



DFG

DFG-Förderakten als Volltexte erschließen

**Ein kleiner Schritt in der Technik,
mit großem Gewinn für die wissenschaftshistorische Forschung**

29. Archivwissenschaftliches Kolloquium, Marburg, 20.-21. Mai 2025

Jürgen Güdler / DFG

- ▶ KI braucht Daten
- ▶ Die DFG und ihr Informationssystem GEPRIS Historisch
- ▶ Digitalisierung des B-227-DFG-Förderaktenbestands des Bundesarchivs
- ▶ KI in der Erschließung von Förderakten
- ▶ Exkurs: KI in der Analyse von Förderakten
- ▶ Fazit

Künstliche Intelligenz (KI) braucht Daten

▶ KI-Definitionsversuche (zit. nach Wikipedia):

- *Die künstliche Intelligenz [...] ist ein Teilgebiet der Informatik, welches sich mit der Erforschung von Mechanismen des intelligenten menschlichen Verhaltens befasst.* Spektrum der Wissenschaft, Lexikon der Neurowissenschaften
- *„Unter künstlicher Intelligenz (KI) verstehen wir Technologien, die menschliche Fähigkeiten im Sehen, Hören, Analysieren, Entscheiden und Handeln ergänzen und stärken.“* Microsoft Corp.[4]

▶ Diesen und weiteren Definitionen fehlt ein wichtiger „Nebensatz“, nämlich *„...unter Zugriff auf Daten“*

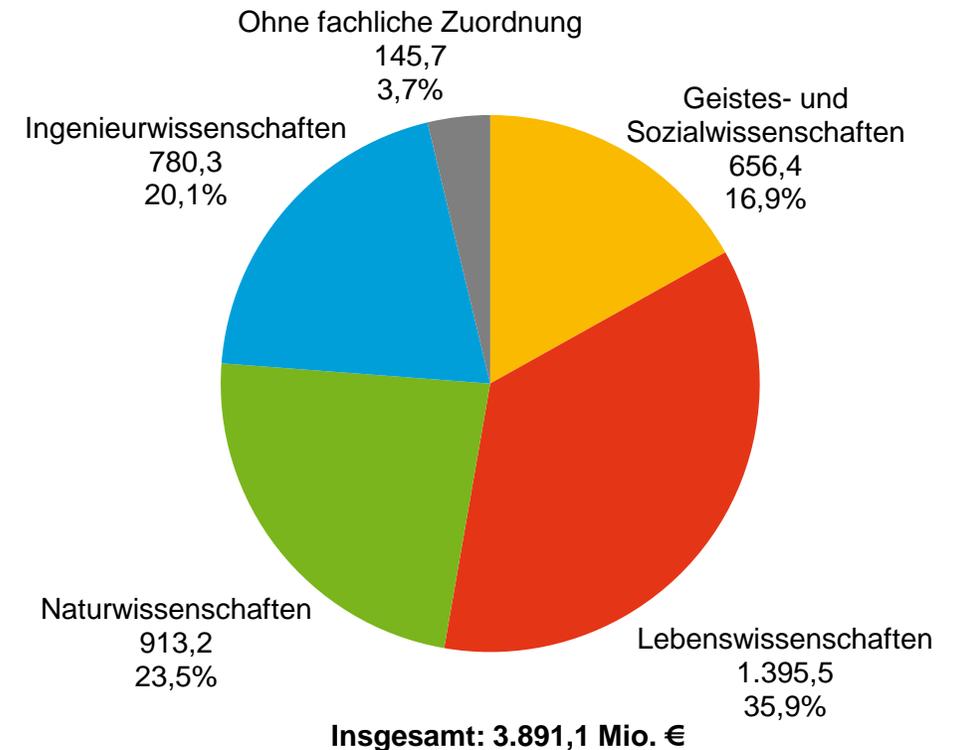
▶ „Daten“ in der Welt der Archive finden sich etwa

- In Form von Katalogeinträgen, Bestandsübersichten, Findbüchern, detaillierten Beschreibungen von häufig genutzten Einzeldokumenten,
- als filterbare Datenbanken mit Ordnungssystemen und Suchfunktionen in Bezug auf Sachaspekte, Personen, Provenienz der Quelle etc.
- und schließlich In Form von **Digitalisaten**, also bild- und – so in diesem Vortrag – volltexterschlossenen Quellen

Die Deutsche Forschungsgemeinschaft

- ▶ Die DFG ist ein deutscher Förderer der Grundlagenforschung, der als Verein organisiert ist
- ▶ Mitglieder sind hauptsächlich Forschungsuniversitäten
- ▶ Mit einem Jahresbudget von rund 3,9 Milliarden Euro ist die DFG ein wichtiger Akteur in der deutschen Forschungslandschaft. Im Jahr 2024 förderte die DFG fast 31.000 neue und laufende Projekte
- ▶ Die DFG ist bestrebt, größtmögliche Transparenz über ihre Aktivitäten zu schaffen. Daten spielen dabei eine wichtige Rolle

Jahresbezogene Bewilligungssumme für 2024 nach Wissenschaftsbereichen (in Mio. € und %)



GEPRIS Historisch

Ausgangspunkt

- ▶ GEPRIS Historisch macht etwa 50.000 Projekte von 13.000 Wissenschaftler*innen recherchierbar
- ▶ Ausgangsbasis: Von einer Forschungsgruppe zur DFG-Geschichte erhobene Metadaten zu diesen Projekten, die auch die Grundlage für das BArch-Findbuch zum Bestand bildeten
- ▶ Eine Stärke des Systems ist, dass es
 - Zugang zum BArch-Findbuch zu diesem DFG-Aktenbestand (R-73) bietet und
 - mit Hilfe eines systematischen Identifier-Systems eine Vielzahl weiterer Quellen erschließt

DFG HISTORISCH 1920 - 1945

SUCHE GEOGRAPHISCHE SUCHE PERSONEN EINRICHTUNGEN THEMEN ÜBER GEPRIS-HISTORISCH

FORSCHUNGSFÖRDERUNG VON 1920 BIS 1945

Entdecken Sie Pionierarbeiten und zu Unrecht Vergessenes, aber auch dunkle Kapitel der Wissenschaftsgeschichte

Suchbegriff eingeben...

1920 1925 1930 1935 1940 1945

PERSONENINDEX

Katalog der in GEPRIS Historisch verzeichneten Wissenschaftlerinnen und Wissenschaftler

Zum Personenindex

EINRICHTUNGSINDEX

Katalog der in GEPRIS Historisch verzeichneten Forschungsstätten

Zu den Einrichtungen

THEMENSEITEN

Themenseiten beleuchten in kurzer Form besondere Aspekte der damaligen Forschungsförderung

Zu den Themenseiten

GEOGRAPHISCHE VERTEILUNG

Finden Sie die Forschungsstätten, an denen DFG-Geförderte damals in Deutschland und weltweit tätig waren.

Zur Karte

SPURENSUCHE IN DER FORSCHUNGSFÖRDERUNG

GEPRIS Historisch erschließt wissenschaftliche Projektbeiträge zwischen 1920 und 1945. Der Film zeigt, wie und warum die Plattform entstand.

Video auf DFG bewegt abspielen

DFG GEPRIS Historisch: Spurensuche in der Forsch... Link kopieren...

Ansehen auf YouTube

© DFG 2024

Startseite | Über GEPRIS Historisch | Impressum | Datenschutz

GEPRIS Historisch

Erschließung des DFG-Aktenbestands (R-73) via Link auf BArch-Findbuch...

- ▶ Beispiel: Akten zum Projekt “Deutsches Volksliedarchiv”, Freiburg (R 73/15401)

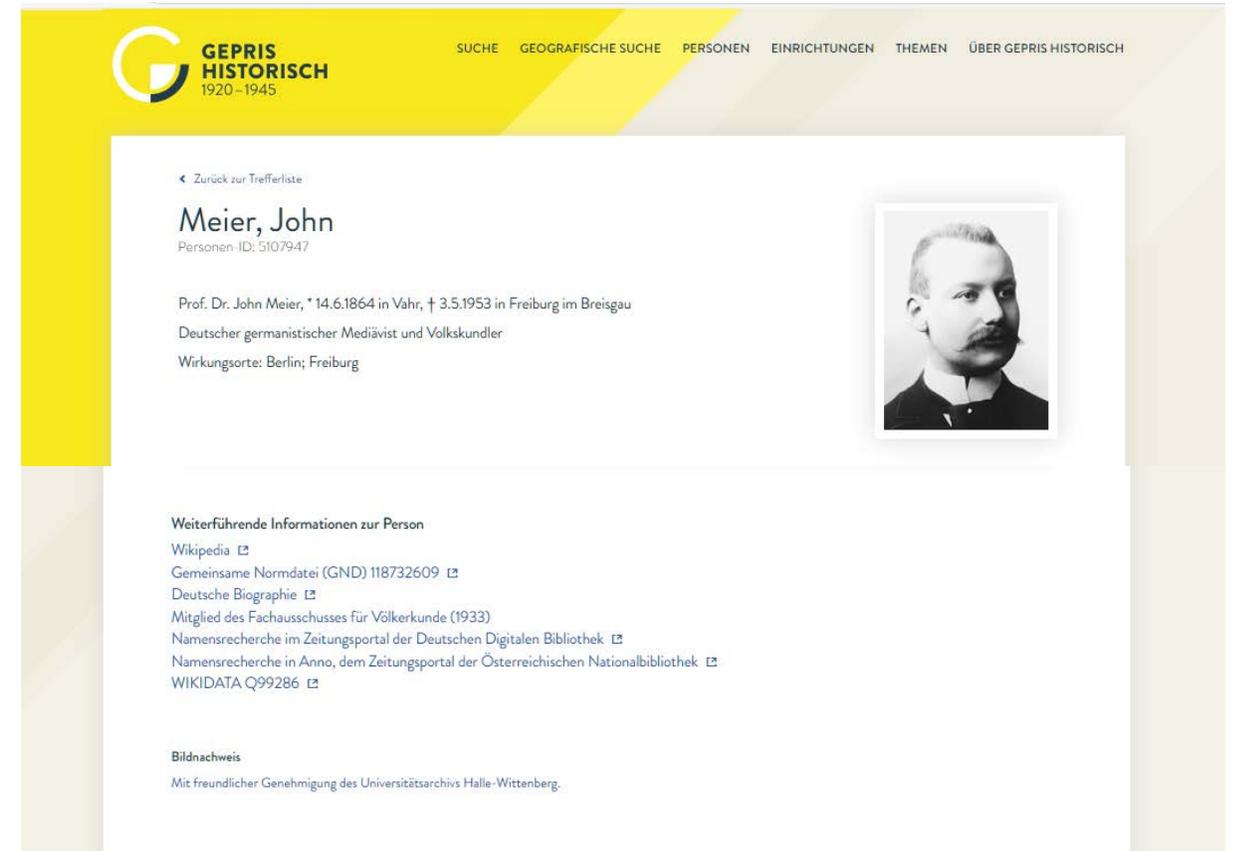
The screenshot shows the search results for 'volkslied' in the GEPRIS Historisch system. The search bar at the top contains 'volkslied' and shows 0/95 results. The results are displayed in a table with columns for 'Archivsignatur', 'Titel und weitere Informationen', and 'Laufzeit'. The first result is checked and shows the archival number R 73/15401. The title is 'Deutsches Volksliedarchiv, Freiburg' and the run dates are 1935 - 1939. The 'Enthält' section provides a detailed description of the contents, including references to various documents and projects related to the 'Deutsches Volksliedarchiv' and the 'Deutsches Volksliedwerk'.

Archivsignatur	Titel und weitere Informationen	Laufzeit
<input checked="" type="checkbox"/> R 73/15401	Deutsches Volkslied archiv, Freiburg Enthält : Mittel für die Arbeiten am Deutschen Volkslied werk (Das Bewilligungsschreiben fehlt, vgl. jedoch Brief der DFG an Prof. Dr. John Meier (Freiburg), 24.7.1935 und Brief von Meier an die DFG, 4.2.1936)., Deutsches Volkslied archiv Freiburg, Prof. Dr. John Meier (geb. 14.06.1864), Vorsitzender des Verbandes deutscher Vereine für Volkskunde, Neuere Philologie.- Sachbeihilfe, 02.05.1935; Forschungsstipendium für den Lehramtsassessor Richard Volk für die Zeit vom 1.9.1935 (?) bis zum 31.3.1936 für die Mitarbeit am Deutschen Volkslied werk, Deutsches Volkslied archiv Freiburg, Prof. Dr. John Meier (geb. 14.06.1864), Vorsitzender des Verbandes deutscher Vereine für Volkskunde, Neuere Philologie.- Sachbeihilfe, 09.09.1935; Mittel für die musikwissenschaftlichen Arbeiten am Deutschen Volkslied werk, Deutsches Volkslied archiv Freiburg, Prof. Dr. John Meier (geb. 14.06.1864), Vorsitzender des Verbandes deutscher Vereine für Volkskunde, Neuere Philologie.- Sachbeihilfe, 23.05.1936; Mittel für die musikwissenschaftlichen Arbeiten am Deutschen Volkslied werk (Gleichzeitig wurde ein von Meier beantragtes Forschungsstipendium für Walter Wiora abgelehnt, da dieser noch nicht promoviert hatte. Die Bewilligungssumme wurde von Meier jedoch zur Bezahlung Wioras vermutlich für die Zeit vom 1.1.1936 bis zum 31.3.1936 verwendet. Vgl. auch R 73/ 15747)., Deutsches Volkslied archiv Freiburg, Prof. Dr. John Meier (geb. 14.06.1864), Vorsitzender des Verbandes deutscher Vereine für Volkskunde, Neuere Philologie.- Sachbeihilfe, 04.11.1935; Mittel für das deutsche Volkslied archiv und alle damit zusammenhängenden Arbeiten, sowie für die volkskundliche Bibliographie (Rechnungsjahr 1937/38). U.a. dienten die Mittel zur Bezahlung der Hilfskräfte Dr. Erich Seemann (geb. 15.1.1888) und Fräulein Dr. Sascha Wigenroth (geb. 23.2.1910). Darüber hinaus wurden aus dieser Summe die Stipendien für Meier, Quellmalz, Wilhelm Heiske (geb. 19.10.1904), Walter Wirora (geb. 30.12.1906), Helbron und Drüner bezahlt., Deutsches Volkslied archiv Freiburg, Prof. Dr. John Meier (geb.	1935 - 1939

GEPRIS Historisch

.... und andere Quellen etwa zu beteiligten Personen via Identifier-Schnittstellen

- ▶ Beispiel: Akten zum Projekt “Deutsches Volksliedarchiv”, Freiburg (R 73/15401)
- ▶ Zu beteiligten Personen erschließt GEPRIS Historisch direkt eine Reihe weiterer Quellen (hier am Beispiel von John Meier):
 - Wikipedia und Wikidata
 - Gemeinsame Normdatei (GND) (Deutsche Nationalbibliothek)
 - Deutsche Biographie (der [Historischen Kommission](#) bei der [Bayerischen Akademie der Wissenschaften](#) und der Bayerischen Staatsbibliothek ([BSB](#)))
 - Zwei Zeitungs-Portale (Anno/Österreich und Zeitungportal der Deutschen Digitalen Bibliothek)



The screenshot shows the GEPRIS Historisch website interface. At the top, there is a navigation bar with the logo and the text "GEPRIS HISTORISCH 1920-1945". To the right of the logo are navigation links: "SUCHE", "GEOGRAFISCHE SUCHE", "PERSONEN", "EINRICHTUNGEN", "THEMEN", and "ÜBER GEPRIS HISTORISCH". Below the navigation bar, the main content area features a profile for "Meier, John" with a "Personen-ID: 5107947". A black and white portrait of John Meier is displayed on the right. The profile text includes: "Prof. Dr. John Meier, * 14.6.1864 in Vahr, † 3.5.1953 in Freiburg im Breisgau", "Deutscher germanistischer Mediävist und Volkskundler", and "Wirkungsorte: Berlin; Freiburg". Below the profile, there is a section titled "Weiterführende Informationen zur Person" with links to "Wikipedia", "Gemeinsame Normdatei (GND) 118732609", "Deutsche Biographie", "Mitglied des Fachausschusses für Völkerkunde (1933)", "Namensrecherche im Zeitungportal der Deutschen Digitalen Bibliothek", "Namensrecherche in Anno, dem Zeitungportal der Österreichischen Nationalbibliothek", and "WIKIDATA Q99286". At the bottom, there is a "Bildnachweis" section with the text "Mit freundlicher Genehmigung des Universitätsarchivs Halle-Wittenberg."

GEPRIS Historisch

.... und andere Quellen etwa zu beteiligten Personen via Identifier-Schnittstellen

- ▶ Über die GND- und Wikidata-ID werden Personen eindeutig identifizierbar und via Wikidata und Deutsche-Biographie indirekt dutzende weitere Quellen erschlossen:
- ▶ Am Beispiel John Meier:
 - Etwa 40 Referenzen bei Wikidata, z.B.
 - Deutsches Literaturarchiv Marbach (Kallias)
 - Catalogus Professorum Halensis ID
 - Historisches Lexikon der Schweiz
 - Open Library (Publikationen von JM online)
 - DigitaltMuseum (schwedisch-norwegisches Dig.-Projekt)
 - Etwa 20 Referenzen bei deutsche-biographie, z.B.
 - Nachlassdatenbank beim Kalliope Verbund
 - Verschiedene Literatur-Kataloge

deutsches
literatur
archiv marbach

MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG



idium Forschung Weiterbildung Karriere Presse International

Catalogus Professorum Halensis

Der Catalogus Professorum Halensis ist ein biographisches Onlinelexikon zu den Professoren der Universität Halle. Zurzeit befinden sich über 1.050 Einträge in der Datenbank. Für die Jahre von 1918 bis 1945 sind alle Professoren dokumentiert. Die Zeiträume davor und danach sind teilweise bearbeitet und befinden sich im Aufbau. Für einen systematischen Zugriff ist das Lexikon in Zeitepochen eingeteilt. Über die Suchfunktion haben Sie die Möglichkeit in allen Einträgen nach Namen oder Begriffen zu recherchieren.



HISTORISCHES
LEXIKON
der SCHWEIZ
HLS



D M DigitaltMuseum

 KALLIOPE-VERBUND

KI braucht Daten

Zwischenfazit

- ▶ Archive stellen keine Dateninseln dar
- ▶ Ihre Bestände lassen sich mit Persistenten Identifiern (PID) in Beziehung setzen
- ▶ Eine Investition in die digitale Erschließung einer Dokumentensammlung eines Archivs ist eine Investition in eine vernetzte Datenwelt

Jüngere DFG-Förderakten im Bestand des Bundesarchivs (BArch)

Der B-227-Bestand

- ▶ Auf Basis eines 1987 geschlossenen Depositatvertrags archiviert das BArch in den Jahren 1988 bis 2005 übergebene DFG Förderakten zu bewilligten und abgelehnten Anträgen
- ▶ Das BArch-Findbuch fußt auf Daten der DFG, macht Informationen wie Name und Ort des Antragstellenden, Titel des Projekts, Entscheidungs-Jahr recherchierbar
- ▶ Mit der Umstellung auf rein elektronische Aktenführung bei der DFG 2005 wurde die Lieferung eingestellt

The screenshot shows the Invenio search interface. At the top, there is a navigation bar with links: Navigierende Suche, Suche, Suchergebnis (highlighted), Merkzettel/Bestellen, Benutzerdaten, Hilfe, and Kontakt. Below the navigation bar, there is a search bar with the text 'sortieren nach: Klassifikationsvorgabe (Standard)'. The main content area is divided into two columns: 'Archivsignatur' and 'Titel und weitere Informationen'. The first entry is a checkbox followed by the text 'Statistische Physik, Weiche Materie, Biologische Physik, Nichtlineare Dynamik'. Below this, it says 'Unterlagenart: Sachakte' and provides links for 'Im Kontext anzeigen' and 'Link kopieren'. The second entry is a checkbox followed by the text 'B 227/65656'. Below this, it says 'Bd. 888 / 154: wachstum und verfestigung physikalischer strukturen'. Below this, it provides detailed information: 'Bandnummer: 154', 'Aktenzeichen/Registriernr.: HA 332/38-1', 'Bemerkung: Ort der Antragsstellung: Stuttgart', 'Name: Haken, Hermann', 'Geburtsdatum: 1927', 'Bemerkung: Förderung: bewilligt', 'Unterlagenart: Sachakte', 'Benutzungsort: Koblenz', and 'Recherchebeschränkungen: Personenbezogene Schutzfrist (Name)'. Below this, it provides links for 'Im Kontext anzeigen' and 'Link kopieren'.

Anlass der Digitalisierung

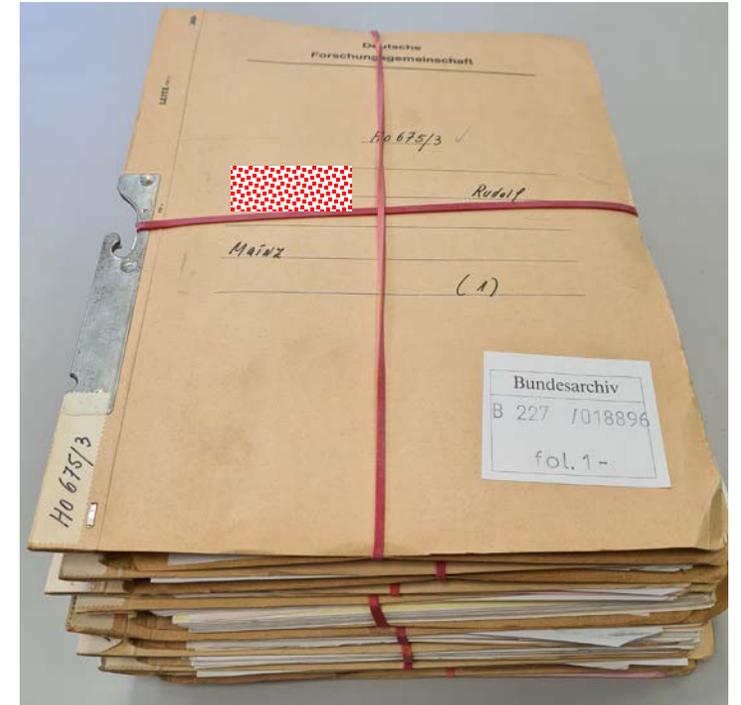
- ▶ Der Bestand umfasst ca. 300.000 Akten auf 4 Regalkilometern, Das Verhältnis von Lagerkosten zur Nutzungshäufigkeit erschien zunehmend unverhältnismäßig. Lösungsszenarien:
 - Idee I: Ausschließliche Archivierung einer Zufalls-Stichprobe (5 Prozent) an Akten
 - Idee II: Digitalisierung der Akten in Form des ersetzenden Scannens
- ▶ Motive BArch:
 - Passung zur allgemeinen Digitalisierungs-Strategie des BArch
 - Perspektive Online-Bereitstellung macht Vor-Ort-Besuche nicht mehr zwingend
 - Raumgewinn
- ▶ Motive DFG:
 - Bewahrung der Akten für die wissenschaftshistorische Forschung
 - Potenzierte Nutzbarkeit der Akten via Volltexterschließung

- ▶ DFG trägt die Kosten der Digitalisierung und managt den Prozess (in Rücksprache mit BArch) (Ausschreibung, Anbieterauswahl, Logistik, Qualitätssicherung etc.)
- ▶ Prozess (Skizze):
 - Chargenweise Abholung der Akten beim BArch
 - In DFG-GS Aufarbeitung der Akten und ihrer Daten
 - Entfernen nicht regelkonform beigefügter Dokumente (insb. Kopien von Zeitschriftenaufsätzen)
 - Aussondern minderwichtiger Anträge (insb. Kleinförderung, Druckbeihilfen)
 - Prüfung von Stammdaten zwecks Verbesserung des Findbuchs (Lückenschließung, Korrekturen)
 - Abholung und Digitalisierung durch DL (in Berlin)
 - Stichprobenartige QS-Prüfung der Digitalisate durch Projektteam und Abnahme
 - Chargenweise Auslieferung der Digitalisate an das BArch, dort eigenständige weitere Erschließung und Bereitstellung

Inhalt einer Akte

► Typische Inhalte einer DFG-Förderakte

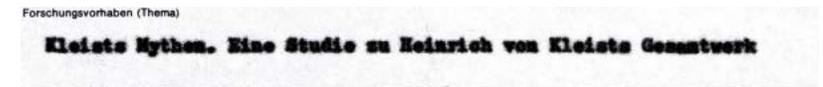
- Einleitendes Datenblatt (Stammdaten)
 - Antrag
 - Korrespondenz
 - Gutachten (idR zwei)
 - GS-interne Prozessdokumente (Entscheidungsvorlagen, Beteiligung ausgewählter Abteilungen etc.)
 - Abschlussbericht
- Nicht Teil der Akte sind Abrechnungs-Dokumente (in sog. Nebenakten (nicht archiviert))
- Die Digitalisierung resultiert in aktenspezifischen Ordnern mit je einer PDF- und TXT-Datei (für den OCR-erkannten Text)



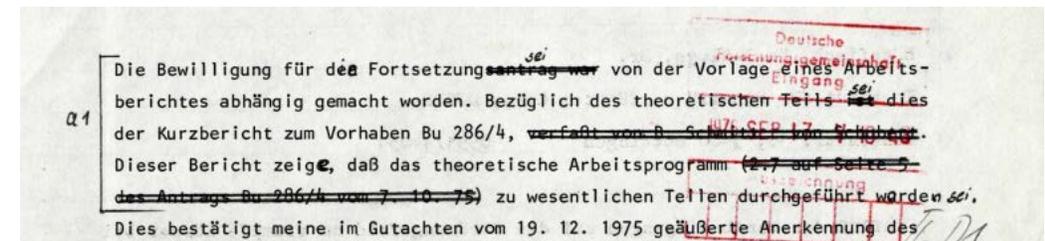
KI in der Erschließung

Überblick

- ▶ Der Fokus des Projekts liegt auf der Volltexterschließung der Akten
- ▶ Herausforderungen
 - Maschinell schlecht lesbare Vorlagen (dritter Schreibmaschinen-Durchschlag auf dünnem Papier, Handschriften, unterbrochener Text (Stempel, texttangierende Unterstreichungen etc.)
 - Mängel einfaches OCR: Formatierungen ((Spalten)-Satz, Überschriften, Silbentrennung, Handschriften etc.) nicht erkannt, mathematische und chemische Formeln, Überschreibungen u.a.)
- ▶ Wie kann KI hier helfen?



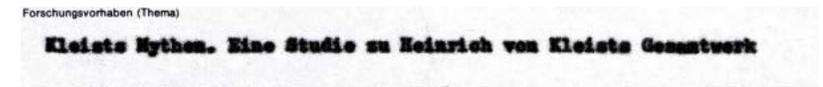
Quelle: 4220



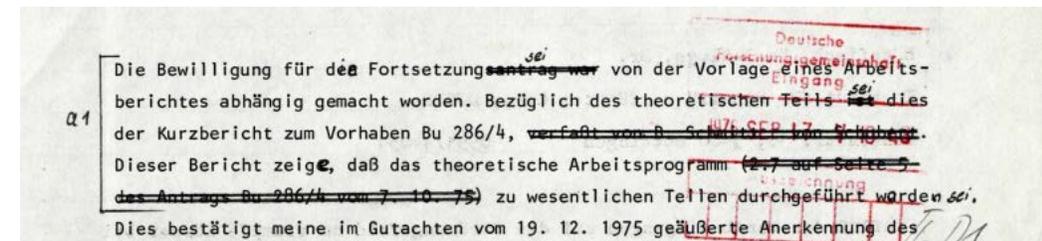
Quelle: 4627

Bildgebung am nächsten am Original

- ▶ Der Fokus des Interesses der DFG liegt auf der Volltexterschließung der Akten...
- ▶ ...aber dem Bild kommt gleichwohl das höchste Gewicht zu:
 - Ist am nächsten am Original
 - Lässt sich dank „menschlicher Intelligenz“ auch unter komplexen Herausforderungen lesen
 - Erschließt Fotos, Zeichnungen und Formeln (chemisch, mathematisch) etc.
- ▶ Ist für die Qualitätsbewertung der Volltexterkennung unverzichtbar
- ▶ Und ist das maßgeblich relevante Datenmaterial für die KI



Quelle: 4220

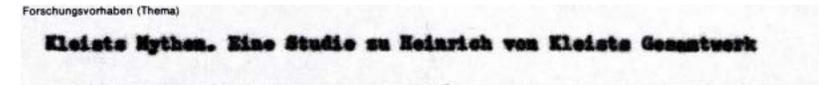


Quelle: 4627

Bildgebung am nächsten am Original

Unterschied zwischen einfacher OCR- und KI-Texterkennung (am Beispiel ChatGPT)

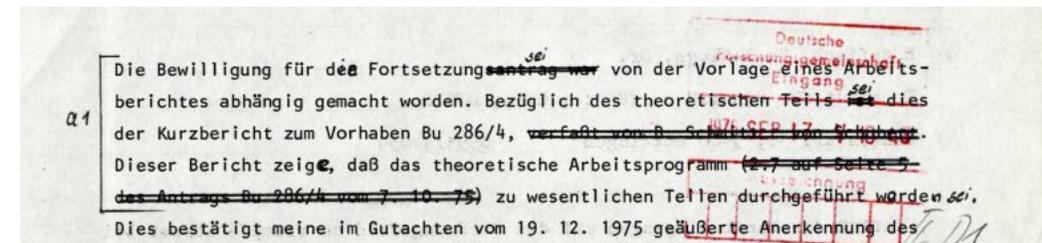
- ▶ Einfaches OCR: itt
- ▶ ChatGPT: Kleists Mythos. Eine Studie zu Heinrich von Kleists Gesamtwerk



Quelle: 4220

- ▶ Einfaches OCR: Die Bewilligung für die Fortsetzung von der Vorlage:efres`Ärmitts-berichtetes abhängig gemacht worden. Bezüglich des theoretisiererr Te`ii.es a?
der Kurzbericht zum Vorhaben Bu 286/4, .,
Dieser Bericht zeige, daß das theoretische Arbeitsprogartinr zu wesentl icken Teil 1-en ,dur--chge L .ra, rde d61. .
, i 4
bestätigt meine im Gutachten vom 19. 12. 1975 geaußß r1e 5nerkenuti d`s` 27,

- ▶ ChatGPT: Die Bewilligung für die Fortsetzung sei war von der Vorlage eines Arbeitsberichtes abhängig gemacht worden. Bezüglich des theoretischen Teils sei dies der Kurzbericht zum Vorhaben Bu 286/4, verfasst am 01. SEP. 74, hinreichend
Dieser Bericht zeige, dass das theoretische Arbeitsprogramm (s. auf Seite des Antrags Bu 286/4 vom 10.10.73) zu wesentlichen Teilen durchgeführt worden sei.
Dies bestätigt meine im Gutachten vom 19.12.1975 geäußerte Anerkennung des



Quelle: 4627

Vergleich Vorlage und einfache OCR-Erkennung

Die Förderungsmittel sind möglichst im Wege der Verwaltungshilfe über die zuständige Amtskasse der Hochschulverwaltung oder anderer staatlicher Stellen oder Kassen der MPG oder Fraunhofer-Gesellschaft abzuwickeln -: Die Mittel sind', bei Bedarf mit DPG-Vordruck 141;03 abzurufen Für die Verwendung und Abrechnung gelten die beiliegenden Richtlinien,,,(DFG7Vordruck 2.'01), die Bestandteil dieser Bewilligung sind

Die Deutsche Forschungsgemeinschaft legt Wert '-darauf, daß wichtige Ergebnisse veröffentlicht werden, und erbittet ein Belegexemplar. Werden die Ergebnisse nicht in einer regulären-Veröffentlichung; die über den Buchhandel zugänglich, ist, sondern in Form eines Forschungsberichts (Report) bekanntgegeben, so bittet die Deutsche Forschungsgemeinschaft, je ein Exemplar an die zentrale Sammelstelle für Forschungsberichte bei der Teöhniscnen,Informationsbibliothek, 'Weif engarten 1 B, 3 Hannover, und an die zuständige Hochschulbibliothek abzuliefern Liegt ein solcher Forschungsbericht,:oder eine reguläre Veröffentlichung vor,' so braucht der Bericht an die Deutsche Forschungsgemeinschaft nur solche Mitteilungen über .Gang und Erfolg der Arbeiten zu enthalten, die für .eine.Veröffentlichung nicht oder, noch nicht ,geeignet sind

Die Annahme dieser Beihilfe verpflichtet "Sie der Deutschen schaft Mitteilung zu machen, wenn •Sie atis ;:der;unterstützten.Forschungsarbeit unmittelbar wirtschaftlichen Gewinn ziehen, Oder' wenn Sie die Ergebnisse 'der Forschung zum -Patent oder zur Erlangung,:anderer-,gewerblicher :Schutzrechte:an+. ,i. melden'. Als wirtschaftlicher Gewinn gelten nicht Einnahmen aus Publikationen (Vorträge, Aufsätze, Bücher usw..).

91249001 0-5 101 11 0030A

Die Förderungsmittel sind möglichst im Wege der Verwaltungshilfe über die zuständige Amtskasse der Hochschulverwaltung oder anderer staatlicher Stellen oder Kassen der MPG oder Fraunhofer-Gesellschaft abzuwickeln. Die Mittel sind bei Bedarf mit DPG-Vordruck 141.03 abzurufen. Für die Verwendung und Abrechnung gelten die beiliegenden Richtlinien (DFG-Vordruck 2.01), die Bestandteil dieser Bewilligung sind.

Die Deutsche Forschungsgemeinschaft legt Wert darauf, daß wichtige Ergebnisse veröffentlicht werden, und erbittet ein Belegexemplar. Werden die Ergebnisse nicht in einer regulären Veröffentlichung, die über den Buchhandel zugänglich ist, sondern in Form eines Forschungsberichts (Report) bekanntgegeben, so bittet die Deutsche Forschungsgemeinschaft, je ein Exemplar an die zentrale Sammelstelle für Forschungsberichte bei der Technischen Informationsbibliothek, Weifengarten 1 B, 3 Hannover, und an die zuständige Hochschulbibliothek abzuliefern. Liegt ein solcher Forschungsbericht oder eine reguläre Veröffentlichung vor, so braucht der Bericht an die Deutsche Forschungsgemeinschaft nur solche Mitteilungen über Gang und Erfolg der Arbeiten zu enthalten, die für eine Veröffentlichung nicht oder noch nicht geeignet sind.

Die Annahme dieser Beihilfe verpflichtet Sie, der Deutschen Forschungsgemeinschaft Mitteilung zu machen, wenn Sie aus der unterstützten Forschungsarbeit unmittelbar wirtschaftlichen Gewinn ziehen, oder wenn Sie die Ergebnisse der Forschung zum Patent oder zur Erlangung anderer gewerblicher Schutzrechte anmelden. Als wirtschaftlicher Gewinn gelten nicht Einnahmen aus Publikationen (Vorträge, Aufsätze, Bücher usw.).

Die Deutsche Forschungsgemeinschaft ist berechtigt, aus den erzielten Gewinnen die Rückzahlung ihrer Beihilfe zuzüglich eines angemessenen Zinsausgleichs zu verlangen.

Die Deutsche Forschungsgemeinschaft behält sich vor, die Sachbeihilfe zu widerrufen, wenn wichtige Gründe dazu Anlaß geben.

Die Beihilfe wird zurückgezogen, wenn sie ein Jahr nach Bewilligungsdatum weder ganz noch teilweise in Anspruch genommen wurde.

Die Deutsche Forschungsgemeinschaft wünscht Ihnen für Ihre Arbeit guten Erfolg.

2. Gerätebeschaffung durch Beihilfempfänger
 Gerätebeschaffung oder -übertragung durch I 6

3. Fachreferat

Vermerk der Entscheidung 26.8.79 C | 2518 JV
 Durchschrift der Bewilligung an Gutachter usw. absenden
 Anlagen zurück an Antragsteller
 Mitteilung der Gutachterhinweise

4. Wv am 10. November 1979 (Berichtstermin) 2 d. A.

Wiederholungsantrag
mündl. Erinnerung 11/12 1979 (Wiederholung)
Kurz vor 3 Monate 28.11.79
T 22.11.79
X keine wv. 8 Monate 29.11.79

gez. Petersen
Dr. Petersen

Vergleich einfache OCR-Erkennung und KI-Texterkennung

Einlesen des OCR-Textes mit dem Auftrag, die zerstückelten Elemente zu reparieren

Die Förderungsmittel sind möglichst im Wege der Verwaltungshilfe über die zuständige Amtskasse der Hochschulverwaltung oder anderer staatlicher Stellen oder Kassen der MPG oder Fraunhofer7-GeSellsChaft abzuwickeln -: Die Mittel sind' . , bei Bedarf mit DPG-Vordruck 141.;03 abzurufen Für die Verwendung Una Abrechnung gelten ,die beiliegenden:RiChtlinien,,,.(DFG7Vordruck 2.'01), die Bestandteil dieser Bewilligung sind

Die Deutsche Forschungsgemeinschaft legt Wert '-darauf, daß wichtige Ergebnisse veröffentlicht werden, und erbittet ein Belegexemplar. Werden die Ergebnisse nicht in einer regulären-Veröffentlichung; die `über den Buchhandel zugänglich, ist, sondern in Form eines Forschungsberichts (Report) bekanntgegeben, so bittet die Deutsche Forschungsgemeinschaft, je ein Exemplar an die zentrale Sammelstelle für Forschungsberichte .bei der Teöhniscnen,Informationsbibliothek, ' Weif engarten 1 B, 3 Hannover, und an die zuständige Hochschulbibliothek abzuliefern Liegt ein solcher Forschungsbericht,:oderceine reguläre Veröffentlichung vor,` so braucht der Bericht an die Deutsche Forschungsgemeinschaft nur. solche Mitteilungen über .Gang und Erfolg der Arbeiten zu enthalten, die für .eine.Veröffent- lichtung nicht oder, noch nicht ,geeignet sind

Die Annahme dieser Beihilfe verpflichtet "Sie der Deutschen schaft Mitteilung zu machen, wenn •Sie atis_::der;unterstützten.Forschungsarbeit unmittelbar wirtschaftlichen Gewinn ziehen, Oder' wenn Sie die Ergebnisse 'der Forschung zum -Patent oder zur Erlangung,:anderer-,gewerblicher :Schützrechte:an+.,.i. melden'. Als wirtschaftlicher Gewinn gelten nicht Einnahmen aus Publikationen (Vorträge, Aufsätze, Bücher usw..).

Die Förderungsmittel sind möglichst im Wege der Verwaltungshilfe über die zuständige Amtskasse der Hochschulverwaltung oder anderer staatlicher Stellen oder Kassen der MPG oder Fraunhofer-Gesellschaft abzuwickeln. Die Mittel sind bei Bedarf mit DFG-Vordruck 14.103 abzurufen. Für die Verwendung und Abrechnung gelten die beiliegenden Richtlinien (DFG-Vordruck 2.01), die Bestandteil dieser Bewilligung sind.

Die Deutsche Forschungsgemeinschaft legt Wert darauf, daß wichtige Ergebnisse veröffentlicht werden, und erbittet ein Belegexemplar. Werden die Ergebnisse nicht in einer regulären Veröffentlichung, die über den Buchhandel zugänglich ist, sondern in Form eines Forschungsberichts (Report) bekanntgegeben, so bittet die Deutsche Forschungsgemeinschaft, je ein Exemplar an die zentrale Sammelstelle für Forschungsberichte bei der Technischen Informationsbibliothek, Welfengarten 1 B, 3 Hannover, und an die zuständige Hochschulbibliothek abzuliefern. Liegt ein solcher Forschungsbericht oder eine reguläre Veröffentlichung vor, so braucht der Bericht an die Deutsche Forschungsgemeinschaft nur solche Mitteilungen über Gang und Erfolg der Arbeiten zu enthalten, die für eine Veröffentlichung nicht oder noch nicht geeignet sind.

Die Annahme dieser Beihilfe verpflichtet Sie, der Deutschen Forschungsgemeinschaft Mitteilung zu machen, wenn Sie aus der unterstützten Forschungsarbeit unmittelbar wirtschaftlichen Gewinn ziehen, oder wenn Sie die Ergebnisse der Forschung zum Patent oder zur Erlangung anderer gewerblicher Schutzrechte anmelden. Als wirtschaftlicher Gewinn gelten nicht Einnahmen aus Publikationen (Vorträge, Aufsätze, Bücher usw.).

► Handschriftliches Original:

Zur Frage der klinischen Relevanz dieses Verfahrens sind im Antrag auf Seite 7, 2. Absatz einige Möglichkeiten angedeutet. Entscheidend ist nach meiner Auffassung die Absicherung, daß die im Arbeitsprogramm (Antrag, S. 9 + 10) genannten Maßnahmen zur maximalen Meßgenauigkeit führen. Die jeweilige Überprüfung muß entsprechend dem Versuchsplan (Antrag, Anlage) die Genauigkeit abgrenzen lassen.

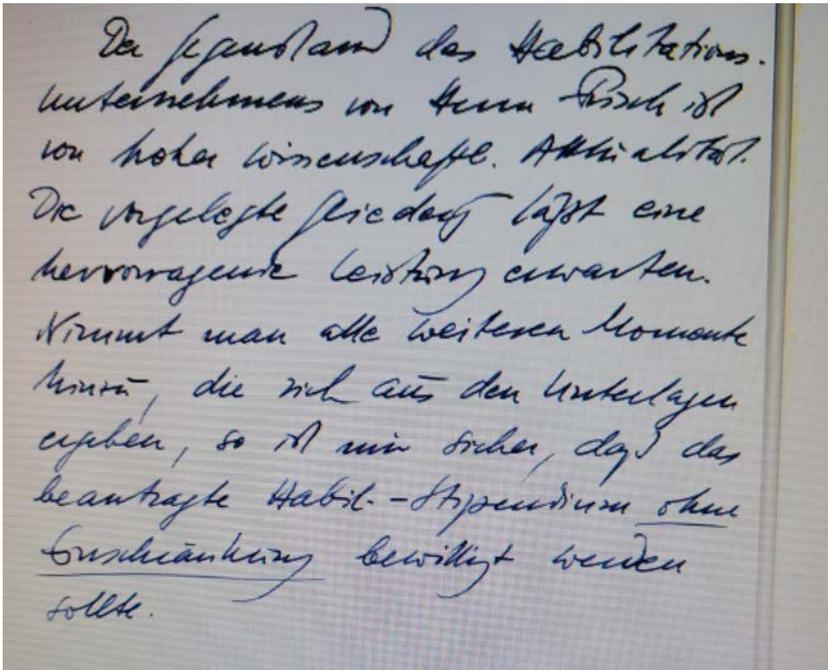
Quelle: 6626

► Übertragung ChatGPT:

Zur Frage der klinischen Relevanz dieses Verfahrens sind im Antrag auf Seite 7, 2. Absatz einige Möglichkeiten angedeutet. Entscheidend ist nach meiner Auffassung die **Klärung**, daß die im Arbeitsprogramm (Antrag S. 9 + 10) genannten Maßnahmen zur maximalen Meßgenauigkeit führen. Die jeweilige Überprüfung muß entsprechend dem Versuchsplan (Antrag, **Beleg**) die Genauigkeit abgrenzen lassen.

KI in der Erschließung

► Handschriftliches Original:



Quelle: 8879

► ...aber einmal auf der falschen Spur (Kommentar ChatGPT: „Der Text scheint sich auf die wirtschaftliche Situation eines Unternehmens zu beziehen, möglicherweise im Kontext einer Genehmigung oder finanziellen Bewertung“), wird diese konsequent verfolgt.

► Lösung:

Wird im Prompt erklärt, dass es sich um einen Text handelt, in dem es um die Förderung wissenschaftlicher Projekte geht, verbessert sich die Texterkennung.

► Übertragung ChatGPT:

Der Zustand des Geschäftsbetriebs unternehmens um Herrn Bisch ist von hoher wirtschaftl. Aktivität. Die vorgelegte Bilanz lässt eine hervorragende Leistung erwarten. Nimmt man alle weiteren Momente hinzu, die wir aus den Unterlagen erheben, so ist mit Sicherheit, dass die beantragte Habib-Bewilligung ohne Einschränkung bewilligt werden sollte.

KI in der Erschließung

Mathematische Formeln (aus Abschlussberichten)

Bild

$$\phi_1(x,y) = \sum_{n=1}^{\infty} \frac{1}{\alpha_n^2} (\Lambda_n^* e^{-\eta_1 \alpha_n y} + \Xi_n^* e^{-\eta_2 \alpha_n y}) \sin \alpha_n x$$
$$\text{mit } \eta_1 = \sqrt{\frac{D_x}{D_{xy}} + \sqrt{\frac{D_x^2}{D_{xy}^2} - \frac{D_x}{D_y}}}$$
$$\eta_2 = \sqrt{\frac{D_x}{D_{xy}} - \sqrt{\frac{D_x^2}{D_{xy}^2} - \frac{D_x}{D_y}}}$$

Quelle: 15930

Transkription ChatGPT

$$\phi_1(x,y) = \sum_{n=1}^{\infty} \frac{1}{\alpha_n^2} (\Lambda_n^* e^{-\eta_1 \alpha_n y} + \Xi_n^* e^{-\eta_2 \alpha_n y}) \sin(\alpha_n x)$$

$$\eta_1 = \sqrt{\frac{D_x}{D_{xy}} + \sqrt{\frac{D_x^2}{D_{xy}^2} - \frac{D_x}{D_y}}}$$

$$\eta_2 = \sqrt{\frac{D_x}{D_{xy}} - \sqrt{\frac{D_x^2}{D_{xy}^2} - \frac{D_x}{D_y}}}$$

Prompt: Hast Du eine Idee, was mit der Formel berechnet wird?

Antwort: Ja! Die Struktur der Formel deutet darauf hin, dass sie eine Lösung für eine partielle Differentialgleichung (PDE) beschreibt, die häufig in der Elastizitäts- oder Diffusionsmechanik vorkommt.

Es folgt: „Ich erkläre meine Interpretation Schritt für Schritt: [...]"

KI in der Erschließung

Potentielle Maßnahmen zur Anreicherung des Findbuchs

- ▶ Bisher enthält das BArch-Findbuch zu den Akten nur wenige, von der DFG bereitgestellt Stammdaten (s.o.)
- ▶ Mit KI ist es möglich, diese Daten um eine Vielzahl von Informationselementen zu erweitern:
 - Erstellung eines Abstracts zu den Zielen eines Forschungsvorhabens (deutsch und englisch)
 - Verschlagwortung und thematische Klassifikation von Digitalisaten
 - Recherche und Zuordnung GND-ID und Wikidata-ID zu Antragstellenden
 - Named entity recognition, etwa
 - Benennung der wiss. Mitarbeiter*innen (erweitert: mit ihren Aufgaben)
 - Angaben zu international beteiligten weiteren Wissenschaftler*innen (mit Herkunftsländern)
 - Angaben zu den begutachtenden Personen
 - Formale Hinweise (z.B. Enthält die Akte einen Abschlussbericht?)
 - Listen der aus einem Projekt hervorgegangen Publikationen (Teil des Abschlussberichts)
 - uvam.

KI in der Erschließung

Potentielle Maßnahmen zur Anreicherung des Findbuchs

- ▶ Bisher enthält das BArch-Findbuch zu den Akten nur wenige, von der DFG bereitgestellt Stammdaten (s.o.)
- ▶ Mit KI ist es möglich, diese Daten um eine Vielzahl von Informationselementen zu erweitern:
 - Erstellung eines Abstracts zu den Zielen eines Forschungsvorhabens (deutsch und englisch)
 - Verschlagwortung und thematische Klassifikation von Digitalisaten
 - Recherche und Zuordnung GND-ID und Wikidata-ID zu Antragstellenden
 - Named entity recognition, etwa
 - Benennung der wiss. Mitarbeiter*innen (erweitert: mit ihren Aufgaben)
 - Angaben zu international beteiligten weiteren Wissenschaftler*innen (mit Herkunftsländern)
 - Angaben zu den begutachtenden Personen
 - Formale Hinweise (z.B. Enthält die Akte einen Abschlussbericht?)
 - Listen der aus einem Projekt hervorgegangen Publikationen (Teil des Abschlussberichts)
 - uvam.

KI in der Erschließung

Automatische Generierung von Abstracts

- ▶ Die Antragsunterlagen umfassen im Regelfall zwischen 30 und 80 Seiten, teilweise aber auch weit darüber hinaus (etwa bei beigefügten längeren Abschlussberichten)
- ▶ Der einfache Zugang zu diesen Texten besteht in der Stichwortsuche im vollständigen Textmaterial. Problem: ein hohes Maß an „white noise“, also eine Unzahl an mehr oder weniger einschlägigen Treffern, in denen sich die Archivnutzer*innen nur schwer zurechtfinden würden
- ▶ Vielversprechender: Recherche zunächst nur innerhalb maschinell erstellter Abstracts
- ▶ Abstracts können zum einen von Nutzenden direkt gelesen werden und/oder sie unterstützen die Vor-Recherche im Gesamtbestand der DFG-Akten

KI in der Erschließung

Automatische Generierung von Abstracts mit Hilfe von ChatGPT

- ▶ **Beispiel-Prompt:** *„In der angehängten Datei sind verschiedene Dokumente enthalten, die den Prozess der Beantragung von Fördermitteln für ein wissenschaftliches Projekt betreffen. Kannst Du aus dem Material eine maximal halbseitige Zusammenfassung generieren, die in leicht verständlicher Form (aber gerne unter Verwendung einschlägiger Fachbegriffe) erklärt, worum es in dem Projekt inhaltlich geht?“*
- ▶ **Der stichprobenartige Vergleich von Autoren-Abstracts und KI-Abstracts ergibt gute Befunde:**
 - Hoher Grad an Verständlichkeit
 - Vernünftige Erschließung der wichtigsten Forschungsfragen
 - Anschauliche Aufbereitung (Überschrift in Fettsatz, Bulletpoints, zentrale Stichwörter hervorgehoben)

Zusammenfassung des Forschungsprojekts „Thermoadaptation“

Das Projekt untersucht die Temperaturempfindung, den thermischen Komfort und die Temperaturregelung beim Menschen, insbesondere im Kontext einer langfristigen Anpassung an unterschiedliche Umweltbedingungen. Ziel ist es, zu verstehen, wie der menschliche Körper und das subjektive Temperaturempfinden auf längere Kälte- oder Hitzeeinwirkungen reagieren.

Im Fokus stehen folgende Fragen:

1. **Unabhängigkeit von Verhalten und autonomer Thermoregulation** – Wie sehr hängt das individuelle Temperaturempfinden von physiologischen Prozessen wie Schweißsekretion oder Durchblutung ab?
2. **Mechanismen des thermischen Komforts** – Welche physiologischen und sensorischen Faktoren bestimmen, wann sich Menschen thermisch wohlfühlen?
3. **Veränderungen durch Langzeitadaptation** – Wie verändert sich das Temperaturempfinden nach mehrtägiger Anpassung an extreme Temperaturen?

Die Forschenden führen kontrollierte Experimente in Klimakammern durch und messen dabei Hauttemperatur, Schweißsekretion, Sauerstoffverbrauch sowie die subjektive Einschätzung von Temperatur und Komfort. Diese Erkenntnisse sollen nicht nur die Grundlagenforschung zur Thermoregulation voranbringen, sondern auch praktische Anwendungen ermöglichen – etwa in der Gestaltung klimatisierter Räume oder Schutzkleidung.

Quelle: 16293

- ▶ ...mit gewissen Stolperfallen. Das eben angezeigte Abstract endet auf: *„Diese Erkenntnisse sollen nicht nur die Grundlagenforschung zur Thermoregulation voranbringen, sondern auch praktische Anwendungen ermöglichen – etwa in der Gestaltung klimatisierter Räume oder Schutzkleidung.“*
- ▶ Auf die Nachfrage nach einem Beleg für die Aussage antwortet ChatGPT: *„Die ursprüngliche Aussage zum praktischen Nutzen (Gestaltung klimatisierter Räume oder Schutzkleidung) habe ich aus allgemeinen Überlegungen zur Relevanz von Forschungsarbeiten zur Temperaturregulation abgeleitet. Sie ist jedoch nicht explizit aus der Quelle hervorgegangen.“*
- ▶ Es bestehen also vor allem zwei Herausforderungen:
 - a. Das Programm ist explizit anzuweisen, nichts dazu zu erfinden.
 - b. Andererseits ist das Erwartungs-Management gefordert: Bis zu welchem Grad ist man bereit, bei der Abwägung zwischen Service-Nutzen und Quellengenauigkeit bzgl. letzterer Abstriche in Kauf zu nehmen?

- ▶ Die DFG setzt in ihrer Geschäftsstelle Methoden des Natural Language Processings (NLP) sowie des Machine Learnings (ML) ein, um Abstracts und Antragstexte zu analysieren
- ▶ Bisher erschlossen wurden ca. 120.000 Anträge und Abstracts der Jahre 2017 bis 2024
- ▶ Die Aufbereitung für NLP/ML-unterstützte Analysen zerlegt das Textmaterial in 1,4 Milliarden sog. tokens. Das sind kleinste sinntragende Elemente (z.B. middle ages; medieval; italien; history)
 - Größenordnung: je Antrag (Durchschnitt: 25 Seiten) werden pro Seite knapp 500 tokens generiert
- ▶ Diese tokens und ihre Beziehungen lassen mit speziellen Programmen (R, Python) analysieren

Ausblick: KI in der Analyse

Nutzung von Textdaten als Basis für (statistische) Analysen

- ▶ Bestimmte Textabschnitte von Anträgen (z.B. Abschnitt zum Thema „Ethikvotum“) lassen sich gezielt auslesen und bzgl. ihrer Inhalte klassifizieren
- ▶ Formale Aspekte (z.B. Sprache eines Dokuments) lassen sich maschinell ermitteln
- ▶ Anträge zu einer bestimmten Ausschreibung lassen sich etwa mit Hilfe von topic modelling aufgrund jeweiliger Schwerpunktsetzungen bestimmten Kategorien zuordnen (vgl. [DFG - Deutsche Forschungsgemeinschaft - Das DFG-Fördergeschehen im Kontext der COVID-19-Pandemie](#))
- ▶ Die Fachlichkeit eines Antrags lässt sich aus dem für ein Fach (der DFG-Systematik, aktuell: 214 Fächer) „typischen“ Vokabular ableiten (auch: mehrere Fächer (Interdisziplinarität erkennbar))
- ▶ Gutachter*innen-Suche: Per Einlesen eines Antrags-Abstracts Generierung einer Liste von Personen, die zu „ähnlichen“ Fragestellungen DFG-Förderungen erhalten haben
- ▶ Uvam.

Fazit

- ▶ KI braucht Daten. Eine wichtige Voraussetzung für den Einsatz von KI in Archiven stellt daher die Generierung hochwertiger Digitalisate inhaltlich gehaltvoller Archivalien dar.
- ▶ KI entwickelt sich rasant. Schon die heute erreichten Möglichkeiten bieten vielfachen Nutzwert für die Erschließung archivierter Digitalisate.
- ▶ Mit dem Einsatz von Persistenten Identifiern (PID) können Archive zur Vernetzbarkeit ihrer Wissensbestände beitragen.
- ▶ Indem der Forschung digitalisierte Archivalien langfristig auch für quantitative Analysen zur Verfügung gestellt werden, ergeben sich faszinierende neue Möglichkeiten ihrer wissenschaftlichen Nutzung.



Vielen Dank für Ihre Aufmerksamkeit

Weitere Informationen

- ▶ zur DFG: <https://www.dfg.de>
- ▶ zum Förderatlas: <https://www.dfg.de/foerderatlas>
- ▶ zu allen geförderten Projekten: <https://www.dfg.de/gepris>
- ▶ zu den deutschen Forschungseinrichtungen: <https://gerit.org>
- ▶ zu historischen DFG-Projekten: <https://gepris-historisch.dfg.de>