

Digitalisierung

BIRGIT SEIDERER

Die Digitalisierung von Zeitungen im deutschsprachigen Raum – ein Zustandsbericht

Zeitungsdigitalisierung steht im Fokus vieler Digitalisierungsaktivitäten im Ausland. Große nationale Zeitungsportale belegen dies eindrücklich. Im deutschsprachigen Raum widmet man der Zeitungsdigitalisierung erst langsam stärkere Aufmerksamkeit. Der Beitrag versucht in ausgewählten Schlaglichtern Probleme und Herausforderungen zu skizzieren sowie das bereits vorhandene, sehr heterogene Angebot gebündelt zu betrachten. Abschließend wird der Frage der Erschließung und damit der Präsentation von digitalisierten Zeitungen Raum gegeben; dies nicht zuletzt deswegen, da die Wahrnehmung und Akzeptanz der Angebote bei den Nutzern dadurch in nicht geringem Maße beeinflusst wird.

Newspaper digitisation is the focus of many digitisation projects abroad. Major national newspaper portals supply impressive proof for this. In German-speaking countries this area of endeavor is only gradually gaining attention. This article selectively highlights the problems and challenges involved and surveys the current, very heterogeneous projects as a whole. In conclusion, the issue of access and, with it, the presentation of digitalized newspapers is considered; this is significant, if for no other reason than because the perception and acceptance of such a service among the users will be influenced by it to no small degree.

ZEITUNGEN – ZUNEHMENDER BRENNPUNKT DER DIGITALISIERUNG?

Fokussiert man den Blick nicht sofort auf den deutschsprachigen Raum, so fallen die regen und umfassenden Aktivitäten auf dem Gebiet der Zeitungsdigitalisierung auf.¹ Dies gilt nicht nur für die USA; insbesondere die Nationalbibliotheken Australiens und Neuseelands haben mit ihren Zeitungsportalen hohe Maßstäbe gesetzt (Abb. 1).² Komfortable Browsingfunktionen und Volltextrecherche werden hier in Präsentationssystemen angeboten, die nicht nur eine beeindruckende Anzahl an Zeitungstiteln integrieren, sondern vor allem der Nutzerfreundlichkeit («usability») einen großen Stellenwert einräumen.

Auch in Europa hat die Zeitungsdigitalisierung Hochkonjunktur. Drei Beispiele seien kurz angeführt: (1) Mit dem 2009 abgeschlossenen Projekt »The British Newspapers 1620–1900« hat die British Library in einer Public Private Partnership mit Gale Cengage Learning wichtige regionale und lokale Zeitungen digitalisiert und, wenngleich mit Zugangsbeschränkungen, im Volltext zugänglich gemacht.³ (2) Die finnische Nationalbibliothek hat mittlerweile sämtliche Zeitungen, die in Finnland zwischen 1771 und 1900 erschienen sind, digitalisiert, auf einem hohem Niveau erschlossen und präsentiert diese in dem Zeitungportal »Historical Newspaper Library«.⁴ (3) Als letztes Beispiel kann man die Nationalbibliothek der Niederlande (KB) anführen, die 2008 damit begonnen hat, ihre histori-

schen Zeitungsbestände, rd. 8 Mio. Seiten, zu digitalisieren.⁵ Allen diesen Projekten ist gemein, dass hinter ihnen nationale Anstrengungen mit hohem finanziellen Input stecken.

Der Stellenwert der Zeitungsdigitalisierung lässt sich auch am »EDL Report on Digitisation in European National Libraries« von 2008 ablesen; bezogen sich 2006 bereits 30 % sämtlicher Digitalisierungsvorhaben der Nationalbibliotheken Europas auf Zeitungen, so werden es 2012 entsprechend der bereits begonne-



Birgit Seiderer

The screenshot shows the 'PAPERS PAST' search interface. At the top, there are navigation tabs: 'Papers Past Home', 'Introduction', 'Search', and 'Browse'. Below this is a search bar with the text 'Papers Past > Search' and 'Enter search criteria'. The search options are 'Search using keywords:' with radio buttons for 'Exact phrase', 'Any of your words', and 'All of your words'. There is a search input field and a 'Search' button. Below the search bar are 'Optional Filters'. The first filter is 'Filter by newspaper, date or content type: (to select more than one option hold down "ctrl")'. Under 'Newspaper:', there is a dropdown menu with options: 'All newspapers', 'Ashburton Guardian (Canterbury)', 'Bay Of Plenty Times (Bay Of Plenty)', 'Bruce Herald (Otago)', and 'Bush Advocate (Hawke's Bay)'. Under 'Date:', there are 'From:' and 'To:' sections, each with dropdowns for 'Day', 'Month', and 'Year'. Under 'Content type:', there is a dropdown menu with options: 'All content types', 'Articles', 'Advertisements', and 'Illustration captions'. At the bottom of the filters, there is a checkbox for 'Show preview images' and 'Reset' and 'Search' buttons. The footer of the interface includes the text 'Te Puna Mātauranga o Aotearoa NATIONAL LIBRARY OF NEW ZEALAND' and links for 'About this site', 'Site map', 'Accessibility', and 'Contact us'.

Abb. 1: Das Zeitungportal »Papers Past« der neuseeländischen Nationalbibliothek mit vielfältigen Suchoptionen

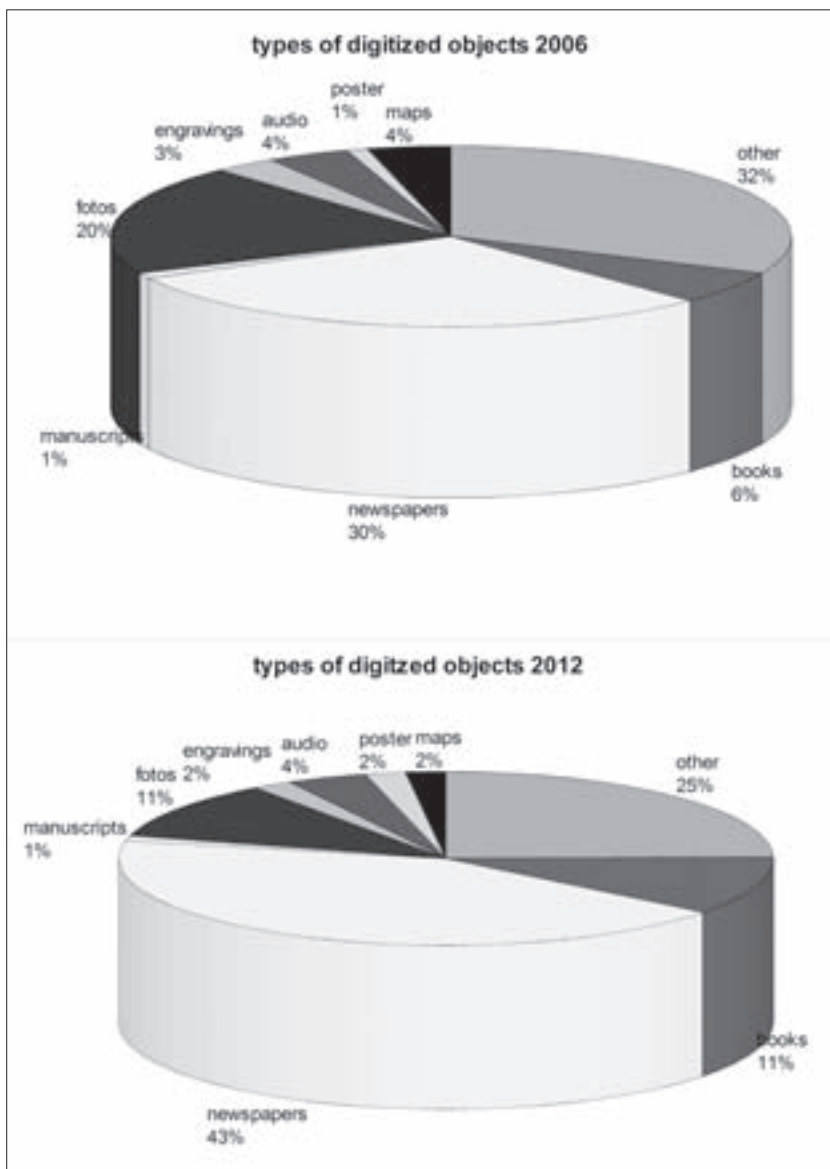


Abb. 2: EDL Report on Digitisation in European National Libraries 2006-2012, 29th February 2008

nen und geplanten Projekte nach diesem Bericht voraussichtlich rund 43 % sein (Abb. 2).⁶

Verglichen mit dem angloamerikanischen Raum und auch manchen europäischen Nachbarländern, steckt die Zeitungsdigitalisierung in Deutschland hingegen tendenziell noch in den Kinderschuhen. Die Gründe hierfür sind mannigfaltig. So erfordern die traditionsreiche, regional dichte Zeitungslandschaft Deutschlands einerseits und das föderal differenzierte Bibliothekswesen andererseits in nicht geringem Umfang konzeptionelle Anstrengungen. Noch schwerwiegender ist allerdings die Frage der Finanzierung, um digitalisierte Zeitungsangebote nachhaltig und unter Berücksichtigung wesentlicher Erschließungs- und Präsentationsanforderungen auf die Beine stellen zu können.

traditionsreiche, regional dichte Zeitungslandschaft

In der Zeitschriftendatenbank sind summa summarum rund 460 digitalisierte Zeitungstitel nachgewiesen. Zählt man nur die Titel, die von Bibliotheken in Deutschland retrodigitalisiert wurden, reduziert sich die Anzahl auf rund 210.⁷ Rein quantitativ ist das eine durchaus beachtliche Anzahl. Die Vielzahl an Präsentationsorten für einzelne Zeitungen sowie Titel, die nur für einzelne Jahrgänge, manchmal auch nur für einzelne Nummern vorliegen, erschweren allerdings die wissenschaftliche Nutzung und schöpfen den Mehrwert von digitalisierten Zeitungen nur unvollständig aus. Die Anzahl der Digitalisierungsakteure aus dem Bibliothekswesen in Deutschland liegt insgesamt bei 24 Einrichtungen. Trotz aller Schwierigkeiten zeigt das letztlich aber auch, dass die sich aus der Zeitungsdigitalisierung ergebenden Vorteile durchaus gesehen werden.

ZEITUNGEN ALS DIGITALISIERUNGSGEGENSTAND – »NOT FOR THE FAINT OF HEART«⁸

Zeitungen bergen schon allein in ihrer Materialität Herausforderungen für die Digitalisierung in sich. Holzschliffpapier und Säurefraß haben vielen Zeitungsoiginalen schwer zugesetzt.⁹ Für die Benutzung sind sie oft nur noch eingeschränkt oder über Mikrofilme zugänglich. Eine Digitalisierung vom Mikrofilm kann durchaus kostengünstig gelingen, wenn die Filmqualität ausreichend ist.¹⁰ Gerade bei historischen Zeitungen in Frakturschrift ist das Ergebnis für eine OCR-Erkennung bei der Digitalisierung vom Mikrofilm allerdings bisher oftmals schlechter als beim Scannen vom Original. Auch die Großformate der Zeitungen sowie die nicht selten zu engen Bindungen machen das Digitalisieren zu einer Herausforderung. Ein schneller und günstiger Durchlaufscan ist nur möglich, wenn die Zeitungsbände aufgeschnitten werden. Bei besonders wertvollen oder gar unikatlen Beständen kommt dies eigentlich nur dann in Frage, wenn bereits eine Beschädigung des Originals vorliegt und weitergehende Maßnahmen der Bestandserhaltung der Digitalisierung folgen.¹¹ Das aufwendigere, manuelle Aufsichtscannen ist daher das vorherrschende Verfahren zur Digitalisierung der großformatigen, fragilen Zeitungen. Aufgrund des Aufwandes, der Tücken sowie v.a. der enormen Masse ist es nicht verwunderlich, dass Zeitungsdigitalisierungsprojekte nicht nur im Bereich der Erschließung und Präsentation, sondern auch im Scanstadium mehrheitlich outgesourct und an spezialisierte Dienstleister vergeben werden.

Als Herausforderungen für die Konzeption größerer Digitalisierungsprojekte können des Weiteren exemplarisch die oftmals nicht ausreichende Er-

schließungssituation für historische Zeitungen, die schwierige Rechtlklärung für nicht zweifelsfrei urheberrechtsfreie Jahrgänge und der nicht geringe Personalaufwand für die Vorarbeiten der Digitalisierung selbst bei Vergabe an einen Dienstleister (z. B. das Kollationieren) genannt werden.

Die Vorteile der Zeitungsdigitalisierung liegen allerdings deutlich auf der Hand: (1) Mühsam benutzbare Zeitungsbestände, sei es im Original oder auf Mikrofilm, werden durch die Digitalisierung ortsunabhängig, weltweit und jederzeit online zugänglich gemacht. Die Benutzerfreundlichkeit von Zeitungsdigitalisaten ist ungleich höher als von Mikrofilmen, selbst bei einer reinen Imagedigitalisierung mit niedriger Erschließungstiefe. (2) Die Digitalisierung bietet des Weiteren auch die Chance des retrospektiven Bestandsaufbaus, da durch Kooperation mit anderen Bibliotheken hinsichtlich fehlender oder zu stark beschädigter Jahrgänge oder Nummern digital eine vollständige Ausgabe gebildet werden kann. Zeitungsportale unterschiedlichster Fassung können darüber hinaus verteilte Bestände virtuell zusammenführen. Für die wissenschaftliche Forschung ist das ein ungeheurer Zugewinn. (3) Die Zeitungsoriginale selbst können nach der Digitalisierung geschont werden; die Digitalisierung ist insofern – bei allen Ambivalenzen, z. B. der Materialbelastung beim Scannvorgang selbst – ein wichtiger Beitrag zum Bestandserhalt.

SKIZZE DES ANGEBOTS IM DEUTSCHSPRACHIGEN RAUM: PROJEKTKONTEXTE UND ZUGÄNGE

Versucht man die verteilten Angebote strukturiert zu betrachten und aus der Vogelperspektive zu konturieren, ist es hilfreich, unterschiedliche Projektkontexte zu unterscheiden, die letztlich für den Nutzer auch unterschiedliche Zugänge zu den Zeitungsangeboten repräsentieren.

— (1) Einzelprojekte:

Einige Bibliotheken haben Zeitungen von lokaler oder regionaler Bedeutung bereits digitalisiert und stellen damit wichtige geschichts- und kulturwissenschaftliche Quellen zur Verfügung, nicht selten in Eigenleistung, teilweise in Kooperation mit der kommunalen Ebene oder mit Verlagen. An vorderster Stelle zu nennen ist hier der älteste überlieferte Jahrgang der »Relation«, der ersten Zeitung der Welt, der von der UB Heidelberg digitalisiert wurde.¹² Exemplarisch kann man auch auf die UB Freiburg verweisen, die in ihren digitalisierten historischen Beständen die »Freiburger Zeitung« (1784–1943) und das »Staufener Wochenblatt« (1875–1968) anbietet.¹³ Ein weiteres Beispiel

aus dem deutschsprachigen Raum, aber mit Blick über die Grenze wäre die Universität Bern, die über die Seiten von »DigiBern« das »Intelligenzblatt für die Stadt Bern« (1834–1922) bereitstellt.¹⁴ Präsentiert werden diese Einzelprojekte zumeist zusammen mit anderen Objekten in den Digitalen Sammlungen der jeweiligen Bibliothek. Inwieweit innerhalb der digitalen Sammlungen entsprechende Sucheinstiege für die Publikationsform Zeitung vorhanden sind, wird dabei von Bibliothek zu Bibliothek unterschiedlich gehandhabt.

— (2) Projekte mit thematischen Schwerpunkten

a) Insbesondere für kurzlebige, in spezifischen historischen Kontexten zu verortende Titel finden sich mitunter thematische Zugänge. So stellt die UB Heidelberg den Einstieg zu »Feldzeitungen aus dem 1. Weltkrieg – online« zur Verfügung.¹⁵ Die »Bayerische Landesbibliothek Online« fasst 15 ausgewählte Zeitungen, die nur in der Revolutionszeit existierten, mit dem Projekt »Bayerische Zeitungen 1848–1850« zusammen (Abb. 3).¹⁶ Eine Quellenarbeit mit diesen flüchtigen, inhaltlich aber zusammengehörenden Presseerzeugnissen wird für den Wissenschaftler durch den thematischen Einstieg mit Sicherheit erleichtert, dies umso mehr, wenn wie in diesen Fällen charakterisierende Erläuterungen oder weiterführende bibliographische Hinweise zur Verfügung gestellt werden. Das Angebot der »Bayerischen Landesbibliothek Online« wird zudem durch das »Historische Lexikon Bayerns« flankiert.¹⁷

b) Von diesen thematischen Schwerpunkten abzugrenzen sind groß angelegte Projekte, allesamt von der DFG gefördert, die ein eigenes, inhaltlich spezifisch ausgerichtetes Portal bilden. »Compact Memo-

Vorteile der Zeitungsdigitalisierung

unterschiedliche Zugänge zu den Zeitungsangeboten



Abb. 3: Zeitungstitel der »Bayerischen Landesbibliothek Online« aus dem Schwerpunkt »Revolutionszeitungen 1848 bis 1850«. Der Bildausschnitt weist zugleich auf eines der größten Probleme der Erschließung mittels OCR hin: die Frakturschrift.



Abb. 4: Eines der ersten großen Digitalisierungsprojekte für Periodika im deutschsprachigen Raum: »Exilpresse digital. Deutsche Exilzeitschriften 1933–1945«

Digitalisierungsleistungen einzelner Institutionen

ry« kann als Paradebeispiel angeführt werden;¹⁸ das Wissenschaftsportal für Jüdische Studien umfasst rund 118 Zeitungen und Zeitschriften und ist entstanden in Kooperation des Lehr- und Forschungsgebiets Deutsch-jüdische Literaturgeschichte der RWTH Aachen mit der UB Frankfurt am Main und der Kölner Bibliothek Germania Judaica.

Die wissenschaftlich bedeutenden Zeitungen und Zeitschriften waren vor der Digitalisierung schwer zugänglich, und vollständige Jahrgänge waren nur an wenigen Bibliotheken zu finden. Wie »Compact Memory« war es auch dem zeitlich früh anzusiedelnden Projekt »Exilpresse digital: Deutsche Exilzeitschriften 1933–1945« ein wichtiges Anliegen, die im Original nur selten an einem Ort vollständig vorhandenen, zudem schlecht erhaltenen Zeitungen virtuell zusammenzuführen und leicht zugänglich zu machen (Abb. 4).¹⁹ Im Projekt »Exilpresse digital« wurden zwischen 1998 und 2003 30 ausgewählte Titel der Exilpresse aus dem Bestand des Deutschen Exilarchivs 1933–1945 Frankfurt am Main und der Sammlung Exil-Literatur Leipzig in der Deutschen Bibliothek Frankfurt am Main digitalisiert. Die hier genannten Großprojekte haben zudem etwas gemeinsam, was bei thematischen Zugängen nicht selten ist: Zeitschriften und Zeitungen werden als Publikationsform nicht unterschieden.²⁰ Der thematische Projektkontext begründet dies eindeutig; das jeweils eigene Portal ermöglicht zudem eine geschlossene Präsentation und bietet einen komfortablen Zugang für den Nutzer. Das sollte aber nicht darüber hinwegtäuschen, dass die Zeitung als serielle Quelle mit häufiger Periodizität und einer Vielzahl an

Präsentationslösungen insgesamt erst im Aufbau

Autoren und Rubriken außerhalb dieser thematischen Projektzusammenhänge spezifische Erschließungs- und Präsentationsanforderungen mit sich bringt, die bei einer Verquickung der Publikationsformen in Portalen zu bedenken sind.

(3) Regionale Portale und Zeitungsportale als einheitliche Präsentationsform für Zeitungsprojekte mit unterschiedlichstem Entstehungskontext

a) Zusätzlich zum Nachweis in den digitalen Sammlungen einer einzelnen Institution kann man digitalisierte Zeitungen auch als Modul in regionalen Portalen finden, die verschiedenartige kulturwissenschaftliche Angebote in sich vereinen. Diese regionalen Portale haben nicht nur die Möglichkeit, die differenzierte Presselandschaft sehr gut widerzuspiegeln, zumindest wenn eine ausreichende kritische Masse an digitalisierten Zeitungen vorliegt. Der Mehrwert liegt hier in der Chance, Digitalisierungsleistungen einzelner Institutionen integriert anzubieten. Als Beispiele für derartige Portale kann man »Sachsen.digital«²¹, das rheinland-pfälzische Portal »dilibri«²² oder die »Bayerische Landesbibliothek Online«²³ anführen. In »Sachsen.digital« bindet u. a. die SLUB Dresden Digitalisate ein. In »dilibri« findet man momentan 33 Zeitungen (davon 23 Fastnachtzeitungen) des Landesbibliothekszentrums Rheinland-Pfalz, der UB Trier sowie der Stadtbibliothek Mainz. Die »Bayerische Landesbibliothek Online« bietet 27 Zeitungen an (davon die o. g. 15 Revolutionszeitungen 1848/49); enthalten sind Einzelprojekte der Bayerischen Staatsbibliothek, der UB Augsburg sowie Kooperationsprojekte der Bayerischen Staatsbibliothek mit weiteren Einrichtungen.²⁴

b) Während im Ausland die Digitalisate der großen Zeitungsprojekte oftmals in eigenen Portalen präsentiert werden, befinden sich derartige Zeitungsportale in Deutschland insgesamt erst im Aufbau.²⁵ Als übergreifendes Zeitungsportal ist »ZEFYS« geplant, das Zeitungsinformationssystem der Staatsbibliothek zu Berlin, das Module Virtueller Fachbibliotheken enthalten wird und Nachweis und Zugriff auf aktuelle und historische Zeitungen aus Deutschland und dem Ausland bieten soll.²⁶

Ein bereits etabliertes, nationales Portal im deutschsprachigen Raum findet man, wenn man über die Grenze nach Österreich blickt: »ANNO« (AustriaN Newspapers Online), der virtuelle Zeitungslesesaal der Österreichischen Nationalbibliothek, weist mittlerweile über rund 4.760.000 digitalisierte Seiten aus 115 verschiedenen Periodika auf.²⁷ Das Portal wird ständig erweitert und konnte Anfang 2010 den Spitzenwert von mehr als 2.000 Lesern an einem Tag aufwei-

sen. Der Erfolg von »ANNO« liegt vermutlich gerade auch im Verzicht auf eine enge thematische Beschränkung, was für das Erreichen einer kritischen Masse mit Sicherheit förderlich war; eine Gesamtschau der historischen Periodika wird damit sukzessive ermöglicht. Hinzu kommt die durchdachte Erschließung und Präsentation der Digitalisate, die die spezifischen Eigenheiten der Zeitung als Publikationsform klar berücksichtigen.

»UND DANN AUCH NOCH DIESE LÄSTIGE FRAKTUR...«²⁸ – ZUR ERSCHLIESSUNGSTIEFE UND USABILITY DER VORHANDENEN ANGBOTE

Eine der größten Herausforderungen der Zeitungsdigitalisierung ist die Frage der Erschließung, da sie unmittelbare Konsequenzen für die Retrievalfähigkeit und die Präsentationsmöglichkeiten hat; sie stellt zudem einen nicht unwesentlichen Kostenfaktor dar. Grundsätzlich können zwei Ebenen der Erschließung unterschieden werden, die entweder (1) eine Browsingfunktion in den Images ermöglichen oder aber (2) zusätzlich die Volltextrecherche anbieten. Bei beiden Optionen gibt es mehrere Abstufungsmöglichkeiten hinsichtlich der Erschließungstiefe.

Am weitesten verbreitet ist im deutschsprachigen Raum die Blätterfunktion im Image. Hierbei stößt

man auf unterschiedliche Varianten. Nicht selten findet man Angebote, die jahrgangs- oder monatszusatmenfassende PDFs zur Verfügung stellen, ohne weitere Strukturdaten anzubieten;²⁹ etwas häufiger ist die Erfassung auf Nummernebene einer Zeitung und die Anzeige als Bilddatei mit der Option eines PDF-Downloads.

Eine Kalenderfunktion als Grundstruktur ist dabei selten implementiert. Wird das Datum als zentrales Strukturmerkmal erfasst, ist ein Browsing in Zeitungsdigitalisaten für den Nutzer wesentlich komfortabler und übersichtlicher. Ein Grund für das geringe Angebot an dieser Funktionalität liegt mit Sicherheit darin, dass der Mehrwert erst bei einer quantitativ größeren Menge an Zeitungen deutlich sichtbar wird. Über eine Jahresübersicht aller verfügbaren Titel kann z. B. schrittweise bis zur Kalenderübersicht aller erschienenen Zeitungstitel an einem Tag eingeschränkt werden (Abb. 5).

Das ist insbesondere auch aus wissenschaftlicher Sicht begrüßenswert, da Recherchen in Zeitungsquellen oftmals zu einem gewissen Ereignis oder Thema über mehrere Titel hinweg durchgeführt werden. Bei »ANNO« und dem digitalen Zeitungsarchiv der Landesbibliothek Dr. Friedrich Tessimann, Südtirol, lässt sich diese Art der Erschließung und Präsentation exemplarisch nachvollziehen.³⁰ Neben dem Browsing

Kalenderfunktion als Grundstruktur



© Österreichische Nationalbibliothek

Abb. 5: Letzte Ebene der Kalenderfunktion bei »ANNO«: Die Tagesübersicht aller erschienenen Zeitungen mit Anzeige von thumbnails.

nach Zeitungstiteln ist bei entsprechend strukturierten, größeren digitalisierten Mengen eine Browsingoption nach Regionen oder Erscheinungsorten ebenfalls sehr sinnvoll, wie sie z. B. im neuseeländischen Portal »Papers Past« oder auch im Portal »British Newspapers 1800–1900« der British Library implementiert ist.³¹ Derart ausgebaute Browsing-Funktionalitäten bieten bereits sehr komfortable Recherchemöglichkeiten und sind als ein wesentlicher Strang der Erschließung wünschenswert.

»State of the Art« ist allerdings die Digitalisierung unter Verwendung von OCR und das Anbieten von Portallösungen mit komplexen Suchmöglichkeiten sowie komfortablen, leicht zu wechselnden Ansichten.

automatische Texterkennung

Die automatische Texterkennung stößt bei der Zeitungsdigitalisierung dabei aber insbesondere, wenn gleich bei weitem nicht ausschließlich, auf zwei Probleme: Dies sind (1) die Frakturschriften der meisten historischen Zeitungen im deutschsprachigen Raum, die technisch eine größere Herausforderung darstellen als Antiquaschriften. Kommen Papier- oder Bindungsprobleme zur Frakturschrift hinzu, kann die Erkennungsgenauigkeit der OCR schnell unter 70 % fallen.³² (2) Auch das unregelmäßige Layout der historischen Tageszeitungen, insbesondere im 19. und Anfang des 20. Jahrhunderts, muss als eigenes Problem erwähnt werden. Spalten- oder auch seitenüberspringende Artikelplatzierungen erschweren automatische Verfahren der Segmentierung als Grundlage einer Volltexterschließung auf Artekelebene.

Eine Volltextsuche wird im deutschsprachigen Raum daher nahezu ausschließlich in den thematischen Portalen großer Einzelprojekte wie »Compact Memory« und »Exilpresse digital« angeboten.³³ Auch die »Amtspresse Preußens«, ein ebenfalls von der DFG gefördertes Digitalisierungsprojekt der Staatsbibliothek zu Berlin, ist im Volltext durchsuchbar. Gerade letzteres verweist aber auf die Problematik der Erschließungstiefe: Die »Amtspresse Preußens«, welche die Digitalisierung der »Provinzial-Correspondenz« (1863–1884) und der »Neuesten Mittheilungen« (1882–1894) umfasst, orientiert sich an der gedruckten wissenschaftlichen Edition.³⁴ Die »Normalansicht« bei der digitalisierten Amtspresse ist der transkribierte Text und nicht das Zeitungsimago. Die Aufbereitung für die Volltextsuche erfolgte dabei über das sogenannte Double-Keying-Verfahren, d. h. die Texte wurden von zwei Schreibkräften getrennt abgetippt, in zwei Dateien gespeichert und am Ende elektronisch miteinander verglichen. Abweichungen wurden manuell überprüft und Fehler ggf. ausgebessert. Mit diesem Verfahren kann eine sehr hohe Treffergenauigkeit von rd. 99,9 % erzielt werden.³⁵ Verbunden ist mit diesem

Problematik der Erschließungstiefe

Verfahren allerdings ein Aufwand, der in diesem Umfang nur für streng ausgewählte Projekte zu realisieren ist.

Ein Verzicht auf OCR-Verfahren wäre deshalb und auch aufgrund der Erwartungen der Nutzer problematisch.³⁶ Die erweiterten Recherchemöglichkeiten überwiegen die technisch bedingten Nachteile »schmutziger« OCR um einiges. Auch die Österreichische Nationalbibliothek beginnt trotz des sehr komfortablen Browsing-Portals »ANNO« die Volltexterschließung anhand eines Pilotprojektes zu erproben.³⁷ Ein Mittelweg zwischen unkorrigiertem, schmutzigem OCR und dem Transkribieren könnte die automatische Texterkennung bei gleichzeitiger intellektueller Korrektur der Überschriften darstellen.³⁸ Dieser Korrekturvorgang bietet die Chance, die Treffergenauigkeit der OCR derer des Transkribierens deutlich anzugleichen. Mittels automatischer Rankingverfahren und Drilldowns, die bei sehr hohem Erschließungsniveau auch Rubriken (z. B. Werbung, Karikaturen etc.) enthalten können, besteht des Weiteren die Möglichkeit, dem Entstehen von unbezwingbar erscheinenden Datengebirgen entgegenzuwirken. Konkrete Praxisbeispiele für diese umfassende Form der Volltexterschließung in einer (Zeitung-)Portallösung findet man in größerem Umfang allerdings noch nicht im deutschsprachigen Raum.³⁹

QUO VADIS?

Bestätigt letzter Befund zwar in gewissem Sinne die eingangs erwähnte Feststellung, dass die Zeitungsdigitalisierung in Deutschland noch in den Kinderschuhen steckt, so dürfte andererseits deutlich geworden sein, dass in den letzten Jahren viel in Bewegung gekommen und weiterhin im Entstehen ist. Eine wirkliche Breite in der Digitalisierung wird zukünftig vermutlich auch nur dann erreicht werden können, wenn weiterhin lokale und regionale Initiativen einerseits und große, inhaltlich strukturierte Digitalisierungsprojekte andererseits als wichtige Standbeine der Zeitungsdigitalisierung gleichermaßen verfolgt werden, das aber mit größerer Kraftanstrengung als bisher. Die Heterogenität und unterschiedliche Verortung der bereits vorhandenen Angebote fordern gleichwohl einen guten Orientierungssinn der Nutzer im Geflecht der digitalen Angebote. Hilfreich scheint deshalb als nächster Schritt eine Einigung auf (Minimal-)Standards der Erschließung und Präsentation sowie insbesondere eine konstruktive Diskussion und Definition der langfristig anzustrebenden Ziele im Bereich der Zeitungsdigitalisierung. Dieser Prozess hat in Ansätzen schon begonnen und wird, so steht zu hoffen, auch manifeste Ergebnisse zeitigen.

¹ Sämtliche Links wurden am 1.3.2010 zuletzt aufgerufen und geprüft.

Einem guten, wenngleich bei weitem nicht vollständigen Überblick über Zeitungsdigitalisierungsprojekte weltweit bietet: icon.crl.edu/digitization.htm; auch auf delicious.com/tag/digi_zeitungen. Übergreifende Literatur zur Zeitungsdigitalisierung ist knapp bemessen. Hilfreich sind daher die Sammelbände der IFLA zu Zeitungen, z. B.: Hartmut Walravens (Hrsg.), *Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul 2006*, München 2006 (IFLA Publications, 122), oder ders. (Hrsg.), *Newspapers collection management: printed and digital challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3–5, 2007*, München 2008 (IFLA Publications 133).

² Vgl. das Portal der Neuseeländischen Nationalbibliothek: paperspast.natlib.govt.nz/cgi-bin/paperspast, sowie das Portal »Historic Australian Newspapers 1803–1954«: newspapers.nla.gov.au/ndp/del/home (beide Portale bieten weiterführende Informationen).

³ Vgl. statt aller: JISC, *British Newspapers 1620–1900, Final Report*, August 2009. www.jisc.ac.uk/media/documents/programmes/digitisation/blfinal.pdf.

⁴ digi.lib.helsinki.fi/sanomalehti/secure/main.html.

⁵ Vgl. www.kb.nl/hrd/digi/ddd/index-en.html; s. a. Neitzel, Jürgen, *Europas größtes Digitalisierungsprojekt für Zeitungen setzt auf Zeutschel Buchscanner. 400 Jahre Zeitgeschichte ins 21. Jahrhundert transferieren*, in: B.I.T. online 12 (2009), S. 429 f.

⁶ EDL Report on Digitisation in European National Libraries 2006–20012, 29th February 2008, cenl.net/docs/Report_digitisation_NLs.pdf, S. 169. Bei diesen Graphiken werden keine expliziten Angaben zur Bezugsgröße genannt; sieht man sich das zugrundegelegte statistische Material an (S. 163–167), so werden augenscheinlich nicht die digitalisierten Seiten, sondern die Zeitungsummern gezählt.

⁷ Zugrunde gelegt wurde eine ZDB-Auswertung mit Stand vom 1. März 2010.

⁸ Nach Lynda James-Gilboe, *The Challenge of Digitization: Libraries Are Finding That Newspaper Projects Are Not for the Faint of Heart*, in: *The Serials Librarian* 49 (2005) 1/2, S. 155–163.

⁹ Vgl. allgemein zu dieser Problematik: Zukunft bewahren. Eine Denkschrift der Allianz zur Erhaltung des schriftlichen Kulturguts, 2009. www.allianz-kulturgut.de/fileadmin/user_upload/Allianz_Kulturgut/dokumente/2009_Allianz-Denkschrift_nurText.pdf; weiterführend zu Zeitungen, mit Blick aber v. a. noch auf der Mikroverfilmung als erhaltender Maßnahme den Schwerpunkt im Forum Bestandserhaltung unter Papierrestaurierung/Sonderprobleme: www.forum-bestandserhaltung.de/.

¹⁰ Bei großen Digitalisierungsprojekten wurde dabei aber oftmals neu verfilmt und nicht auf »alte« Sicherungsfilme zurückgegriffen. Vgl. z. B. JISC, *British Newspapers 1620–1900, Final Report*, August 2009. www.jisc.ac.uk/media/documents/programmes/digitisation/blfinal.pdf, u. a. S. 8, 14.

¹¹ Vgl. z. B. das Projekt »Stabilisierung zerfallgeschädigter Zeitungsseiten« der Staatsbibliothek zu Berlin, zefys.staatsbibliothek-berlin.de/index.php?id=bestandserhaltung.

¹² diglit.uni-heidelberg.de/diglit/relation1609.

¹³ »Freiburger Zeitung digital«: www.ub.uni-freiburg.de/index.php?id=117; »Staufener Wochenblatt digital«: www.ub.uni-freiburg.de/index.php?id=118.

¹⁴ intelligenzblatt.unibe.ch/Default/Skins/Berna/Client.asp?Skin=Berna&enter=true&AW=1268063975511&AppName=2.

¹⁵ www.ub.uni-heidelberg.de/helios/digi/feldzeitungen.html.

¹⁶ www.bayerische-landesbibliothek-online.de/bayerische-zeitungen-1848-1850.

¹⁷ www.historisches-lexikon-bayerns.de/base/start.

¹⁸ www.compactmemory.de/ (mit ausführlicher Dokumentation zum Projekt). Vgl. z. B. aber auch: Hans Otto Horch, *Compact Memory – ein DFG-Projekt zur retrospektiven Digitalisierung jüdischer Periodika im deutschsprachigen Raum*, in: *ZfBB* 53 (2006), S. 77–80; Till Schicketanz/Kay Heiligenhaus, »Inseln im Meer des Beliebigen«. Architektur und Implementierung eines Internetportals Deutsch-jüdische Periodika, in: *Jahrbuch für Computerphilologie* 5 (2003), S. 67–96.

¹⁹ deposit.d-nb.de/online/exil/exil.htm. Hierzu auch: Robert Zepf, *Exilpresse digital. Ein gelungenes Digitalisierungsprojekt Der Deutschen Bibliothek*, in: *ZfBB* 47 (2000), S. 283–289; Renate Seib, *Exilpresse digital. Deutschsprachige Exilzeitschriften 1933–1945*, in: *Fundus* 5 (2003), S. 173–186, webdoc.sub.gwdg.de/edoc/p/fundus/5/seib.pdf.

²⁰ Dies gilt z. B. auch für das von der DFG geförderte Projekt »Retrospektive Digitalisierung wissenschaftlicher Rezensionenorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum« der UB Bielefeld (in Kooperation mit dem Georg Olms Verlag und der Akademie der Wissenschaften zu Göttingen), das zwar mehrheitlich Zeitschriften, aber auch drei Zeitungen beinhaltet. Vgl. www.ub.uni-bielefeld.de/diglib/aufklaerung/. *Exilpresse digital* umfasst, entsprechend der ZDB-Codierung, acht Zeitungen, *Compact Memory* 14 Zeitungen.

²¹ www.sachsendigital.de.

²² www.dilibri.de. Vgl. auch Barbara Koelges/Hagen Reinstein/Elmar Schackmann/Hans-Ulrich Seifert, *dilibri – ein einheitliches Di-*

gitalisierungsportal für Rheinland-Pfalz, in: B.I.T. online 3/2008, www.b-i-t-online.de/heft/2008-03/nach3.htm.

²³ www.bayerische-landesbibliothek-online.de. Vgl. auch Stephan Kellner, *Vom Landesgeschichtlichen zum kulturwissenschaftlichen Portal – Die Bayerische Landesbibliothek Online*, in: *Historisches Forum*, Bd. 10 (2007) 11, edoc.hu-berlin.de/histfor/10_11/.

²⁴ Vgl. z. B. den Schwerpunkt »Coburger Zeitungen«, www.bayerische-landesbibliothek-online.de/coburger-zeitung, ein Gemeinschaftsprojekt der Bayerischen Staatsbibliothek mit der Landesbibliothek Coburg.

²⁵ »Journals@UrMEL«, das Portal der Thüringer Universitäts- und Landesbibliothek Jena, ist primär als Zeitschriftenportal konzipiert, wenngleich auch einige retrodigitalisierte Zeitungen zu finden sind, vgl.zs.thulb.uni-jena.de/content/below/index.xml.

²⁶ zefys.staatsbibliothek-berlin.de/.

²⁷ anno.onb.ac.at/. Vgl. auch Christa Müller, *ANNO – Austrian Newspapers Online. Zielsetzungen, Methoden und Ergebnisse*, in: *Archives et Bibliothèques de Belgique – Archief en Bibliotheekwezen in België*, 78 (2007) 1–4, S. 175–183.

²⁸ Nach Richard Friebe, *Im Zeitungsarchiv: Und dann auch noch diese lästige Fraktur*, *FAZ.net*, 31.20.2009, www.faz.net/s/RubF4BEE0E0C39429A8565089709B70C44/Doc-E3504698548384CED8D45DC01728DA3BB-ATpl-Ecommon-Scontent.html.

²⁹ Ein Beispiel für eine Zeitung, die pro Jahrgang als PDF bereitgestellt wird, ist die »Buddelmeyer-Zeitung« 1849–1952 auf den Seiten der Zentral- und Landesbibliothek Berlin, opus.kobv.de/zlb/abfrage_collections.php?coll_id=392&la=de.

³⁰ dza2.tessmann.it/tessmannPortal/Portal.po?lid=de_DE.

³¹ *Papers Past*, s. Anm. 2; »British Newspapers 1800–1900«: news.papers.bl.uk/blcs/.

³² Das gilt schon für Antiquaschriften. Vgl. hierzu und zu weiteren Aspekten der Volltexterschließung: Klijn, Edwin, *The Current State-of-art in Newspaper Digitization. A Market Perspective*, in: *D-Lib Magazine*, 14 (2008) 1/2, www.dlib.org/dlib/january08/klijn/01klijn.html, sowie: Rose Holley, *How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs*, in: *D-Lib Magazine* 15 (2009) 3/4, www.dlib.org/dlib/march09/holley/03holley.html. Holley geht hier am Ende auch auf die Korrekturen der OCR durch die Nutzer bzw. die social community ein; in Australien ein Erfolgsmodell. Ob diese Form der »Erschließung 2.0« im deutschsprachigen Raum eine Zukunft hat, bleibt abzuwarten.

³³ Bei »Compact Memory« sind allerdings nur einige ausgewählte Titel im Volltext durchsuchbar; zudem wird darauf hingewiesen, dass es sich um eine Testversion handelt. Bei »Exilpresse digital« werden die Treffer im Volltext der Seite leider nicht markiert, was vermutlich den technischen Möglichkeiten der Implementationszeit geschuldet ist.

³⁴ amtspresse.staatsbibliothek-berlin.de/index.html. Mittlerweile liegt auch das »Teltower Kreisblatt« (1856–1882) digitalisiert vor.

³⁵ Vgl. hierzu mit weiterführender Beschreibung des gesamten Projekts: Albrecht Hoppe / Rudolf Stöber, *DFG-Projekt Digitalisierung der Amtspresse Preußens in der zweiten Hälfte des 19. Jahrhunderts und Erstellung eines Sachkommentars*, in: *Jahrbuch für Kommunikationsgeschichte* 8 (2006), S. 220–246, hier v. a. 231 f.

³⁶ Vgl. auch exemplarisch die Digitalisierungsstrategie der Schweizer Nationalbibliothek 2009–2015, die den Mehrwert der Volltextsuche klar hervorhebt, als PDF abrufbar auf www.nb.admin.ch/themen/02074/02076/index.html?lang=de.

³⁷ Bisherige Präsentation der »Wiener Zeitung«: [anno.onb.ac.at/cgi-content/anno?aid=wrz;temporärer_Zugang_bei_CCS\(Content_Conservation_Specialists\)_zur_Volltextsuche_in_der_»Wiener_Zeitung«:_onb-ccs.dlconsulting.com/](http://anno.onb.ac.at/cgi-content/anno?aid=wrz;temporärer_Zugang_bei_CCS(Content_Conservation_Specialists)_zur_Volltextsuche_in_der_»Wiener_Zeitung«:_onb-ccs.dlconsulting.com/).

³⁸ Diesen Schritt hat z. B. auch das JISC-Projekt in Großbritannien für die Qualitätssicherung festgelegt: JISC, *British Newspapers 1620–1900, Final Report*, August 2009. www.jisc.ac.uk/media/documents/programmes/digitisation/blfinal.pdf, u. a. S. 15; vgl. auch Klijn, *Current State-of-art* (wie Anm. 32).

³⁹ Vgl. neben den eingangs erwähnten Portalen Neuseelands und Australiens (s. Anm. 2), auch das finnische Portal »Historical Newspaper Library« (digi.lib.helsinki.fi/sanomalehti/secure/query.html?language=en) sowie »British Newspapers 1800–1900« der British Library (newspapers.bl.uk/blcs/).

DIE VERFASSERIN

Birgit Seiderer ist Mitarbeiterin der Bayerischen Staatsbibliothek im Referat Zeitschriften und Elektronische Medien / Zeitungsreferentin, Ludwigstraße 16, 80539 München, Mail: birgit.seiderer@bsb-muenchen.de