

b

Staats- und
Universitätsbibliothek
Bremen

Themenwoche *Digital Humanities*



Word Cloud, Quelle: CC0,
Creator: CRDA,
Owner: NFDI4Culture

Themenwoche *Digital Humanities* – 06.02. bis 10.02.2023 – Dr. Manfred Nölte

Themenwoche Digital Humanities

- Montag
 - Überblick und Einstieg in DH-Datenquellen und DH-Methoden
- **Dienstag**
 - **DH-Datenquellen und die ersten Schritte der Verarbeitung**
- Mittwoch
 - Einstieg in DH-Methoden, Agnes Brauer, UB Frankfurt
- Donnerstag
 - Fortgeschrittene DH-Methoden
- Freitag
 - Ada Lovelace Center for Digital Humanities, Dr. Dennis Mischke

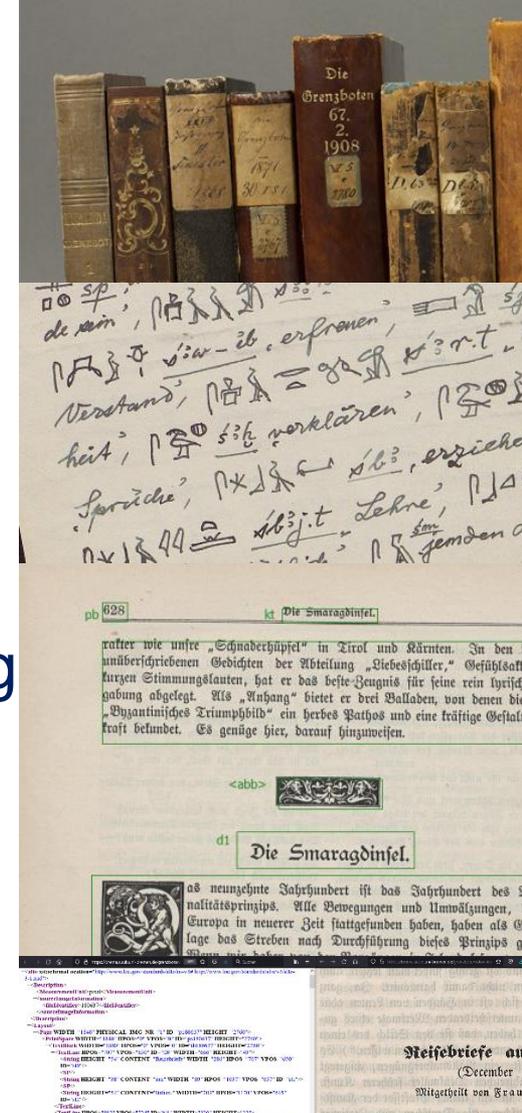


Themenwoche *Digital Humanities* – DH-Datenquellen und die ersten Schritte der Verarbeitung



Gliederung

- Datenquellen – Vielfalt, Umfang, Qualität
- Datenquellen – Zugang, Formate, Struktur
- Einstieg in automatisierte Datenprozessierung
- Beispiel – Shakespeare Korpus
- Ausblick, Diskussion



Datenquellen – Vielfalt, Umfang, Qualität

- Welche Quellen von digitalen Daten gibt es?

- Digitale Editionen – in guter Qualität
- Digitalisate – Digitale Sammlungen von digitalisierenden Bibliotheken -- in großen Mengen
- Digitale Bilddatenbanken (D. Diatheken)
- Digitale Audio- und Video-Archive
- Wikimedia-Projekte: Wikisource, Wikidata, Wikipedia, etc.
- Daten gibt es überall – ...

- Was erhalte ich?

- Text ← zumeist große digitale Textkorpora
- Bild
- Audio
- Animation, Video
- Neue Medien: 3D Medien, 360° Medien, Software, etc.

- In welcher Qualität und Form erhalte ich Daten?

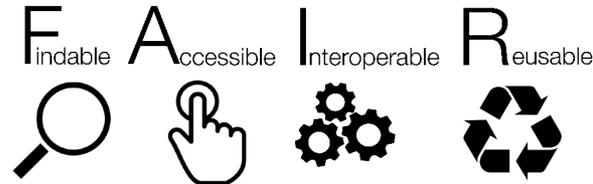
- → Metadaten, Datenstruktur, Dateiformate, SW-Tools

- Twitter
- Blogs
- Kochrezepte
- Zeitungsportale
- Open GLAM
- Citizen Science Projekte
- Aktuell an der Uni Bremen:
Memes (Bildwitze)
- Likes / Dislikes

Datenquellen – Daten erstellen und aufbereiten

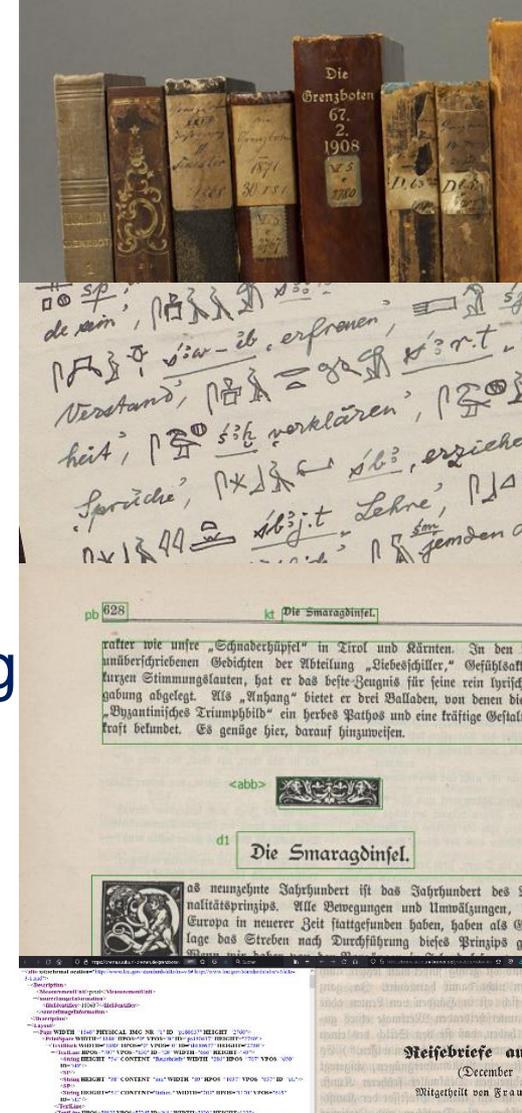
- OCR – Optical Character Recognition
- Transkription – aktuell: Transcriptions
- Annotation (generell: Metadaten ergänzen)
- Daten-Vorverarbeitung, Data Refinement
- „Laborbuch“ – Protokoll der durchgeführten Schritte
 - Sogar dafür gibt es Tools
- Auch hier! : FORMATE, Dokumentation, FAIR und FDM (Forschungsdatenmanagement)

*Aktuell, nächste Woche –
13.02. bis 17.02.2023
Love Data Week*



Gliederung

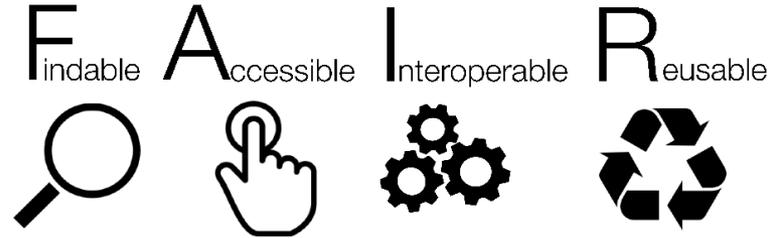
- Datenquellen – Vielfalt, Umfang, Qualität
- Datenquellen – Zugang, Formate, Struktur
- Einstieg in automatisierte Datenprozessierung
- Beispiel – Shakespeare Korpus
- Ausblick, Diskussion



Datenquellen – Zugang, Formate, Struktur

- Guter Zugang zu Daten

- FAIR
- „FAIR Data“ →
Maschine-Maschine-Kommunikation → Effizienz
- DraCor ("drama corpora", <https://dracor.org/shake>) → mehr im nächsten Abschnitt
- Deutsches Textarchiv (DTA), Wikipedia vollständiger Download, CLARIN (Forschungsinfrastruktur, <https://vlo.clarin.eu>)

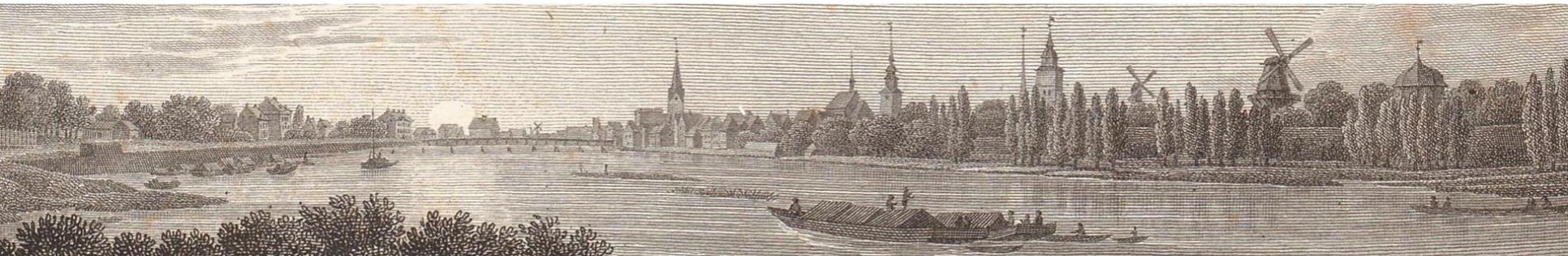


- Mäßiger Zugang zu Daten

- Einzelseiten Download
- Kein Twitter-Download → Stream mitschneiden
- Data Scraping
- Email an eine Kontaktperson
- Google Books !?

Datenquellen für die Digital Humanities

- Vielfältige Präsentation von digitalen Datenquellen
 - Digitalisierende Bibliotheken mit Volltextzugriff
 - Library of Congress (LOC), <https://www.loc.gov/preservation/digital/>
<https://chroniclingamerica.loc.gov/lccn/sn83030311/>
 - TROVE Australian Newspapers
<https://trove.nla.gov.au/newspaper/>
<https://trove.nla.gov.au/newspaper/article/229435103?browse=ndp%3Abrowse%2Fdate%2F1890%2F01%2F01%2Ftitle%2F1150%2Fpage%2F23330986%2Farticle%2F229435103#>
 - Digitale Sammlungen der Universitätsbibliothek der LMU München
<https://digi.ub.uni-muenchen.de/start/> → 1904 → Wüste | Protestantismus
 - DFG Viewer – Volltext-Downloads (Per Email angefragt)



The herald. [volume] (New York X +

https://chroniclingamerica.loc.gov/lccn/sn83030311/1836-01-01/ed-1/seq-1/

LIBRARY LIBRARY OF CONGRESS

ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS Search Search Loc.gov GO

The Library of Congress > Chronicling America > The herald. > January 01, 1836 > Image 1

Humanities NATIONAL ENDOWMENT FOR THE

CHRONICLING AMERICA Historic American Newspapers Search America's historic newspaper pages from 1777-1963 or use the U.S. Newspaper Directory to find information about American newspapers published between 1690-present. Chronicling America is sponsored jointly by the National Endowment for the Humanities and the Library of Congress. [Learn more](#)

Search Pages Advanced Search All Digitized Newspapers 1777-1963 US Newspaper Directory, 1690-Present

All states + from 1777 to 1963 + enter one or more search words GO

Print Subscribe Share/Save Give Feedback

The herald. [volume], January 01, 1836, Image 1
About The herald. [volume] (New York [N.Y.]) 1835-1837

Image provided by: Library of Congress, Washington, DC

Image: 1 of 4. Page All Pages Issues All Issues Text PDF JP2 (3.0 MB)

THE HERALD

VOLUME I. NEW-YORK, FRIDAY, JANUARY 1, 1836.

PUBLISHED DAILY BY **JAMES GORDON BENNETT.**
Office, No. 141 Nassau street, between Beekman and Spruce streets, opposite Dr. Spring's Church, basement story.

TERMS FOR ADVERTISING.

1 square a year,	\$20.00	1 square 2 weeks,	\$2.25
" 3 months,	6.00	" 1 week,	1.75
" 2 months,	4.00	" 3 days,	1.00
" 1 month,	2.00	" 1 insertion,	0.50

Station lines make a square.

TO THE PUBLIC—ENLARGEMENT OF THE HERALD.

Having determined as soon as arrangements can be made, to enlarge the HERALD to at least one third over its present size, we shall, with the permission of our kind and numerous patrons, say a few words on this subject, more interesting to us than to them, and then start afresh on our delightful and pleasant course.

It is yet hardly four months since the Herald rose from the Ann street conflagration, with impaired means, but undying energy of mind and spirit. During this brief period, we have struggled with competition as numerous and powerful as the elements themselves. Yet the independent, honorable, and liberal plan on which the Herald has been conducted, has attained a complete victory over interested opponents and private foes. The generous public has nobly sustained us.

In the short period of three months and a few days, the

NEW-YORK LIFE INSURANCE & TRUST COMPANY.—Persons may effect insurances with this Company on their own lives, or the lives of others, and either for the whole duration of life, or for a limited period. The payments of premium may be either made annually, or in a gross sum:

Age.	1 year.	1 year—Age.	1 year.	1 year.	1 year.
14	0 72	26	1 07	32	1 18
15	0 77	27	1 12	38	1 57
16	0 84	28	1 20	40	1 69
17	0 96	29	1 28	41	1 73
18	0 98	30	1 31	42	1 83
19	0 90	31	1 32	43	1 89
20	0 91	32	1 33	44	1 90
21	0 92	33	1 34	45	1 91
22	0 94	34	1 35	46	1 92
23	0 97	35	1 36	47	1 93
24	0 99	36	1 39	48	1 94
25	1 00	37	1 43	49	1 95

Money will be received in deposit by the Company, and held in trust, upon which interest will be allowed as follows:

Upon any sum over \$100, irrevocable for 1 year,	4 1/2 per cent.
" " " 100, " for 5 months,	4 per cent.
" " " 100, " for 2 months,	3 per cent.

TRUSTEES.

Wm. Bard,	Saml. Thompson,	H. C. De Rham,
Thomas W. Ludlow,	Isaac Bronson,	Jonathan Goodhue,
Wm. B. Lawrence,	Peter Remond,	James McBride,
Isaac Lorillard,	Stephen Warren,	John Rathbone, Jr.,
John Doer,	James Kent,	P. G. Staybrook,
Peter Harmany,	Nathaniel Prime,	Thomas J. Oakley,
S. Van Rensselaer,	N. Devereux,	Steph'n Whitney,
John G. Costar,	Benj. Knicker,	John J. Astor,
Thos. Saffers,	Guilias C. Verplanck,	Benj. L. Swan,
Jos. Mason,	Cornelius W. Lawrence,	

R. A. NICOLL, Secretary. **WM BARD,** President.
F. U. JOHNSTON, Physician to the Company

TO THE TRADE.—CHEMICALS.—The undersigned takes pleasure in offering a very extensive assortment of the

NEW ESTABLISHMENT FOR THE SALE OF CABINET FURNITURE.—H. BAUNSWICK returns thanks for the liberal patronage he has received since the opening of his establishment for the sale of Cabinet Furniture, both new and second hand, and has the pleasure to inform his customers and others that he has removed to the spacious building, corner of Bowery and Division streets, (up stairs, entrance 73 Bowery,) where he keeps a general assortment of new and good articles of second hand furniture, such as Bureaus, Sides, Secretaries, Book Cases, Sideboards, Tables, Looking Glasses, Brackets, Piers, Centre and Work Tables, Marble Slabs, Bedsteads, Trunks, Trunk Cases, Washstands, Carpets, Rugs, Feather Beds, Mattresses, &c.

He will buy or exchange household furniture or any article in his line of business.

WANTED IMMEDIATELY TWO FIRST RATE HANDS; to such the highest wages will be given and constant employment. Apply at the manufactory of **S. W. BRYHAM,** 298 Bowery, between Spring and Prince streets.

TO JOURNEYMEN CONFECTIONERS.—Wanted immediately two first rate hands; to such the highest wages will be given and constant employment. Apply at the manufactory of **S. W. BRYHAM,** 298 Bowery, between Spring and Prince streets.

EVERING SCHOOL FOR YOUNG GEN- TLEMEN.—The subscriber has removed to the elegant and spacious rooms of the Columbia School, 109 Canal street, entrance on the corner of Varick street.

A select class of young gentlemen will be instructed in the usual branches of education, also in the Classics and Modern Languages. The French department is under the immediate direction of A. Fournier, Professor of the French Language.

N. B. Particular attention will be devoted to Book-keeping and Penmanship.

Terms moderate. Apply as above. **C. C. JENNINGS,** dec 24

COUGH DROPS.—A superior article for colds, coughs and all affections of the respiratory organs. Preps. and sold only at 145 Broadway.

BOOK ESTABLISHMENT FOR SALES.—That excellent stand, Clinton Hall, corner of Nassau and Bowery, has been removed to the new building, corner of Nassau and Bowery, where he keeps a general assortment of new and good articles of second hand furniture, such as Bureaus, Sides, Secretaries, Book Cases, Sideboards, Tables, Looking Glasses, Brackets, Piers, Centre and Work Tables, Marble Slabs, Bedsteads, Trunks, Trunk Cases, Washstands, Carpets, Rugs, Feather Beds, Mattresses, &c.

He will buy or exchange household furniture or any article in his line of business.

I'm happy, my friends, that you're Once again on the affairs of the Before you, I'll now faithfully set ("A duty annex'd to my Station An account of all that's occur'd Since from here you got under Of most, you've undoubtedly heat For none of you, surely, do fail, To read my Newspaper, "the Globe (The only one fit for your eyes, For it requires the patience of Job To read others—so d—d full For I've always made it a rule To have Blair at my House ev' (And Mrs. B. too—who's no fool, To be sure that his "Squibs" v Some have the great boldness to I do wrong to edit the Globe, And, receiving friend Blair ev'ry Is soling my "President's robe But the Devil a bit do I care, And what am I President for, If I can't consult with friend Blair Against it I'm sure there's no lo

Titel: Der Niedergang des Protestantismus**Autor:** Schmidt, Ferdinand Jakob**Einrichtung:** Universitätsbibliothek der Ludwig-Maximilians-Universität München (DE-19)**Signatur:** 0001/4 H.lit. 3035e(1904)**Erscheinungsort:** Berlin**Erscheinungsjahr:** 1904**Titel:** Textanfang

Einzelseite herunterladen

AD BUBL.
UNIVERS.
MONAC.

Motto: ὁ θεὸς κέρως τὸ πνεῦμα ἔσται. 2. Cor. 3, 17.



Mit Staunen gewahren wir heut, dafs nach der ruhmreichen Erhebung der Reformation und nach den Tagen der klassischen deutschen Literatur und Philosophie seit geraumer Zeit wieder der Katholizismus die bedeutendste Macht im öffentlichen und geistigen Leben geworden ist. Mag der Protestant auch noch so bekümmert darüber sein, so kann doch ein ängstliches Vertuschen dieses Tatbestandes um keinen Deut weiter helfen, und was not tut, ist vor allen Dingen die Einsicht und der energische Wille, dafs wir der Wahrheit erst einmal mutig ins Angesicht schauen müssen. Es ist ja unmöglich, dafs die universelle Idee des Protestantismus wieder an sich von dem Katholizismus überholt zu werden vermag, denn das wäre geradeso, als ob die Planeten aus ihren Bahnen geraten sollten; wohl aber ist es möglich, dafs bei einem einzelnen Volk, wie etwa dem unsrigen, die Bewegung einer Kulturidee derartig in Stagnation gerät, dafs jenes darüber die geistige Führung unter den vorwärts strebenden Nationen wieder verliert. Noch vor einem halben Jahrhundert konnten hervorragende englische und französische

Der Niedergang des Protestantismus [1]

Titelseite [1]





01 Jan 1890 - Zum Neujahrs Mo...

https://trove.nla.gov.au/newspaper/article/229435103?browse=ndp%3Abrowse%2Fdate%2F1890%2F01%2F01%2FTitle%2F1150%2Fpage%2F23330986%2Fartide%2F229435103#

Search Newspapers and Gazettes Filter Advanced Search



ABOUT HELP NEWS PARTNERS SIGN UP LOGIN

Home / Newspapers & Gazettes / Browse / Australische Zeitung (Adelaide, SA: 1875 - 1916) / Mi 1 Jan 1890 / Page 1 / Zum Neujahrs Morgen 1890.

Match text

, Zum Neujahrs Morgen 1890.
Hoffnungsvoller hat wohl kaum ein Jahr begonnen,
MS das neue, das Du fröhlich heut' begrüßt,
Wo, nachdem der früheren Jahre Schmerz zerronnen,
Nun so manche DankeSträh'n dem Angl entfließt.
Deine Felder stehen reich im Erntesegen, 7
Alle Deme GorrätSräume find gefüllt,
^ Und allüberall auf Demen Lebenswegen
- Siehst Du aufgestellt ein schönes Hoffaugsbild. .

Australische Zeitung

Diese Zeitung erscheint jeden Mittwoch in Adelaide und wird an bestimmten Tagen auch allen Orten Australiens wie der übrigen Erde telegraphisch versandt.

This Journal is published every Wednesday in Adelaide, and having a very extensive circulation among the German inhabitants of those colonies, offers a most eligible medium for advertising. All letters to be addressed to Messrs. Woodrow, Elmer, & Co.

Abonnement: 6 Schill. per Quartal per Jahr. - Bei Vorauszahlung Quartal über 30 Schill. vor

Redacteurs: G. Woodrow und Dr. C. Müller.

N^o. 1.

Adelaide, Mittwoch, den 1. Januar 1890.

XII

Zum Neujahrs Morgen 1890.

Hoffnungsvoller hat wohl kaum ein Jahr begonnen,
Wo das neue, das Du fröhlich heut' begrüßt,
Wo, nachdem der früh'n Jahre Schmerz zerronnen,
Nun so manche Dankesträh'n dem Angl' entfließt.
Deine Felder stehen reich im Erntesegen,
Alle Deme GorrätSräume find gefüllt,
Und allüberall auf Demen Lebenswegen
- Siehst Du aufgestellt ein schönes Hoffaugsbild. .

die sie gelüschet und sich anfänglich als Mittelst...
Einem zweiten großen Uebelstand in der Kolonie, welcher die...
Ein ungerechtes System ist ein Hinderniß für das...
Es ist dieser Uebelstand bereits sehr lebhaft erkannt...

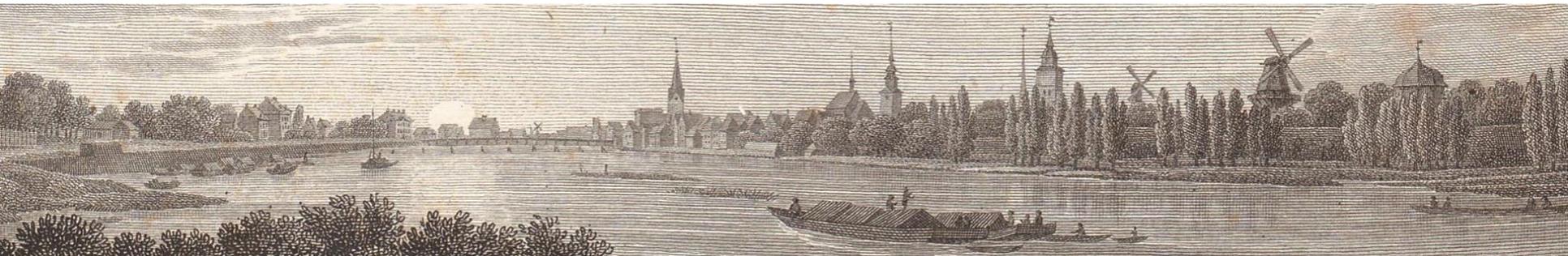
worden. Purple Straw und vor allem Steinwedel-Weizen...
Im Hundert Weizen sind circa 12 000 Aker gerechnet...
Der bösliche Wind in den letzten Monaten hat häufig...
Purple Straw hat sich als für diese Distrikte am geeignetsten bewährt.

Botanische wie auch der Zoologische Garten...
Little Lord Kaurteroy" eine überaus warme...
Ausgang aus Feed. Dring's...
Sowohl die Ernterträge als...
Willing Company hat in...
Purple Straw hat sich als für diese Distrikte am geeignetsten bewährt.



(Daten)Quellen – Zugang, Formate, Struktur

- **Beispiel: Eine Quelle – Drei Plattformen**
 - Wikisource, SuUB Bremen, DTA (BBAW Berlin)
 - Transkription – zeilengetreu?
- **Wikisource, <https://de.wikisource.org/>**
 - „Die Grenzboten“: https://de.wikisource.org/wiki/Die_Grenzboten
https://de.wikisource.org/wiki/Deutsche_Eisenbahnen_und_belgische
 - **Zum Vergleich:** <https://brema.suub.uni-bremen.de/grenzboten/periodical/pageview/179460>
DTA: https://www.deutschestextarchiv.de/dtaq/book/download_html/grenzboten_179382_282158





Deutsche Eisenbahnen und belgische

https://de.wikisource.org/wiki/Deutsche_Eisenbahnen_und_belgische

150%

Leesezeichen importier... Mensa Timeos Home | Staats- und U... Deutsche Eisenbahnen...



WIKISOURCE

- Hauptseite
- Systematischer Einstieg
- Themenübersicht
- Autorenindex
- Zufällige Seite
- Mitmachen
- Letzte Änderungen
- Neuer Artikel
- Korrekturen des Monats
- Gemeinschaftsportal
- Skriptorium
- Auskunft
- Hilfe
- Spenden

Nicht angemeldet [Diskussionsseite](#) [Beiträge](#) [Benutzerkonto erstellen](#) [Anmelden](#)

Quellentext [Index](#) [Diskussion](#) [Lesen](#) [Bearbeiten](#) [Versionsgeschichte](#)

Deutsche Eisenbahnen und belgische

[Herunterladen](#)

[69]

Deutsche Eisenbahnen und belgische.

Der Sommer zeigte sich günstig für die Eisenbahnen auf dem Continente. Da haben wir schon wieder drei, die Schlag auf Schlag dem Verkehr Preis gegeben wurden: eine von Düsseldorf nach Elberfeld, eine andere von Straßburg nach Basel, und jetzt die zwischen Cöln und Aachen. Ich kenne die erstere nicht, habe aber die große, eiserne Straße, welche das Elsaß der Ausdauer eines einfachen Bürgers (des Herrn Köchlin) verdankt, voriges Jahr gesehen, als sie noch im Baue begriffen war. Es ist ein schönes Kunstwerk, das von den Sachverständigen als eines der bestangelegten in Europa betrachtet wird. Wie in Belgien, kam hier das natürliche **Gleichmaß** des Bodens, in der ganzen ungeheuern Fläche, die sich von der Veresen nach dem Rheine zu erstreckt, den Arbeiten der

Textdaten	
<<<	>>>
Autor:	E. R.
Titel:	Deutsche Eisenbahnen und belgische
aus:	Die Grenzboten (1841/1842), 1. Jg., Band 1, S. 69-77
Erscheinungsdatum:	1841
Verlag:	Herbig
Erscheinungsort:	Leipzig

Sammlung Grenzboten

https://breme.suu.uni-bremen.de/grenzboten

Staats- und
Universitätsbibliothek
Bremen

Über uns Services & Beratung Kataloge & Sammlungen Fachinformationen Standorte Kontakt Literaturverwaltung & publizieren

Sie sind hier: Home — Kataloge & Sammlungen — Digitale Sammlungen — Sammlung Grenzboten — Home

aaa

Suche: z.B. Titel, Thema, Person ...

Digitale Sammlungen

Projektinformationen

Neuzugänge

Projekte

Sammlung Grenzboten

Sammlung Zeitschriften des 17. Jahrhunderts

Sammlung Breitsche Zeitsungen

Einiger Theaterzettel

Die Bremer Papyri

Mittelalterliche Handschriften

Digitale Sammlung Deutscher Kolonialismus

Rachlässe und Autographen

Historische Karten

Die Grenzboten

Eine digitalisierte Zeitschrift

Durchgeführt mit Mitteln der Deutschen Forschungsgemeinschaft (DFG)

Inhaltsverzeichnis

In der Digitalen Sammlung "Die Grenzboten" bieten wir Ihnen den kostenfreien Zugang zur digitalisierten Zeitschrift "Die Grenzboten" (1841-1922) aus dem Bestand der Staats- und Universitätsbibliothek Bremen. Alle Titel- und Strukturdaten sowie der Volltext sind durchsuchbar. Die einzelnen Artikel stellen sich als PDF-Dokumente kostenlos zur Verfügung.

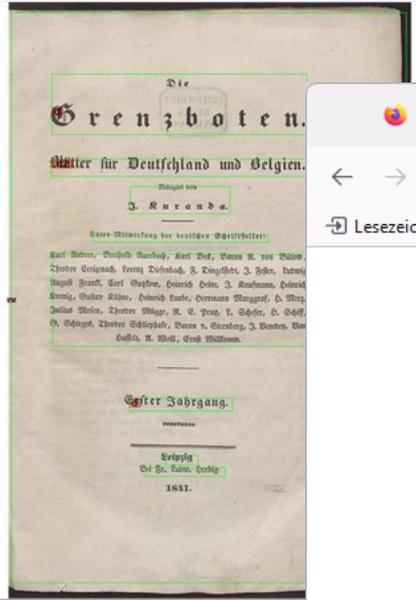
Die zwischen 1841 und 1922 wöchentlich, teilweise zweisprachlich erscheinende national liberale Zeitschrift "Die Grenzboten" hatte das Ziel, die gesamte bürgerliche Lebenswelt abzubilden. Ab 1871 gibt der Leitartikel über die Inhalte zutreffend Auskunft: "Zeitschrift für Politik, Literatur und Kunst". Beiträge aus Politik und Geschichte über Weltanschauung bis hin zu schillerndsten Themen aus Kunst, Musik und Literatur zeigen die immense Themenvielfalt der Zeitschrift, die das bürgerliche

Deutsche Eisenbahnen und belgische.

Sommer zeigte sich günstig für die Eisenbahnen auf dem Continente. wir schon wieder drei, die Schlag auf Schlag dem Verkehr ben wurden: eine von Düsseldorf nach Elberfeld, eine andere von nach Basel, und jetzt die zwischen Cöln und Aachen. Ich kenne nicht, habe aber die große, eiserne Straße, welche das Elsaß der eines einfachen Bürgers (des Herrn Köchlin) verdankt, voriges Jahr gesehen, als sie noch im Baue begriffen war. Es ist ein schönes Kunstwerk, das von den Sachverständigen als eines der bestangelegten in Europa betrachtet wird. Wie in Belgien, kam hier das natürliche Gleichmaß des Bodens, in der ganzen ungeheuern Fläche, die sich von den Bergen nach dem Rheine zu ausdehnt, den Arbeiten der Ingenieure zu Hülfe, und gestattete ihnen, auf die Vollendung der Details mehr Fleiß zu verwenden. Es ist in der That eine con amore gebaute Eisenbahn, mit Schienen, so stark und fest, um unsere spätesten Enkel in Erstaunen zu setzen; mit prächtigen Stationen, in rothem Sandstein, die dereinst in den Augen der Nachwelt die Stelle der römischen Ruinen einnehmen können, wenn es mit unseren Eisenbahnen gehen wird, wie jetzt mit den, von dem erhabendsten der Völker, für die Ewigkeit erbauten Brücken, Straßen und Wasserleitungen, und wenn die Forscher eines andern Zeitalters im verwachsenen

Die Grenzboten. Erster Jahrgang. Leipzig, 1841.

Informationen **Ansichten**
Quelle: Die Grenzboten
• Korrekturumgebung



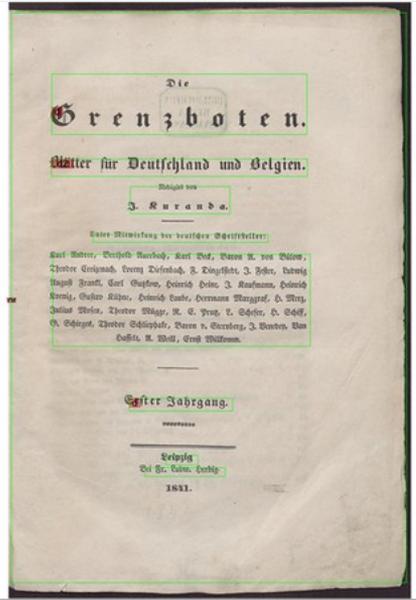
Cover in hoher Auflös...
Buch in Korrekturumgebung

[0077:](#)

Deutsche Eisenbahnen und belgische.

Der Sommer zeigte sich günstig für die Eisenbahnen auf dem Continente. Da haben wir schon wieder drei, die Schlag auf Schlag dem Verkehr Preis gegeben wurden: eine von Düsseldorf nach Elberfeld, eine andere von Straßburg nach Basel, und jetzt die zwischen Cöln und Aachen. Ich kenne die erstere nicht, habe aber die große, eiserne Straße, welche das Elsaß der Ausdauer eines einfachen Bürgers (des Herrn Köchlin) verdankt, voriges Jahr gesehen, als sie noch im Baue begriffen war. Es ist ein schönes Kunstwerk, das von den Sachverständigen als eines der bestangelegten in Europa betrachtet wird. Wie in Belgien, kam hier das natürliche **Gleich-**maß des Bodens, in der ganzen ungeheuern Fläche, die sich von den Vo-gesen nach dem Rheine zu ausdehnt, den Arbeiten der Ingenieure zu Hülfe, und gestattete ihnen, auf die Vollendung der Details mehr Fleiß zu ver-wenden. Es ist in der That eine con amore gebaute Eisenbahn, mit Schie-nen, so stark und fest, um unsere spätesten Enkel in Erstaunen zu setzen; mit prächtigen Stationen, in rothem Sandstein, die dereinst in den Augen

Die Grenzboten. Erster Jahrgang. Leipzig, 1841.



Cover in hoher Auflösung

[Buch in Korrekturumgebung öffnen ...](#)

Informationen

Quelle: Die Grenzboten
Publikationstyp: Zeitschrift/Zeitung
Umfang: 324 Scans
ca. 804823 Zeichen
ca. 116208 Wortformen
/[:alnum:]]+/
ca. 20983 Oberflächentypes

Schriftart: Fraktur
Genre: Gebrauchsliteratur :: Zeitschrift
im DTA seit: 2013-11-19 18:30:31
zuletzt geändert: 2021-05-13 12:28:24
Verfügbarkeit: Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): **Dieses Werk ist gemeinfrei.**
Weitere Informationen: [Nutzungsbedingungen.](#)

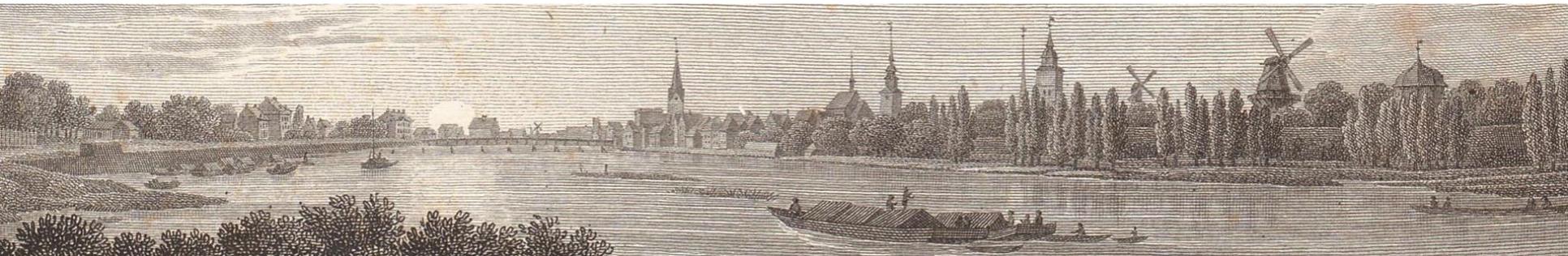
Grundlage dieses Digitalisats: Staats- und Universitätsbibliothek (SuUB) Bremen: Bereitstellung der Texttranskription. (2013-11-19T17:23:38Z) *Bitte beachten Sie, dass die aktuelle Transkription (und Textauszeichnung) mittlerweile nicht mehr dem Stand zum Zeitpunkt der Übernahme des Digitalisats entspricht.*

Ansichten

- Korrekturumgebung
- Webversion
- Faksimiles (DFG-Viewer)
- Download:
 - Text (UTF-8, Seitenumbrüche als ASCII \014)
 - TEI/XML (mit Silbentrennung)
 - HTML (mit Silbentrennung)
 - TCF (text annotation layer)
 - TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
 - TEI/XML (inkl. att.linguistic)
 - TEI-Header
 - CMDI
 - Dublin Core
- Lemmata:
 - nach Frequenz
 - nach Frequenz (nur Nomen)
 - Wortwolke
 - Wortwolke (nur Nomen)
- Wortformen (Types):
 - nach Frequenz
 - nach Frequenz (nur Nomen)
 - Wortwolke
 - Wortwolke (nur Nomen)
- Voyant Tools 🤖?:
 - transliterierter Text
 - normalisierter Text

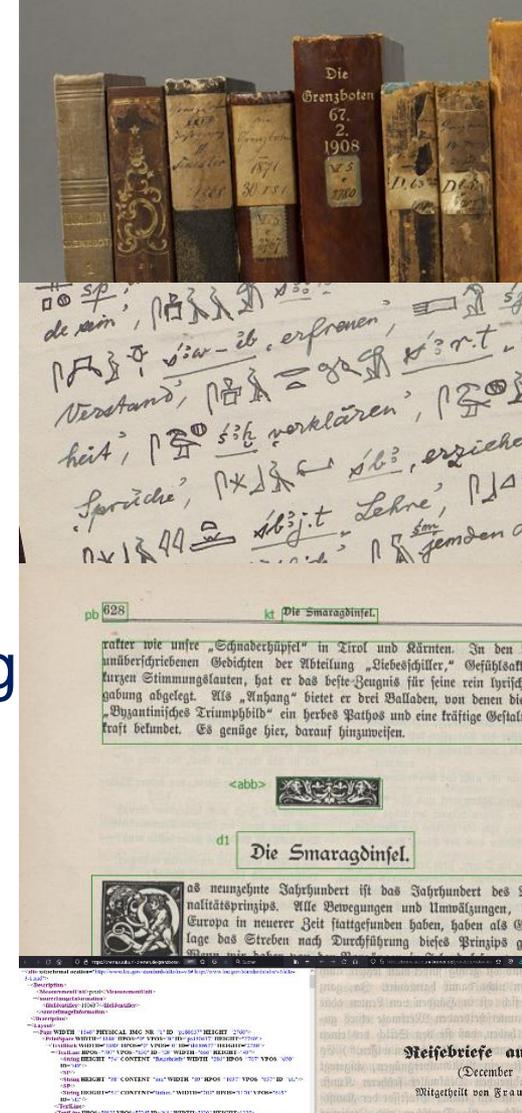
Datenquellen – Zusammenfassung

- Text-Kodierung, Struktur und Metadaten werden ermöglicht durch die Dateiformate
 - hOCR, PAGE, ALTO, TEI, METS/MODS, plain text, html, xhtml, ...
 - DH-Tools ~ Dateiformate
- Kritischer Blick auf die Quellen
 - Charakterisieren: wissenschaftliches Niveau, originäre Quellen („as is“) etc.
 - Kritisch: kommerzielle Angebote
 - Image-Digitalisat < Volltext-Digitalisat < Korpus < Edition
- Kritischer Blick auf Dateiformate
 - Häufigkeit; empfehlenswert; wenig geeignet; „inakzeptabel“



Gliederung

- Datenquellen – Vielfalt, Umfang, Qualität
- Datenquellen – Zugang, Formate, Struktur
- Einstieg in automatisierte Datenprozessierung
- Beispiel – Shakespeare Korpus
- Ausblick, Diskussion

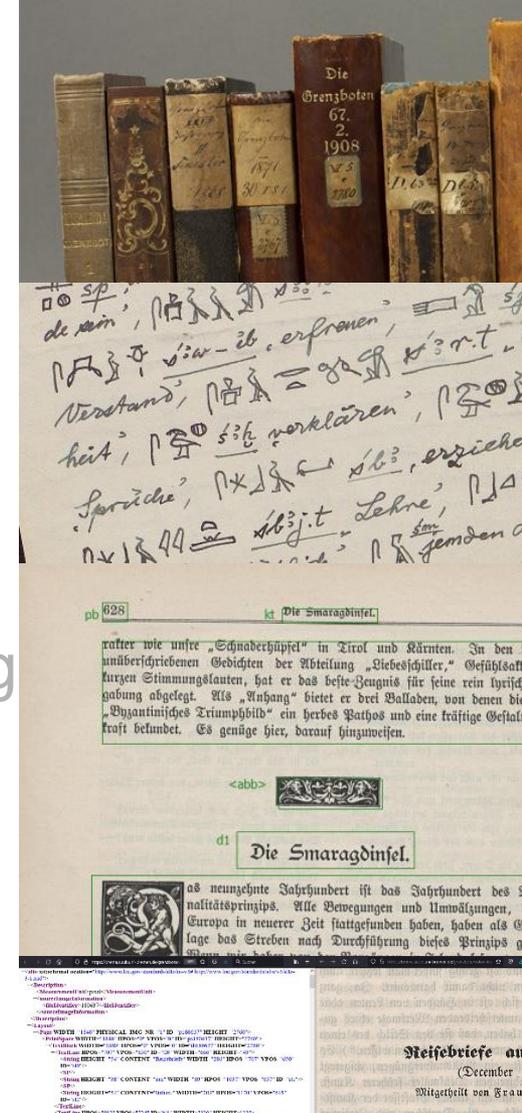


Einstieg in automatisierte Datenprozessierung

- Automatisierung selbst beim Zugang zu Daten (Download)
- Niederschwellige Methoden
 - Kommandozeile (nicht nur unter Linux)
 - *curl*-Skript – automatisierter Download
 - Plain text & Regex (Reguläre Ausdrücke);
gerne in der Diskussion:
Was ist geeigneter für den Einstieg: *grep*, *notepad++* oder *OpenRefine*?
 - Visualisierungen erstellen
- Infrastruktur am *Data Science Center* (DSC) der Universität Bremen
- „Laborbuch – Dokumentation der einzelnen Schritte“
 - Download, Datenvorverarbeitung, Analysen, Visualisierungen

Gliederung

- Datenquellen – Vielfalt, Umfang, Qualität
- Datenquellen – Zugang, Formate, Struktur
- Einstieg in automatisierte Datenprozessierung
- Beispiel – Shakespeare Korpus
- Ausblick, Diskussion



Beispiel – Shakespeare Korpus

- DraCor, <https://dracor.org/shake>
- Orientierung auf der dracor-Webseite
 - Mehrere Titel – alle von Shakespeare
 - Gut! – „Download a comprehensive table with metadata on all plays in the corpus: JSON, CSV“
- Korpus-Metadaten in den Formaten JSON und CSV
 - Es sind 37 Titel - einen Titel per „normal-Klick“ herunterladen
 - Einen weiteren mit „Ziel speichern unter ...“
- Bei hunderten von Titeln würde das **manuelle** Herunterladen jedoch **unangenehm, fehleranfällig** und **undokumentiert** sein
 - Beispielsweise haben der „Calderón Drama Corpus“ hat **205 Titel** und „gerdracor“ hat **605 Titel**

Beispiel – Shakespeare Korpus

- DraCor, <https://dracor.org/shake>
- Orientierung auf der dracor-Webseite
 - Mehrere Titel – alle von Shakespeare
 - Sehr schön! – „Download a comprehensive table with metadata on all plays in the corpus: JSON, CSV“
- JSON und CSV herunterladen und öffnen
 - Aha – es sind 37 Titel - einen Titel per „normal link“ herunterladen
 - Einen weiteren mit „Ziel speichern unter“
- Bei hunderten von Titeln würde das Herunterladen jedoch unangenehm, fehleranfällig und undokumentiert sein
 - Der „Calderón Drama Corpus“ hat 205 Titel und „gerdracor“ hat 605 Titel

Erstellung eines Download-Skripts (curl-Skript) und automatisierter Download

Beispiel – Shakespeare Korpus

- Erstellung eines Download-Skripts und automatisierter Download:
- Was habe ich für die folgenden Schritte benutzt?
 - Notepad++ - Reguläre Ausdrücke (Regex)
 - Schritt 1: Download-Skript (curl-Skript) erstellen
 - curl (Client for URLs) auf der Kommandozeile
 - Schritt 2: Gesamten *Shakespeare Drama Corpus* herunterladen
 - antillesXML – tei2txt.xslt
 - Schritt 3: TEI-Dateiformat in *plain text* konvertieren
 - Notepad++ – Regex Prozessor für ganze Verzeichnisse
 - Schritt 4: Text-Dateien – Datenvorverarbeitung

$^"([\^,]*)".*$
 $[\t\r\n]^+$
 $([.!?]) \rightarrow \t\r\n$

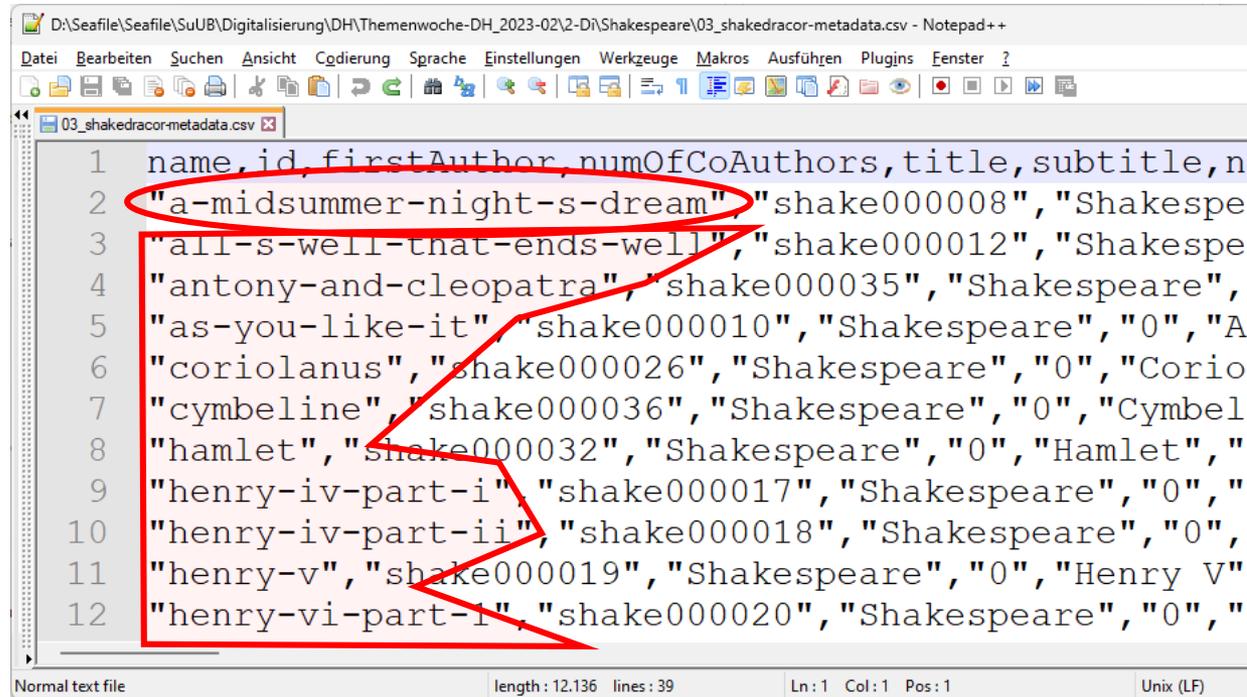
- Es ist eine Sprache für Textmuster mit Hilfe von Zeichenkürzeln, die in der Informatik eingesetzt wird
- Reguläre Ausdrücke (Regex) ermöglichen ein sehr mächtiges und flexibles Suchen und Ersetzen

Beispiel – Shakespeare Korpus – Schritt 1

Download-Skript (curl-Skript) erstellen

- Download-Skript (curl-Skript) auf Basis der csv-Metadaten erstellen
 - Notepad++ - Reguläre Ausdrücke (Regex)
- URL-Muster: `https://dracor.org/api/corpora/shake/play/<NAME>/tei` mit `<NAME>` = Korpus-Kurztitel (Beispiel: `henry-vi-part-1`)

- csv-Metadaten
- Erstes Feld ausschneiden:
 - `^"([\^,]*)".*` → `\1`



```
D:\Seafiler\Seafiler\SuUB\Digitalisierung\DH\Themenwoche-DH_2023-02\2-Di\Shakespeare\03_shakedracor-metadata.csv - Notepad++
Datei Bearbeiten Suchen Ansicht Codierung Sprache Einstellungen Werkzeuge Makros Ausführen Plugins Fenster 2
03_shakedracor-metadata.csv x
1 name,id,firstAuthor,numOfCoAuthors,title,subtitle,n
2 "a-midsummer-night-s-dream","shake000008","Shakespe
3 "all-s-well-that-ends-well","shake000012","Shakespe
4 "antony-and-cleopatra","shake000035","Shakespeare",
5 "as-you-like-it","shake000010","Shakespeare","0","A
6 "coriolanus","shake000026","Shakespeare","0","Corio
7 "cymbeline","shake000036","Shakespeare","0","Cymbel
8 "hamlet","shake000032","Shakespeare","0","Hamlet","
9 "henry-iv-part-i","shake000017","Shakespeare","0","
10 "henry-iv-part-ii","shake000018","Shakespeare","0",
11 "henry-v","shake000019","Shakespeare","0","Henry V
12 "henry-vi-part-1","shake000020","Shakespeare","0",
Normal text file length: 12,136 lines: 39 Ln: 1 Col: 1 Pos: 1 Unix (LF)
```

Beispiel – Shakespeare Korpus – Schritt 1

- DraCor, <https://dracor.org/shake>
- Download-Skript (curl-Skript) auf Basis der csv-Metadaten erstellen
 - Notepad++ - Reguläre Ausdrücke (Regex)
- URL-Muster: `https://dracor.org/api/corpora/shake/play/<NAME>/tei`
mit `<NAME>` = Korpus-Kurztitel (Beispiel: `henry-vi-part-1`)

• csv-Metadaten

• Erstes Feld

• `^"([\^,]*`

The screenshot shows a Notepad++ window with a CSV file named "03_shakedracor-metadata.csv". The file contains a list of Shakespeare plays with their IDs and author information. A red box highlights the first column of the CSV, which contains the play titles. A search and replace dialog box is open, showing the search pattern `^"([\^,]*` and the replacement `|1`. The dialog box also shows various search options and a status bar at the bottom indicating that 1 occurrence was replaced.

name	id	firstAuthor	numOfCoAuthors	title	subtitle	normalizedGe
a-midsummer-night-s-dream						
all-s-well-that-ends-well						
antony-and-cleopatra						
as-you-like-it						
coriolanus						
cymbeline						
"hamlet"	"shake000032"	"Shakes				
"henry-iv-part-i"	"shake000017					
"henry-iv-part-ii"	"shake000018					
"henry-v"	"shake000019"	"Shake				
"henry-vi-part-1"	"shake000020					



Beispiel – Shakespeare Korpus – Schritt 1

The screenshot shows a Notepad++ window with a CSV file named '03_shakedracor-metadata.csv'. The file contains metadata for Shakespeare plays, including titles like 'a-midsummer-night-s-dream', 'all-s-well-that-ends-well', 'antony-and-cleopatra', 'as-you-like-it', 'coriolanus', 'cymbeline', 'hamlet', 'henry-iv-part-i', 'henry-iv-part-ii', 'henry-v', and 'henry-vi-part-1'. A search and replace dialog box is open over the file. The search criteria are set to 'Suchen nach: ^"([\^,]*)".*' and the replacement is 'Ersetzen durch: curl -o Korpus\\1.tei -J -L http://dracor.org/api/corpora/shake/play/1/tei'. The dialog also shows options for search direction, word boundaries, case sensitivity, and search mode (Normal, Extended, Regular expressions). The status bar at the bottom indicates 'Ersetzen: 1 Vorkommen wurde ersetzt. Nächster Treffer gefunden'.

- Erstes Feld ausschneiden durch curl-Aufruf ersetzen:
 - `^"([\^,]*)".*` →
`curl -o Korpus\\1.tei -J -L http://dracor.org/api/corpora/shake/play/1/tei`



Beispiel – Shakespeare Korpus – Schritt 2

Korpora herunterladen

```

D:\Seafile\Seafile\SuU8\Digitalisierung\DH\Themenwoche-DH_2023-02\2-D\Shakespeare\download-37.bat - Notepad++
Datei Bearbeiten Suchen Ansicht Codierung Sprache Einstellungen Werkzeuge Makros Ausführen Plugins Fenster ?
download-37_2.bat
1 curl -o Korpus\a-midsummer-night-s-dream.tei -J -L http://dracor.org/api/corpora/shake/play/a-midsummer-night-s-dream/tei
2 curl -o Korpus\all-s-well-that-ends-well.tei -J -L http://dracor.org/api/corpora/shake/play/all-s-well-that-ends-well/tei
3 curl -o Korpus\antony-and-cleopatra.tei -J -L http://dracor.org/api/corpora/shake/play/antony-and-cleopatra/tei
4 curl -o Korpus\as-you-like-it.tei -J -L http://dracor.org/api/corpora/shake/play/as-you-like-it/tei
5 curl -o Korpus\coriolanus.tei -J -L http://dracor.org/api/corpora/shake/play/coriolanus/tei
6 curl -o Korpus\cymbeline.tei -J -L http://dracor.org/api/corpora/shake/play/cymbeline/tei
7 curl -o Korpus\hamlet.tei -J -L http://dracor.org/api/corpora/shake/play/hamlet/tei
8 curl -o Korpus\henry-iv-part-i.tei -J -L http://dracor.org/api/corpora/shake/play/henry-iv-part-i/tei
9 curl -o Korpus\henry-iv-part-ii.tei -J -L http://dracor.org/api/corpora/shake/play/henry-iv-part-ii/tei
10 curl -o Korpus\henry-v.tei -J -L http://dracor.org/api/corpora/shake/play/henry-v/tei
11 curl -o Korpus\henry-vi-part-1.tei -J -L http://dracor.org/api/corpora/shake/play/henry-vi-part-1/tei
12 curl -o Korpus\henry-vi-part-2.tei -J -L http://dracor.org/api/corpora/shake/play/henry-vi-part-2/tei
13 curl -o Korpus\henry-vi-part-3.tei -J -L http://dracor.org/api/corpora/shake/play/henry-vi-part-3/tei
14 curl -o Korpus\henry-viii.tei -J -L http://dracor.org/api/corpora/shake/play/henry-viii/tei
15 curl -o Korpus\julius-caesar.tei -J -L http://dracor.org/api/corpora/shake/play/julius-caesar/tei
16 curl -o Korpus\king-john.tei -J -L http://dracor.org/api/corpora/shake/play/king-john/tei
17 curl -o Korpus\king-lear.tei -J -L http://dracor.org/api/corpora/shake/play/king-lear/tei
18 curl -o Korpus\love-s-labor-s-lost.tei -J -L http://dracor.org/api/corpora/shake/play/love-s-labor-s-lost/tei
19 curl -o Korpus\macbeth.tei -J -L http://dracor.org/api/corpora/shake/play/macbeth/tei
20 curl -o Korpus\measure-for-measure.tei -J -L http://dracor.org/api/corpora/shake/play/measure-for-measure/tei
21 curl -o Korpus\much-ado-about-nothing.tei -J -L http://dracor.org/api/corpora/shake/play/much-ado-about-nothing/tei
22 curl -o Korpus\othello.tei -J -L http://dracor.org/api/corpora/shake/play/othello/tei
23 curl -o Korpus\pericles.tei -J -L http://dracor.org/api/corpora/shake/play/pericles/tei
24 curl -o Korpus\richard-ii.tei -J -L http://dracor.org/api/corpora/shake/play/richard-ii/tei
25 curl -o Korpus\richard-iii.tei -J -L http://dracor.org/api/corpora/shake/play/richard-iii/tei
26 curl -o Korpus\romeo-and-juliet.tei -J -L http://dracor.org/api/corpora/shake/play/romeo-and-juliet/tei
27 curl -o Korpus\the-comedy-of-errors.tei -J -L http://dracor.org/api/corpora/shake/play/the-comedy-of-errors/tei
28 curl -o Korpus\the-merchant-of-venice.tei -J -L http://dracor.org/api/corpora/shake/play/the-merchant-of-venice/tei
29 curl -o Korpus\the-merry-wives-of-windsor.tei -J -L http://dracor.org/api/corpora/shake/play/the-merry-wives-of-windsor/tei
30 curl -o Korpus\the-taming-of-the-shrew.tei -J -L http://dracor.org/api/corpora/shake/play/the-taming-of-the-shrew/tei
31 curl -o Korpus\the-tempest.tei -J -L http://dracor.org/api/corpora/shake/play/the-tempest/tei
32 curl -o Korpus\the-winter-s-tale.tei -J -L http://dracor.org/api/corpora/shake/play/the-winter-s-tale/tei
33 curl -o Korpus\timon-of-athens.tei -J -L http://dracor.org/api/corpora/shake/play/timon-of-athens/tei
34 curl -o Korpus\titus-andronicus.tei -J -L http://dracor.org/api/corpora/shake/play/titus-andronicus/tei
35 curl -o Korpus\troilus-and-cressida.tei -J -L http://dracor.org/api/corpora/shake/play/troilus-and-cressida/tei
36 curl -o Korpus\twelfth-night.tei -J -L http://dracor.org/api/corpora/shake/play/twelfth-night/tei
37 curl -o Korpus\two-gentlemen-of-verona.tei -J -L http://dracor.org/api/corpora/shake/play/two-gentlemen-of-verona/tei
38

```

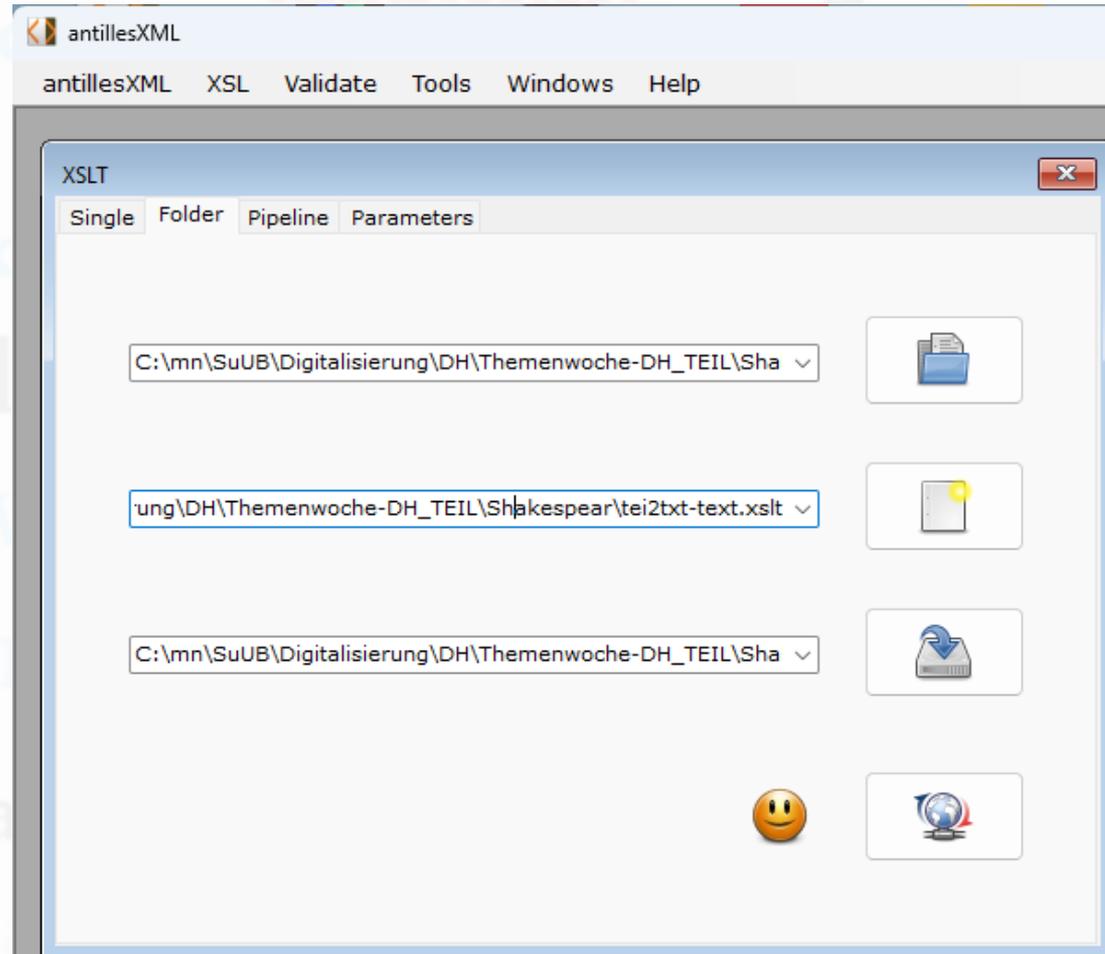
Dieses Skript

- hat 32 Sekunden gedauert
- hat ca. 100 MB heruntergeladen
- ist weniger fehleranfällig als ein manuelles Herunterladen
- und dokumentiert diesen Schritt der Daten-Akquisition

Beispiel – Shakespeare Korpus – Schritt 3

TEI-Dateiformat in *plain text* konvertieren

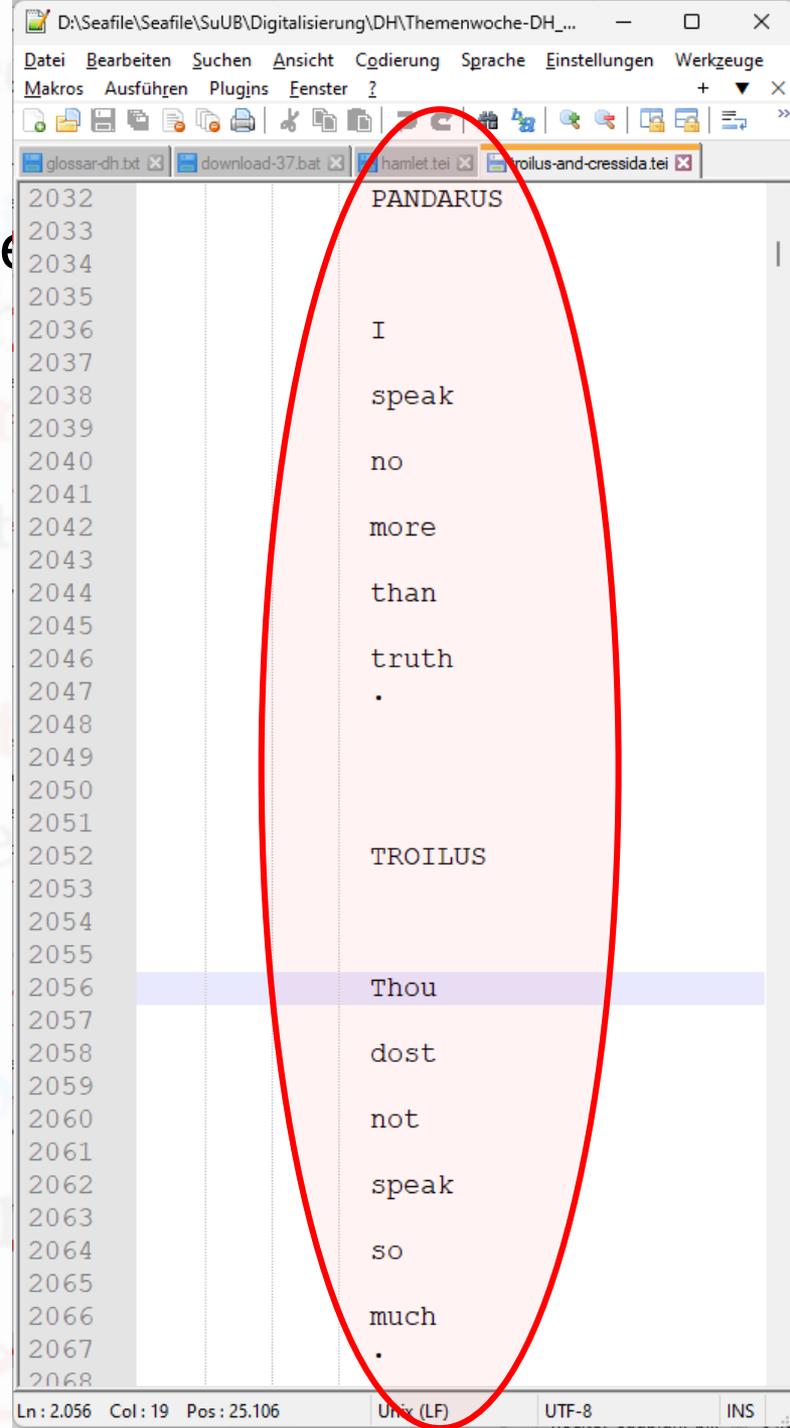
- antillesXML
 - tei2txt.xslt



Beispiel – Shakespeare

Text-Dateien – Datenvorverarbeitung

- Ziel: Text-Dateien in eine „geeigneterere“ Form bringen
- [\t\r\n]+ → „ “ und ([.!?]) → \1\r\n



The screenshot shows a text editor window with a menu bar (Datei, Bearbeiten, Suchen, Ansicht, Codierung, Sprache, Einstellungen, Werkzeuge) and a toolbar. The main text area contains a list of lines numbered 2032 to 2068. A red oval highlights the text from line 2032 to 2068. The text is as follows:

```
2032 PANDARUS
2033
2034
2035
2036 I
2037 speak
2038 no
2039 more
2040 than
2041 truth
2042 .
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052 TROILUS
2053
2054
2055
2056 Thou
2057
2058 dost
2059 not
2060 speak
2061 so
2062 much
2063 .
2064
2065
2066
2067
2068
```

The status bar at the bottom shows: Ln: 2056 Col: 19 Pos: 25.106 Unix (LF) UTF-8 INS



Beispiel – Shakespeare

Text-Dateien – Datenvorverarbeitung

- Ziel: Text-Dateien in eine „geeigneterere“ Form bringen
- [\\t\\r\\n]+ → „ “ und ([.!?]) → \\1\\r\\n
- Regex Prozessor für ganze Verzeichnisse

```
2032 PANDARUS
2033
2034
2035
2036 I
2037
2038 speak
2039
2040 no
2041
2042 more
2043
2044 than
2045
2046 truth
2047 .
```

```
43 PANDARUS I speak no more than truth .
44 TROILUS Thou dost not speak so much .
45 PANDARUS Faith , I'll not meddle in it .
46 Let her be as she is .
47 If she be fair , 'tis the better for her ; an she be not , she has
the mends in her own hands .
48 TROILUS Good Pandarus – how now , Pandarus ?
49 PANDARUS I have had my labor for my travail , ill thought on of her
, and ill thought on of you ; gone between and between , but small
```

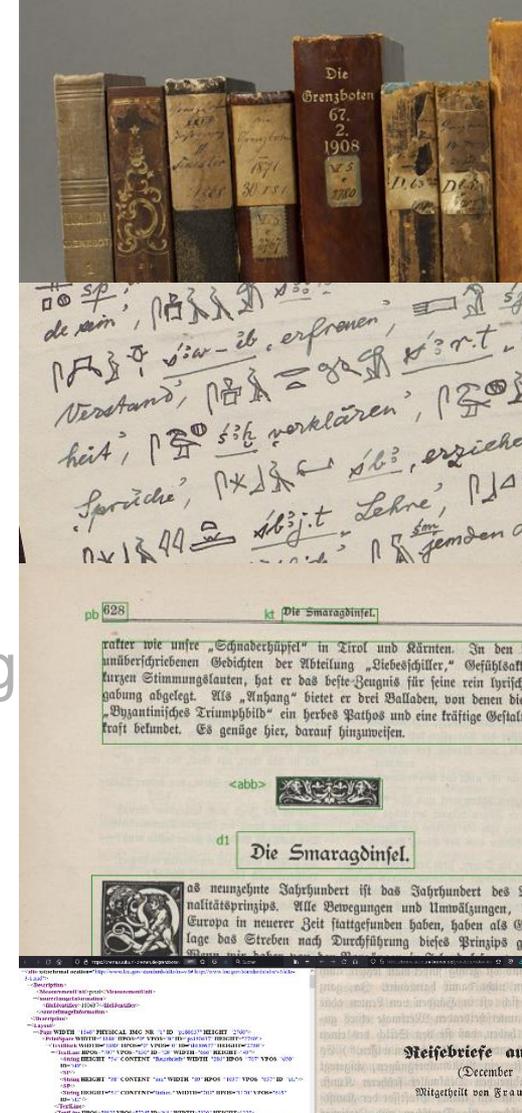
Beispiel – Shakespeare Korpus

Möglichkeiten weiterer Untersuchungen

- „moon“ – Häufigkeit, Kollokationen, Sentiment, Kontext, Position im Text, etc.
- Es kann untersucht werden, ob die Verwendung der Pronomina *you / thou, thy thine, ye* und *you* in Shakespeare's Dramen erkennbaren Gesetzmäßigkeiten unterliegen
 - Wortfrequenzen? / Visualisierung
- Länge der Sätze / der Beiträge im Dialog
 - Visualisierung → Muster?

Gliederung

- Datenquellen – Vielfalt, Umfang, Qualität
- Datenquellen – Zugang, Formate, Struktur
- Einstieg in automatisierte Datenprozessierung
- Beispiel – Shakespeare Korpus
- **Ausblick, Diskussion**



Ausblick auf die folgenden Termine der Themenwoche

- Montag
 - Überblick und Einstieg in DH-Datenquellen und DH-Methoden
- Dienstag
 - DH-Datenquellen und die ersten Schritte der Verarbeitung
- Mittwoch
 - Einstieg in DH-Methoden, Agnes Brauer, UB Frankfurt
- Donnerstag
 - Fortgeschrittene DH-Methoden
- Freitag
 - Ada Lovelace Center for Digital Humanities, Dr. Dennis Mischke

SAVE THE date

Themenwoche
DIGITAL HUMANITIES

Immer um 12:00 Uhr
für ca. 30 Minuten
online via Zoom

Mo, 06.02.	Überblick und Einstieg in DH-Datenquellen & DH-Methoden
Di, 07.02.	DH-Datenquellen und die ersten Schritte der Verarbeitung
Mi, 08.02.	Einstieg in DH-Methoden und DH-Aktivitäten an der UB Frankfurt
Do, 09.02.	Fortgeschrittene DH-Methoden
Fr, 10.02.	Ada Lovelace Center for Digital Humanities

b Staats- und Universitätsbibliothek Bremen



Staats- und
Universitätsbibliothek
Bremen

Vielen Dank für die Aufmerksamkeit!

Wenden Sie sich gerne für eine DH Beratung an:

Dr. Manfred Nölte, noelte@suub.uni-bremen.de, 0421 218 59416

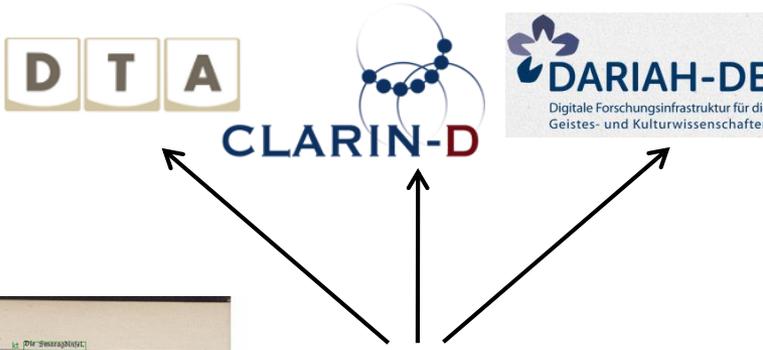
<https://www.suub.uni-bremen.de/fachinformationen/digital-humanities/>



Professional Imaging and Archiving
H+K
Herrmann und Kraemer



ABBYY



```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader> ...
<orgName>Staats- und Universitätsbibliothek (S
...
<sourceDesc><bibl type="J">Die Grenzboten.
...
<p xml:id="ID_8" next="#ID_9">
```

Ich lasse Ihnen das Glauben und Wissen vor
andere Richterstühle gehörig hier zur Se
eigenthümlichen Stellung dieser drei Männ
Volksstämmen. Strauß, der Schwabe mit s
Urkraft, die mit einem Hiebe durch Reiter u
tief in Pferdes Rücken schlug; Ludwig Feu
zornesschnaubend und gewandt auch das

Die jüngste päpstliche Bulle und die künftige Papstwahl.
Bedenken über die Sprachverbesserung.
Eine neue Sentenz.
Das Atelier eines französischen Bildhauers.
Die jüngste Papiere.
Der Vertrag über den Papiere.
Bedenken über die Sprachverbesserung.