GEORG ECKERT INSTITUT
Leibniz-Institut für internationale Schulbuchforschung

gei.Digital

Kenne Dein Korpus –

Computerlinguistische Analysen in 5000+ Werken des „GEI-Digital-2020 Korpus“

Dr. Maret Nieländer, nielaender@leibniz-gei.de

Digital Turn und Historische Bildungsforschung – Bestandsaufnahme und Forschungsperspektiven
Online-Forschungstag, Freitag, 18. Juni 2021

WWU Münster, Institut für Erziehungswissenschaft

WWU MÜNSTER

Institut für Erziehungswissenschaft

Delphische Sibylle, Ausschnitt aus dem Fresko von Michelangelo in der Sixtinischen Kapelle, CC0

Agenda

Das ideale Korpus

GEI-Digital: Digitale Erschließung historische Schulbücher

Das GEI-Digital-2020 Korpus:

- Kuration der Daten
- Demo D* und DiaCollo:
- Einfache und komplexe Suchen
- Diachrone Kollokationen

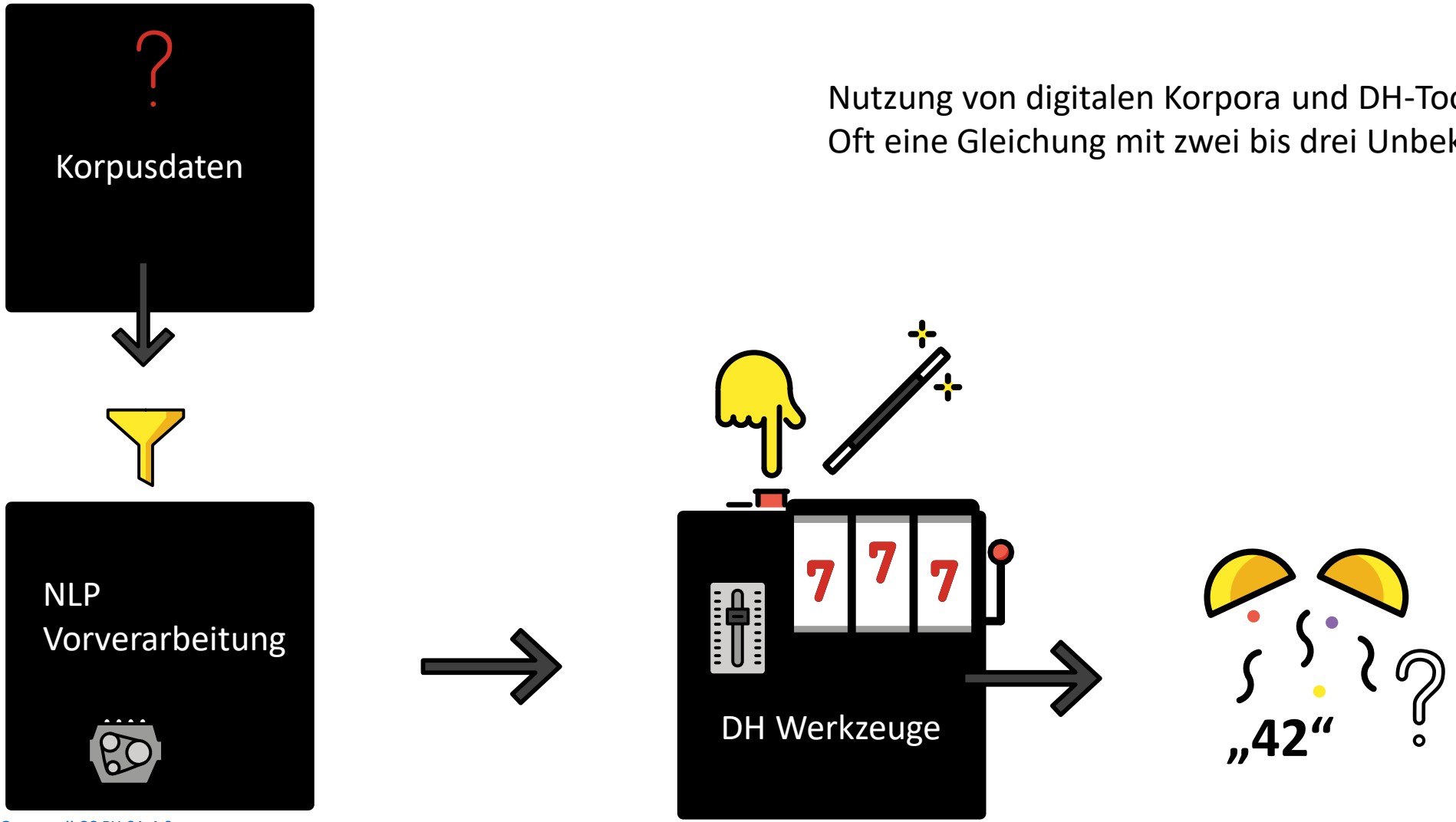
Befunde und ihre erkenntnistheoretischen Implikationen

Das ideale Textkorpus

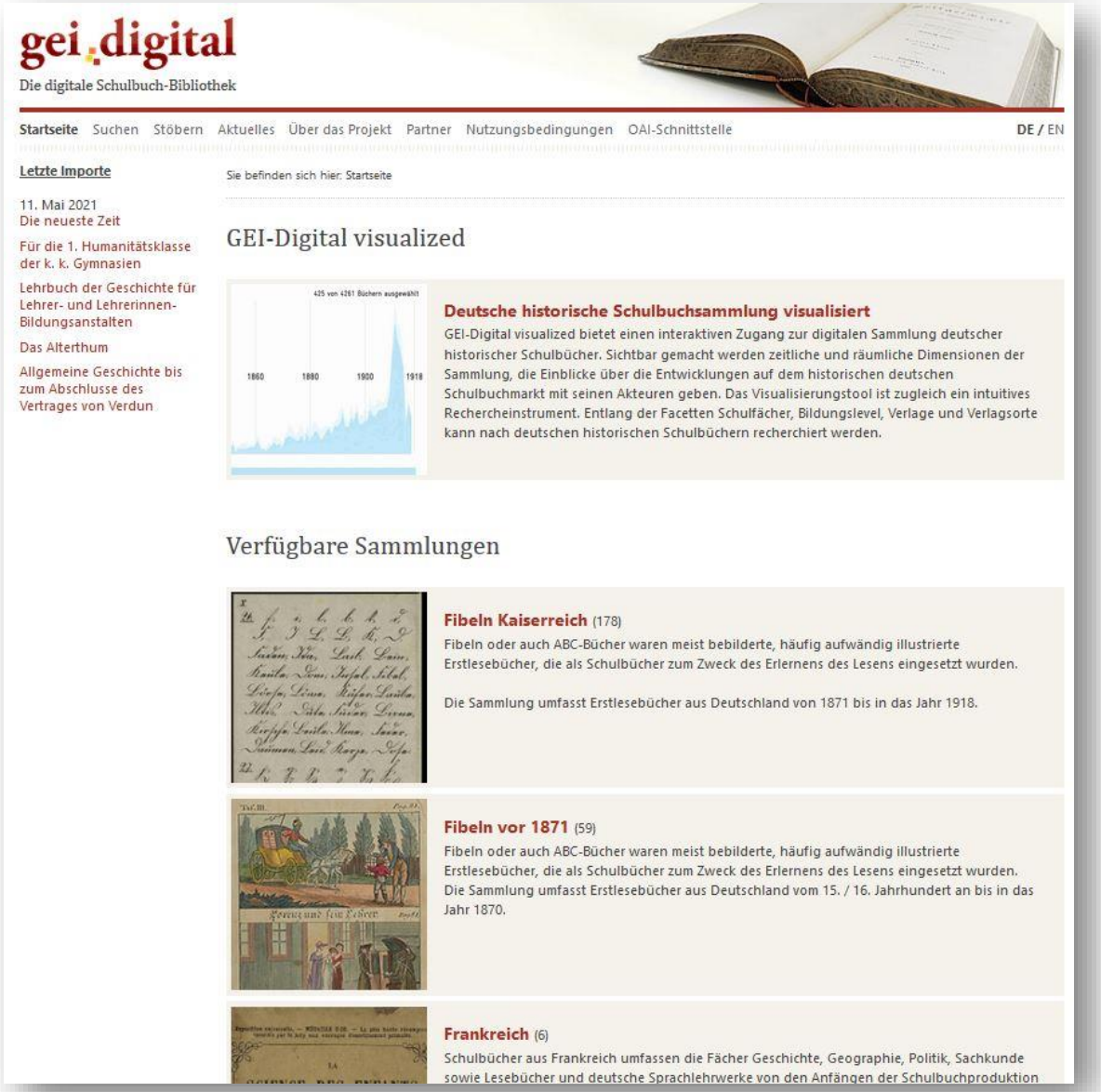
- vollständig oder repräsentativ (relevant und aussagekräftig)
- Open Access
- Eindeutige ID
- Versionierung
- klein genug (close reading / intensive Lektüre), ODER groß genug (statistische Analysen)
- originalgetreue Volltexte
- Metadaten korrekt, standardisiert, detailliert
- Annotation/Vorverarbeitung/Indexierung angemessen
- Verbindung zu passenden digitalen Werkzeugen

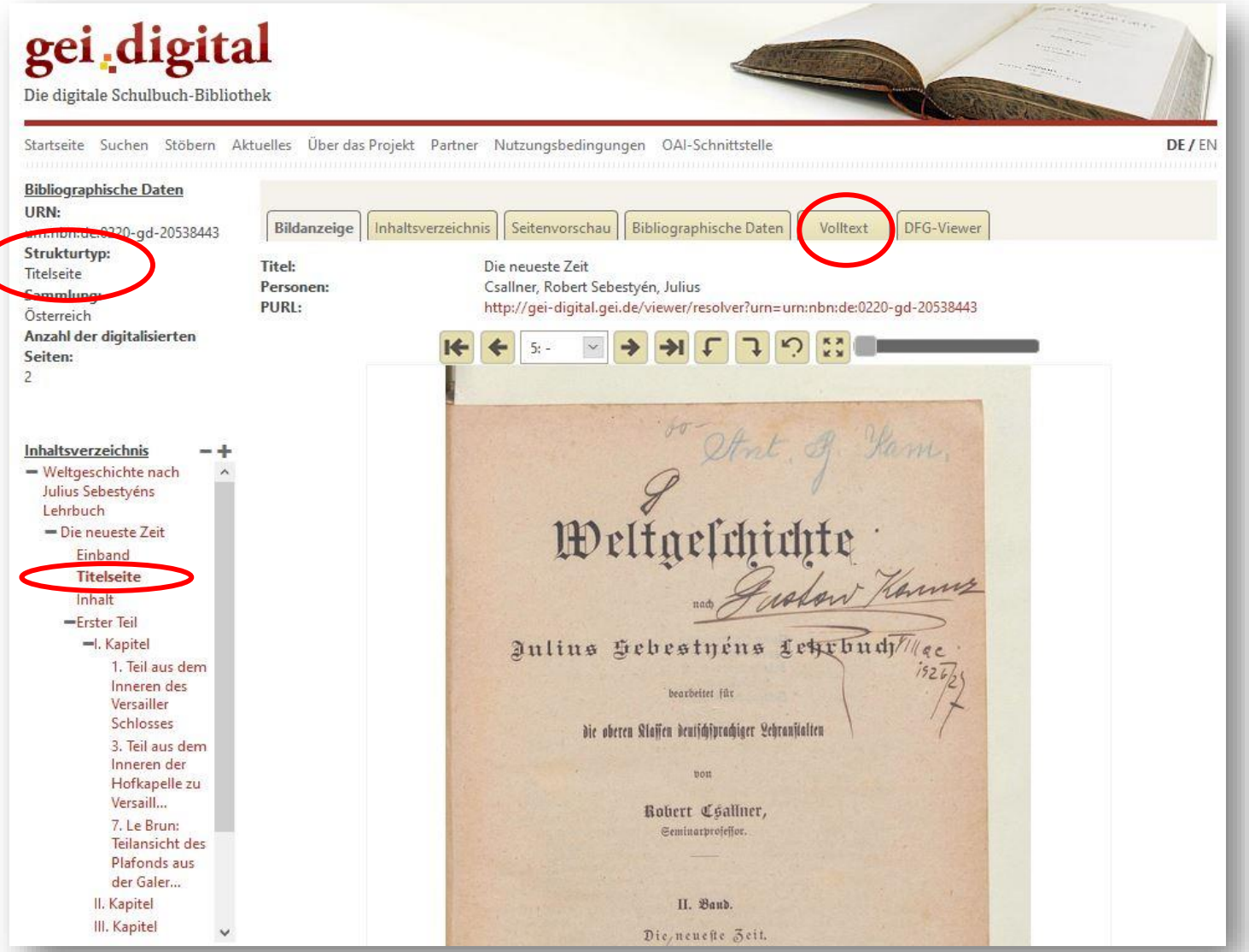
- Dokumentation!

Nutzung von digitalen Korpora und DH-Tools:
Oft eine Gleichung mit zwei bis drei Unbekannten



Openmoji CC BY-SA 4.0


gei.digital
Die digitale Schulbuch-Bibliothek
Startseite Suchen Stöbern Aktuelles Über das Projekt Partner Nutzungsbedingungen OAI-Schnittstelle DE / EN
Letzte Importe
11. Mai 2021
Die neueste Zeit
Für die 1. Humanitätsklasse der k. k. Gymnasien
Lehrbuch der Geschichte für Lehrer- und Lehrerinnen-Bildungsanstalten
Das Alterthum
Allgemeine Geschichte bis zum Abschlusse des Vertrages von Verdun
Sie befinden sich hier: Startseite
GEI-Digital visualized
425 von 4261 Büchern ausgewählt
1860 1880 1900 1918
Deutsche historische Schulbuchsammlung visualisiert
GEI-Digital visualized bietet einen interaktiven Zugang zur digitalen Sammlung deutscher historischer Schulbücher. Sichtbar gemacht werden zeitliche und räumliche Dimensionen der Sammlung, die Einblicke über die Entwicklungen auf dem historischen deutschen Schulbuchmarkt mit seinen Akteuren geben. Das Visualisierungstool ist zugleich ein intuitives Rechercheinstrument. Entlang der Facetten Schulfächer, Bildungslevel, Verlage und Verlagsorte kann nach deutschen historischen Schulbüchern recherchiert werden.
Verfügbare Sammlungen
Fibeln Kaiserreich (178)
Fibeln oder auch ABC-Bücher waren meist bebilderte, häufig aufwändig illustrierte Erstlesebücher, die als Schulbücher zum Zweck des Erlernens des Lesens eingesetzt wurden.
Die Sammlung umfasst Erstlesebücher aus Deutschland von 1871 bis in das Jahr 1918.
Fibeln vor 1871 (59)
Fibeln oder auch ABC-Bücher waren meist bebilderte, häufig aufwändig illustrierte Erstlesebücher, die als Schulbücher zum Zweck des Erlernens des Lesens eingesetzt wurden.
Die Sammlung umfasst Erstlesebücher aus Deutschland vom 15. / 16. Jahrhundert an bis in das Jahr 1870.
Frankreich (6)
Schulbücher aus Frankreich umfassen die Fächer Geschichte, Geographie, Politik, Sachkunde sowie Lesebücher und deutsche Sprachlehrwerke von den Anfängen der Schulbuchproduktion



Nutzung der Metadaten und Volltexte für die Suche

Das ideale Textkorpus

gei.digital

- vollständig oder repräsentativ (relevant und aussagekräftig) ?
- Open Access ☑
- Eindeutige ID ☑
- Versionierung → to do
- klein genug (close reading / intensive Lektüre), ODER groß genug (statistische Analysen) ☑
- originalgetreue Volltexte → to do
- Metadaten korrekt, standardisiert, detailliert ☑ (jedoch immer ausbaufähig)
- Annotation/Vorverarbeitung/Indexierung angemessen → to do
- Verbindung zu passenden digitalen Werkzeugen → to do

- Dokumentation!

gei.digital
Die digitale Schulbuch-Bibliothek

+

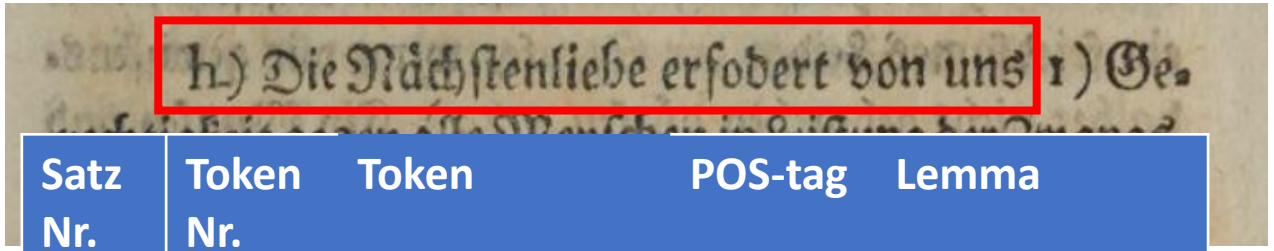
DTA

=>

gei.Digital

Maret Nieländer (GEI)
Christian Scheel (GEI)
Bryan Jurish (BBAW)

GEI-Digital-2020 Korpus: NLP/ Sprachverarbeitung



Satz Nr.	Token Nr.	Token	POS-tag	Lemma
1	1	h	NN	h
1	2	)	$(	)
1	3	Die	ART	d
1	4	Nächstenliebe	NN	Nächstenliebe
1	5	erfo[r]dert	VVFIN	erfordern
1	6	von	APPR	von
1	7	uns	PPER	wir

Wort- und Satzsegmentierung (Tokenisierung):

- ✓ Jedes Token erhält eine adressierbare ID
- ✓ Absolute und relative Häufigkeiten können ermittelt werden

Normalisierung:

- ✓ Suche nach „normalisierter“ Form findet auch historische Formen (z. B. „Tatsachen“ > „Thatsachen“)

Wortartenerkennung (PoS-Tagging):

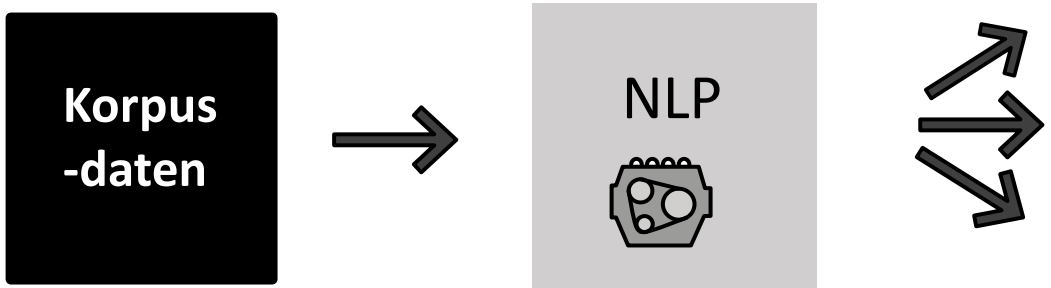
- ✓ Suchen lassen sich eingrenzen auf bestimmte Wortarten

Lemmatisierung:

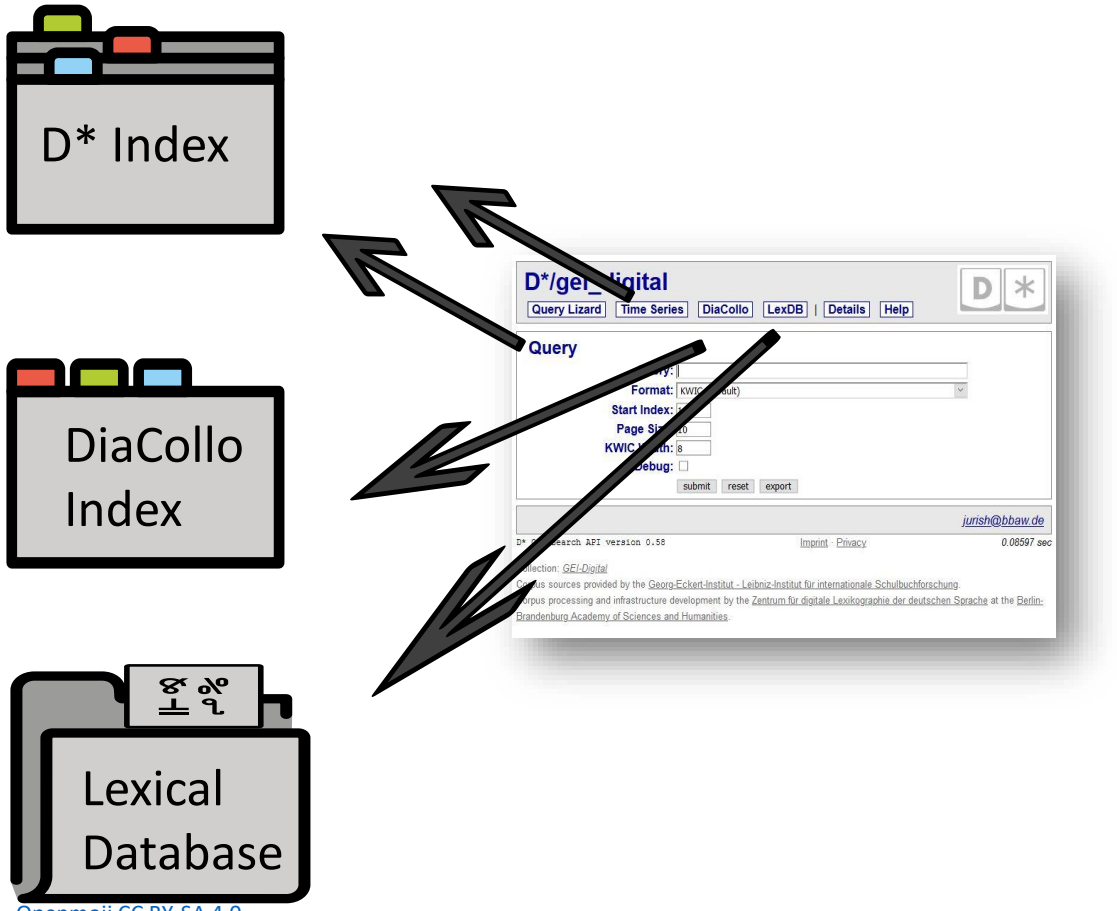
- ✓ Suche nach Lemma findet auch verwandte Formen (z. B. „sehen“ findet „gesehen“, „sah“ usw.)

GEI-Digital-2020 Korpus: Indexierung

Die mit NLP generierten Zusatzinformationen (Token-Attribute) wurden zusammen mit den Daten und Metadaten in verschiedenen Indizes und Datenbanken gespeichert.



Openmoji CC BY-SA 4.0



Openmoji CC BY-SA 4.0

Dort stehen sie für Suchen und Analysen mit unterschiedlichen Werkzeugen zur Verfügung, die auf diese Indizes zugreifen.

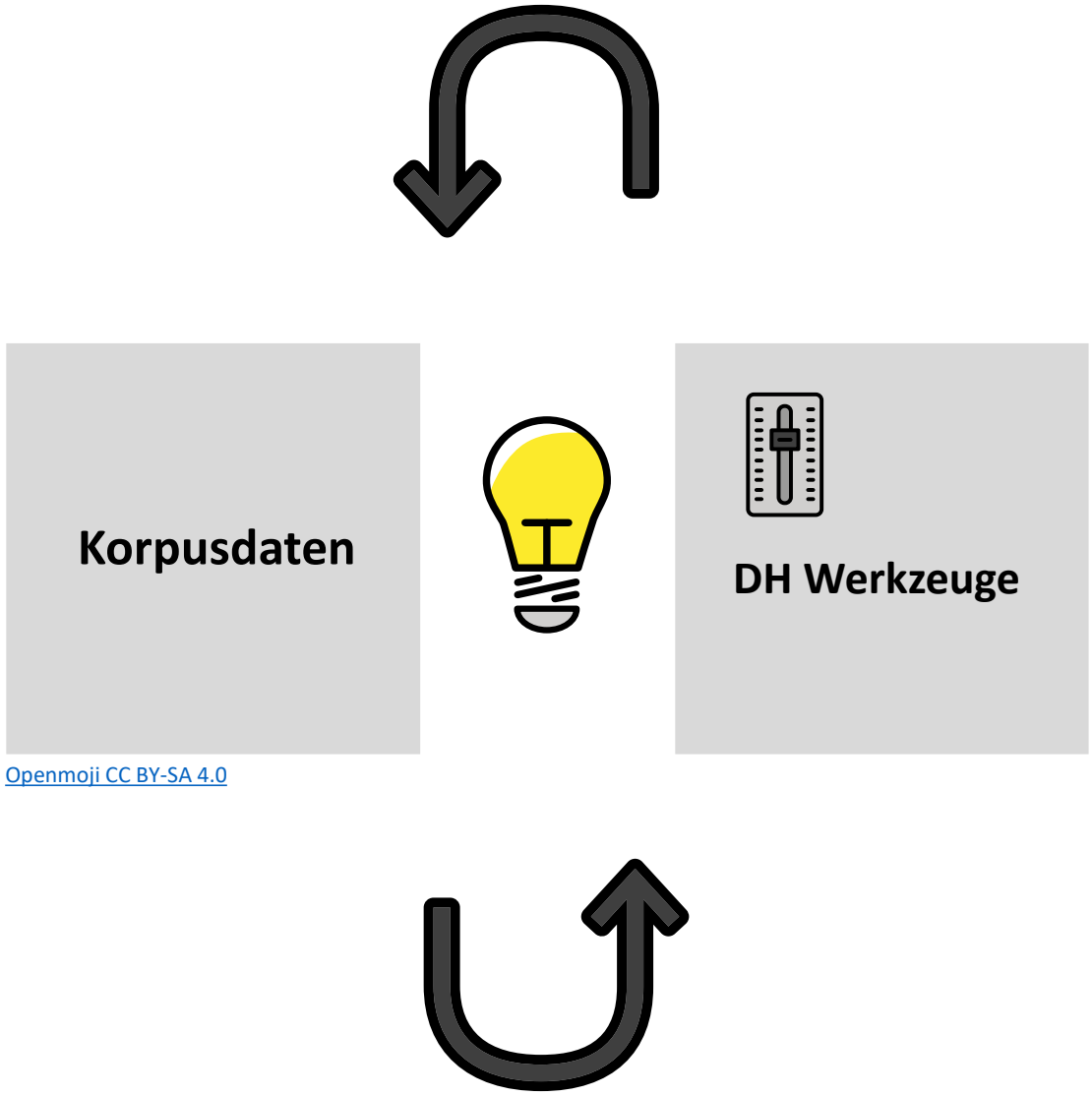
Anwendungsbeispiele: Einfache und komplexe Suchen im Korpus mit D* Query

https://diacollo.gei.de/ → https://diacollo.gei.de/gei-digital-2020/dstar.perl

NEAR(Kind, müssen, 5) #HAS[geiclass, 'Lesebuecher vor 1871'] #asc_date

Diachrone Kollokationsanalysen mit DiaCollo:

https://diacollo.gei.de/ → https://diacollo.gei.de/gei-digital-2020/diacollo/



Openmoji CC BY-SA 4.0

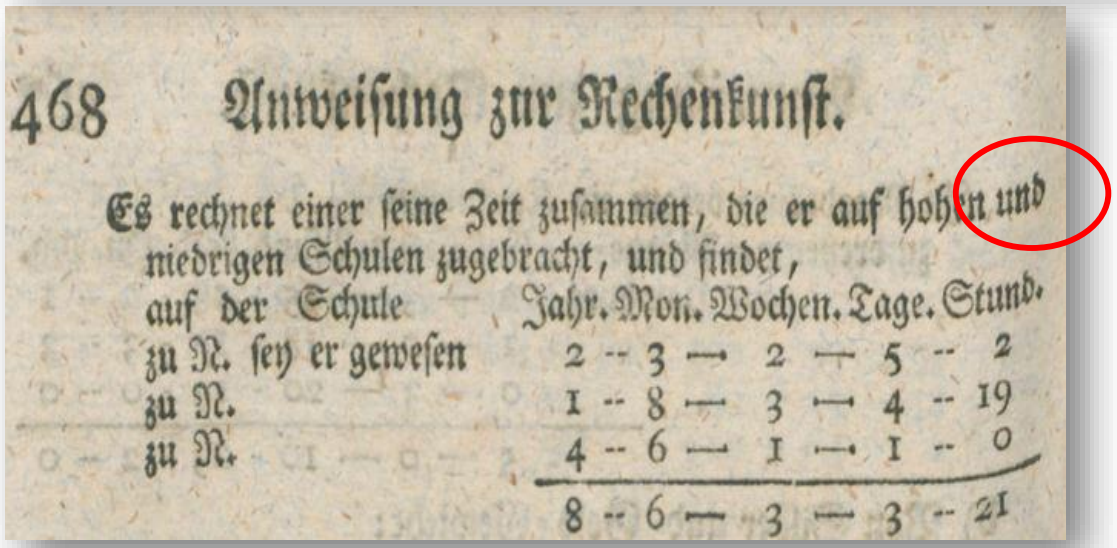
Qualität der Daten:

Fehler in der Volltexterkennung → Fehler bei der Vorverarbeitung → Werkzeuge können keine vollständigen + korrekten Ergebnisse garantieren

Merke:
„Es kann nur gefunden werden, was auch indexiert wurde.“

623310	529	9	2	6.9285	1760	Uiw	NN	KWIC

 468 Anweisung zur Rechenkunst.
 Es rechnet einer leine Zeit jusnnmen, die er auf hohen uiw <br> niedrigen Schulen zugebracht, und findet, <br> auf der Schule, Jahr. Mon. Wochen. Tage. Stuttv'
 zu R. sey er gewesen 2-3 — 2 — 5 -- 2 <br> zu N. i .. g — 3 — 4 -- 19 <br> zu N. 4 -- 6 — j — i - ^
 ' ; 8 - 6 — 3 — 3 - äl



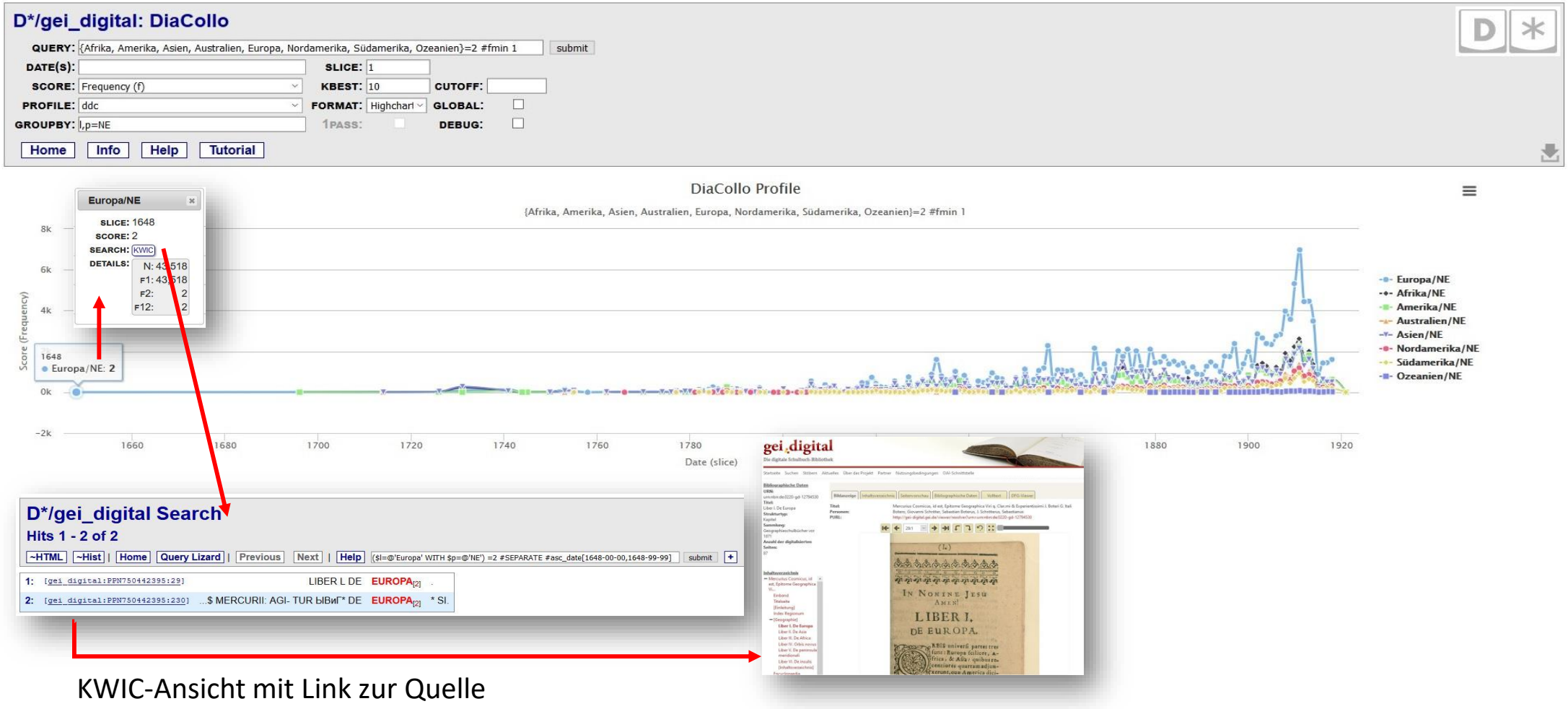
468 Anweiſung zur Rechenkunſt.

Es rechnet einer ſeine Zeit zuſammen, die er auf hohen und niedrigen Schulen zugebracht, und findet,

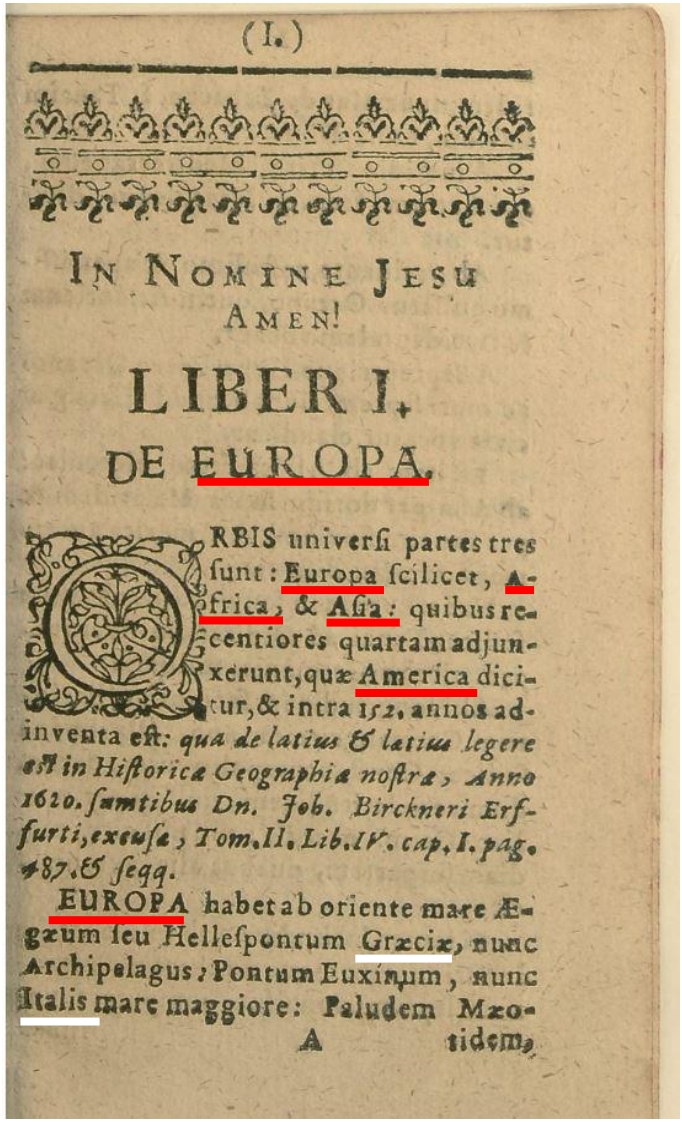
	Jahr.	Mon.	Wochen.	Tage.	Stund.
auf der Schule zu N. ſey er geweſen	2	3	2	5	2
zu N.	1	8	3	4	19
zu N.	4	6	1	1	0
	8	6	3	3	21

http://gei-digital.gei.de/viewer/resolver?urn=urn:nbn:de:0220-gd-11608143

Frequenzvergleich in DiaCollo: Nennungen von Kontinenten



KWIC-Ansicht mit Link zur Quelle



(I.)

IN NOMINE JESU
AMEN!

LIBER I.
DE EUROPA.

RBIS univerſi partes tres ſunt: Europa ſcilicet, Africa, & AGa: quibus recentiores quartam adjunxerunt, quæ America dicitur, & intra 152. annos adinventa eſt: *qua de latius & latius legere eſt in Hiſtorica Geographiæ noſtræ, Anno 1620. ſumtibus Dn. Joh. Birckneri Erfurti, excuſæ, Tom. II. Lib. IV. cap. I. pag. 487. & ſeqq.*

EUROPA habet ab oriente mare Ægæum ſeu Helleſpontum Græcix, nunc Archipelagus; Pontum Euxinum, nunc Italis mare maggiore: Paludem Mæotidem,

A

http://gei-digital.gei.de/viewer/resolver?urn=urn:nbn:de:0220-gd-12794530

 (Ю
o O_O O O O O o oo o
I ^ Nomine Jesu
Am en!
LIBER L
DE EUROPA.
RBIS univcrii partes tres
funt: Europa fcilicet, A,-
frica> &AG‘a: quibus rc-
centiores quartam adjun-
xerunt,quae America dici¬
tur,& intra i/j, anno* ad-
inventa eft; qua ¿e latitu & Utua legere
в Я in Hift or i c л Geograf bi л noftra , y.inno
iCio.fumtibu* Dn. Job. Birckneri Erf-
furti,ex tu fa , Tomoli» Lib.lK, cap<, I. pag,
feqq.
EUROPA habet ab oriente ma-re ЛЕ-
gzum feu Hellefpontum Grzcix; mmc
Archipelagus : Pontum Euxinam , nunc
Stalis шare maggiore: Paludem M*o«
A tidem#

Besonderheiten des Korpus:

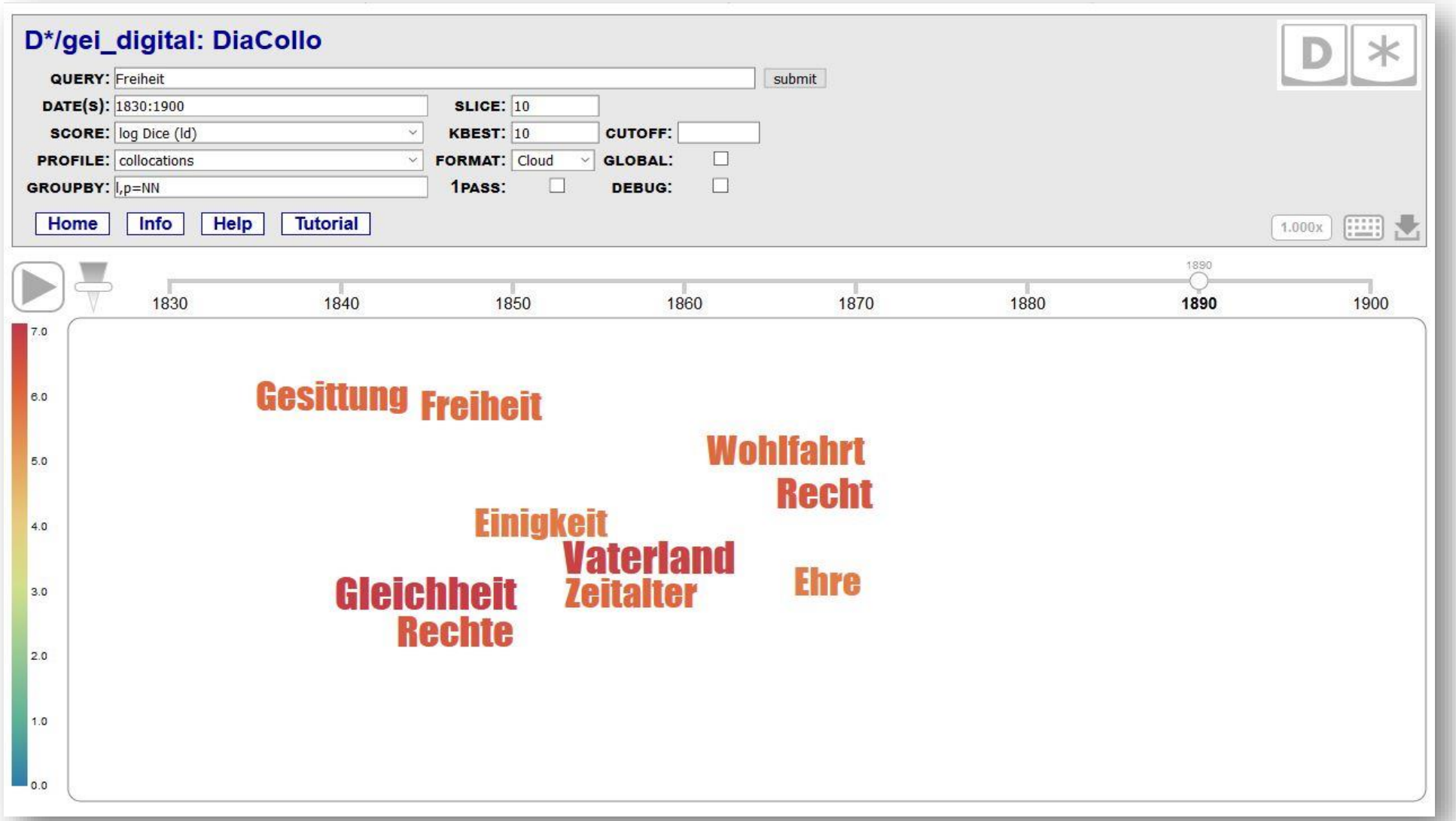
Nutzungskontexte:

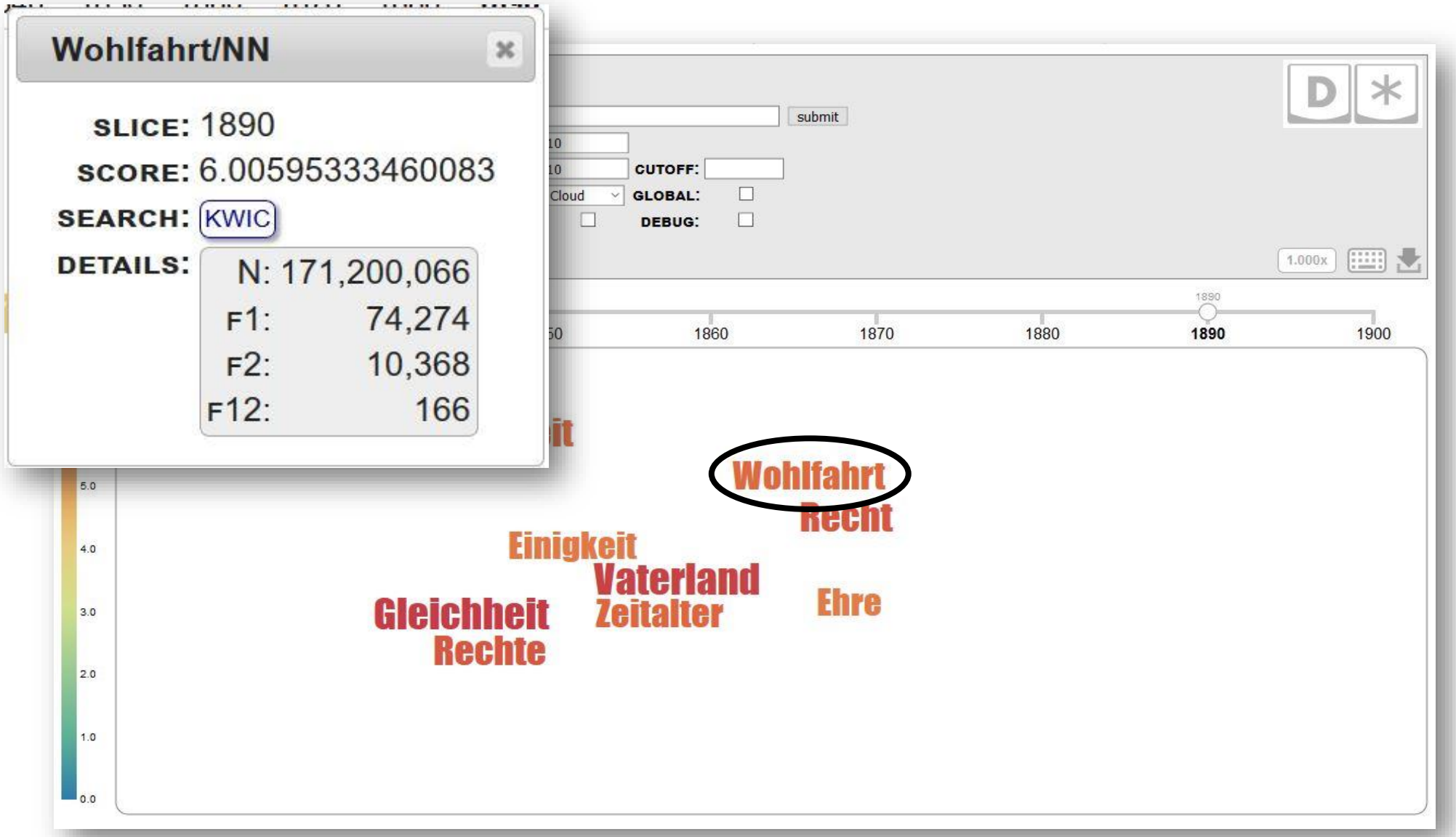
Lehrbücher für verschiedene

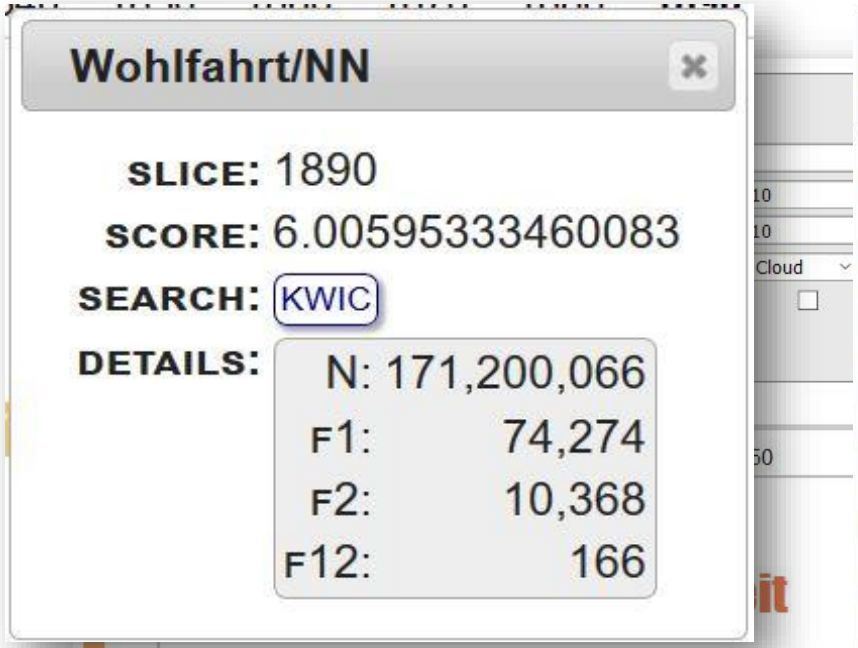
Regionen,
Schulformen,
Schulstufen,
Schulfächer,
Konfessionen,
Gender,
….,

… und teils in verschiedenen Auflagen.

Sprachen: Fremdsprachliche Einschübe/Werke
Typographie: Verschiedene Schriftarten
Layout: Gliederungen, Tabellen, Grafiken, Abbildungen
Autorenschaften: Texte verschiedener Autor:innen
(übernommen, zitiert, adaptiert,…)


D*/gei_digital: DiaCollo
QUERY: Freiheit
submit
DATE(S): 1830:1900
SLICE: 10
SCORE: log Dice (ld)
KBEST: 10
CUTOFF:
PROFILE: collocations
FORMAT: Cloud
GLOBAL:
GROUPBY: l,p=NN
1PASS:
DEBUG:
Home
Info
Help
Tutorial
1.000x
1830
1840
1850
1860
1870
1880
1890
1900
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Gesittung
Freiheit
Wohlfahrt
Recht
Einigkeit
Vaterland
Gleichheit
Zeitalter
Ehre
Rechte


Wohlfahrt/NN
SLICE: 1890
SCORE: 6.00595333460083
SEARCH: KWIC
DETAILS:
N: 171,200,066
F1: 74,274
F2: 10,368
F12: 166
submit
CUTOFF:
GLOBAL:
DEBUG:
1.000x
1860
1870
1880
1890
1900
5.0
4.0
3.0
2.0
1.0
0.0
Wohlfahrt
Recht
Einigkeit
Vaterland
Gleichheit
Zeitalter
Ehre
Rechte

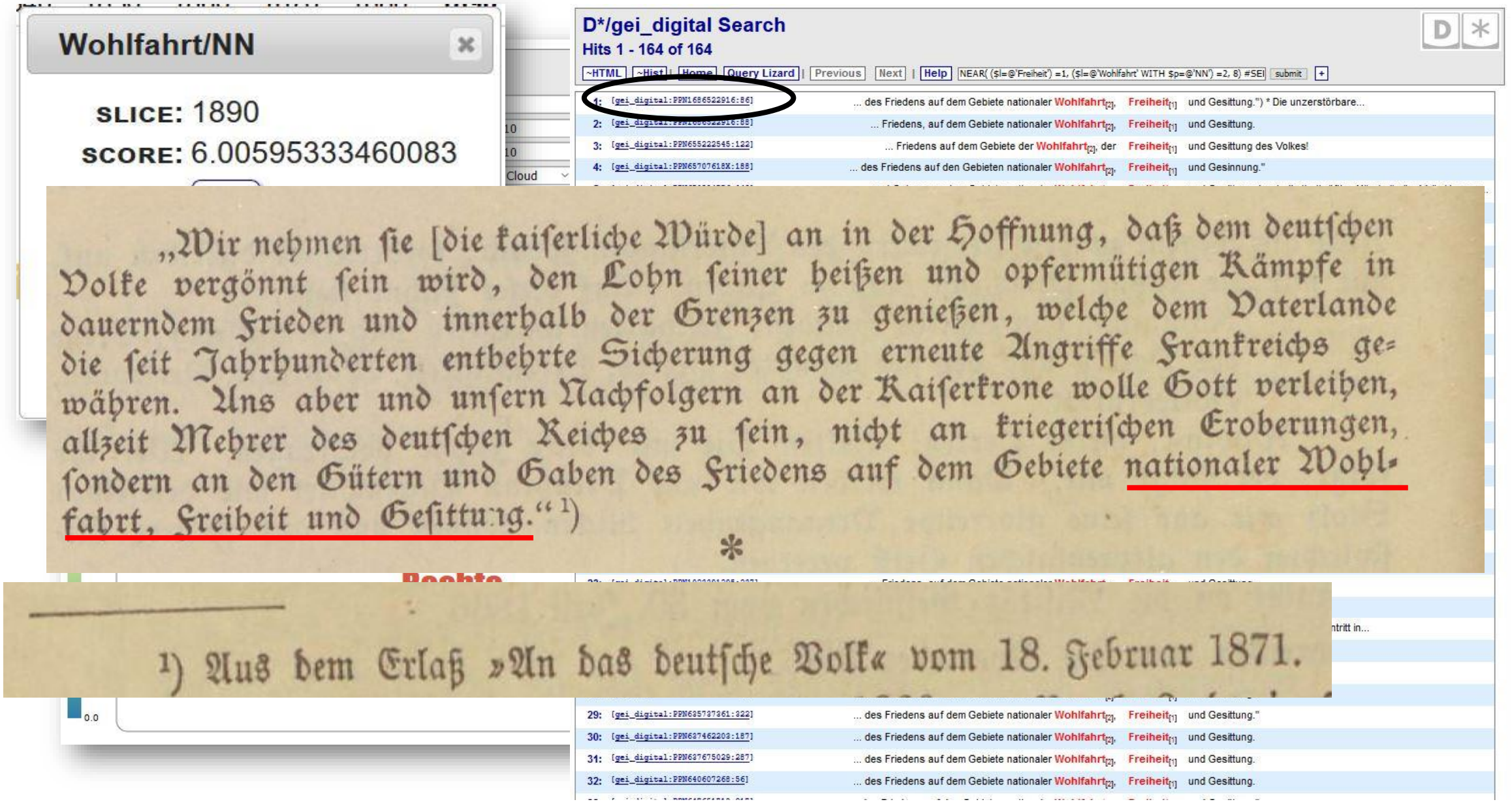


D*/gei_digital Search

Hits 1 - 164 of 164

~HTML | ~Hist | Home | Query Lizard | Previous | Next | Help | NEAR(($l=@'Freiheit') =1, ($l=@'Wohlfahrt' WITH $p=@'NN') =2, 8) #SEI | submit | +

1:	[gei_digital:PPN1686522916:86]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.") * Die unzerstörbare...
2:	[gei_digital:PPN1686522916:88]	... Friedens, auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
3:	[gei_digital:PPN655222545:122]	... Friedens auf dem Gebiete der Wohlfahrt[2], der	Freiheit[1]	und Gesittung des Volkes!
4:	[gei_digital:PPN65707618X:188]	... des Friedens auf den Gebieten nationaler Wohlfahrt[2],	Freiheit[1]	und Gesinnung."
5:	[gei_digital:PPN670594776:342]	... und Gaben ans dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung durch die thatkräftige Mitarbeit aller Volksklassen...
6:	[gei_digital:PPN68056277X:95]	... Gaben Friedens auf deni Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung."
7:	[gei_digital:PPN737822244:233]	... Gedeihen der Saaten, Wohlfahrt[2] und Glück,	Freiheit[1]	und Unabhängigkeit dankte.
8:	[gei_digital:PPN1041407394:451]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
9:	[gei_digital:PPN1046184741:336]	... und auf dem Gebiete nationaler Wohlfahrt[2], der	Freiheit[1]	und Gesittung.
10:	[gei_digital:PPN622812599:348]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Ge-sittung.
11:	[gei_digital:PPN639578136:230]	... Friebens, auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung."
12:	[gei_digital:PPN639578136:241]	... Friedens, ans dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung," das hat er bis...
13:	[gei_digital:PPN640356850:56]	...mlich den Frieden zu wahren und Wohlfahrt[2],	Freiheit[1]	und Ge-sittung zu f$\square$rdern, das...
14:	[gei_digital:PPN653904401:12]	... Friedens, auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
15:	[gei_digital:PPN656483148:197]	... des Friedens aus dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
16:	[gei_digital:PPN664242294:178]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung!
17:	[gei_digital:PPN664244254:140]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung!
18:	[gei_digital:PPN799339857:228]	... des Friedens auf den Gebieten nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung."
19:	[gei_digital:PPN799339857:229]	... des Friedens ans dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung", das hat er in...
20:	[gei_digital:PPN821277286:251]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung," war in den blutigen...
21:	[gei_digital:PPN826843123:537]	... Friedens, auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
22:	[gei_digital:PPN1012659593:471]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
23:	[gei_digital:PPN1023391295:237]	... Friedens, auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
24:	[gei_digital:PPN620767480:191]	... des Friedens auf dem Gebiet nationaler Wohlfahrt[2],	Frei-heit[1]	und Gesittung.
25:	[gei_digital:PPN622888714:191]	... bei streng verfassungsm$\square\square$iger Regierung, Englands Wohlfahrt[2] und	Freiheit[1]	; auch den Juden wurde der Eintritt in...
26:	[gei_digital:PPN622888714:220]	... des Friedens aus dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
27:	[gei_digital:PPN630626928:32]	... des Friedens auf den Gebieten nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung!
28:	[gei_digital:PPN631115951:29]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung."
29:	[gei_digital:PPN635737361:322]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung."
30:	[gei_digital:PPN637462203:187]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
31:	[gei_digital:PPN637675029:287]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.
32:	[gei_digital:PPN640607268:56]	... des Friedens auf dem Gebiete nationaler Wohlfahrt[2],	Freiheit[1]	und Gesittung.


Wohlfahrt/NN
SLICE: 1890
SCORE: 6.00595333460083
D*/gei_digital Search
Hits 1 - 164 of 164
~HTML ~Hist Home Query Lizard Previous Next Help NEAR(($l=@'Freiheit') =1, ($l=@'Wohlfahrt' WITH $p=@'NN') =2, 8) #SEI submit
1: [gei_digital:PPN1686522916:86] ... des Friedens auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung.") * Die unzerstörbare...
2: [gei_digital:PPN1686522916:88] ... Friedens, auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung.
3: [gei_digital:PPN655222545:122] ... Friedens auf dem Gebiete der Wohlfahrt[2], der Freiheit[1] und Gesittung des Volkes!
4: [gei_digital:PPN65707618X:188] ... des Friedens auf den Gebieten nationaler Wohlfahrt[2], Freiheit[1] und Gesinnung."
„Wir nehmen sie [die kaiserliche Würde] an in der Hoffnung, daß dem deutschen Volke vergönnt sein wird, den Lohn seiner heißen und opfermütigen Kämpfe in dauerndem Frieden und innerhalb der Grenzen zu genießen, welche dem Vaterlande die seit Jahrhunderten entbehrte Sicherung gegen erneute Angriffe Frankreichs gewähren. Uns aber und unsern Nachfolgern an der Kaiserkrone wolle Gott verleihen, allzeit Mehrer des deutschen Reiches zu sein, nicht an kriegerischen Eroberungen, sondern an den Gütern und Gaben des Friedens auf dem Gebiete nationaler Wohlfahrt, Freiheit und Gesittung.“ [1])
*
[1]) Aus dem Erlaß »An das deutsche Volk« vom 18. Februar 1871.
29: [gei_digital:PPN635737361:322] ... des Friedens auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung."
30: [gei_digital:PPN637462203:187] ... des Friedens auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung.
31: [gei_digital:PPN637675029:287] ... des Friedens auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung.
32: [gei_digital:PPN640607268:56] ... des Friedens auf dem Gebiete nationaler Wohlfahrt[2], Freiheit[1] und Gesittung.

GEI-Digital-2020: Das ideale Textkorpus?

- vollständig oder repräsentativ (relevant und aussagekräftig) ?
- Open Access ☑
- Eindeutige ID ☑
- klein genug (close reading / intensive Lektüre), ODER groß genug (statistische Analysen) ☑
- originalgetreue Volltexte → to do
- Metadaten korrekt, standardisiert, detailliert ☑ (jedoch immer ausbaufähig)

- Versionierung
- Annotation/Vorverarbeitung/Indexierung angemessen
- Verbindung zu passenden digitalen Werkzeugen
- Dokumentation: 50-Seiten-Tutorial
  Schulungsangebote
  Metadaten als Tabelle
  Visualisierung der Metadaten

gei.Digital

Vielen Dank für Ihre Aufmerksamkeit!
Haben Sie Fragen?

www.diacollo.gei.de

Kontakt:
nielaender@leibniz-gei.de
scheel@leibniz-gei.de

www.gei.de/abteilungen/difi/leistung-im-historischen-schulbuch.html

Clipart auf diesen Folien stammt von
Openmoji CC BY-SA 4.0