

What's in the news? (Erfolgs-)Rezepte für das wissenschaftliche Arbeiten mit digitalisierten Zeitungen

Einleitung

Why newspapers? Wie geht man methoden-kritisch mit digitalisierten Zeitungskorpora um? Welcher technische Aufwand ist nötig? Wie gestaltet sich die Zusammenarbeit mit Bibliotheken zum Thema Verfügbarkeit und Lizenzierung? Die zahlreichen Panels, Präsentationen und Poster auf der DH2019 haben gezeigt, dass ein großes Interesse seitens der Wissenschaft und auch der Bibliotheken besteht umfangreiche Zeitungs- und Zeitschriftensammlungen für die Öffentlichkeit und Wissenschaft digital verfügbar zu stellen (vgl. ADHO). Dabei konzentrierten sich die Debatten aber überwiegend auf die Provider-Perspektive, d.h. welche Herausforderungen sind im Rahmen des Digitalisierungsprozesses und der Zusammenarbeit mit wissenschaftlichen Institutionen zu überwinden. Im Spielraum der Wissenschaft bedeutet diese Erzeugung der Forschungsdaten, Zeitungen als Primärtexte aufgrund ihrer Menge, Zeitspannen und textuellen und thematischen Vielfalt nutzen zu können, um temporale, grenzübergreifende, multilinguale und -modale Querbezüge zu erstellen. Diese Forschung bietet die Möglichkeit operative, methodische und organisatorische Herausforderungen anzugehen, um innovative computergestützte Modelle, Tools, Codes, Daten und Infrastrukturen zu entwickeln. Die Panelvorträge geben Einblicke in Forschungsprojekte zu digitalisierten Zeitungen, die mit unterschiedlichen Korpora und Fragestellungen, aber teils ähnlichen Verfahren und Forschungsdesigns an umfangreichen historischen Zeitungssammlungen arbeiten. Dabei sollen die folgenden Kategorien angesprochen werden: Herausforderungen, Forschungsfragen, Methoden (inklusive Tools), Teamkomposition, Korpusbeschreibung, Projektformat und Projektdauer. Das Ziel dieses Panels ist es gemeinsam das Generalisierungspotential der Methoden und den wissenschaftlichen Output zu diskutieren, um eine Zutaten- und Werkzeugliste für das wissenschaftliche Arbeiten mit digitalisierten Zeitungen zu generieren. Entsprechend der vorgestellten Beiträge sollen gemeinsame Fragen diskutiert werden wie:

- Wie ist das Verhältnis von Korpuskomposition und Fragestellung in den jeweiligen Projekten?
- Welche forschungsspezifischen Hürden oder Fallstricke gab es in den einzelnen Projekten?

- Welche geistes- und kulturwissenschaftlichen Fragen können in Anforderungen in digitale Methoden übersetzt werden und wie werden diese im digitalen Umfeld operationalisiert?
- Wird die Fragestellung ggf. durch ein Interface beeinflusst, oder andersherum: lässt sich das Interface der Fragestellung anpassen?
- Wie können wir sicherstellen, dass wir nicht nur projektspezifische Tools bauen und Methoden entwickeln, sondern diese auch für weitere Anwendungen in der Wissenschaft und Öffentlichkeit nachnutzbar machen?

Panelvorträge

NewsEye Case Study: Rückkehrmigration in österreichischen Tageszeitungen zwischen 1850 und 1950

Sarah Oberbichler

Ein nicht unbedeutender Teil jener Menschen, die freiwillig oder unfreiwillig zwischen 1850 und 1950 ihre Heimat verlassen hatten, kehrten in ihr Ursprungsland zurück. In österreichischen Tageszeitungen wurde regelmäßig über die Rückkehr von freiwilligen Auswander*innen, in Gefangenschaft geratenen Soldaten oder Flüchtlingen berichtet, weshalb dieses Medium eine geeignete Quelle für die Erforschung des bis dato vernachlässigten Themas der Rückkehrmigration darstellt. Im Rahmen des NewsEye Projektes wurden deshalb folgende Forschungsfragen aufgegriffen: Wie und in welchem Kontext wurde in österreichischen Tageszeitungen über Heimkehrer*innen berichtet und wie hat sich die Berichterstattung im Laufe der Zeit verändert? Das zur Beantwortung der Fragestellung notwendige Korpus wird mithilfe des Online Zeitungsarchives der Österreichischen Nationalbibliothek ("ANNO") erstellt und für die weitere Anwendung von Text-Mining Methoden und qualitativen, diskursanalytischen Analysen aufbereitet. Gerade aber dieser erste Schritt – die Bildung des Korpus – bringt eine Reihe von Herausforderungen mit sich. Nicht alle Suchbegriffe führen zu eindeutigen Ergebnissen und fehlende Speicher- und Download-Optionen führen zu langen und aufwendigen "Copy und Paste"-Verfahren. Vor diesem Hintergrund stellt sich die Frage, welche Methoden und Ansätze (beispielsweise Article Separation, Topic Modeling oder Wort Embeddings) für die Bildung eines Korpus herangezogen werden können, wenn manuelle Vorgehen einen zu großen Zeitaufwand darstellen. Ebenfalls stellt sich die Frage, wie viel Spielraum zwischen vorgegeben Funktionen und individuellen Einstellungen zielführend ist.

More than a Feeling: Media Sentiment as a Mirror of Investors' Expectations at the Berlin Stock Exchange, 1872-1930

Lino Wehrheim / Bernhard Liebl

Das Verhalten von Finanzinvestoren wird nicht nur durch Fundamentalwerte wie etwa künftige Zahlungsströme, sondern auch durch „weiche“ Faktoren wie Stimmungen, Launen und Gefühle beeinflusst. Entsprechend hat sich in der Finanzmarktforschung das Konzept des „Investor Sentiment“ etabliert, was als eine (individuelle oder kollektive) Einstellung in Bezug auf künftige Marktentwicklungen verstanden werden kann, die nicht auf rationaler Abwägung basiert. Das Ziel des Projekts ist es, die Bedeutung von Sentiment für die Berliner Börse zwischen 1872 und 1930 zu erfassen, dem bedeutendsten deutschen Finanzplatz dieser Zeit. Inwieweit beeinflussten historische Erfahrungen wie Kriege oder politische Ereignisse die Stimmung von Finanzinvestoren, und welchen Einfluss übte Sentiment auf die Entwicklung historischer Börsenkurse aus? Hat sich dieser Einfluss im Zeitverlauf verändert? Um die Stimmung an der Berliner Börse zu quantifizieren, wird auf Basis historischer Zeitungsartikel ein Sentiment-Index erstellt, ein Ansatz, der seit der Arbeit von Tetlock (2007) verbreitet Anwendung findet, so etwa bei Ferguson et al. (2015), García (2013) und Hanna et al. (2017). Konkret werden wörterbuchbasierte Verfahren sowie Ansätze des maschinellen Lernens herangezogen. Besondere Bedeutung erhält die Generierung eines domain-spezifischen Sentiments-Wörterbuchs (Finanzmarktdeutsch des 19. Jahrhunderts). Um den Sentiment-Index um die Komponente medialer Narrative zu ergänzen, werden die Zeitungsartikel zusätzlich mit Topics Models ausgewertet. Das zugrundeliegende Korpus besteht aus Artikeln der Berliner Börsen-Zeitung, die in täglichen Marktberichten über das Geschehen und die Stimmung am Finanzplatz Berlin berichtete.

"Horizontales Lesen" als digitale Analyseverfahren von Zeitungskritiken

Torsten Roeder

Zeitungen und Zeitschriften sind spätestens seit dem 19. Jahrhundert - bis in die Jetztzeit - ein Medium für öffentliche Debatten über Kunst und Kultur. Während historische Untersuchungen sich vor nicht allzu langer Zeit noch aufgrund der diffusen Quellenlage auf Einzelfalluntersuchungen beschränken mussten, macht aktuell die immense Menge an Material, das durch die Digitalisierung verfügbar ist und wird, die Entwicklung und Anwendung neuer Verfahren notwendig. Ein Ansatz besteht in dem Verfahren des "horizontalen" Lesens, mit dem sich thematisch zusammenhängende Texte zu einem (historischen) Meinungsspektrum anordnen lassen. Den "Named Entities" fällt dabei eine Schlüsselrolle

zu, da diese die zentralen Vergleichspunkte liefern. Anhand einer ausgewählten Entity (z.B. der Titel eines musikalischen Werkes) und aller Textausschnitte, die sich direkt darauf beziehen, kann manuell und ggf. mithilfe übertragener Anwendung von Sentimentanalyse oder ggf. Topic Modeling die gesamte Bandbreite zu einem historischen Zeitpunkt oder an einem historischen Publikationsort abgebildet werden. Voraussetzungen dafür sind jedoch eine hochwertige Texterschließung und semantische Annotationen, bestenfalls mit Normdaten versehene Eigennamen von Personen, Orten, Werken etc. p.p. Während diese Methode durchaus verwertbare Ergebnisse produziert, bleibt für alle Anwendungsfälle, an welcher Stelle für die notwendige Qualität der Daten Sorge zu tragen ist: Kann dies bereits durch Provider geschehen, oder muss dies notwendigerweise - ggf. auch abgestuft - im jeweiligen Forschungsprojekt geschehen?

Oceanic Exchanges: Transnationale Textmigration

Jana Keck

Im 19. Jahrhundert entstand die Massenpresse. Die technischen Innovationen der Druckpressen, fehlende Regulierungen der Gesetzgebung und Durchsetzung des Urheberrechts und das wachsende Interesse der Bevölkerung an Informationen weltlicher, sensationeller und politischer Natur schuf eine globale Kultur reichhaltiger, schnell zirkulierender Informationsquellen. Auch wenn dies auf den grenzübergreifenden Charakter der Presse hinweist, wurde die Zeitungsforschung bisher weitgehend in Metropol- und nationalen Räumen definiert. Im Rahmen des internationalen Forschungsprojektes "Oceanic Exchanges" wurden über 100 Millionen digitalisierte Zeitungsseiten aus mehr als 7 Nationen gesammelt, um den transnationalen Charakter der Presse im 19. Jahrhundert zu untersuchen. Welche Texte und Ideen – literarisch, politisch, wissenschaftlich, wirtschaftlich, religiös – zirkulierten im öffentlichen Raum? Wie wurden diese Texte in dem jeweiligen Raum und Sprache übersetzt und modifiziert? Text Reuse Detection hat sich schon trotz „noisy“ OCR als erfolgreiche Text-Mining Methode erwiesen, um ähnliche Textpassagen in umfangreichen Datenbanken digitalisierter englischsprachiger Zeitungen zu erkennen und zu modellieren (Viral Texts Project). Das Ziel dieses Beitrages ist zu zeigen, welche methodischen Möglichkeiten, aber auch Herausforderungen beim Erkennen und Modellieren von Reprints in multi-lingualen Sammlungen und der zeitlichen Klassifizierung entstehen. Damals wie heute, nutzen diese Reprinting-Praktiken die Paradoxien eines jeden internationalen Mediensystems: scheinbare Verbundenheit und doch beständige Distanzen.

Digitale Ideengeschichte: der antimoderne Diskurs über Europa in der schweizer Presse (1900-1945) (Estelle Bunout / Marten Düring)

Estelle Bunout / Marten Düring

Die Diskussionen über Europa haben sich in der Schweiz in der ersten Hälfte des 20. Jahrhunderts auf mehreren Ebenen entwickelt: über die Initiativen zum Aufbau eines institutionellen Europas oder noch über die Rolle, die die Schweiz dabei spielen sollte. Ein Blick auf die Presse gibt eine breitere Perspektive auf diese Diskussionen. Vertreter aus unterschiedlichen Bereichen wie der Mathematik, mit Sophie Piccard (1904–1990), Verfechterin der Paneuropa-Union und der Literatur, mit Gonzague de Reynold (1880–1970) verteidigten jeweils eine eigene Vision für Europa. De Reynold, ein bekannter Antimodernist, versuchte ein Europa der Aristokratie und des Korporatismus gegenüber dem liberalen, demokratischen Europa zu verteidigen und war in den Zeitungsredaktionen gut eingebunden. In diesem Zusammenhang gehen wir der Frage nach ob es besondere Bemühungen von Antimodernisten gab, die Diskussion über Europa wieder anzueignen und die Assoziation Europas mit Frieden, Progressivismus und Aufklärung im Kontext ihrer Weltanschauung neu zu gestalten. Diese Frage wird anhand der digitalisierten Schweizer Presse bearbeitet, die im Rahmen des impresso Projekts mit NLP bereichert wurde und durch eine forschungsorientierte Oberfläche zugänglich gemacht wurde. Diese Infrastruktur und weitere NLP-Methoden helfen bei der Erstellung einer Sammlung von Artikeln, die diesen antimodernen Diskurs zu Europa beinhalten, was durch gewöhnlichen Stichwortsuche nicht möglich wäre.

Koenen, Erik (2018): “Digitale Perspektiven in der Kommunikations- und Mediengeschichte”, in: *Publizistik* 63(4): 535–56 <https://doi.org/10.1007/s11616-018-0459-4> [letzter Zugriff 31. August 2019].

NewsEye: <https://www.newseye.eu/> [letzter Zugriff 31. August 2019].

Oceanic Exchanges: <http://oceanicexchanges.org/> [letzter Zugriff 31. August 2019].

Tetlock, PC (2007): “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”, in: *The Journal of Finance* 62 (3): 1139–1168.

Viral Texts: <https://viraltexts.org/> [letzter Zugriff 31. August 2019].

Wijffjes, Huub (2017): “Digital Humanities and Media History. A Challenge for Historical Newspaper Research”, in: *Tijdschrift Voor Mediageschiedenis* 20(1): 4–24–24 <https://doi.org/10.18146/tmg20277> [letzter Zugriff 31. August 2019].

Bibliographie

ADHO: <https://dh2019.adho.org/conftool/> [letzter Zugriff 31. August 2019].

ANNO (AustriaN Newspapers Online): <http://anno.onb.ac.at/> [letzter Zugriff 31. August 2019].

Crymble, Adam (2016): “Digital Library Search Preferences amongst Historians and Genealogists: British History Online User Survey”, in: *Digital Humanities Quarterly* 10(4) <http://www.digitalhumanities.org/dhq/vol/10/4/000270/000270.html> [letzter Zugriff 31. August 2019].

Ehrmann, Maud / Bunout, Estelle / Düring, Marten (2019): “Historical Newspaper User Interfaces: A Review”, in: *IFLA* <http://library.ifla.org/2578/1/085-ehrmann-en.pdf> [letzter Zugriff 31. August 2019].

Ferguson, NJ et al. (2015): “Media Content and Stock Returns: The Predictive Power of Press”, in: *Multinational Finance Journal* 19(1): 1–31.

García, D. (2013): “Sentiment during Recessions”, in: *The Journal of Finance* 68(3): 1267–1300.

Hanna, Alan J. / Turner, John D. / Walker, Clive B. (2017): “News Media and Investor Sentiment Over the Long Run”, in: QUCEH Working Paper Series, Working Paper No. 2017-06.

impresso: <https://impresso-project.ch/> [letzter Zugriff 31. August 2019].