

Veranstaltungsreihe «Aus der digitalen Werkstatt»



# Digitale Methoden der Zeitungsanalyse

Summer School

- Mittwoch bis Donnerstag, 11. bis 12. September 2024
- Zentralbibliothek Zürich, Hermann-Escher-Saal
- Anmeldung erforderlich: [t.zbzuerich.ch/summerschool24](https://t.zbzuerich.ch/summerschool24)

## Konzept und Programm

Zeitungen und Zeitschriften sind wichtige Medien für die öffentliche Meinungsbildung und Teil des offenen demokratischen Diskurses. In historischer Perspektive gehören sie deshalb zum überlieferungswürdigen Kulturerbe. Für Forschung und Wissenschaft sind sie wichtige Quellendokumente, die mit vielfältigen Fragestellungen erforscht und analysiert werden können. Die Summer School der Zentralbibliothek Zürich möchte Forschenden eine Einführung in Praxis und Theorie der digitalen Analysemethoden von Zeitungen geben. Es werden keine entsprechenden Erfahrungen vorausgesetzt; die Summer School richtet sich sowohl an Anfänger als auch an Fortgeschrittene. Sie verfolgt folgende Zielsetzungen:

- Vermittlung von digitalen Analysekompetenzen in Theorie und Praxis: Analyse von Volltexten, Bildern und Metadaten
- Reflexion digitaler Analysemethoden (digitale Quellenkritik, Herangehensweisen/Methodik, Möglichkeiten und Grenzen von Tools, mögliche Alternativen)
- Kennenlernen zentraler Zeitungsportale aus dem deutschsprachigen Raum (ANNO, Deutsches Zeitungsportal, Impresso-Projekt, etc.)

### Mittwoch, 11. September

<p>9.00 - 12.00</p>	<p><b>Historische Medienanalyse mit Impresso</b> <i>Dozierende: Simon Clematide, Juri Alexander Opitz (Universität Zürich)</i></p> <p>In dieser Unterrichtseinheit erfahren die Teilnehmenden Hintergrundinformationen zur technischen Aufbereitung von Volltexten für Information Retrieval (IR) und erlernen die Anwendung von IR-Techniken zur Bearbeitung von Forschungsfragen am Beispiel des Impresso-Projekts. Zu Beginn demonstrieren wir in der Impresso Web-App Techniken für Suchanfragen, auch unter spezialisierter Anwendung von Filtern aus Metadaten und Text-Mining-Resultaten. Anschließend verschaffen wir uns einen Überblick über Methoden, die derzeit für die Named Entity Erkennung und und das Named Entity Linking angewendet werden (Personen, Toponyme), und diskutieren die damit verbundenen Möglichkeiten und Herausforderungen. Es folgt eine Hands-On-Session, in der wir thematische Textsammlungen extrahieren/erstellen, mithilfe von maschineller und menschlicher Intelligenz. Danach möchten wir uns über die gesammelten Erfahrungen austauschen, zum Beispiel indem wir Resultate teilen und Fehler, bzw. Fehlerquellen, besprechen und analysieren. Die Einheit endet mit einer allgemeinen Reflektion zu datenbasierten Methoden und der Verwendung von maschineller Inhaltsanalyse von Texten.</p>
	<p>Mittagspause</p>
<p>13.30 - 17.00</p>	<p><b>Digitales Arbeiten mit Bildern in Zeitungen und Zeitschriften</b> <i>Dozierende: Johanna Störiko (Georg-August-Universität Göttingen)</i></p> <p>In dieser Einheit lernen die Teilnehmenden, automatisiert auf digitale Bildersammlungen mit IIF-Schnittstellen zuzugreifen und mithilfe von Image Embeddings explorative Untersuchungen auf den Bildern durchzuführen. Die Einheit besteht aus zwei Teilen: Die erste Hälfte beginnt mit einer Einführung in den Aufbau von IIF-Schnittstellen. Es wird vermittelt, wie mithilfe der Programmiersprache Python automatisiert eine Linkliste zum Herunterladen der Bilder erstellt werden kann.</p>

	<p>Hierfür stehen vorbereitete Jupyter-Notebooks zur Verfügung, mit denen die Bilder auf die eigene Festplatte heruntergeladen werden können. Im zweiten Teil erkunden wir unsere erstellte Bildersammlung mithilfe von Neuronalen Netzen. Dafür gibt es zuerst eine kurze Einführung in die Grundidee von Image Embeddings. Anschließend wird gezeigt, wie vortrainierte Netze wie CLIP oder ResNet über die Python-Schnittstelle der Plattform Huggingface mit wenigen Zeilen Code in das eigene Projekt eingebunden werden können. Mit diesen Netzen können wir die zuvor erstellte Bildersammlung nach Stichworten durchsuchen und Bilder entdecken, die einander ähnlich sind. In einer abschließenden Diskussion werden die Möglichkeiten und Limitationen dieses Ansatzes besprochen.</p>
18.15 - 19.30	<p><b>Öffentlicher Abendvortrag von Steven Claeysens (Koninklijke Bibliotheek (KB), National Library of the Netherlands):</b> (weitere Angaben folgen)</p>

### Donnerstag, 12. September

9.00 - 12.00	<p><b>ANNO – von Daten zur Forschung. Arbeiten mit dem Zeitschriftenportal der Österreichischen Nationalbibliothek</b></p> <p><i>Dozierende: Volker Haider, Johannes Knüchel, Simon Mayer (Österreichische Nationalbibliothek)</i></p> <p>In diesem Unterrichtsblock wird ein umfassender Überblick über das Datenangebot aus dem ANNO-Portal der Österreichischen Nationalbibliothek (ÖNB) gegeben. Die Vortragenden erkunden mit den Teilnehmenden Möglichkeiten mit über 27 Millionen Zeitungsbzw. Zeitschriftenseiten aus dem Portal zu arbeiten. Die Teilnehmenden lernen so unterschiedliche Aspekte aus dem Bestand kennen und haben dabei selbst die Möglichkeit, mit vorbereiteten Jupyter-Notebooks auf die Daten aus dem Portal zuzugreifen und mit computerunterstützten Methoden zu verarbeiten. Ein Einblick hinter die Kulissen von ANNO soll Forschenden Kontext zur Entstehungsgeschichte der Daten und des Portals geben. Im Anschluss soll mit besagten Notebooks das Gesamtkorpus anhand von Metadaten (Erscheinungsjahre, Themenschwerpunkte, ...) explorativ erschlossen werden. Ein weiterer Schwerpunkt umfasst die Volltexte aus dem Bestand. Es wird skizziert bzw. erarbeitet, welche Formate vorhanden sind, welche Qualität zu erwarten ist und wie Volltexte zu einzelnen Zeitungsseiten selbst mit OCR neu erstellt werden können. Ein weiterer Fokus des Unterrichtsblocks wird auf Forschungstools an der ÖNB liegen, die über die ÖNB Labs angeboten werden. Das aktuelle Angebot soll präsentiert werden, um darauf gemeinsam diskutieren zu können, wie die Tools in Zukunft ausgebaut werden können.</p>
	<p>Mittagspause</p>
13.30 - 17.00	<p><b>Historische Forschung digital: ein Workshop zum Deutschen Zeitungsportal</b></p> <p><i>Dozierende: Michael Büchner (Deutsche Digitale Bibliothek), Franziska Fuchs (Deutsche Nationalbibliothek), Stephanie Nitsche (Deutsche Nationalbibliothek)</i></p> <p>Das Deutsche Zeitungsportal – ein Subportal der Deutschen Digitalen Bibliothek – ist mit knapp vier Millionen Ausgaben aus über 1.800 Zeitungstiteln, die fast vollständig mit Volltext vorliegen, der größte Anbieter für historische, digitalisierte Zeitungen in Deutschland. Es stellt damit eine wertvolle Quelle für geschichtlich arbeitende Wissenschaften dar.</p>

	<p>Ziel des dreistündigen Workshops ist es, den Teilnehmenden einen umfassenden Überblick darüber zu vermitteln, wie das Deutsche Zeitungsportal für die Forschung, insbesondere in den Digital Humanities, genutzt werden kann. Die Teilnehmenden erhalten dazu zunächst eine kurze Einführung in das Deutsche Zeitungsportal, seine Entstehungsgeschichte, die angebotenen Funktionalitäten und die Vielfalt der verfügbaren Inhalte. Anschließend lernen die Teilnehmenden, wie sie die Programmierschnittstelle (API) der Deutschen Digitalen Bibliothek nutzen können, um Datensets aus dem Deutschen Zeitungsportal herunterzuladen. Der Fokus liegt dabei auf der Vermittlung grundlegender Kenntnisse, die es den Teilnehmenden ermöglichen, in Zukunft eigenständig Datenabfragen zu erstellen. Sie lernen die Funktionsweise der Schnittstellen kennen und erfahren, wie sie Abfragen mithilfe der Dokumentation anpassen und erweitern können. Nach diesem Einblick führen die Teilnehmenden, begleitet vom DNBLab-Team, eine Datenanalyse auf Basis der gemeinsam heruntergeladenen Daten durch. Hierfür werden Jupyter Notebooks mit Python-Programmcode genutzt. Durch das gemeinsame Live-Coding werden auftretende Fragen und Probleme direkt und interaktiv gelöst. Am Ende werden die gemeinsam analysierten Daten in geeigneten Visualisierungen dargestellt, die Aufschlüsse über die Datenzusammensetzung und mögliche Forschungsansätze geben.</p>
18.15 - 19.30	<p><b>Öffentlicher Abendvortrag von Louise-Anne Charles (Bibliothèque nationale de France DataLab):</b> (weitere Angaben folgen)</p>

## VeranstalterInnen/OrganisatorInnen

- Estelle Bunout (Université du Luxembourg)
- Nina C. Rastinger (Austrian Centre for Digital Humanities and Cultural Heritage/Österreichische Akademie der Wissenschaften)
- Sarah Oberbichler (Leibniz-Institut für Europäische Geschichte Mainz)
- Johanna Störiko (Institut für Digital Humanities der Universität Göttingen)
- Jesko Reiling (Zentralbibliothek Zürich)

Die Summer School ist eine Kooperation der AG «Zeitungen & Zeitschriften» des Verbandes Digital Humanities im deutschsprachigen Raum (DHD) und der Zentralbibliothek Zürich.

## Kosten

Die Summer School ist für die Teilnehmenden kostenlos. Anreise und Unterkunft müssen die Teilnehmenden selbst organisieren/bezahlen. Den Teilnehmenden wird während der Summer School ein Lunch sowie morgens und nachmittags Kafi/Gipfeli offeriert.

Es besteht in Einzelfällen die Möglichkeit, bei der AG «Zeitungen & Zeitschriften» einen Antrag auf Erstattung der Reisekosten zu stellen. Weitere Informationen dazu finden Sie im Anmeldeformular (siehe unten).

## Anmeldung

Bitte melden Sie sich bis zum 11. August 2024 hier an: [t.zbzuerich.ch/summerschool24](https://t.zbzuerich.ch/summerschool24)

Sie werden nach dem 11. August 2024 so schnell wie möglich über die Aufnahme benachrichtigt. Bei Fragen wenden Sie sich bitte an [summerschool@zb.uzh.ch](mailto:summerschool@zb.uzh.ch).