

Montags 14:15 - 15:45 Uhr
Unitobler, Raum F-121



^b
**UNIVERSITÄT
BERN**

Ringvorlesung

Einblicke in die Digital Humanities

HS 2024 / Autumn 2024

Lecture Series

Insights into the Digital Humanities



ADOBE STOCK, AI GENERATED BY MAKSIM.

Veranstaltungen

- 16.09.2024** **Einführung/Introduction [DE/EN]**
Informationen zur Ringvorlesung und zum Studienprogramm.
- Workshop** **Einführung ins Bloggen [DE/EN]**
16:00-17:30 mit Dominik Kilchmann
Vorbereitung zum Leistungsnachweis
- 30.09.2024** **Elena Spadini (Universitäten Basel und Bern) – Materiality and the digital for scholarly editing: from medieval manuscripts to electronic writing [EN]**
This talk will delve into how the material aspects of textual works are taken into account in scholarly editing. By examining examples from medieval and twentieth-century literature, we will explore how the physical properties of documents, alongside their social, historical, and technological contexts of production and transmission, influence interpretation and inform editorial strategies. The digital adds further complexity, offering new methods for producing and publishing editions, while also introducing new types of sources, as archives increasingly preserve digital storage devices like computers and hard drives from writers and public figures.
- 07.10.2024** **Andreas Weilinghof (Universität Koblenz) – AI vs. Human? Cutting-Edge Tech Transforming Linguistic Research [EN]**
The transcription of sound data is an essential yet time-consuming and labour-intensive part of almost all linguistic research projects. This is especially true for corpus linguistics, as only well-transcribed and carefully annotated corpora provide a reliable basis for subsequent analyses. As recent years have seen great advancements in the field of Automatic Speech Recognition (ASR) (Radford et al. 2022; Jurafsky and Martin 2023), a central question is how the latest ASR models can be effectively used to enhance corpus transcription. More specifically, how do the latest ASR models perform in terms of accuracy and speed when compared to human transcribers?
This talk will address this fundamental question. I will focus on the end-to-end ASR system OpenAI Whisper (Radford et al. 2022) and demonstrate its capabilities for different languages and sound files. Furthermore, I will also present a study that evaluates the accuracy and speed of the system when applied to sociolinguistic datasets. For this purpose, the spoken components of the corpora ICE Nigeria (Wunder et al. 2008) and ICE Scotland (Schützler et al. 2017) are re-transcribed with different Whisper models and the resulting transcriptions are then compared to the manual reference transcriptions via Word Error Rate (WER) metrics.

To find out what significantly influences the performance of Whisper, the analysis applies linear mixed effects modelling of WER with the lme4 (Bates et al. 2015) and lmerTest (Kuznetsova et al. 2017) packages in R (R Core Team 2023). The individual speakers are treated as random factors. The findings show that the best Whisper models perform well on both varieties. However, there is huge variability depending on individual sound files, with generally worse performance on ICE Nigeria due to the overall poorer quality of recordings. A crucial issue is that Whisper automatically deletes hesitations, repetitions, and interruptions, which can pose a challenge for sociolinguistic data transcription. Based on the findings, I will discuss key opportunities and challenges of implementing Whisper for (socio-)linguistic data preparation workflows.

Workshop **Praktische Einführung in die automatische Spracherkennung mit**
16:00-17:45 **OpenAI Whisper [DE/EN]**

In diesem Workshop lernen Sie, wie Sie Whisper effektiv für ihre Transkriptionsarbeiten und Forschung nutzen können. Der Kurs richtet sich an Personen mit unterschiedlichem technischen Hintergrund, einschließlich Anfänger ohne Vorkenntnisse in der Programmierung. Die Teilnehmer müssen ihren eigenen Laptop mit Administratorenrechten mitbringen, um die erforderliche Software installieren zu können. Am Anfang des Workshop gibt es die Möglichkeit, Probleme bei der Installation auszuräumen

Die Agenda des Workshops umfasst:

1. Installation von Python, Whisper sowie allen notwendigen Paketen.
2. Einführung in die Bedienung von Whisper über das Terminal.
3. Nutzung von Whisper mittels Python-Skripten.
4. Praktische Anwendung von Whisper zur Transkription eigener Mediendateien.
5. Einführung in grundlegende Programmierkonzepte und -techniken in Python für die Arbeit mit Sprachdaten.

Ziel des Workshops ist es, den Teilnehmenden praktische Kenntnisse und Fähigkeiten zu vermitteln, die sie befähigen, Whisper in verschiedenen Kontexten selbstständig einzusetzen.

21.10.2024

Tessa Gengnagel (Universität Köln) – Zum Scheitern verurteilt? Über Big-Tech-Blühtträume, Silicon Valley und die Niederungen der digitalen Geisteswissenschaften [DE]

In den letzten Jahren hat ein Diskurs in den Digital Humanities an Fahrt aufgenommen, der sich um die Frage des Scheiterns dreht. Von dem Mislingen millionenschwerer Großprojekte über fehlende Perspektiven für die Langzeitbetreuung von Ressourcen bis hin zu lokalen Initiativen, die sich schwertun, Fuß zu fassen: Das Spektrum an Fehlentwürfen ist groß und wächst stetig an. Gleichzeitig werden die Rufe nach Transparenz und Aufarbeitung lauter. Dabei geht es nicht nur um eine offene Auseinandersetzung mit Risikofaktoren. Vielmehr gilt die Forderung, man müsse Scheitern als Chance begreifen. In dieser positiven Umdeutung zeigt sich ein Unternehmergeist, der die Digital Humanities seit ihren Anfängen (im Bannkreis von IBM) begleitet hat und heute dem Sloganeering des Silicon Valley nahesteht.

Der Vortrag zeichnet diese Entwicklung von den 1940ern bis in die Gegenwart nach und wirft einen Blick hinter die Kulissen jener Technikphilosophien, deren Innovationskraft die Arbeit in den Digital Humanities immer erst ermöglicht, zugleich aber ihren Anspruch an das eigene Gelingen rhetorisch verzerrt haben.

28.10.2024

Katarzyna Anna Kapitan (École Nationale des Chartes, Paris) – Digital approaches to loss of medieval literature [EN]

How much of medieval literature has been lost over time? This question has fascinated scholars for generations, leading to the development of various methods to estimate the loss of medieval manuscripts and literary works. In Icelandic saga studies, for example, Judith Jesch's (1984) examined formulaic references in *Íslendingasögur* to uncover lost sagas. However, this method has limitations when applied to other genres, such as romances, as the references to these works are sparse. In this lecture, I will present the results of digital analysis of the corpus of medieval Icelandic romances obtained by applying methods borrowed from ecology and statistics. This approach enables scholars to analyse large datasets, compare loss rates across different genres and linguistic traditions, and provide insights into patterns of literary survival and loss.

11.11.2024 Archim Rabus (Universität Freiburg i.Br.) – ‚Smarte‘ Handwritten Text Recognition (HTR) und Large Language Models (LLMs): Potenzierter KI als Verheißung oder Büchse der Pandora? [DE]

Verfahren der Handwritten Text Recognition (HTR) haben in den letzten Jahren durch neuronale Netze und das Aufkommen benutzerfreundlicher Plattformen wie Transkribus eine enorme Entwicklung durchlaufen. Damit stehen mittlerweile für zahlreiche Sprachen und Schriften frei verfügbare Transkriptionsmodelle zur Verfügung, die die Transkription historischer Texte mit einer niedrigen Fehlerrate in hoher Geschwindigkeit und somit Massendigitalisierung ermöglichen. HTR kann damit als ein mittlerweile gut etabliertes Verfahren in den textbezogenen Digital Humanities gelten.

Large Language Models (LLMs) wie ChatGPT und dessen Nachfolger GPT-4 wiederum haben die Art, wie wir KI zur Manipulation sprachlicher Daten einsetzen, revolutioniert und bieten vielfältige Perspektiven, für Aufgaben im Bereich der Digital Humanities eingesetzt zu werden.

Im Vortrag lerne ich anhand von Beispielen aus unterschiedlichen Sprachen und Schriftsystemen Möglichkeiten und Grenzen des Zusammenspiels der beiden KI-Technologien HTR und LLMs aus. Dabei spielen unter anderem ‚smarte‘ Eigenschaften von HTR-Modellen eine Rolle, weiterhin der Einsatz von LLMs zur Post-Korrektur von HTR-Output. Neben konkreten Anwendungsbeispielen liegt ein Fokus auf der Frage, inwieweit der potenzierte Einsatz zweier KIs in einem Workflow zu unerwünschten Ergebnissen führen und wie dem gegebenenfalls entgegengewirkt werden kann..

18.11.2024 Claus Beisbart (Universität Bern) – Artificial humanities? What the use of AI means for the humanities [EN]

The use of the digital computer, in particular AI, in the humanities and other disciplines has a profound impact on research. The aim of this lecture is to discuss this impact with philosophical means. Topics covered include: paradigms of AI, classifying uses of AI in research, the issue of justification, the opacity of AI, and repercussions for humanistic understanding.

25.11.2024

**Moritz Feichtinger (Universität Basel) – De-Construction – Re-Construction:
Reverse engineering als Methode und Paradigma [DE]**

Daten sind immer konstruiert, historisch situiert und lokal. Sie entstehen in einer technisch-sozialen und epistemologischen Umgebung, von der sie geprägt werden und innerhalb der sie bestimmte Funktionen erfüllen. Eine kritische Genealogie von Daten muss deren Konstruktionsprozess rückverfolgen, um die Situietheit und Ursprungsumgebung kenntlich zu machen und in die Analyse mit einfließen zu lassen. Bei ausreichender Dokumentation und Überlieferung der technisch-informationellen Ursprungsumgebung sollte es sogar möglich sein, den Prozess der Produktion/Konstruktion von Daten nachzuvollziehen, also von einer De-Konstruktion der Daten und ihrer Entstehung zu einer Re-Konstruktion dieses Prozesses zu gelangen. Dies bedeutet, die epistemologischen Vorannahmen und Entscheidungen der Datafizierung zu untersuchen, die Architektur und Funktionsweise der Datenbanken, in denen Daten abgelegt wurden, sowie Code und Algorithmen, mit denen sie weiter prozessiert wurden. Es kommen dabei also Methoden der Datafizienzanalyse, der Code Studies, Reverse Engineering, Computer-Forensik und der kritischen Untersuchung von Datenbanken zum Einsatz. Der Vortrag gibt Einblick in ein laufendes Forschungsprojekt, in dem eine solche Dekonstruktion und kritisch Rekonstruktion versucht wird. Ein Datenbanksystem, das ab 1966 von der US-Armee während der Vietnamkrieges zur Bevölkerungskontrolle entwickelt und eingesetzt wurde, ist im US-amerikanischen Nationalarchiv nur noch in Form der durch das System hervorgebrachten Datensätze sowie einer umfangreichen Dokumentation erhalten. Das Forschungsprojekt «Computing the Social» versucht, aus der Struktur der erhaltenen Datensätze Rückschlüsse über die Funktionsweise des Datenbanksystems zu gewinnen. Mithilfe von zusätzlicher Dokumentation wie Handbüchern und Korrespondenzen wird zudem eine Rekonstruktion des Ursprungssystems unternommen. Durch dieses Re-enactment soll letztendlich eine kritische Bewertung der Daten ermöglicht werden, nicht nur was ihren (semantischen) Wahrheitsgehalt anbelangt, sondern auch ihre formale Konsistenz und Integrität. Ein Schwerpunkt des Vortrags liegt auf den Schwierigkeiten und Erkenntnissen beim Einsatz von Reverse Engineering zur Daten-Kritik und mündet in einem Plädoyer für eine breite Auseinandersetzung mit solchen Zugangsweisen in den DH und (kritischen) Datenwissenschaften.

02.12.2024 Noah Bubenhofer (Universität Zürich) – Sprachliche Muster in Texten: Warum sie entstehen, was sie bedeuten und wie sie korpuslinguistisch untersucht werden können [DE]

Unsere Sprache lebt von Routineformeln und sprachlichen Mustern: Doch warum eigentlich? Der Grund liegt in der engen Verknüpfung von handeln und sprechen. Die so entstehenden Sprachgebrauchsmuster können untersucht werden, um eine ganze Reihe von interessanten Erkenntnissen über die Verknüpfung von Sprache und Gesellschaft zu gewinnen. Das ist das Ziel von korpuslinguistischen / korpuspragmatischen Methoden, wie am Vortrag plausibilisiert werden soll.

Workshop Hands-on Korpuslinguistik [DE]

16:15-17:45 Im Workshop können erste korpuslinguistische Erfahrungen gewonnen werden, indem bestehende Korpusdatenbanken recherchiert und grosse Textsammlungen analysiert werden.

09.12.2024 DH und dann? Ein Roundtable mit Absolvent*innen anschliessend Semesterendeapéro der Digital Humanities [DE]

**Informa-
tionen für
Studierende**

Die Veranstaltung besteht aus zwei Teilen, die beide eng gekoppelt sind:

1) Im Rahmen der Vorlesung (jeweils von 14:15-15:45) präsentieren die Referierenden ein Forschungsprojekt bzw. eine theoretische Perspektive aus den Digital Humanities

2) Im darauffolgenden Workshop (jeweils 16:15-17:45) wird *hands-on* demonstriert, was mit Tools oder Ansätzen gemacht werden kann. Der Workshop findet an drei Terminen statt.

Beides, Ringvorlesung und Workshop, sind essentielle Teile der Ringvorlesung und wir erwarten den Besuch beider Teile. Die Eintragung in KSL ist etwas irreführend, da KSL auch für die Raumreservation gebraucht wird.

Die Aufnahmen der Vorlesung und des Workshops werden auf Ilias via Link zur Verfügung gestellt.

Der Leistungsnachweis wird während des Semesters erbracht und besteht aus dem Verfassen eines Blogposts auf unserer Veranstaltungsseite «<https://einblicke.hypotheses.org/>». Die Blogposts werden als Einzelarbeiten verfasst. Falls mehrere Personen zum selben Beitrag schreiben (problemlos möglich), bitten wir um Absprache mit den Kommiliton*innen.

Die Eintragung im Ilias-Etherpad «Organisation Blogging» ist notwendig, ansonsten kann der Beitrag nicht als Leistungsnachweis angerechnet werden.

Ziel eines Blogposts ist es einen Aspekt bzw. eine Perspektive auf einen Vortrag oder Workshop herauszuarbeiten. Darunter fällt die Beschäftigung mit einer Methode wie auch die Beschreibung des Umgangs mit dem jeweiligen Material. Ein Blogpost ist ähnlich wie ein Essay aufgebaut. Eine klar formulierte Problemstellung wird hergeleitet, bearbeitet und fließt in eine Schlussfolgerung. Ein guter Blogpost ist bebildert, verlinkt auf weitere Ressourcen und verweist auf weiterführende Literatur. Der Text sollte flüssig zu lesen sein und sprachlich korrekt sein. Das Einbauen von subjektiven Perspektiven ist erwünscht.

Der Blogpost soll mindestens 2'000 Zeichen lang werden.

Die Posts werden vor der Veröffentlichung gegengelesen, kommentiert und müssen unter Umständen überarbeitet werden.

Die Beiträge sollen drei Wochen nach dem jeweiligen Vorlesungstermin zur Veröffentlichung an mich gesandt werden (tobias.hodel@unibe.ch).

Die Benotung der Vorlesung basiert auf den Blogposts.

Zur Organisation der Vergabe der Blogposts finden Sie auf Ilias ein Etherpad.
Die Studierenden schreiben sich bitte bis zum 29.09.2024 für den gewünschten Termin ein.

Zur Erleichterung der Vereinbarkeit von Studium, berufliche Tätigkeiten und Privatleben kann die Ringvorlesung via ZOOM besucht werden. Dafür ist eine Registrierung notwendig: http://bit.ly/rv_2024.

Bei Fragen und Unklarheiten, sowie längeren Absenzen bitte ich um Kontaktaufnahme, präferiert per Mail: tobias.hodel@unibe.ch.

Die Ringvorlesung der Digital Humanities ist ein offenes Format und die Teilnahmen ist jederzeit möglich.

Gerne können Sie auch online via ZOOM teilnehmen (Registrierung erforderlich):

https://bit.ly/rv_2024

The Digital Humanities lecture series is an open format and participation is possible at any time.

You are also welcome to participate online via ZOOM (registration required):

https://bit.ly/rv_2024

WALTER BENJAMIN KOLLEG
DIGITAL HUMANITIES

Universität Bern
Walter Benjamin Kolleg
Muesmattstrasse 45
CH-3012 Bern

www.wbkolleg.unibe.ch
www.dh.unibe.ch

