

Fünf Jahre DFG-Projekt Deutscher Reichsanzeiger – was haben wir gelernt?



Jahrgangsübersicht - Deutscher Reichsanzeiger und Preußischer Staatsanzeiger

Bitte wählen Sie zunächst einen Jahrgang aus.

1892	1897	1898	1907	1908	1909	1910	1911	1912	1913
1914	1915	1916	1917	1918	1919	1920	1921	1922	1923



Worum geht es?



- *Deutscher Reichsanzeiger und Preussischer Staatsanzeiger* (1871–1945) war die amtliche Zeitung des deutschen Reiches
- Mikroverfilmung vermutlich zwischen 1970 und 1985
- Digitalisierung einzelner Ausgaben an der Bayerischen Staatsbibliothek (MDZ)
- 2014 Digitalisierung der Mikrofilme (Dissertation Christoph Kling, Universität Mannheim)
- 2015 Provisorische Online-Präsentation der Digitalausgaben (UB Mannheim)
- 2019 Start des auf drei Jahre angelegten DFG-Projektes an der UB Mannheim

Projektziele

Das DFG-Projekt *Strukturdatenerfassung, ergänzende Digitalisierung und OCR-Erschließung des Deutschen Reichsanzeigers und Preußischen Staatsanzeigers (1871–1945)* hat folgende Ziele:

- Erfassung der Strukturdaten, dabei Erschließung bis auf Artikelebene
- Digitalisierung fehlender Ausgaben / Beilagen, aber auch von mikroverfilmten Seiten mit Qualitätsproblemen
- Verbesserte automatisierte Volltexterkennung (OCR)
- Digitalausgabe ohne Zugangs- und Nutzungsbeschränkungen

2020/21 – da war doch was?

Im März 2020 wurde der Präsenz-Studienbetrieb an der Universität Mannheim von heute auf morgen eingestellt (COVID-19-Pandemie).

Fast alle Beschäftigten der Universitätsbibliothek mussten plötzlich zu Hause bleiben.

13 Festangestellte und 40 Hilfskräfte aus der Universitätsbibliothek und dem Institut für Sport standen so dem Projekt zur Verfügung.


Aufgaben: Qualitätsprüfung 😞, Erfassung des Erscheinungsdatums 👍, Anzahl der erwarteten und tatsächlich verfilmten Beilagen 👍

Auch später durch Abstandsregeln starke Einschränkungen

Strukturdatenerfassung

Erschließung bis auf Artikelebene kostet sehr viel Zeit ohne wirklichen Mehrwert (macht niemand sonst, nicht unterstützt im Deutschen Zeitungsportal, obsolet durch Volltextsuche und zukünftige Möglichkeiten mit KI). 😞

→ Inzwischen Beschränkung auf Zeitungsausgaben und Beilagen (nach Abschluss von 25 Jahrgängen)



The screenshot shows a digital archive interface. At the top, it displays '[1930]' and 'Metadaten für 1930-02-01'. Below this is a table of contents titled 'Inhaltsverzeichnis' for 'Deutscher Reichsanzeiger und Preußischer Staatsanzeiger'. The table lists various sections with expandable/collapsible icons and page counts:

- [-] 1930 [1]
- [-] 01.02.1930 [1]
- [-] Amtliches. [1]
- Bekanntmachung, betreffend die ... [1]
- Bekanntgabe der Reichsindexziffer für die ... [1]
- Nichtamtliches. [1]
- [+] Erste Beilage [1]

To the right, a preview of a newspaper page is shown. The page is titled 'Nr. 27. Reichsbankgironkonto. Be...' and contains the following text:

Inhalt des amtlichen Teiles:
Deutsches Reich.
Bekanntmachung, betreffend die Umsatzsteuerumrechnungssätze auf Reichsmark für die Umsätze im Januar, Februar und März 1930.
Bekanntgabe der Reichsindexziffer für die Lebenshaltungskosten im Januar 1930.

Amtliches.
Deutsches Reich.
Bekanntmachung.
Die Umsatzsteuerumrechnungssätze auf Reichsmark für die Umsätze im Januar, Februar und März (ersten Kalendervierteljahr) 1930 werden auf Grund von § 8 Abs. 8 des Umsatzsteuergesetzes in der Fassung der Bekanntmachung vom 8. Mai 1926 (RGBl. I S. 218) in Verbindung mit § 45 der Durchführungsbestimmungen zum Umsatzsteuer...

Ergänzende Digitalisierung

- Umfang total unterschätzt. Ursprüngliche Planung mit Ergänzung der digitalisierten Mikrofilmen um kleinere Anzahl neuer Scans erwies sich als unbefriedigend, da Qualität zu inhomogen war.
- Zustand der Zeitungsbände sehr unterschiedlich: welliges Papier, brüchiges Papier (Säureschäden), Schimmel, sehr dicke und unhandliche Bände mit enger Bindung erschweren das Scannen
- Nicht jeder Buchscanner kommt mit den schwierigen Vorlagen klar
- Bezugsquellen für Papierbände: UB Tübingen, UB Freiburg, Bibliothek des Geheimen Staatsarchivs Preußischer Kulturbesitz Berlin, UB der LMU München, WLB Stuttgart, ZBW Hamburg 👍, aber auch zwei Paletten von einem Secondhand-Shop in Lippstadt 😞

Zeitungsbände aus dem Secondhand-Shop



Schlechter Zustand der Bände
Wenig verwertbares Material

Nach Entdeckung von 😞
Schädlingsbefall alles entsorgt

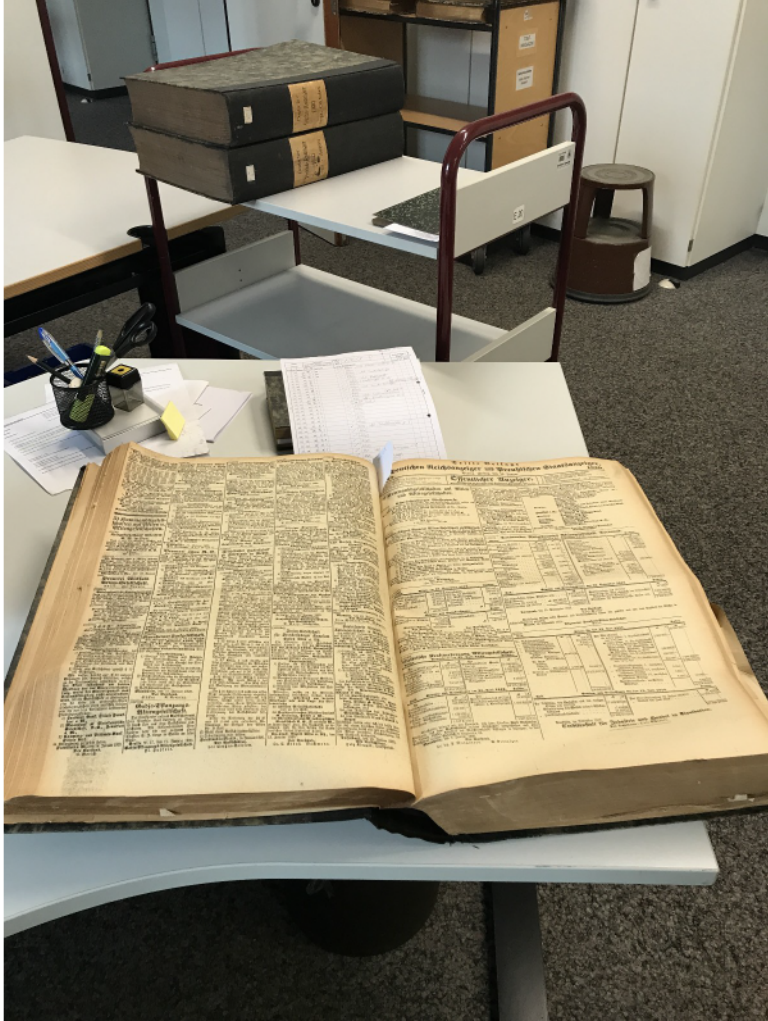
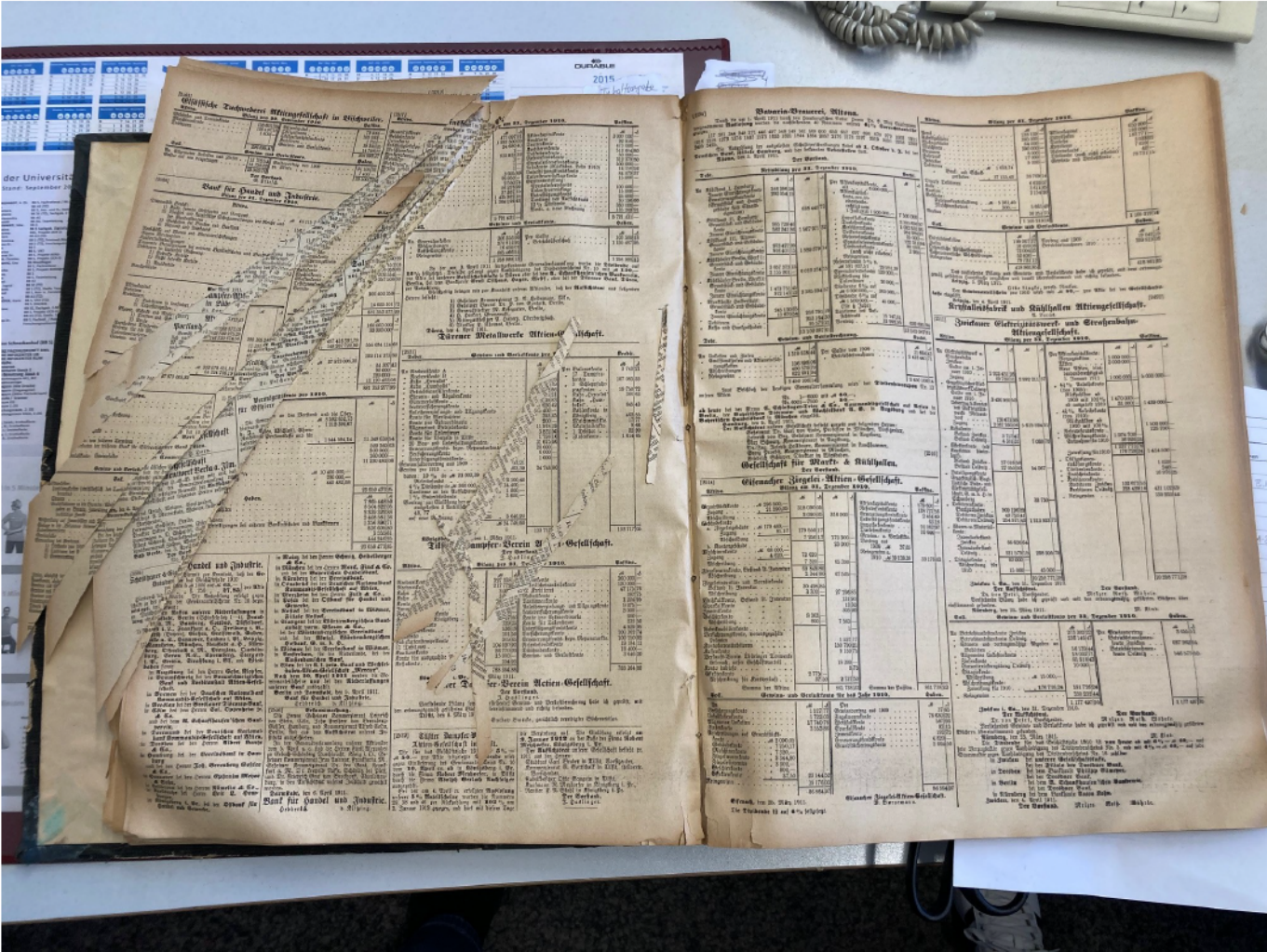


Zeitungsbände aus Berlin



Stefan Weil, Tünde Gottschling (Universitätsbibliothek Mannheim)
5. Juni 2024

Zeitungsbande - Beispiele



Stefan Weil, Tünde Gottschling (Universitätsbibliothek Mannheim)
5. Juni 2024

Scanwerkstatt der UB Mannheim

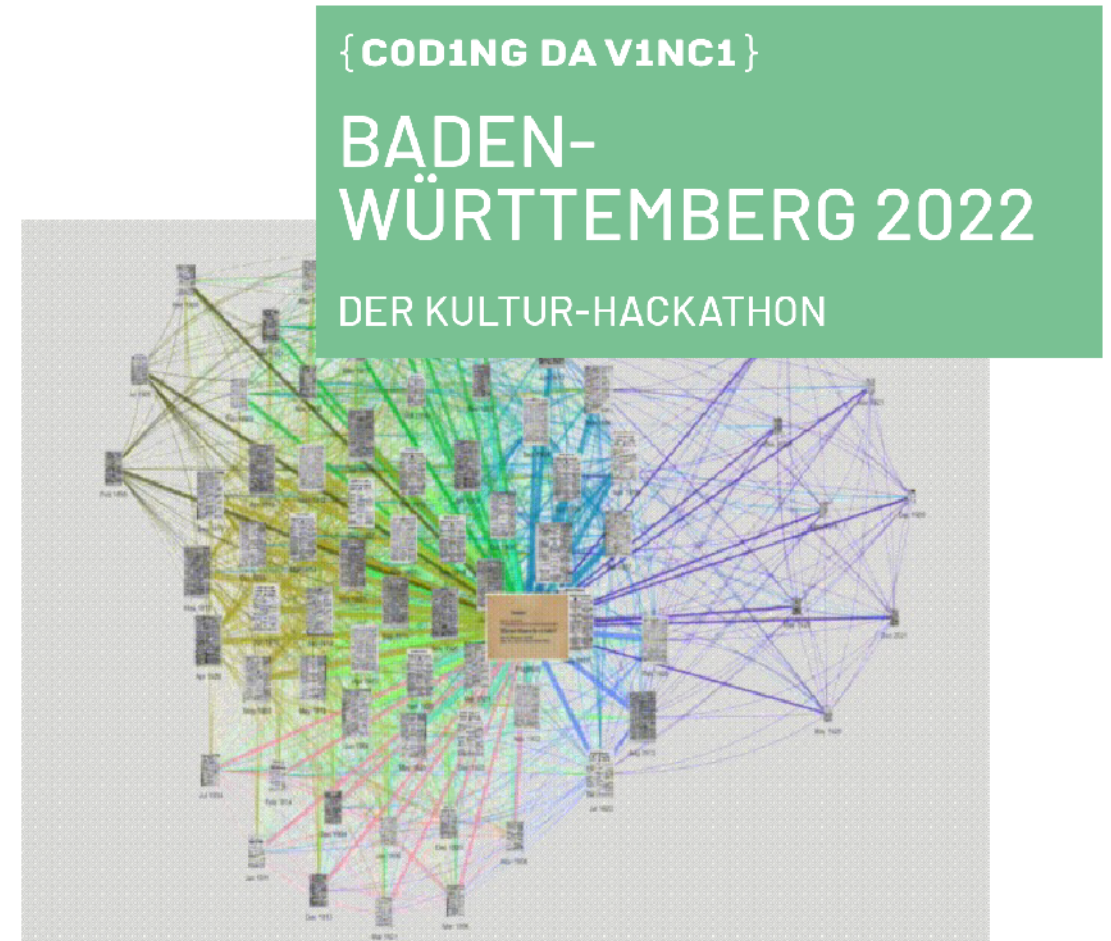


Stefan Weil, Tünde Gottschling (Universitätsbibliothek Mannheim)
5. Juni 2024

Digitalausgabe ohne Zugangs- und Nutzungsbeschränkungen

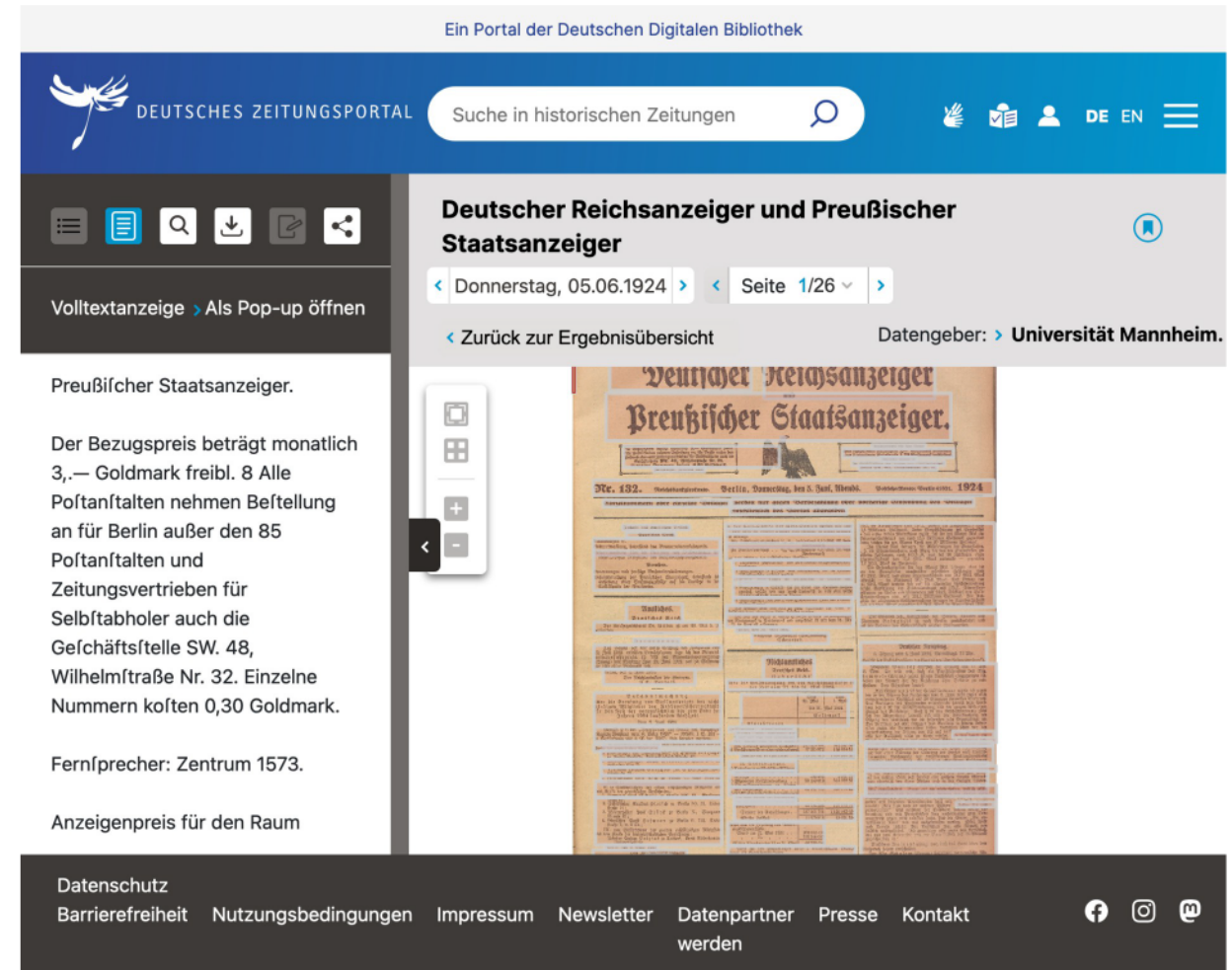
- Digitalisierte Mikrofilmausgabe war nur eingeschränkt zugänglich (“für wissenschaftliche Zwecke”)
- Uneingeschränkte Nutzung dank Vertrag mit Mikrofilmanbieter 👍
- 👍 Anwendungsbeispiel *Coding da Vinci 2022 Baden-Württemberg*: ansights – verknüpft Flug- und Extrablätter aus der Revolutionszeit 1918 bis 1920 mit Inhalten des Deutschen Reichsanzeigers

Stefan Weil, Tünde Gottschling (Universitätsbibliothek Mannheim)
5. Juni 2024



Texterkennung, Deutsches Zeitungsportal

- Automatisierte Texterkennung (OCR) mit Open-Source-Software Tesseract und den Modellen *frak2021 / german_print* aus DFG-Projekt *OCR-D Training* der UB Mannheim: bessere Ergebnisse als mit ABBYY Finereader
- Datenlieferung ans Deutsche Zeitungsportal, dabei ständige Qualitätskontrolle der Metadaten mit Schematron-Validierungen der DDB



Ein Portal der Deutschen Digitalen Bibliothek

DEUTSCHES ZEITUNGSPORTAL Suche in historischen Zeitungen

DE EN

Deutscher Reichsanzeiger und Preußischer Staatsanzeiger

Donnerstag, 05.06.1924 Seite 1/26

Zurück zur Ergebnisübersicht Datengeber: Universität Mannheim.

Preußischer Staatsanzeiger.

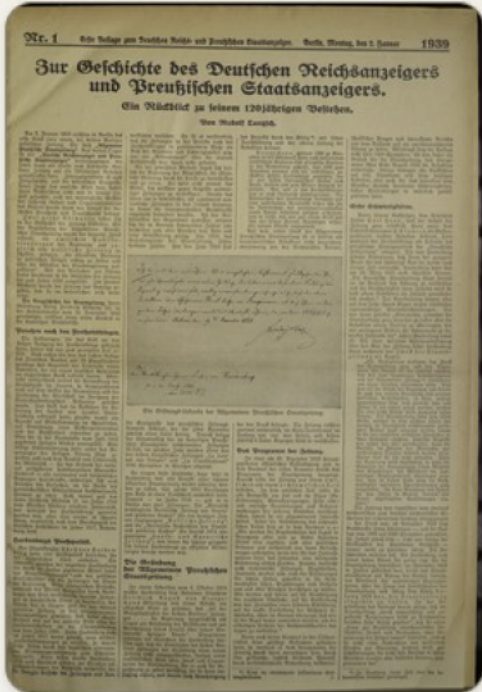
Der Bezugspreis beträgt monatlich 3,— Goldmark freibl. 8 Alle Pofanftalten nehmen Befellung an für Berlin außer den 85 Pofanftalten und Zeitungsvertrieben für Selbstabholer auch die Gefchäftstfelle SW. 48, WilhelmftraÙe Nr. 32. Einzelne Nummern kosten 0,30 Goldmark.

Fernfprecher: Zentrum 1573.

Anzeigenpreis für den Raum

Datenschutz
Barrierefreiheit Nutzungsbedingungen Impressum Newsletter Datenpartner Presse Kontakt werden

Claude.ai – LLMs als Zukunft der Texterkennung?



Der Text ist ein Rückblick auf die Geschichte des Deutschen Reichsanzeigers und des Preussischen Staatsanzeigers anlässlich ihres 120-jährigen Bestehens im Jahr 1939. Er beschreibt die Gründung dieser offiziellen Veröffentlichungsorgane, ihre Entwicklung und Bedeutung für die Verbreitung staatlicher Informationen und Gesetze im Deutschen Reich und in Preußen. Außerdem geht es um die Herausforderungen bei der Produktion und Auslieferung der Zeitungen in früheren Zeiten. Der Autor betont die Wichtigkeit einer zuverlässigen staatlichen Veröffentlichung für die Rechtssicherheit und Transparenz. Der Text vermittelt einen historischen Überblick über diese zentralen Publikationen im Kontext der damaligen staatlichen Strukturen.


S Bitte fasse den Text zusammen.

Abschluss der Arbeiten

- Das Projekt wird Mitte des Jahres 2024 abgeschlossen.
- Die meisten Ausgaben des 20. Jahrhunderts sind dann bearbeitet.
- Die Arbeiten gehen danach noch weiter, bis auch die fehlenden Jahrgänge des 19. Jahrhunderts in der neuen Präsentation online sind.

Verfügbare Ausgaben

19. Jahrhundert

1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811
1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823
1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835
1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847
1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859
1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	
1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883
1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895
1896	1897	1898	1899								

Ausblick

- 1819–1871 erschien der Vorgänger des Deutschen Reichsanzeigers, die Allgemeine Preußische Staats-Zeitung (mehrfach umbenannt, zuletzt Königlich Preussischer Staats-Anzeiger)
- Provisorische Digitalausgabe für diesen Zeitraum ist online (digitalisierte Mikrofilme, bisher nur zwei Jahrgänge mit besseren Scans auf Basis von Papierausgaben)

→ Folgeprojekt notwendig



Herzlichen Dank allen Bibliotheken, die uns ihre Bestände für die Digitalisierung zur Verfügung gestellt haben!

Stefan Weil

stefan.weil@uni-mannheim.de

Tünde Gottschling

tuende.gottschling@uni-mannheim.de

Links

- Provisorische Digitalausgabe
→ <https://digi.bib.uni-mannheim.de/periodika/reichsanzeiger/>
- Neue Digitalausgabe (in Arbeit)
→ <https://digi.bib.uni-mannheim.de/periodika/zeitungen/>
- Provisorische Volltextsuche
→ <https://digi.bib.uni-mannheim.de/fess/>
- Deutscher Reichsanzeiger und Preußischer Staatsanzeiger im Deutschen Zeitungsportal
→ <https://www.deutsche-digitale-bibliothek.de/newspaper/2851723-4>
- Deutscher Reichsanzeiger bei Wikipedia
→ https://de.wikipedia.org/wiki/Deutscher_Reichsanzeiger
- Ground truth for German newspaper "Deutscher Reichsanzeiger und Preußischer Staatsanzeiger"
→ <https://github.com/UB-Mannheim/reichsanzeiger-gt>
- NER/NEL corpus for the German historical newspaper "Deutscher Reichsanzeiger [...]"
→ <https://github.com/UB-Mannheim/reichsanzeiger-nlp>

Links (Fortsetzung)

- Coding da Vinci – Projekt ansights
→ <https://codingdavinci.de/de/projekte/ansights>
- Tesseract OCR und Modell *frak2021* für historische Drucke und Zeitungen
→ <https://github.com/tesseract-ocr/tesseract/>
→ <https://zenodo.org/doi/10.5281/zenodo.10125245>
- Kraken OCR und Modell *german_print* für historische Drucke und Zeitungen
→ <https://github.com/mittagessen/kraken/>
→ <https://zenodo.org/doi/10.5281/zenodo.10519595>
→ <https://ub-backup.bib.uni-mannheim.de/~stweil/tesstrain/>
- Software and data related to "Deutscher Reichsanzeiger und Preußischer Staatsanzeiger"
→ <https://github.com/UB-Mannheim/Reichsanzeiger>
- Schematron-Validierungen der Fachstelle Bibliothek der Deutschen Digitalen Bibliothek
→ <https://github.com/Deutsche-Digitale-Bibliothek/ddb-metadata-schematron-validation>

Abbildungen

- Ctenolepisma longicaudata („Papierfischchen“)
→ https://commons.wikimedia.org/wiki/File:Ctenolepisma_longicaudata_20080805_403.jpg
- Logo für Coding da Vinci Baden-Württemberg 2022
→ <https://codingdavinci.de/presse>
- Animierter Graph zu CdV-Projekt ansights
→ <https://github.com/joelosw/ansights>

- Lantzsch, R.: *Zur Geschichte des Deutschen Reichsanzeigers und Preußischen Staatsanzeigers. Ein Rückblick zu seinem 120jährigen Bestehen.* Deutscher Reichsanzeiger und Preußischer Staatsanzeiger, 1939, Nr. 1, 1. Beilage.
- Kling, C. (2016). *Deutscher Reichsanzeiger und Preußischer Staatsanzeiger : Einleitung zur Veröffentlichung der Digitalausgabe.* <https://madoc.bib.uni-mannheim.de/41378/>
- Weil, S. (2018). *126 Jahre Zeitung online - Fundgrube für historisch Interessierte und Motor für die Bibliotheks-IT : 126 years of the newspaper online.* 107. Deutscher Bibliothekartag, Berlin, Germany. <https://madoc.bib.uni-mannheim.de/46507/>
- Schmidt, T., Kamlah, J., & Weil, S. (2024). *Reichsanzeiger-GT : an OCR ground truth dataset based on the historical newspaper "Deutscher Reichsanzeiger und Preußischer Staatsanzeiger" (German Imperial Gazette and Prussian Official Gazette) (1819–1945).* Data in Brief, 54(Article 110274), 1–7. <https://doi.org/10.1016/j.dib.2024.110274>

Literatur (Fortsetzung)

- Weil, S., Kamlah, J., Schmidt, T. (2024). *Abschlussbericht zu DFG-Projekt "Workflow für werkspezifisches Training auf Basis generischer Modelle mit OCR-D sowie Ground-Truth-Aufwertung"*. <https://madoc.bib.uni-mannheim.de/67174/>
- Kiessling, B. (2022). *The Kraken OCR system (Version 4.1.2) [Computer software]*. <https://kraken.re>