

Historische Zeitungen kollaborativ erschließen: Die älteste, noch erscheinende Tageszeitung der Welt “under annotation”

Resch, Claudia

claudia.resch@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Kampkaspar, Dario

dario.kampkaspar@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Schopper, Daniel

daniel.schopper@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Die digitale Erschließung historischer Zeitungen lag vor wenigen Jahrzehnten noch außerhalb des Vorstellbaren. Inzwischen ist die Zeitung als Forschungsgegenstand etabliert und damit für viele verschiedene Disziplinen ins Zentrum gerückt: NutzerInnen sehen Nachrichtenblätter nicht nur als Quelle, die punktuell und komplementär zu anderen Texten befragt wird, sondern haben auch ein besonderes Interesse an der diachronen Entwicklung von historischen Zeitungen und ihren Themen, von Textsorten und natürlich von Sprache ganz im Allgemeinen. Die Transformation historischer Zeitungen in ein digitales Format, die von Bibliotheken, Archiven und Forschungseinrichtungen in den letzten Jahren stark vorangetrieben worden ist, befördert Fragestellungen dieser Art und überlässt den geistes- und sozialwissenschaftlichen Disziplinen einen reichen Schatz an Daten.

Im Zuge des wachsenden Interesses an historischem Zeitungsmaterial in digitaler Form haben Bibliotheken eigene Portale mit Bild-Digitalisaten und Kalenderübersichten eröffnet (vgl. etwa die Staats- und Universitätsbibliothek Bremen ¹, die Staatsbibliothek zu Berlin ², die Universität Bonn ³ oder die Österreichische Nationalbibliothek ⁴) oder stellen ihre Daten Europeana ⁵ zur Verfügung. Um historische Zeitungen besser durchsuchbar zu machen, wird vereinzelt bereits an der sorgfältigen Volltexterschließung besonderer historischer Zeitungen gearbeitet, vgl. etwa das Projekt „Volltextdigitalisierung der Staats- und Gelehrte[n] Zeitung des Hamburgischen Unpartheyischen Correspondenten und ihrer Vorläufer (1712-1848)“ ⁶ (Schuster, Wille 2016: 7-29) oder das „Mannheimer

Korpus für Historische Zeitungen und Zeitschriften in COSMAS II“ ⁷. Internationale Vernetzungsinitiativen ⁸ machen einerseits auf die Herausforderungen bei der Erschließung historischer Zeitungen aufmerksam (enorme Textmenge, fehlerhafte OCR-Ergebnisse, wenig Trainingsdaten) und lassen andererseits die Entwicklung gemeinsamer Strategien und Standards bei der Aufbereitung erkennen.

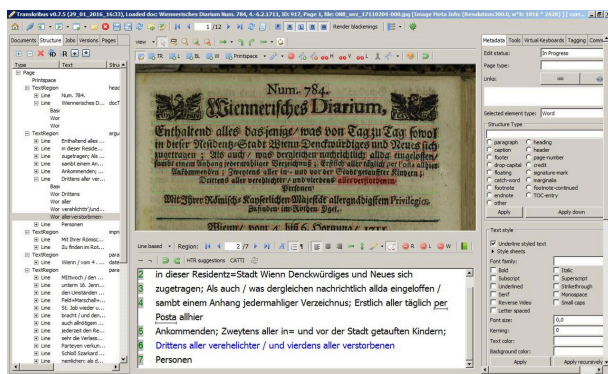
Die AutorInnen des vorliegenden Beitrags gehen davon aus, dass historische Zeitungen von Institutionen zwar „digital verfügbar“ gemacht werden, aber bislang kaum an die Erkenntnisinteressen potentieller NutzerInnen angepasst sind - auch, aber nicht nur aufgrund der oben genannten Herausforderungen und des erheblichen Aufwandes, der mit der Erschließung einer historischen Zeitung verbunden ist. Bevor ein aufwändiges Projekt startet, ist daher zu überlegen, wie eine Erschließung geplant sein muss, sodass sie idealerweise für mehrere Disziplinen von Nutzen ist. Bei der Konzeption ist darauf zu achten, dass keine (für den überwiegenden Teil der antizipierten Disziplinen) relevanten Informationen vernachlässigt werden oder verloren gehen: Die Entscheidungen, die zu treffen sind, beginnen (1) bei der Auswahl der Ausgaben, setzen sich (2) bei den Transkriptionsrichtlinien fort und lassen sich (3) bis zu den Annotationskonzepten weiterführen.

Anhand eines laufenden Projektes, das sich der Volltextdigitalisierung ausgewählter Nummern des „Wien[n]erischen Diariums“ ⁹ aus dem 18. Jahrhundert widmet, sollen all diese Aspekte kritisch hinterfragt und anhand von Beispielen, Erfahrungswerten und Zwischenergebnissen dargestellt werden. Das methodische Konzept, auf dem dieses konkrete Vorhaben beruht, wurde ausgehend von der Überzeugung gestaltet, dass insbesondere bei einer vielseitig nutzbaren Ressource wie dem “Wienerischen Diarium” die einzelnen Fachdisziplinen und NutzerInnen bereits möglichst früh in den interdisziplinären Erschließungsprozess einzubeziehen sind. Im Vortrag soll darüber berichtet werden, welche Maßnahmen bereits gesetzt wurden, um die Fachwissenschaften zu vernetzen, welche Tools im Projekt entwickelt wurden, um Personen außerhalb des Kernteams zu involvieren, und durch welche digitalen Angebote diese Kollaboration ermöglicht wird und gelingen kann.

Auswahl der Ausgaben: Um ein ausgewogenes Korpus von mehreren hundert Ausgaben verteilt über das 18. Jahrhundert zu erstellen, waren sowohl ExpertInnen mehrerer geisteswissenschaftlicher Disziplinen als auch LeserInnen der heutigen Wiener Zeitung dazu aufgefordert, jene Nummer(n) online zu nominieren, die sie als besonders relevant einstufen würden. Bei der Auswertung dieses Calls hat sich erneut bestätigt, wie breit das Themenspektrum und damit die individuellen Erkenntnisinteressen sind: Nominiert und zur Volltextdigitalisierung empfohlen wurden Ausgaben mit Geburten, Taufen und Sterbefällen bekannter Persönlichkeiten, Geburts- und Namenstage, Krönungen und Erbhuldigungen, kirchliche und weltliche

Feste, Ankündigungen, Eröffnungen und Einweihungen sowie die Besuche prominenter Gäste in der Residenzstadt. Errungenschaften im weitesten Sinn – wie die Erklärung der Menschenrechte oder der Beginn der Luftfahrt – waren ebenso in der Auswahl wie das medienimmanente Thema der Herausgeberschaft. Die Ergänzungen, die das Projektteam letztlich vorgenommen hat, betrafen daher weniger die Vielfalt der angesprochenen Themen, sondern waren darauf ausgerichtet, zeitliche Lücken zu füllen. Um das Wien[n]erische Diarium als Korpus in einer kontinuierlich chronologischen Jahresabfolge plausibel dokumentieren zu können, wurden die Nominierungen dahingehend komplettiert, dass sich die Umbrüche und Wendungen in einer Periode sich verändernder politischer, sozialer, wissenschaftlicher und künstlerischer Bedingungen im 18. Jahrhundert, idealerweise auch korpusgestützt nachvollziehen lassen.

Texterstellung und -transkription: Ausgangspunkt für die Weiterverarbeitung sind jene Image-Digitalisate, welche die Österreichische Nationalbibliothek in ANNO (Austrian Newspapers Online) zur Verfügung stellt.¹⁰ Die Erstellung des digitalen Textes für die ausgewählten Nummern erfolgt mit Transkribus¹¹, genauer mittels der Handwritten Text Recognition (HTR). Anhand einer kleinen Zahl an Ausgaben, die im Rahmen von projektbezogenen Vorstudien (vgl. Resch et al. 2016) bearbeitet wurden, konnte ein erstes Modell für das Diarium erstellt werden. Je nach Beschaffenheit der Digitalisate liegt die Genauigkeit hiermit zwischen 70% und 95%. Um die Qualität zu steigern, werden mehrere Tranchen von 40 bis 50 Ausgaben von zwei unterschiedlichen externen Dienstleistern nach den für dieses Projekt erstellten Transkriptionsrichtlinien erfasst, sodass das Modell weiter trainiert und die Erfassungsgenauigkeit erhöht werden kann.



Ein „Reporting-Tool“¹², das direkt auf die Transkribus-Plattform zugreift, dokumentiert den Bearbeitungsstatus der ausgewählten Einzelnummern, gibt Auskunft über deren Umfang (Zählung von Regionen, Zeilen und Wörtern) und informiert die Fachgemeinschaft tagesaktuell über den Fortschritt des Projekts.

Erprobung von Annotationskonzepten: Parallel dazu arbeitet das Kernteam an einer digitalen Arbeits-

Annotationsoberfläche: Für die Präsentation der Texte wird nach derzeitigem Projektstand eine auf eXist basierende Umgebung genutzt, die auch in anderen Projekten des Instituts und in anderen Institutionen Anwendung findet. In die HTML-Präsentation integriert ist die Möglichkeit zur Annotation des Textes. Dabei kann entweder der Text korrigiert, eine Entität ausgezeichnet und/oder identifiziert oder eine Volltextanmerkung geschrieben werden. Die Möglichkeiten der Annotation, die Modellierung der annotierten Entitäten, ihre Verwaltung und Darstellung sollen im Rahmen eines „Annotate-a-thons“ in Zusammenarbeit mit den NutzerInnen weiterentwickelt werden. Für das Projektteam ist es etwa wichtig zu erfragen, welche unterschiedlichen Sichtweisen es seitens der verschiedenen Disziplinen auf den Text gibt, welche Aspekte bei der Erschließung von besonderer Relevanz sind oder ob es so etwas wie einen „kleinsten, gemeinsamen Nenner“ aller Annotationskonzepte geben kann. Die Einbeziehung von ExpertInnen im Annotationsprozess ist dem Projektteam ein besonderes Anliegen: Es begreift das Annotieren als höchst anspruchsvolle Forschungsleistung, die ein profundes historisch-kulturelles Wissen bei der Beurteilung erfordert und für eine Quelle wie dem „Wienerischen Diarium“ bei genauerer Betrachtung eigentlich nur gemeinsam erbracht werden kann. Ein kollaboratives Annotieren technisch vorzusehen und zu ermöglichen, ist hierbei die besondere Herausforderung. Die dafür entstehende benutzerfreundliche Präsentations- wie auch die Annotationsumgebung wird unter einer freien Lizenz zur Nachnutzung zur Verfügung gestellt werden.

Fußnoten

1. Vgl. <https://www.suub.uni-bremen.de/ueber-uns/projekte/alte-zeitungen/> [letzter Zugriff 14. Januar 2018]
2. Vgl. <http://zeffys.staatsbibliothek-berlin.de/> [letzter Zugriff 14. Januar 2018]
3. Vgl. <http://digitale-sammlungen.ulb.uni-bonn.de/ulbbnz/date/list/229854> [letzter Zugriff 14. Januar 2018]
4. Vgl. <http://anno.onb.ac.at/> [letzter Zugriff 14. Januar 2018]
5. Vgl. <http://www.europeana-newspapers.eu/> [letzter Zugriff 14. Januar 2018]
6. Vgl. <https://kw.uni-paderborn.de/institut-fuer-germanistik-und-vergleichende-literaturwissenschaft/germanistische-und-allgemeine-sprachwissenschaft/schuster/forschung/projekte/der-hamburgische-unpartheyische-correspondent-volltextdigitalisierung/> [letzter Zugriff 14. Januar 2018]
7. Vgl. <http://repos.ids-mannheim.de/mkhz-beschreibung.html> [letzter Zugriff 14.01.2018]
8. Vgl. etwa die Einrichtung einer „Special Interest Group Newspapers“ bei der TEI-Konferenz 2016 in Wien, das CLARIN-Vernetzungstreffen „Working with Digital Collections of Newspapers“ 2016 in Leuven oder das

“Transatlantic Digitised Newspaper Symposion” 2017 in London.

9. Das „Wien[n]erische Diarium“ ist am 8. August 1703 erstmals erschienen, als „Wiener Zeitung“ bis heute erhältlich und somit die älteste, noch existierende Tageszeitung der Welt. Insbesondere im 18. Jahrhundert lässt sich die Entwicklung der Zeitung von den Anfängen des modernen Journalismus in einer spannenden Zeit und unter sich verändernden politischen, sozialen und künstlerischen Bedingungen gut nachverfolgen und mitvollziehen.

10. Vgl. <http://anno.onb.ac.at/cgi-content/anno?aid=wrz> [letzter Zugriff 14. Januar 2018]

11. Vgl. <https://transkribus.eu/Transkribus/> [letzter Zugriff 14. Januar 2018]

12. Vgl. <https://www.oeaw.ac.at/acdh/projects/wienerisches-diarium-digital/> [letzter Zugriff 14. Januar 2018]

historic newspaper for interdisciplinary studies with the TEI Guidelines. <https://goo.gl/a9SmLk>

Schuster, Britt-Marie / Wille, Manuel (2016): Von der Kanzlei- zur Bürgersprache? Textsortengeschichtliche Betrachtungen zur „Staats- und gelehrten Zeitung des Hamburgischen unpartheyischen Correspondenten“ im 18. Jahrhundert in: Jahrbuch für Kommunikationsgeschichte 17 Stuttgart: Franz Steiner Verlag 7-29.

ZEFYS Das Zeitungsinformationssystem der Staatsbibliothek zu Berlin: <http://zefys.staatsbibliothek-berlin.de/> [letzter Zugriff 14. Januar 2018]

Bibliographie

ANNO – AustriaN Newspapers Online. <http://anno.onb.ac.at/> [letzter Zugriff 14. Januar 2018]

Harald Burger / Luginbühl, Martin (2015): *Mediensprache*, Berlin / Boston, 2015, 39-45.

Reisner, Andrea / Schiemer, Alfred (2016): „Das Wien(n)erische Diarium und die Entstehung der periodischen Presse“ in: *Österreichische Mediengeschichte 1*: 87-112.

Digitalisierung der vollständigen deutschsprachigen Zeitungsbestände des 17. Jahrhunderts der Staats- und Universitätsbibliothek Bremen: <https://www.suub.uni-bremen.de/ueber-uns/projekte/alte-zeitungen/> [letzter Zugriff 14. Januar 2018]

Europeana Newspapers: <http://www.europeana-newspapers.eu/> [letzter Zugriff 14. Januar 2018]

Mannheimer Korpus für Historische Zeitungen und Zeitschriften in COSMAS II: <http://repos.ids-mannheim.de/mkhz-beschreibung.html> [letzter Zugriff 14. Januar 2018]

Newspapers Sammlung der Universitäts- und Landesbibliothek Bonn: <http://digitale-sammlungen.ulb.uni-bonn.de/ulbbnz/date/list/229854> [letzter Zugriff 14. Januar 2018]

Projekt „Volltextdigitalisierung der Staats- und Gelehrte[n] Zeitung des Hamburgischen Unpartheyischen Correspondenten und ihrer Vorläufer (1712-1848): <https://kw.uni-paderborn.de/institut-fuer-germanistik-und-vergleichende-literaturwissenschaft/germanistische-und-allgemeine-sprachwissenschaft/schuster/forschung/projekte/der-hamburgische-unpartheyische-correspondent-volltextdigitalisierung/> [letzter Zugriff 14. Januar 2018]

Resch, Claudia / Schopper, Daniel / Hanneschläger, Vanessa / Wohlfarter, Eva / Mader, Anna / Fischer, Nora (2016): *Wienerisches Diarium Digital: Unlocking a*