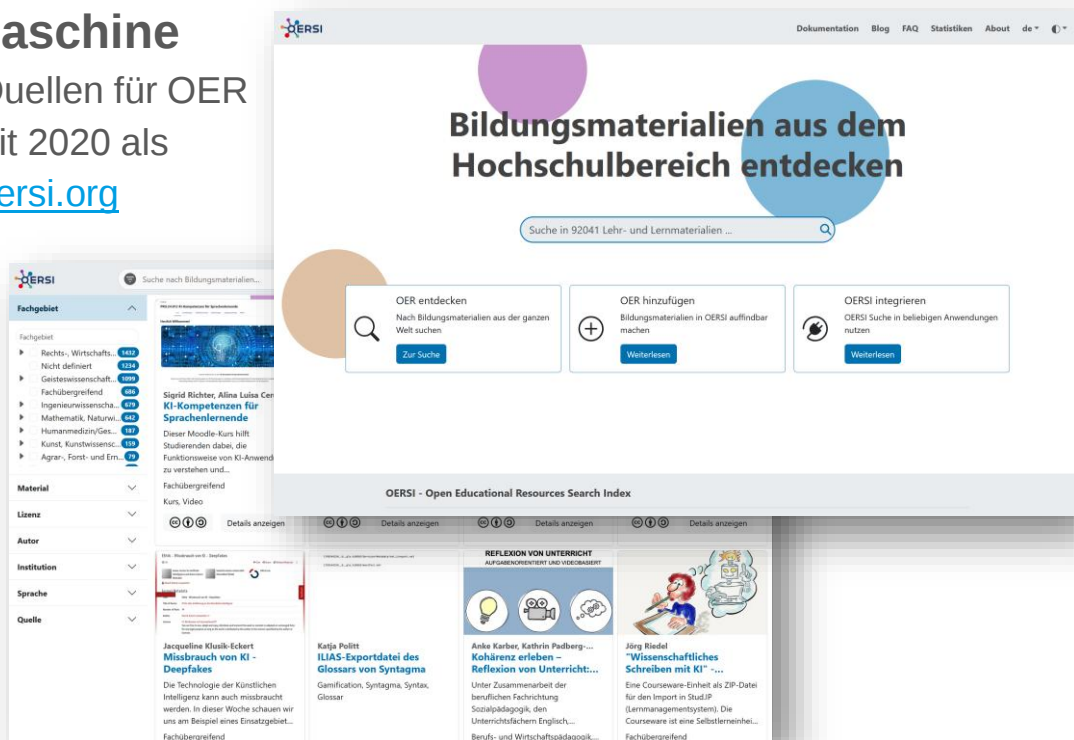
- Zusammenführung verteilter Quellen für OER
- Entwickelt von TIB und hbz seit 2020 als Open Source Lösung <https://oersi.org>

- **Umfang**

- 92K Materialien
- 50 Quellen, u.a. twillo, ZOERR, VCRP, VLC u.v.a.m.

- **Integration**

- Materialsuche u.a. für twillo.de, orca.nrw, oer.eulist.university



Was sind Metadaten?



- **Beschreibung der Materialien**

- Metadaten helfen bei der Auffindbarkeit passender Materialien
- Deskriptiv: Titel, Beschreibung, Zusammenfassung, Schlagwörter
- Administrativ: Autor, Lizenz, Sprache, Fachzuordnung
- Technisch: Format, Version, Änderungsdatum

- **Hier: Fokus auf deskriptiven Metadaten!**

Lea Vilchez Bayer, Lena Mohrschladt...
Unterrichtsentwurf "Die Süßwasserlinse"

Das Unterrichtskonzept „Die Süßwasserlinse - Trinkwasser in Meeresnähe auf Spiekeroog“ ist Teil einer umfassenden Sammlung zu Themen aus der aktuellen...
Mathematik, Naturwissenschaften, Biologi...
Übung, Experiment, Textdokument, Skript...



Details anzeigen

Beschreibung

Das Unterrichtskonzept „Die Süßwasserlinse - Trinkwasser in Meeresnähe auf Spiekeroog“ ist Teil einer umfassenden Sammlung zu Themen aus der aktuellen Meeresforschung und dem Wattenmeer. Es wurde im Rahmen des Seminars „Lehren und Lernen im Lernlabor Wattenmeer“ der Biologiedidaktik an der Carl von Ossietzky Universität Oldenburg gemeinsam von Dozierenden, Studierenden und Schüler:innen entwickelt, erprobt, evaluiert und optimiert. Die Materialien sind für eine 90-minütige Lerneinheit konzipiert und enthalten Experimente, Hintergrundinformationen sowie – falls passend – Elemente der Bildung für Nachhaltige Entwicklung (BNE). Sie richten sich vor allem an Biologielehrkräfte der Sekundarstufen I und II sowie an Fortbildner:innen und Multiplikator:innen. In dieser Serie wird das Thema „Die Süßwasserlinse auf Spiekeroog“ so aufbereitet, dass es besonders für den Unterricht in der Sekundarstufe I geeignet ist. Die Serie umfasst folgende vier Dokumente in jeweils zwei Formaten (jeweils als PDF zum direkten Ausdruck und einmal als Word-Datei zur individuellen Anpassung): - Unterrichtsentwurf zum Thema „Die Süßwasserlinse - Trinkwasser in Meeresnähe auf Spiekeroog“; - ein didaktisch gestaltetes „Forscherheft“ mit detaillierten, verständlichen Aufgabenstellungen für Schüler:innen; - ein „Lösungsheft“ für Lehrkräfte, das die korrekten Antworten bzw. den Erwartungshorizont zu den Aufgaben im Forscherheft enthält; - eine kompakte Sachanalyse des Unterrichtsthemas, die die zentralen fachlichen Hintergründe für eine schnelle Einarbeitung zusammenfasst. Dieses Material ist Bestandteil der Sammlung „Lernlabor Wattenmeer“.

Schlagwort

Lernexperiment (Lernen) Lösungsheft Trinkwasserersparnis (Trinkwasserverbrauch, Trinkwasser)
Aufgabe mit Lösung Unterrichtsmaterial Experiment Meer (Wattenmeer, See)
Watt (Wattenmeer, Intertidalebereich) Spiekeroog <Motiv> Sachanalyse Didaktik (Unterrichtsfächer, Schuldidaktik)
Lernlabor Wattenmeer Forschendes Lernen Trinkwasser <Motiv> Insel (Inseln) Forscherheft
Küste (Meer, Coast) Spiekeroog Süßwasser Süßwasserlinse Wasserverbrauch Lernmaterialien
Bildung für nachhaltige Entwicklung <Unterrichtsfach> Bildung für nachhaltige Entwicklung (BNE) Unterrichtsplanning
Trinkwasserersparnis (Trinkwasser) Arbeitsmaterial Süßwasser (Frischwasser)
Salzwasser-Süßwasser-Grenze (Salzwasser, Süßwasser-Salzwasser-Grenze) Tabellarischer Unterrichtsentwurf
Trinkwasserverbrauch (Trinkwasser)

Institution

Carl von Ossietzky Universität Oldenburg

Sprache

Deutsch

Lizenz



Quelle

twilio

Metadaten in OERSI



- **Zusammentragen von Metadaten aus verschiedenen Quellen**
 - OERSI sammelt die Metadaten aus verschiedenen Quellen für OER ein und vereinheitlicht sie u.a. durch Mapping auf gemeinsame Vokabulare
 - Vollständigkeit und Qualität der Metadaten schwankt zwischen und innerhalb von Quellen
- **Bedeutung guter Metadaten**
 - Besonders wichtig zur inhaltlichen Einordnung sind aussagekräftige Titel und präzise Beschreibungen in Suchergebnissen
 - Qualität bei Freitext-Metadaten wie Beschreibungen sehr unterschiedlich z.B. teilweise Beschreibung von Projektvorhaben oder personeller Beteiligung statt inhaltliche Beschreibung des Materials

Problem: Fehlende Metadaten



- **Fehlende Metadaten erschweren die Auffindbarkeit**

- Materialien werden nicht indiziert
- Beim Filtern werden Materialien nicht angezeigt
- In den Suchergebnissen ist die Relevanz nicht zu erkennen

- **Fehlende Beschreibungen**

- Von 90K Materialien fehlen bei 25K die Beschreibungen
- Von 51K Videos fehlen bei 16K die Beschreibungen
- Von 16K Textbooks fehlen bei 6K die Beschreibungen

- **Weitere Metadaten**

- 11K Materialien haben keine Fachzuordnung
- 6K Materialien haben keine Sprache angegeben

[Vorlesung01](#)

Video



Show Details

[Lars Gadermann, Thomas Maier, Pascal Rommel](#)

[Drehmoment aus der Sicht des Ingenieurwesens](#)

In diesem Video werden Drehmomente aus der Sicht des Ingenieurwesens erklärt. Dabei wird darauf eingegangen, was diese sind, wie diese wirken und wie diese berechnet werden. Das Videotutorial ist als Selbstlernmaterial...

Maschinenbau/-wesen, Verfahrenstechnik,...

Video



Details anzeigen

Ansatz zur Verbesserung der Metadaten



- **Erzeugen beschreibender Metadaten aus Texten via LLM**
 - Fokus auf Open Textbooks und Videos
 - Download der Volltexte von Open Textbooks
 - Transkription der Videos
- **Verwendung von LLMs zum Generieren von**
 - Zusammenfassungen für Detailseiten
 - Kurzbeschreibungen für Suchergebnisse
 - Titeln – zunächst nur versuchsweise
 - Schlagworten für Suche
 - Fachzuordnungen für Filter

Auswahl eines geeigneten LLM



- **Auswahlkriterien**

- Open source
- Bis zu 40B Parameter
- Großes Kontextfenster
- Stark bei Mehrsprachigkeit und Zusammenfassungen

- **Modellvergleich**

- Qwen2.5 7B Instruct
- LLaMA 3.2 8B Instruct
- Phi-3.5

- **Lokaler Betrieb mit Ollama**



Aufbau der Pipeline

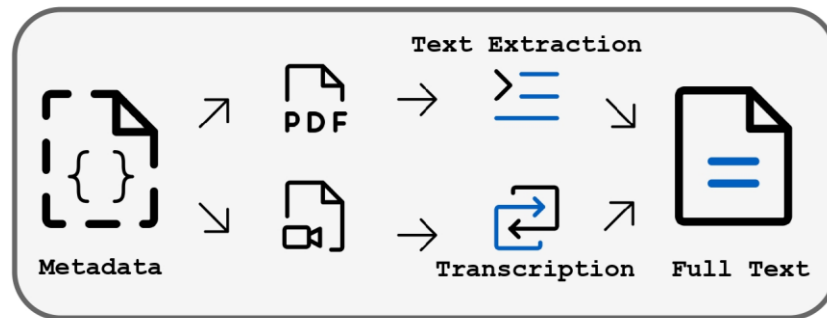


- **Verarbeitungsschritte beim Import von Metadaten**

- Prozesse für Textbooks und Videos
- Wenn bei einem Material beim Import die Beschreibung fehlt
 - Download des Materials
 - Textextraktion bzw. Transkription
 - Generieren von Zusammenfassung, Kurzbeschreibung, Titel, Schlagwörtern, Fachzuordnung

- **Konfigurierbarkeit**

- Anpassung der Selektion
- Anpassung der Prompts
- Anpassung der Modellauswahl



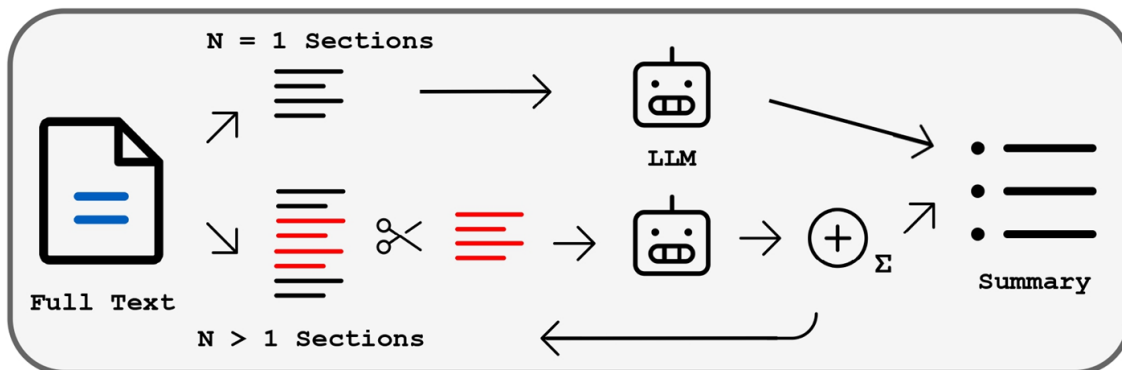
Definition der Testmenge

- **Auswahl einer kleinen Menge an Materialien**
 - Einfach beschreibbare Selektion aus OERSI-Suche z.B. nach Fach, Autor
 - Videos und Textbooks
 - Materialien ohne Beschreibung
 - Ziel: zunächst 20-30 Materialien jeweils von Videos und Textbooks
- **Weitere Aspekte**
 - Länge der Texte und Videos beachten
bzw. herausfinden wo die Grenzen sind
 - Lizenzen berücksichtigen – CC-ND könnte hier ein Problem sein

Anpassung der Verarbeitung

- **Berücksichtigung der Textlänge**

- Zerlegen des Volltextes und Prompts in Token
- Wenn die Anzahl Token von Volltext und Prompt größer als Kontextfenster sind => Zerlegen des Textes in Abschnitte/Kapitel
- Jeden Abschnitt/jedes Kapitel Zusammenfassen
- Alle Zusammenfassungen zu einer Gesamtzusammenfassung vereinen



Halluzinationen und andere Fehler

- **Irreführende oder falsche Informationen**
 - Kann ein einzelnes Beispiel überbewerten
 - Kann wichtige bzw. im Kontext bedeutende Beispiele auslassen
 - Ebenfalls möglich: Probleme bei der Volltext-Extraktion
- **Gelegentliche Rechtschreibfehler im Deutschen**
 - Tritt vor allem in gemischtsprachigen Kontexten auf
 - Falsche Schreibweise von Namen/Begriffen bei Transkripten

Fachzuordnung



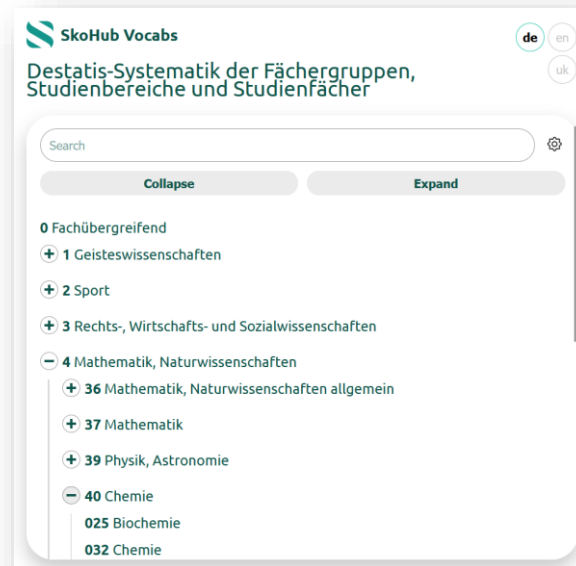
- **Kontrolliertes Vokabular**

- Hochschulfächersystematik DESTATIS (rund 350 Fächer)
- Generator berücksichtigt die ersten beiden Ebenen des Vokabulars

- **Fachzuordnung häufig nicht korrekt**

- Gültige Fächer falsch zugeordnet
- Ungültige Fächer außerhalb des Vokabulars
- Je feiner die Ebene, desto fehlerhafter die Zuordnung

- **Zuordnung zu Studienfächern mit diesem Ansatz bisher nicht ausreichend**



Erste Ergebnisse

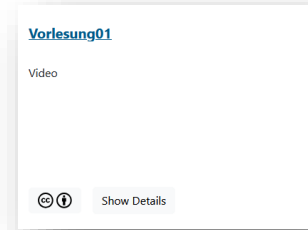


- **Kleinere Mengen bereits mit gutem Ergebnis**

- Rund 40 Textbooks und Videos bereits mit recht guten generierten Beschreibungen
- Generierte Beschreibungen sind als solche gekennzeichnet
- Manuelle angelegte und generierte Beschreibung werden beide untereinander angezeigt
- Anzeige der Titel bisher nur intern zu Testzwecken

- **Qualität**

- Bei gemischtsprachigen Inhalten teils auffällige Rechtschreibfehler bei Eigennamen und Fachbegriffen
- Manuelle Qualitätskontrolle erforderlich



Beschreibung (KI generiert, BETA) "Engineering and Information: Research Skills for Engineers" is an extensive educational resource designed by McMaster University to enhance undergraduate engineering students' research abilities through interactive online modules. These modules cover essential topics such as evidence-based practice, books, web information, journal articles, patents, standards, and citation management. Each module includes engaging elements like videos, quizzes, and practical activities, ensuring a comprehensive understanding of how to effectively find, evaluate, and document various sources in engineering research. This resource is part of the Ontario Ministry of Training, Colleges, and Universities' Open Education Initiative, aiming to support traditional teaching methods with structured, interactive learning experiences that develop both technical and non-technical skills crucial for successful engineering practice.

Kurzbeschreibung (KI generiert, BETA) Engineering and Information: Research Skills for Engineers offers seven interactive online modules to teach essential research skills, including finding, understanding, evaluating, and documenting various sources relevant to engineering practice.

Titel (KI generiert, BETA) Research Skills for Engineers: Navigating Information Sources

Fachzuordnung (KI generiert, BETA) Ingenieurwesen allgemein

Zusammenfassung



- **Aktueller Stand**

- Zusammenfassungen und Kurzbeschreibungen funktionieren schon relative gut, aber Qualitätskontrolle ist sicherzustellen
- Alternative Titel nicht bei allen Fächern sinnvoll
- Insgesamt sehr rechenintensiv, daher Ergebnisse persistent speichern
- Quellcode, Dokumentation und Planung

<https://gitlab.com/oersi/sidre/metadata-optimization>

- **Ausblick**

- Neugenerierung der Metadaten bei geänderten Inhalten
- Skalierung auf größere Mengen an Materialien (derzeit Engpass Infrastruktur)
- Nachnutzung neben vollautomatischen Prozessen bei Massenverarbeitung auch zur Unterstützung bei manueller Einreichung in Repositorien (OA, OER, ...)
- Modulare Open Source Komponente zur einfachen Einbindung in weitere Dienste in Arbeit

Vielen Dank!



Kontakt:

- axel.klinger@tib.eu
- <https://oersi.org/contact>
- <https://openbiblio.social/@oersi>



Creative Commons Attribution 4.0 International
<https://creativecommons.org/licenses/by/4.0/deed.de>