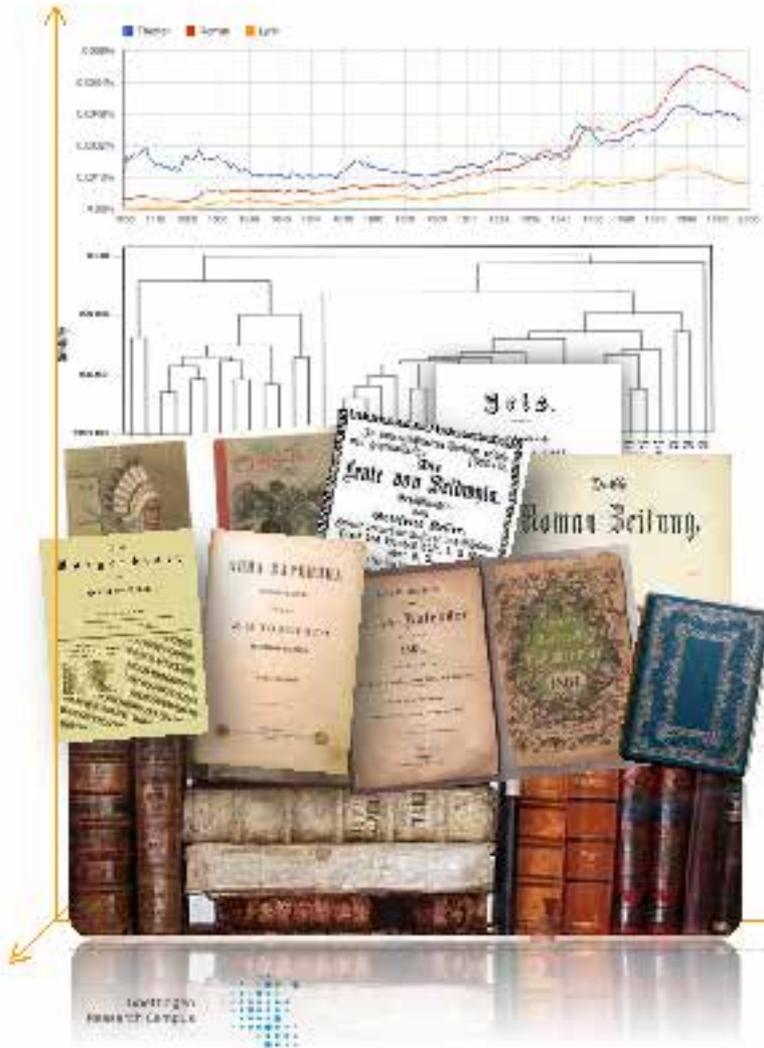




# Digitale Editionen und Korpora

Potenziale und Grenzen  
digitaler Vernetzung





## Bibliotheken, Archive und Editionen

The works touching books are two — first, **libraries**, which are as the shrines where all the relics of the ancient saints, full of true virtue, and that without delusion or imposture, are preserved and reposed; secondly, **new editions of authors**, with more correct impressions, more faithful translations, more profitable glosses, more diligent annotations, and the like.

Francis Bacon, *The Advancement of Learning*, 5th book, 1605





# Arbeitsplatz und Edition



Arbeitszimmer  
von Jacob Grimm



# Arbeitsweisen

- **Discovering / Retrieval** (Sammeln, Zusammenstellen und Sichern)
- **Annotating** (Attributierung)
- **Comparing** (Analogieschluss als wichtigstes hermeneutisches Mittel)
- **Referring** (Verweisen, wiss. Diskussion)
- **Sampling** (Gruppieren, Statistik)
- **Illustrating** (Visualisierung, Bildgebung)
- **Representing** (Präsentation)

John Unsworth (2000), *Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?*





# Index Thomisticus

1949

Roberto Busa SJ (1913-2011) &  
Thomas J. Watson (Gründer von IBM)



Forschung zum Begriff der *praesens / praesentia*

Lemmatisierung des Gesamtwerks des Hl. Thomas

56 Bände, <http://www.corpusthomisticum.org>







# Blumenbach online



*for Blumenbach*  
Blumenbach – online

ICMNE	Sammlungsobjekte	Literatur	Video	Biographie
1800-1801	1790-1800	1780-1790	1770-1780	1760-1770
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	1870-1871
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	1900-1901
1902-1903	1904-1905	1906-1907	1908-1909	1910-1911
1912-1913	1914-1915	1916-1917	1918-1919	1920-1921
1922-1923	1924-1925	1926-1927	1928-1929	1930-1931
1932-1933	1934-1935	1936-1937	1938-1939	1940-1941
1942-1943	1944-1945	1946-1947	1948-1949	1950-1951
1952-1953	1954-1955	1956-1957	1958-1959	1960-1961
1962-1963	1964-1965	1966-1967	1968-1969	1970-1971
1972-1973	1974-1975	1976-1977	1978-1979	1980-1981
1982-1983	1984-1985	1986-1987	1988-1989	1990-1991
1992-1993	1994-1995	1996-1997	1998-1999	2000-2001
2002-2003	2004-2005	2006-2007	2008-2009	2010-2011
2012-2013	2014-2015	2016-2017	2018-2019	2020-2021
2022-2023	2024-2025	2026-2027	2028-2029	2030-2031



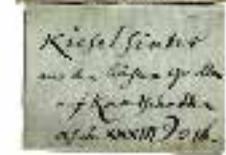
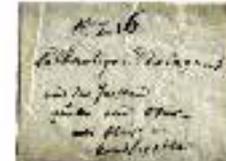
**Malware von Maschinen, Oberlehrer (König Ludwig I von Bayern (1786 – 1848))**

180 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 1870-1871 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 1900-1901 1902-1903 1904-1905 1906-1907 1908-1909 1910-1911 1912-1913 1914-1915 1916-1917 1918-1919 1920-1921 1922-1923 1924-1925 1926-1927 1928-1929 1930-1931 1932-1933 1934-1935 1936-1937 1938-1939 1940-1941 1942-1943 1944-1945 1946-1947 1948-1949 1950-1951 1952-1953 1954-1955 1956-1957 1958-1959 1960-1961 1962-1963 1964-1965 1966-1967 1968-1969 1970-1971 1972-1973 1974-1975 1976-1977 1978-1979 1980-1981 1982-1983 1984-1985 1986-1987 1988-1989 1990-1991 1992-1993 1994-1995 1996-1997 1998-1999 2000-2001 2002-2003 2004-2005 2006-2007 2008-2009 2010-2011 2012-2013 2014-2015 2016-2017 2018-2019 2020-2021 2022-2023 2024-2025 2026-2027 2028-2029 2030-2031



# Blumenbach online

Akademie der Wissenschaften, Projekt mit 15 Jahre Laufzeit  
 Johann Friedrich Blumenbach (1752-1840), Begründer der Zoologie und  
 Anthropologie als wissenschaftliche Disziplinen  
 ca. 10.000 gedruckte Seiten & ca. 8.000 analoge Objekte, die 2D und 3D  
 digitalisiert werden





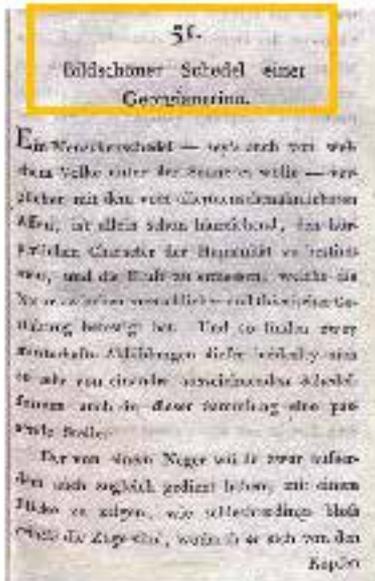
# Blumenbach online

**Digitalisat (Text)**  
mit formaler und  
inhaltlicher  
Erschließung

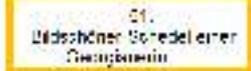
**Transkription** mit  
TEI Metadaten,  
z.B. Titel, Orts-  
namen, Personen-  
namen etc.

**Digitalisat**  
(Abbildung im  
Text) mit  
Metadaten

**3D-Digitalisat** mit  
inhaltlichen und  
fachwissenschaft-  
lichen Metadaten



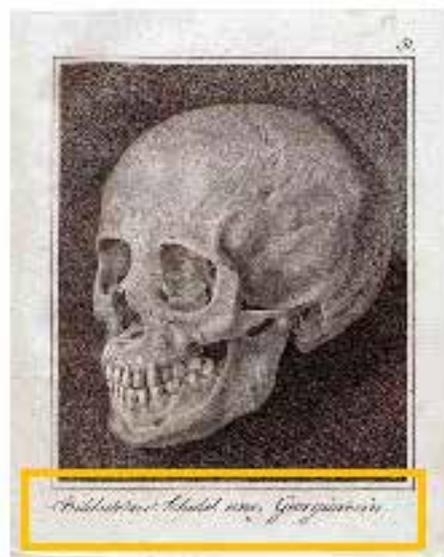
Waldenburger, Carl: *Die Tierwelt Deutschlands* (1857)



51.  
Goldschener Schödel einer  
Geopithecina

Ein Menschenschödel — wie's auch von we-  
chem Velle unter der Bezeichnung wele — ver-  
glichen mit dem von einem menschenähnlichen  
Affen, ist demselben insofern, dass im-  
besonderen Charakter der Humanität zu bestim-  
men, und die Klaffen zu erkennen, welche die  
Natur zwischen menschlicher und thierischer Ge-  
staltung befestigt hat, und so hundert zuey  
mehrfache Abbildungen dieses Schädel, wie  
so sehr von einander abzulehnenen Schödel-  
formen auch in dieser Sammlung eine pas-  
sende Stelle.

Der von einem Nagel wurde zwei weiten-  
dem auch zugleich odern haben mit einer  
Dicke zu zeigen, wie schlechdings dass  
nicht die Zähne sind, wodurch er sich von der  
Kapfer







# Eindeutige Metadaten

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

Réponse n° 1



Domaine **peinture**

Type d'objet **tableau**

Titre **PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE**

Auteur/exécutant **LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)**

Précision auteur/exécutant **Vinci, 1452 ; Amboise, 1519**

Ecole **Italie**

Période création/exécution **1er quart 16e siècle**

Milésime création/exécution **1503 entre ; 1506 et**

Genèse **oeuvre en rapport ; reproduit en gravure**

Historique **commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 ; nombreuses copies dont une conservée au Louvre ; gravé par Feuchery, par Filhol, par Landon**

Matériaux/techniques **peinture à l'huile ; bois**

Mesures **77 H ; 53 L**

Sujet représenté **portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assise, accoudé, loggia, Italien) ; fond de paysage (montagne, rocher, cours d'eau, pont, plaine, route)**

Date sujet représenté **1479-1528**

Lieu de conservation **Paris ; musée du Louvre département des Peintures**



**Musée de France**

**au sens de la loi n°2002-5 du 4 janvier 2002**

Statut juridique **propriété de l'Etat ; musée du Louvre département des Peintures**

Antennes appartenances **François 1er ; Couronné de France**





# Community

Etablierte Anlaufstellen in der Community diskutieren und setzen Standards (Terminologie und Ontologien, Metadaten und Annotationen, Strukturen):

DARIAH (Digital Research Infrastructure for the Arts and Humanities): <https://de.dariah.eu> ,  
<http://www.gwdg.de/index.php?id=dariah>

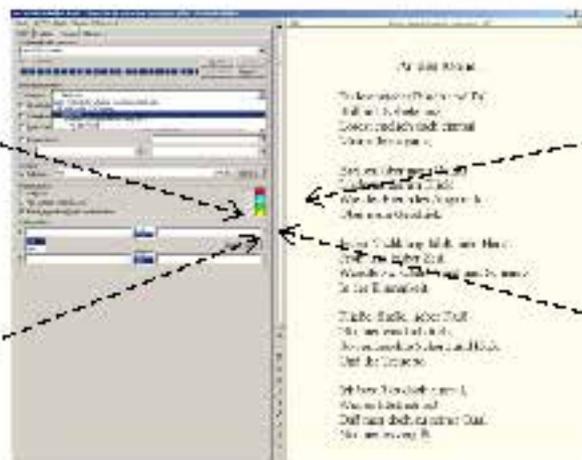
Europeana (virtuelle Bibliothek zum kulturellen Erbe Europas): <http://www.europeana.eu>

CLARIN-D (Common Language Resources and Technology Infrastructure Deutschland):  
<http://de.clarin.eu/de/>





# Werk- oder Archiv-Erschließung



INTERNET ARCHIVE  
**WayBack** Machine





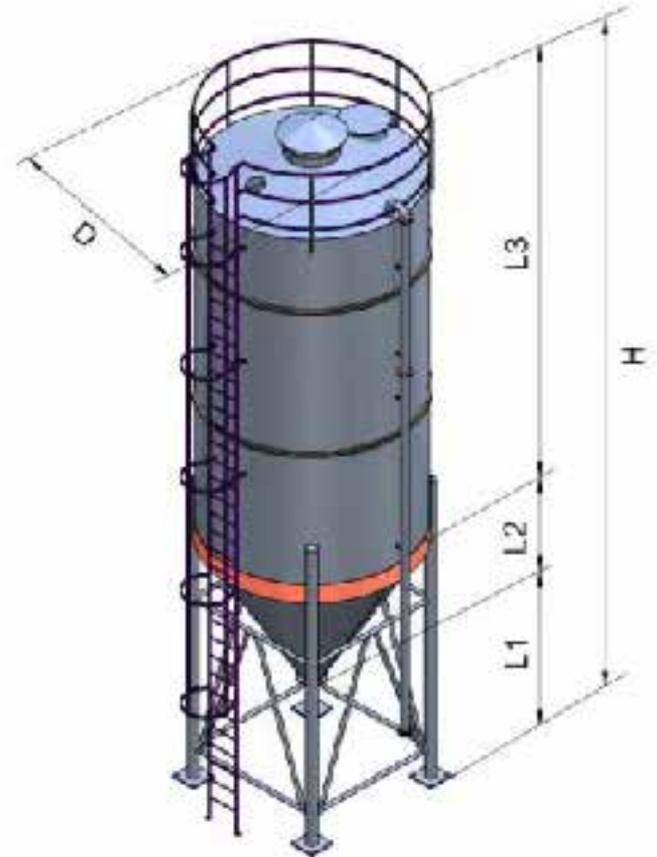
# Achtung: Silo!

Werkcharakter der Editionen

oder

Kulturelle Bedeutsamkeit der Quellen

- je für sich -







# Blumenbach online



## DARWIN ONLINE

Publications Manuscripts Biographies Media About us Search Advanced search

The world's largest and most widely used resource on Darwin. Facebook. What's new?

Edited by John van Wyhe

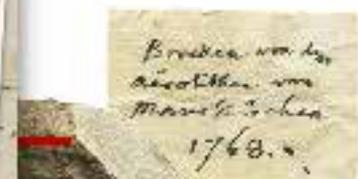
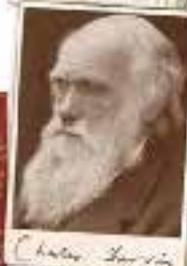
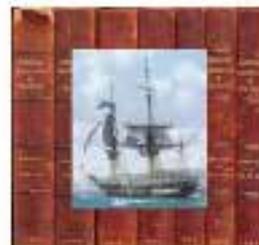
**NEW!** Darwin's *Beagle* Library now online

**Darwin's Complete Publications**

- Books *Origin of Species, Descent of Man, Voyage of the Beagle...*
- Articles Volcanic, Darwin & Wallace paper...
- Published Letters *Life and letters, Darwin & Henslow...*
- Published Manuscripts *Autobiography, Beagle diary* (audio)...

**Darwin's Private Papers & Manuscripts**

- Notebooks, Journal, student bills, marriage notes,



**Manuskript von Maximilian, Oberlehrer (des König Ludwig I von Bayern (1796 – 1848)**





# Mit den Bibliotheken

## Bach, Johann Sebastian 1685-1750

### Overview

<b>Works:</b>	122,768 works in 223,477 publications in 13 languages and 1,245,785 library holdings
<b>Genres:</b>	nomadic catalogs, biography, Manuscripts, Oratorio, interpretation, etc. Musical settings, Suites, Hymns, Juvenile works, Piano music, Music
<b>Subject Headings:</b>	Composers
<b>Roles:</b>	Komponistin, Bearbeiterin, bibliographisches Bezugselement, Urheber, Compiler, Donator, Bewahrerin, Mitarbeiterin, Umkodiererin, Zuschreiberin/Räuber, Instrumentalistin, Leherin, Prologistin, Widmerin, Adapter, Darstellerin, Lexikerin, Lektorin, Autorin unbek., Übersetzerin, Direktorin, s. s., Kuratorin, opm
<b>Classification:</b>	ML410.B1, S27.B

### Publication Timeline



#### Key

- Publications about Johann Sebastian Bach
- Publications by Johann Sebastian Bach
- Publications by Johann Sebastian Bach, published posthumously

[www.worldcat.org](http://www.worldcat.org)





# Korpora: Deutsches Textarchiv



NEAR(Gott, "die Welt", 1)

suchen

Texte ▾

Projekte ▾

In der Thesauriendatenbank In der Korpora In der Dokumentationsdatenbank

## Suche im Deutschen Textarchiv

Treffer 1 - 10 von 456

KWIC • Neue Suche • Vorherige • Nächste • Hilfe | NEAR(Gott, "die Welt", 1)

suchen

10 Treffer pro Seite Suchrichtung: Datum aufsteigend/absteigend • zufällig



- 1: Arndt, Johann: Vom wahren Christenthumb. Bd. 4. Magdeburg, 1610. #244
- Darumb ist das **die** aller herrlichste / heilichste / vnd trostlichste consequenz der Himelischen Discretion: Also hat **Gott die Welt** gelebet / das Er seinen Eingebornen Sohn gab / das ist:

[mehr]

Lesen Sie über die einzelnen Teilkorpora zur Maus, um folgende Informationen zu sehen:

• die Originaldatei, die in kritischer Editionsmorphologie (Lohn + Text) ist (URL-normale in, etc. Wortform)  
 • die Lemmata (URL-feldern) und  
 • die Part-of-Speech-Analyse



- 2: Arndt, Johann: Vom wahren Christenthumb. Bd. 4. Magdeburg, 1610. #244
- Gott** hat **die Welt** auts höchstegelebet / Darumb hat Er ihr seinen Sohn gegeben.

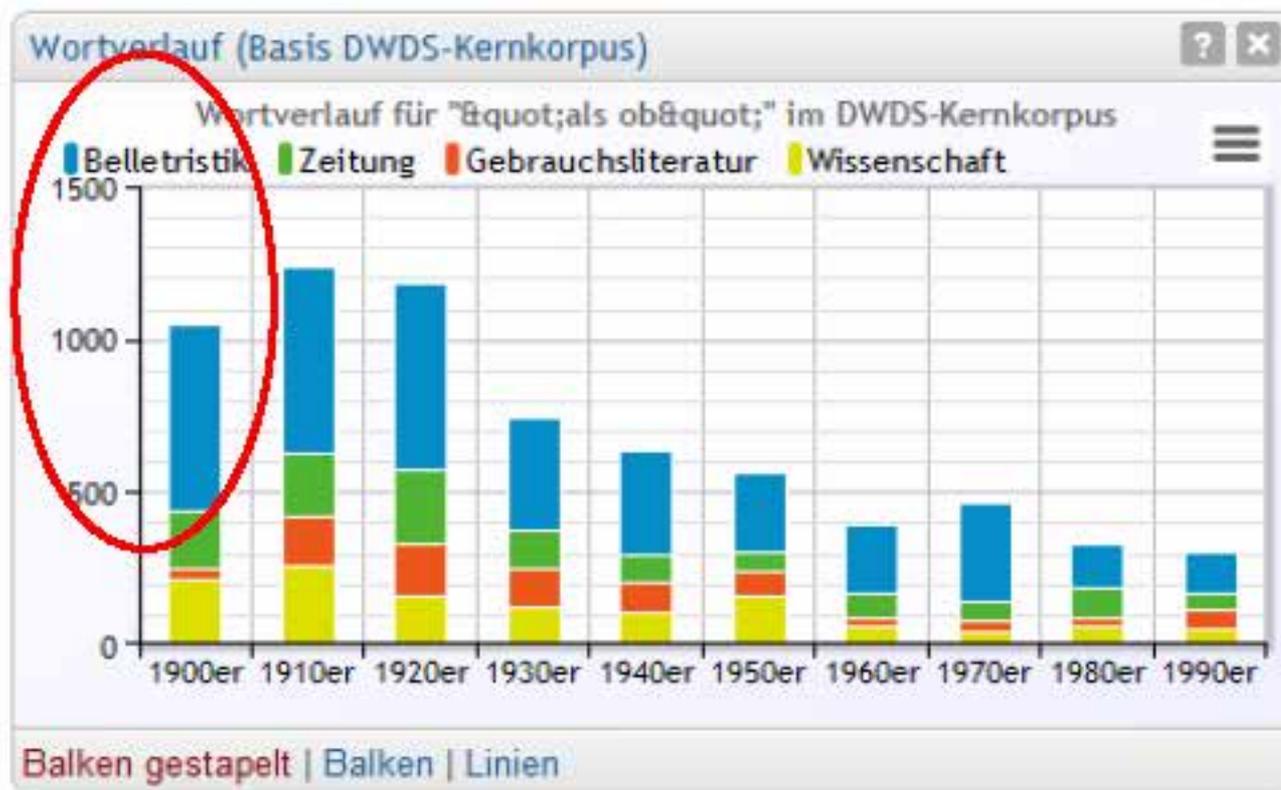
[mehr]

www.deutschestextarchiv.de



## Korpora: DWDS

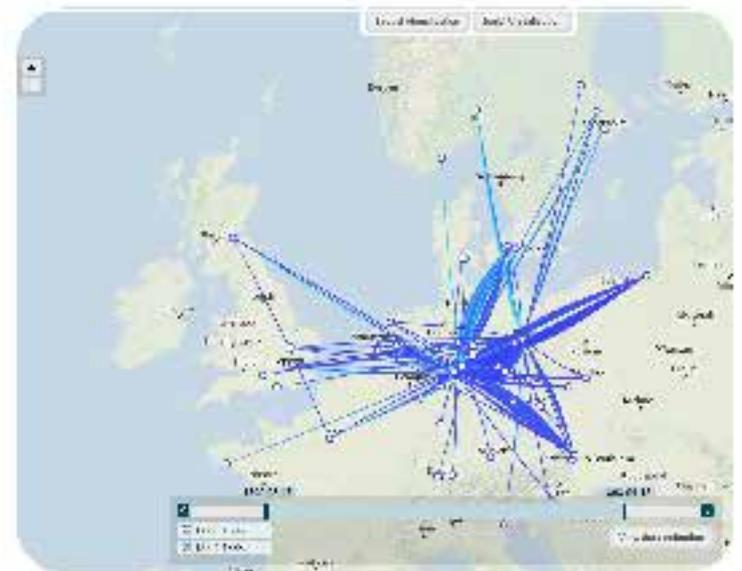
‚als ob‘



[www.dwds.de](http://www.dwds.de)



# Muster erkennen in Korrespondenz-Netzen



<http://toolingup.stanford.edu/rplviz/>

<http://spinn et.humanities.uva.nl/>



# European Holocaust Research Infrastructure

Collections at Yad Vashem

Arie Biegan, Page of Testimony

Personal Tax File, Beest Linnik Municipality, October 1941

Record Group

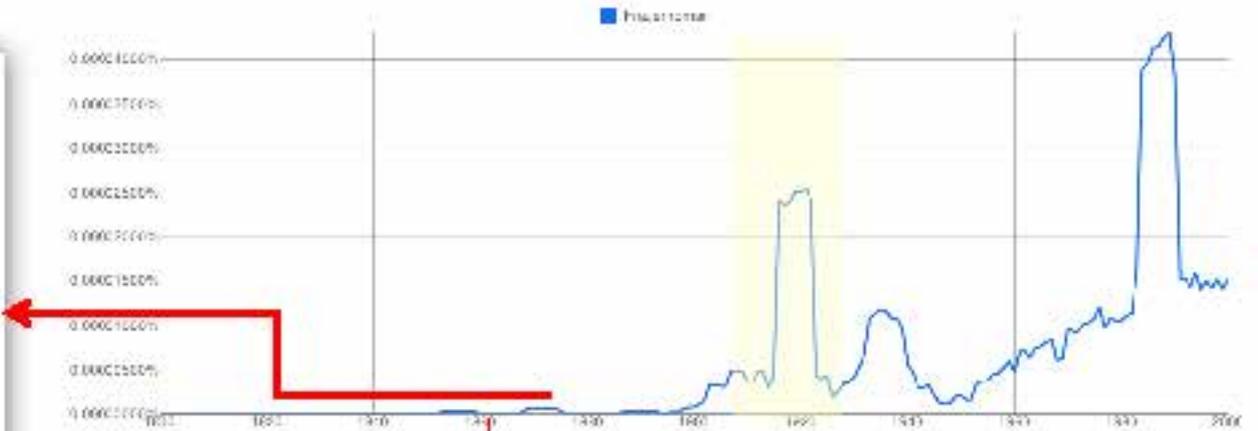
Collection	Collection		
File	File	File	File





# „Frauenroman“

Graph these **case-sensitive** case-insensitive phrases: **Frauenroman**  
 between 1820 and 2020 from the corpus: German with smoothing of 3  
 [Switch line of tracks]



Das mehr naturalistische oder realistische Element werden wir erst während des Jahres 1859 sich mehr entwickeln sehen. Der moderne weibliche Roman wird besonders den Dichtungen von Charlotte Brontë zugeschrieben, an deren eigenste und zuweilen kühne Originalität sich auch viele Nachfolgerinnen anlehnen, so daß es in kritischen Besprechungen Stil ward, im Roman den Einfluß derselben ebenso nachzuweisen, wie in der Lyrik den von Wordsworth.

Google Ngram Viewer

„Frauenroman“, Jahrbuch fuer romanische und englische Sprache und Literatur, 1861

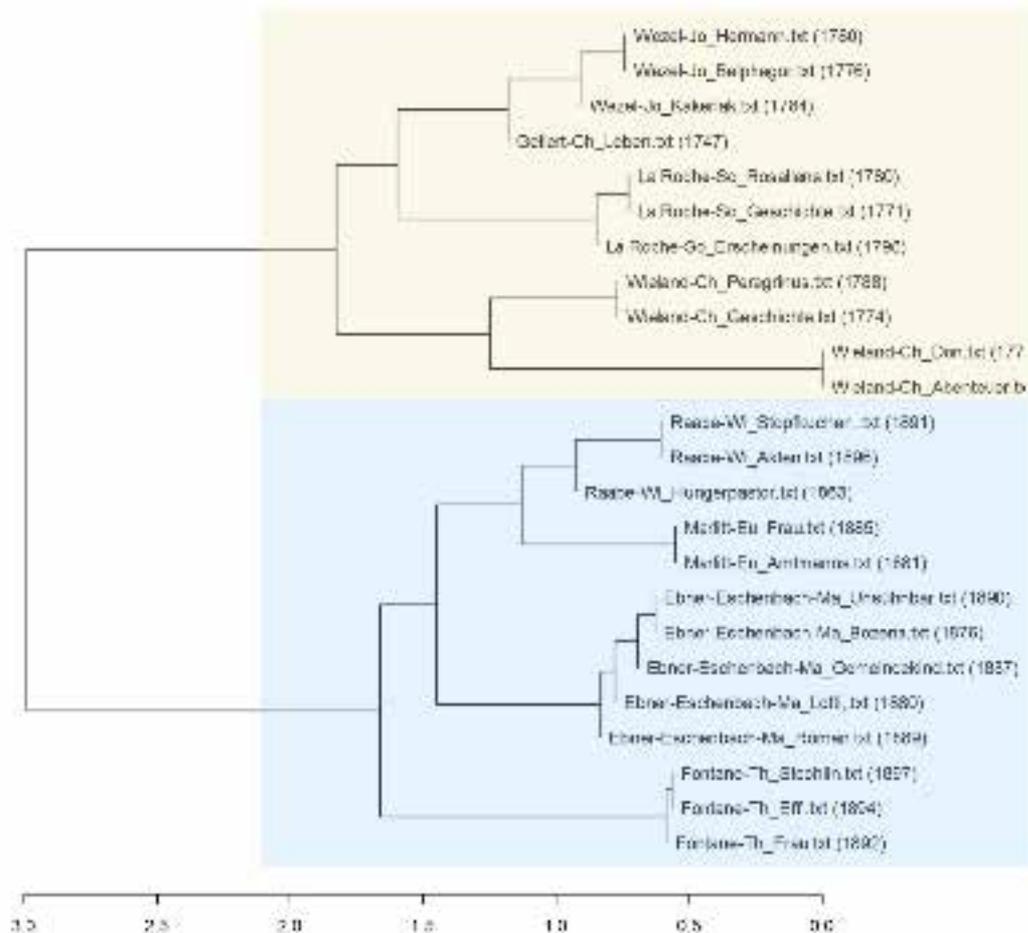


# Epochen

## Stylometry with R

(<https://sites.google.com/site/computationalstylistics/home>)

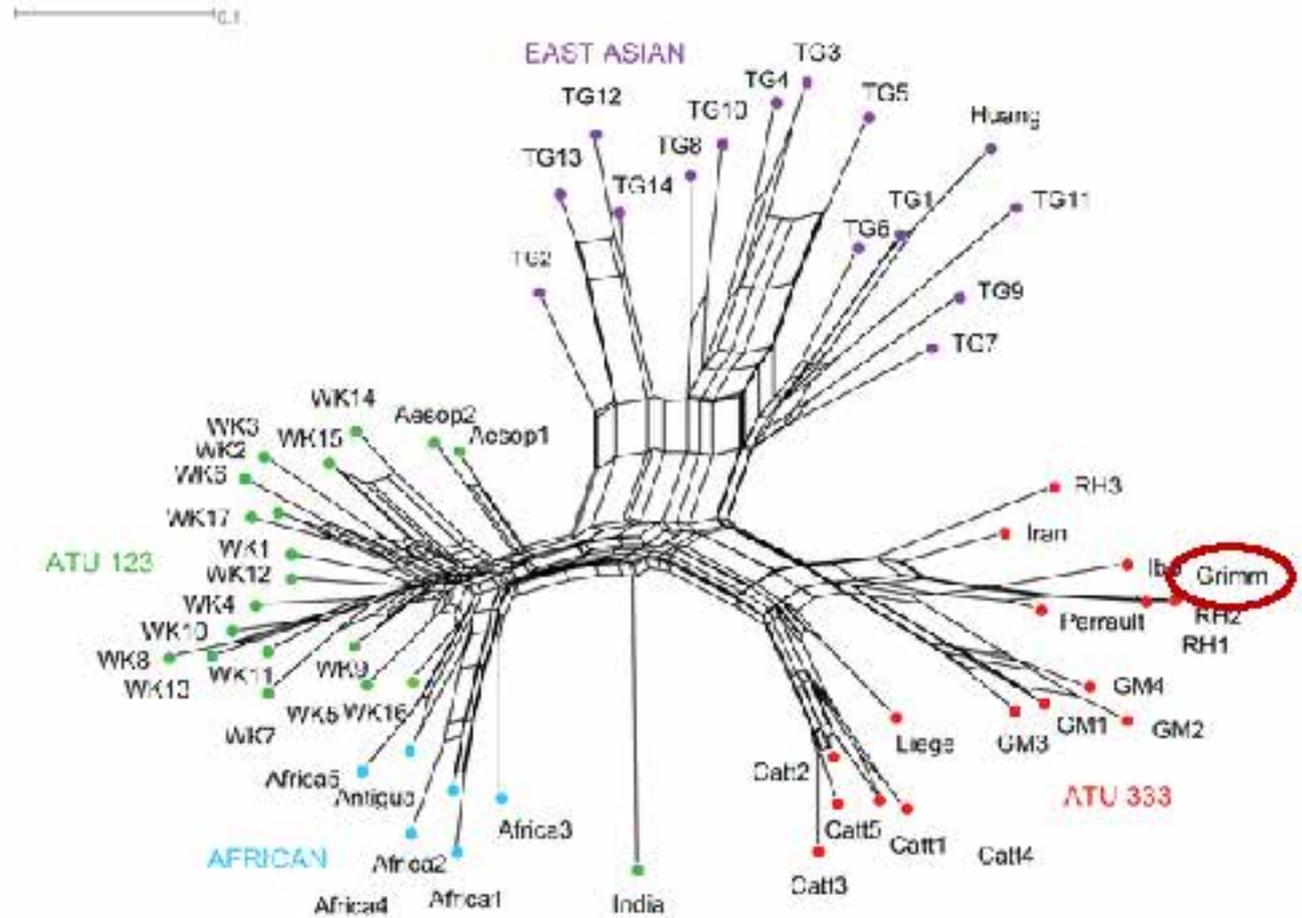
"Distinction between literary epoch:  
Enlightenment vs. Realism":  
Dendrogram, Classic Delta distance,  
300-3000 MFW, consensus 0.5





# Märchen der Welt

## SplitsTrees



Therani,  
J.J. (2013)



# Visualisierung großer Datenmengen

Sentiment  
Topic/Event  
Analysis



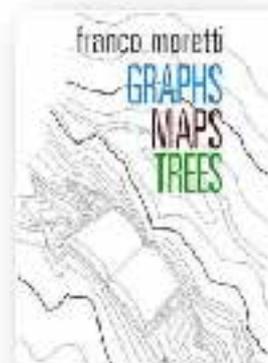
Fig. 5 Age Suitability Visualization of the book *Harry Potter and the Deathly Hallows*. In the upper left corner: distribution of emotions, in the middle: readability overview, upper right side: the topic(s) of the book. The lower part shows the occurrence of the various characters in the book. Each line reflects a character. The topmost line is Harry Potter, the main character in the book. He occurs in almost each text section.

Wanner et  
al. (2011)



## Distant reading: big and dirty

Franco Moretti, *Conjectures on World Literature*,  
New Left Review 1 (2000)



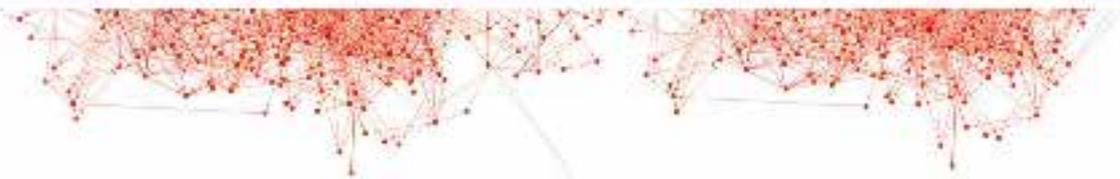
“literary history will quickly become very different from what it is now: it will become ‘second hand’: a patchwork of other people’s research, *without a single direct textual reading*. Still ambitious, and actually even more so than before (world literature!); but the ambition is now directly proportional to the distance from the text: the more ambitious the project, the greater must the distance be”

<http://newleftreview.org/11/1/franco-moretti-conjectures-on-world-literature>  
Crane, G. (2006). What do you do with a million books,  
D-Lib-magazine 12, 3





# Korpus



- Sprachkorpora (British National Corpus u.a.)
- Baumbanken (TIGERCorpus u.a.)
- Wörterbuchnetze ([www.woerterbuchnetz.de](http://www.woerterbuchnetz.de) u.a.)
- Web als Korpus (Internet Archive u.a.)
- Autorenportale (Heine-Portal u.a.)
- Themenportale (Irish History u.a.)
- MetaCrawler (Europeana z.B.)

<http://www.bubenhofner.com/korpuslinguistik>



## Europeana Data Model (EDM)

- Unterscheidung von vorgehaltenem Objekt und seiner digitalen Repräsentation
- Unterscheidung zwischen Objekt und seinen Metadaten
- Erlaubnis verschiedene Beschreibungen des selben Objekts, auch einander widersprechende
- Unterstützung für Objekte, die aus verschiedenen Objekten bestehen
- Spezifikation von Metadaten Formaten (z.B. Dublin Core, Open Archives Initiative Object Reuse and Exchange)
- Kombination standardisierter und kontrollierter Vokabulare (SKOS)







## ... oder nur unstrukturierte Daten?

```
*new 1 - Notepad--
Datei Bearbeiten Suchen Ansicht Kodierung Sprachen Einstellungen Mskro Ausführen Erweiterungen Fenster ?
new 1
1 Jemand mußte Josef K. verleumdet haben, denn ohne daß er etwas Böses getan hätte, wurde er eines Morgens verha
2
3 Noch war er frei. "Erlauben Sie", sagte er und ging silig zwischen den Wächtern durch in sein Zimmer. "Er sch
4
5 Ohne auf dieses Angebot zu antworten, stand K. ein Weilchen lang still. Vielleicht würden ihn die Beiden, wenn
6
7 Er warf sich auf sein Bett und nahm vom Nachttisch einen schönen Apfel, den er sich gestern Abend für das Früh
8
9 Da erschreckte ihn ein Zuruf aus dem Nebenzimmer derartig, daß er mit den Zähnen ans Glas schlug. "Der Aufseher
10
11 Als er vollständig angezogen war, mußte er knapp vor Willem durch das leere Nebenzimmer in das folgende Zimmer
12
13 "Josef K.?" fragte der Aufseher, vielleicht nur um K. 's zerstreute Blicke auf sich zu lenken. K. nickte. "Sie
14
15 K. starrte den Aufseher an. Schelmäßige Lehren bekam er hier von einem vielleicht jüngeren Menschen? Für seine
```





# Open Access



## OpenEdition

Books, journals, research blogs and news  
in the humanities and social sciences

<http://www.openedition.org>



hypotheses

1000000 1000000 1000000

France | Europe | Amérique | Asie | Afrique | Océanie | Méditerranée | Langues | Histoire | Philosophie | Sciences | Arts | Littérature | Musique | Cinéma | Théâtre | Sport | Santé | Environnement | Économie | Droit | Éducation | Technologie | Société | Culture | Religion | Éthique | Écologie | Climat | Énergie | Santé | Éducation | Technologie | Société | Culture | Religion | Éthique | Écologie | Climat | Énergie



1000000 1000000 1000000

**Le futur de la recherche**  
Le futur de la recherche est incertain. Les chercheurs doivent être conscients de ce qui se passe et de ce qui est possible. Les chercheurs doivent être conscients de ce qui se passe et de ce qui est possible.

**La recherche française**  
La recherche française est en crise. Les chercheurs doivent être conscients de ce qui se passe et de ce qui est possible. Les chercheurs doivent être conscients de ce qui se passe et de ce qui est possible.





# Citizen science



The New York Times  
**Madis**

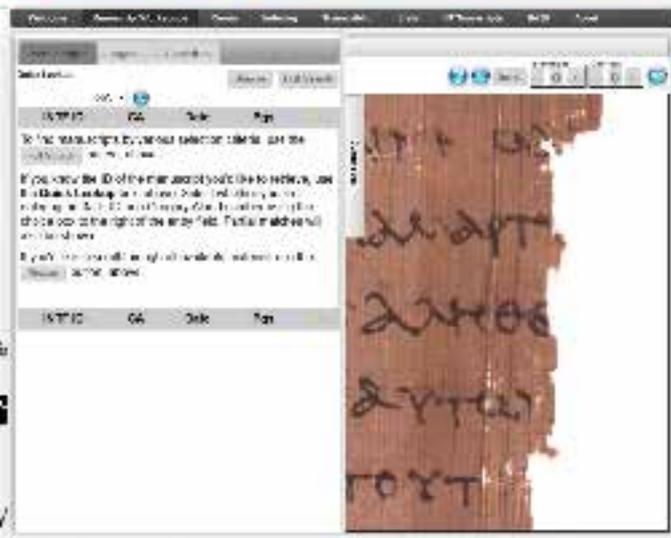
Help preserve history with

The New York Times archives are full of advertisements that give glimpses into daily life and cultural history. Help us digitize our historic ads by answering simple questions. You'll be creating a unique resource for historians, advertisers and the public — and leaving your mark on history.

Get started with our collection of ads from the 1960s (additional decades will be opened later)

Get started

H. Althman's ad, January 25th, 1963



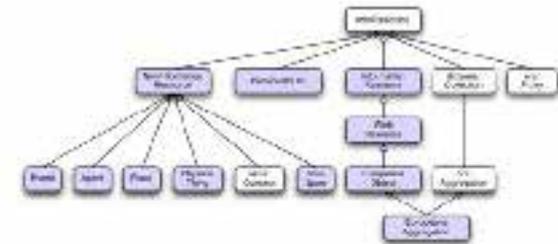
<http://ntvmr.uni-muenster.de/home>

<http://madison.nytimes.com/>



# Größer Denken

- (Normdaten-)Standards, Standards
- Big (dirty) data vernetzen
- Eindeutige Referenzen
- Kollaboratives Arbeiten
- Open Access
- Öffentlichkeit mit einbeziehen





# Edition haben ihre Zukunft vor sich!

