Lamento – Latrine – Leyptzigk

Entitäten-basierte Inhaltssuche in verteilten Ressourcen

Thomas Eckart, Felix Helfer, Uwe Kretschmer und Erik Körner

Agenda

1. Kurzvorstellung

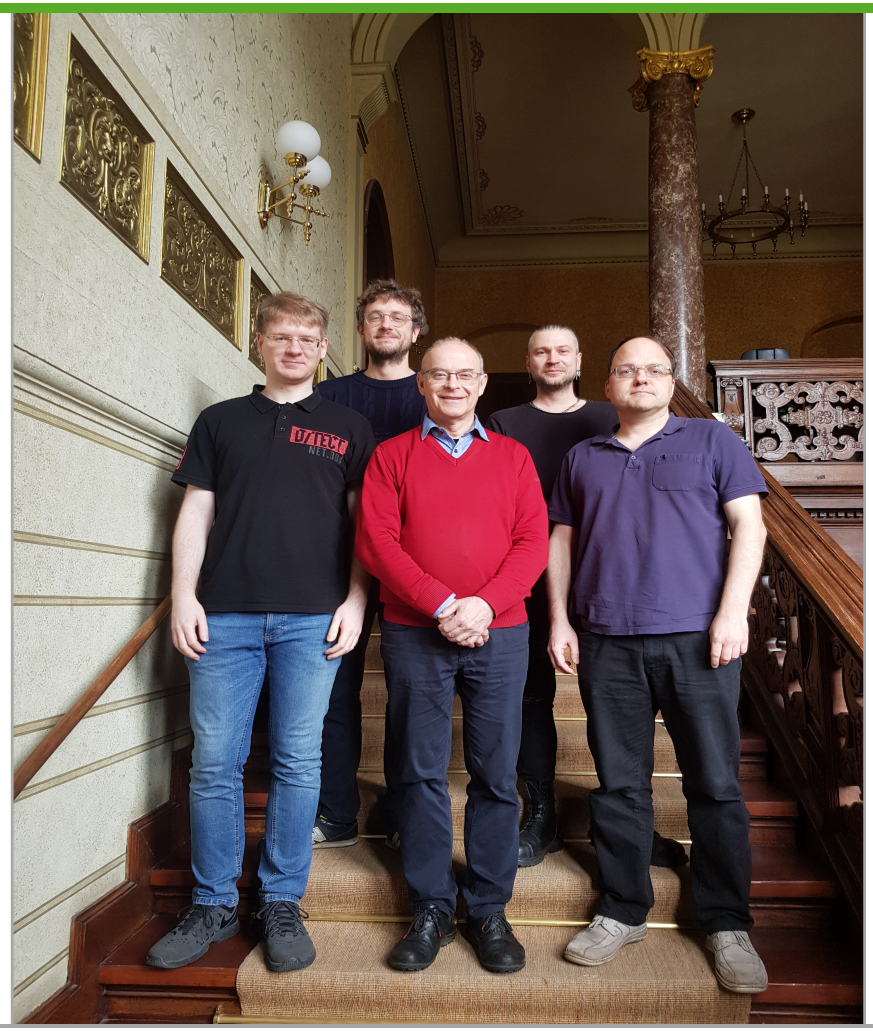
Entitäten-basierte Inhaltssuche

2. Was?
3. Womit?
4. Wie?
5. Wofür?

Foto von Jordan Benton: https://www.pexels.com/de-de/foto/flacher-fokus-der-klaren-sanduhr-1095601/

Das Text+ Team der SAW

- Text+ ist eines der 26 Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI)
- SAW ist partizipierende Einrichtung und tätig in den Task Areas Lexikalische Ressourcen und Infrastruktur / Betrieb
- ein Entwicklungsschwerpunkt ist die föderierte Inhaltssuche (FCS)
- https://www.saw-leipzig.de/de/projekte/nfdi-textplus



Projektteam (v. l.): Erik Körner, Felix Helfer, Gerhard Heyer, Uwe Kretschmer, Thomas Eckart

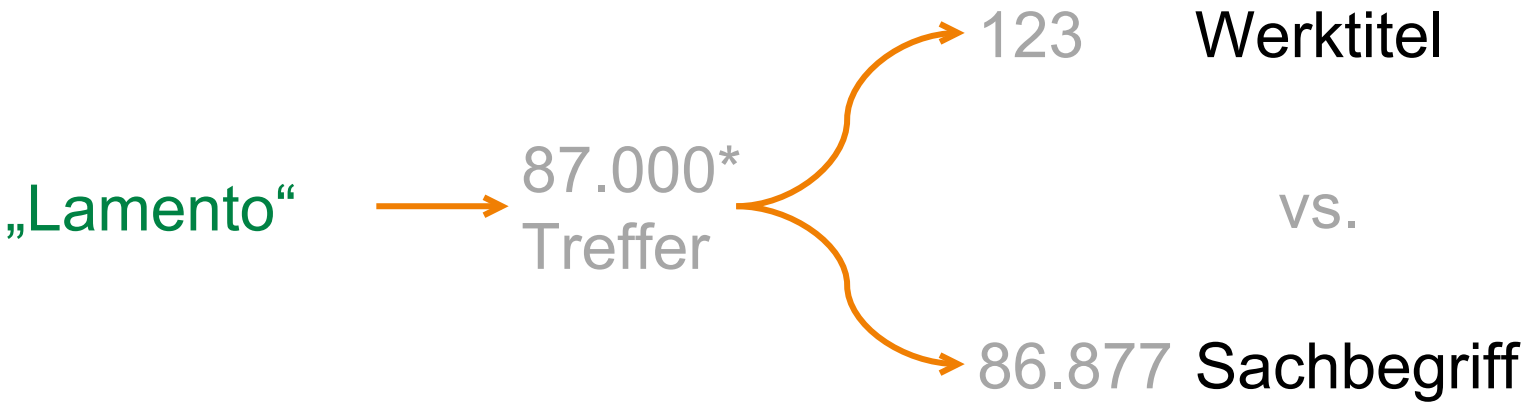
WAS ist das Problem**?**

Beispiel 1: Lamento

„Lamento“ → 87.000 Treffer

Finde Erwähnungen von Rüdiger Blömers „Lamento“ (1997)!

Beispiel 1: Lamento



Finde Erwähnungen von Rüdiger Blömers „Lamento“ (1997)!

- Homonyme
- Disambiguierung
- Spezifizierung

* fiktive Zahlen, zur Verdeutlichung

Beispiel 2: Latrine

„Latrine“

16.259
Treffer

Finde Wörter in deren Definition oder etymologischer Beschreibung „Latrine“ vorkommt!

Beispiel 2: Latrine

„Latrine“

16.259*
Treffer

~~16.245~~ ~~Belegstellen~~

Kein Land zeigte sich bereit, die 60000 dringend benötigten *Latrinen* zu liefern

…Knollenhosen, vor die erheblich gefüllte *Latrine*…

vs.

14 Wörterbucheinträge

Donnerbalken : *NOUN* . *Sitzstange der behelfsmäßigen Latrine*

Latrine : Latrine f. „(behelfsmäßiger) Abort“ wurde im 16. Jh. aus lat. lātrīna, -ae f. „Abort; Kloake“ entlehnt.

Finde Wörter in deren Definition oder etymologischer Beschreibung „Latrine“ vorkommt!

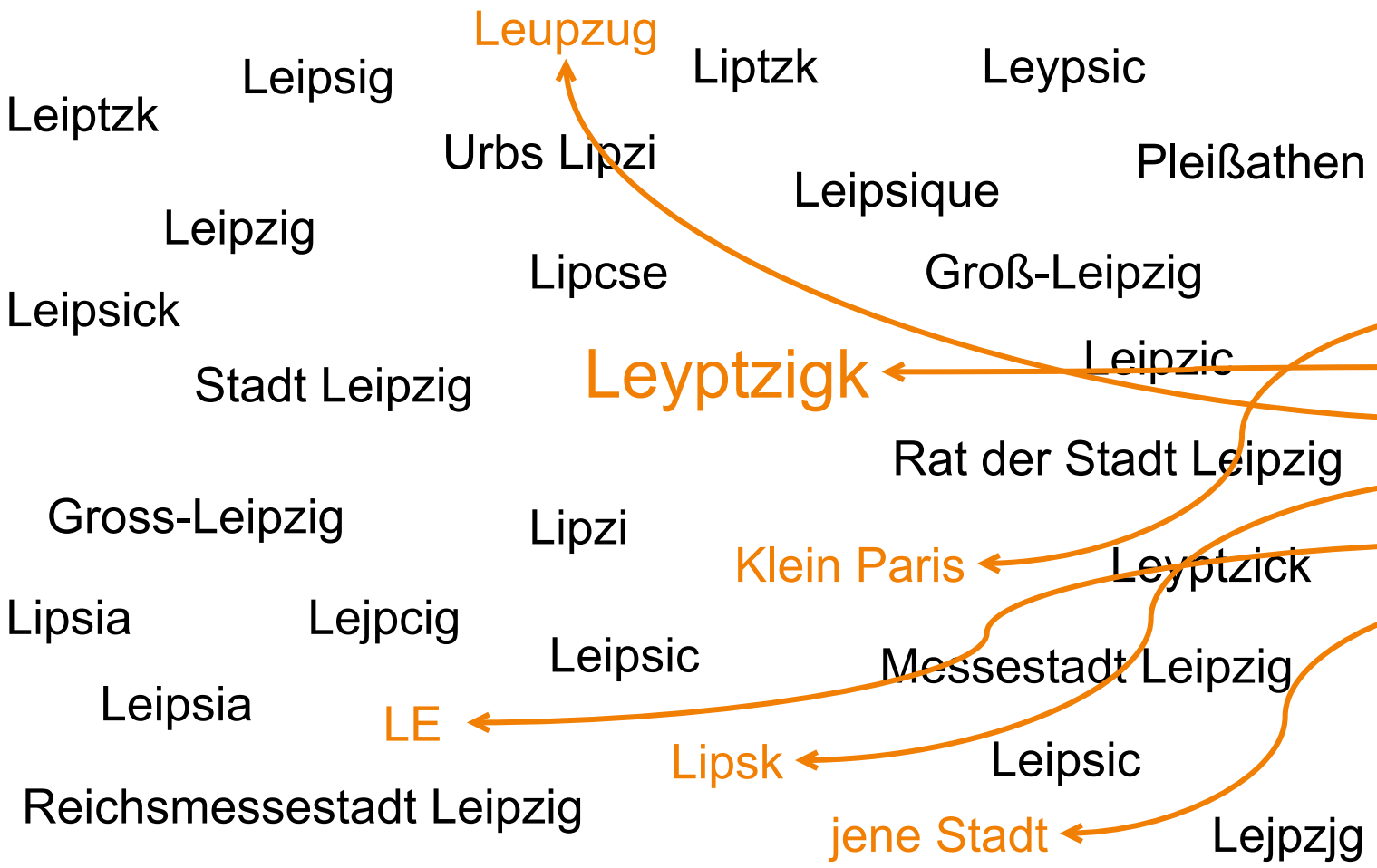
Suche in
- Definitionen
- Synonymen
- Wörterbüchern

* Trefferzahlen über FCS-Demo vom 12.06.2024

Beispiel 3: Leyptzigk

Finde Erwähnungen der Stadt Leipzig!

Beispiel 3: Leyptzigk



Finde Erwähnungen der Stadt Leipzig!

- alternative Bezeichnungen
- historische Schreibweisen
- Schreibfehler
- andere Sprachen
- Abkürzungen
- Relativpronomen

Womit annotieren**?**

Die wunderbare Welt der Entitäten

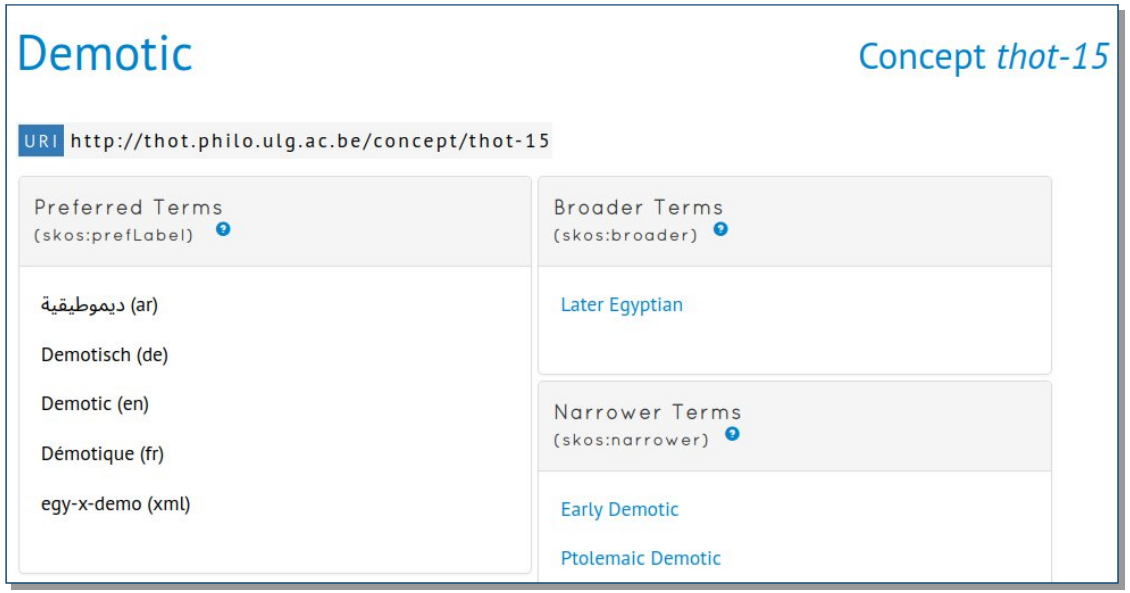
Entität *(Entity)*: *Eine Entität ist ein eindeutig identifizierbares Objekt oder Ding, charakterisiert durch seinen Namen, Typen, Attribute und Beziehungen zu anderen Entitäten.* (nach Balog 2018)

Entitäten

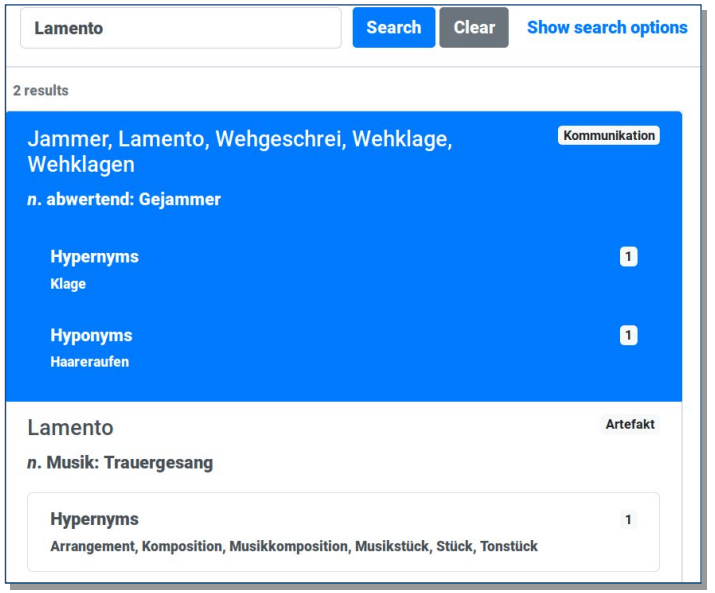
Im Folgenden auch über klassische *Normdaten* hinaus:

- Projektbezogene Normdaten
- Community-getragene Vokabulare und Thesauri
- Lexikalische Bedeutungsdefinitionen (Princeton WordNet, GermaNet, ...)

...



Quelle: Thot - Thesauri & Ontology for documenting Ancient Egyptian Resources



Quelle: GermaNet Rover

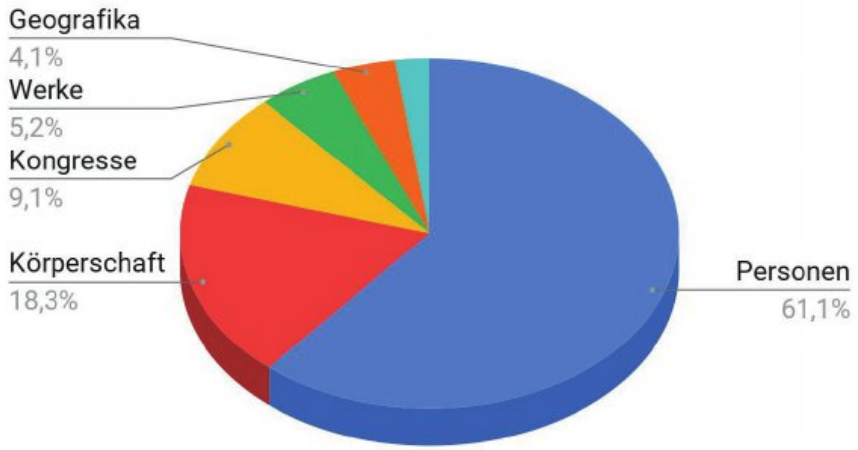
Beispiel: Gemeinsame Normdatei (GND)

- Normdatei für Personen, Organisationen, Geografika, Werke ...
- Verwaltung, Hosting & Pflege u.a. durch die Deutsche Nationalbibliothek DNB
- Kontinuierliche Weiterentwicklung, u.a. über GND-Agenturen

Insgesamt rund 10 Mio. Entitäten

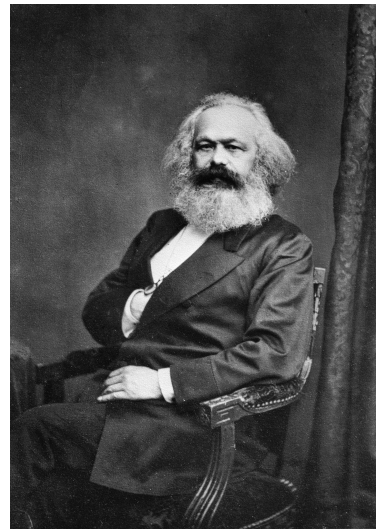
- Personen: 6.000.000
- Körperschaften: 1.800.000
- Kongresse: 890.000
- Werke: 510.000
- Geografika: 400.000
- Sachbegriffe: 220.000

Jährlicher Zuwachs variiert in Abhängigkeit von Projekten



Entitäten nach Typ in der GND, Stand 05/2022
Quelle: DNB / Peter Leinen

GND als Wissensbasis: Karl Marx



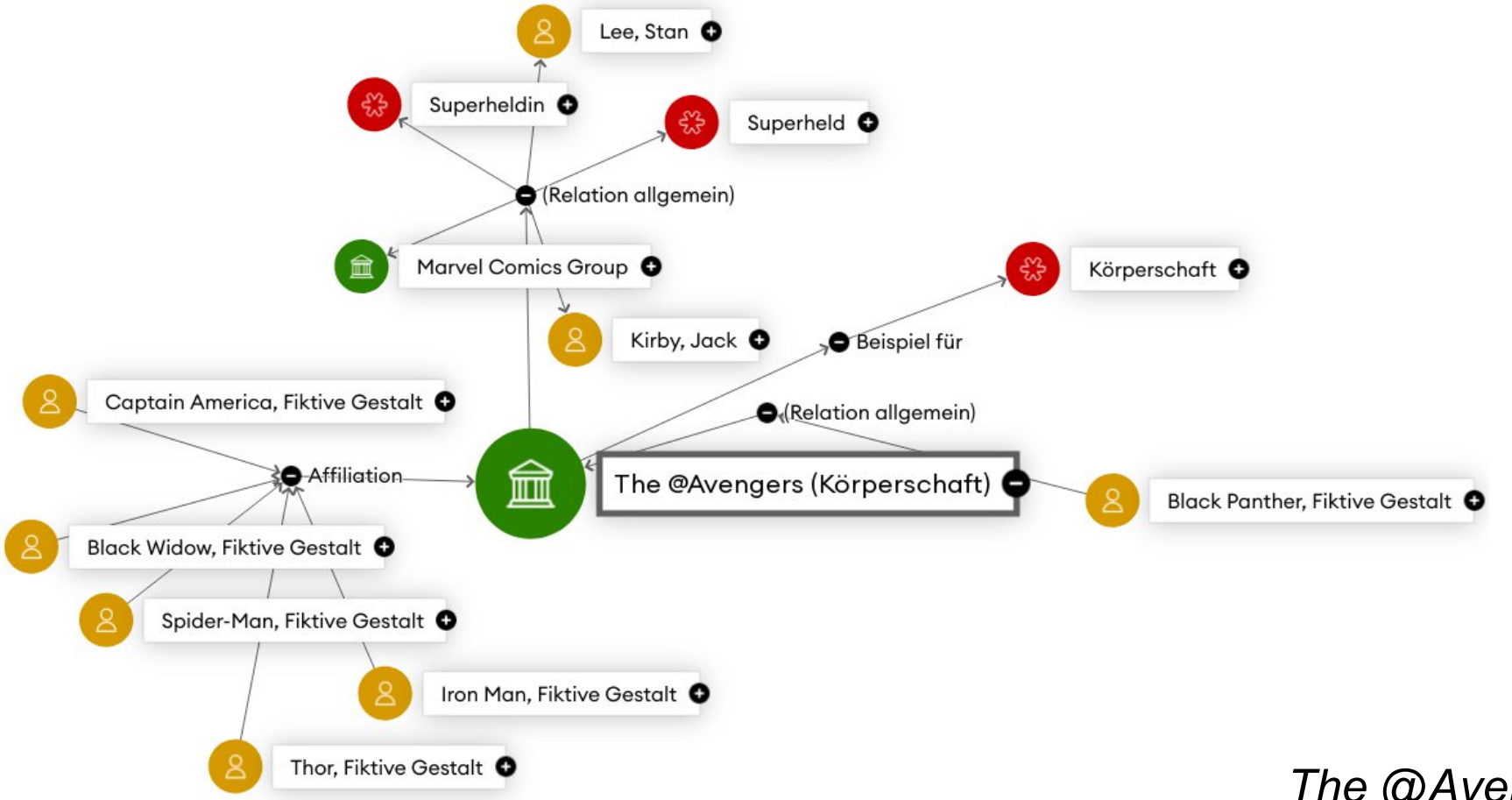
Quelle: Wikimedia Commons

Karl Marx (GND-ID: [118578537](https://explore.gnd.network/gnd/118578537))

„Protagonist der Arbeiterbewegung; Kritiker der bürgerlichen Gesellschaft; Theoretiker des Sozialismus und Kommunismus“

Quelle Wissensgraph: https://explore.gnd.network/gnd/118578537

GND als Wissensbasis: Die Avengers



Quelle: Wikimedia Commons

The @Avengers (GND-ID: [1190237261](https://explore.gnd.network/gnd/1190237261))

Quelle Wissensgraph: https://explore.gnd.network/gnd/1190237261

Wo? Edition „BAKFJ“!¹

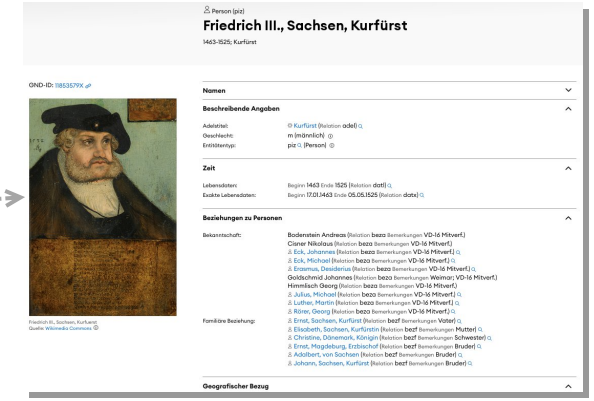
¹ Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johanns des Beständigen 1513 bis 1532

Nr. 125 Kf. Friedrich *an*
Prior [Johann Oertel] des Dominikanerklosters Leipzig

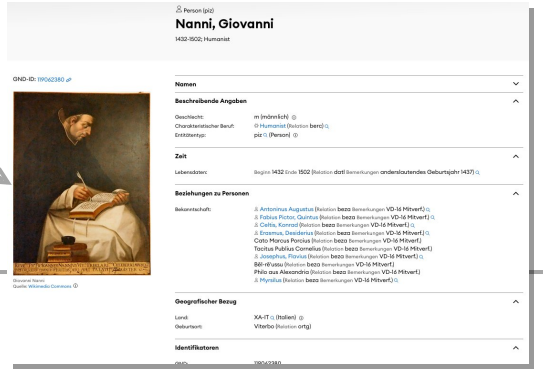
6. April 1514 (Donnerstag nach Judica) · Torgau · Brief · Konzept · deutsch

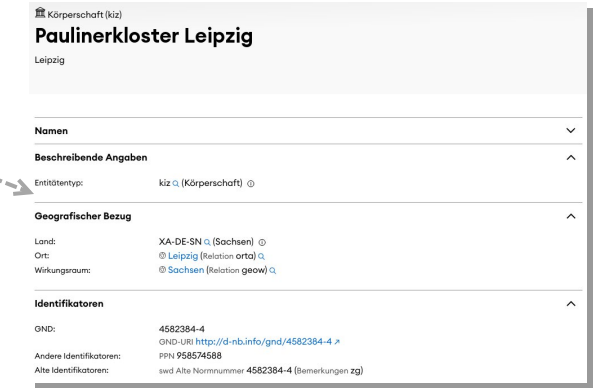
A: LATh – HStA Weimar, EGA, Reg. Kk 747, fol. 1rv (Konzept).

[1] Kf. Friedrich bedankt sich bei dem Prior [Johann Oertel] des Dominikanerklosters St. Paul zu Leipzig für das übersendete Verzeichnis der Klosterbibliothek, um das er gebeten hatte. [2] Darin hat er die „Expositio super apocalypsim“ des Annius von Viterbo entdeckt, die er für die Erarbeitung der gerade entstehenden Chronik benötigt. Um daraus Abschriften anzufertigen, bittet er, dass das Buch einem Terminierer mit einem Begleitschreiben in der kommenden Woche mitgegeben wird. Zur nächsten Leipziger Ostermesse soll das Buch den Mönchen unversehrt zurückgegeben werden. [3] Darüber hinaus sucht der Kf. noch einige Bücher, die in einem beiliegenden Verzeichnis genannt sind.



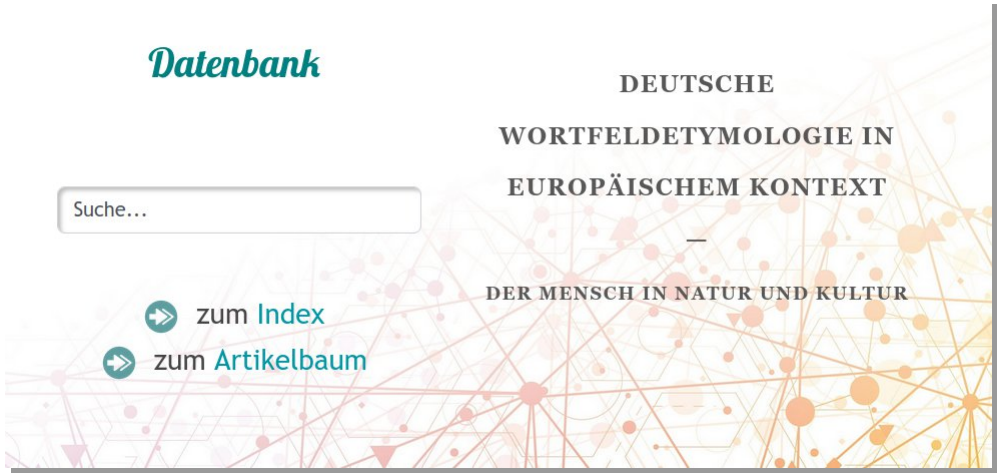
GND GEMEINSAME NORMDATEI





Wo? Wörterbuchprojekt DWEE![1]

[1] Deutsche Wortfeldetymologie in europäischem Kontext



Bedeutung: ***00830365-n***
„the act of dressing and preparing yourself“

Toilette

Toilette f. „Klosett (19. Jh.); Nachttisch, Frisiertisch (18. Jh.); Körperpflege (18. Jh.); elegante Kleidung (18. Jh.)“ ist eine Entlehnung aus frz. *toilette* f. „Tüchlein (zum Einwickeln von Kleidung u.a.) (1352); Gesamtheit der Dinge zum Herrichten (1661); Damenkleidung (1776)“. Das französische Wort ist Deminutiv zu frz. *toile* f. „Tuch“, Fortsetzer von lat. *tēla*, *-ae* f. „Gewebe“, das eine Ableitung zum Verb lat. *texere* „weben, flechten“ ist.

Toilette war im Deutschen wie im zunächst die Bezeichnung für ein Tuch zum Einschlagen von Wäsche oder zum Herrichten des Äußeren. Metonymisch entstand daraus die Bedeutung „Körperpflege“.

Die Bedeutung „Klosett“ entstand im Anschluss an frz. *cabinet de toilette* „Raum für die Körperpflege“ (1740) als verhüllender Ausdruck für „Abort, Klosett“. Im Französischen findet sie sich erst spät (als Plurale tantum): *Il gagna le petit couloir qui desservait la salle de douches, les toilettes et la cuisine* (CAMUS, *Exil et Roy.*, 1957, p. 1648).

Die ältere Bedeutung „Nachttisch, Frisiertisch“ basiert auf französischen Kollokationen wie *coffre de toilette* (1690) und *table de toilette* (1705).

Princeton WordNet

Bedeutung: ***04453410-n***
„a room or building equipped with one or more toilets“

Abort

Die Bezeichnungen für den „Ort zur Verrichtung der Notdurft“ (Pfeifer) sind oft aus Tabugründen verhüllend und werden häufig durch neue Bildungen ersetzt (siehe Kapitel 3.5). Als Erstbeleg für *Abort* m. in der Bedeutung „Toilette“ wird im DWb[2] ein Beleg von 1755 gebucht: *an eenen af-ort gahn* (Richey id. hamb. 3). Das Wort ersetzt dabei älteres *Abtritt*, das insbesondere von der Mitte des 16. bis ins 19. Jh. gebräuchlich war. Die Verwendung nimmt zur Mitte des 19. Jh. zu, allerdings v.a. auf den öffentlichen Raum und damit die Behördensprache bezogen und im nieder- und mitteldeutschen Sprachgebiet (vgl. auch den Erstbeleg, der auf eine Entstehung in diesem Raum verweist). Zum Ende des 20. Jh. wird *Abort* von *Toilette* und *Klo* sowie *WC* (schriftsprachlich) wieder zurückgedrängt. In der Bedeutung „abgelegener Ort“ ist das Wort im Niederdeutschen auch schon früher belegt: *is ist up einem aforde edder lande .. und de möllen kemen to schaden, und de mölenherren fingen nicht to buwende .., dar magh de buwen, wer könde* (Normann rüg. landrecht 108 F; DWb[2] s.v. *Abort*). Benennungsmotiv und zugleich Motiv für die metaphorische Übertragung ist, dass Toiletten früher außerhalb von Wohnungen und Häusern, also weiter entfernt, lagen. Bei *Abort* han-

Klo

Klo n. ist eine Kurzform von *Klosett* n. „Toilettenraum“. Das Wort wurde im 19. Jh. aus engl. *water-closet* *„abgeschlossener Raum mit Wasser“ entlehnt.

Toiletten mit Wasserspülung gab es eigentlich schon bei Römern. Als „Vater des WCs“ gilt aber Sir John Harington, der eine Toilette mit Spülung 1596 in seinem Buch „Metamorphosis of Ajax“ beschrieb. Der Erstbeleg im OED s.v. *water-closet* stammt von 1755 Connoisseur No. 100, *It was always my office ... to attend him in the water-closet when he took a cathartic*. Ein besonders wichtiger Schritt bei der Entwicklung des WCs war die Erfindung des Siphons, des S-förmigen Ausgangs, in dem das frische Wasser einen Geruchsabschluss zum Fallrohr bildete - 1775 als Patent von Alexander Cummings eingereicht. Die Patentschrift zum *water-closet* (18. Jh.) bezog sich im Übrigen auf den Geruchsabschluss durch Wasser (es ist also nicht ein ursprünglich euphemistischer Ausdruck wie Görlach 2001: s.v. *closet* vermutet). In Deutschland setzte sich die Wasserspülung nur langsam durch. Zur Bezeichnung diente *Wasserclosett*, abgekürzt als WC, später mit Vereinfachung des Kompositums, d.h. Wegfall des Bestimmungswort *Klosett* (vgl. Sanders s.v.

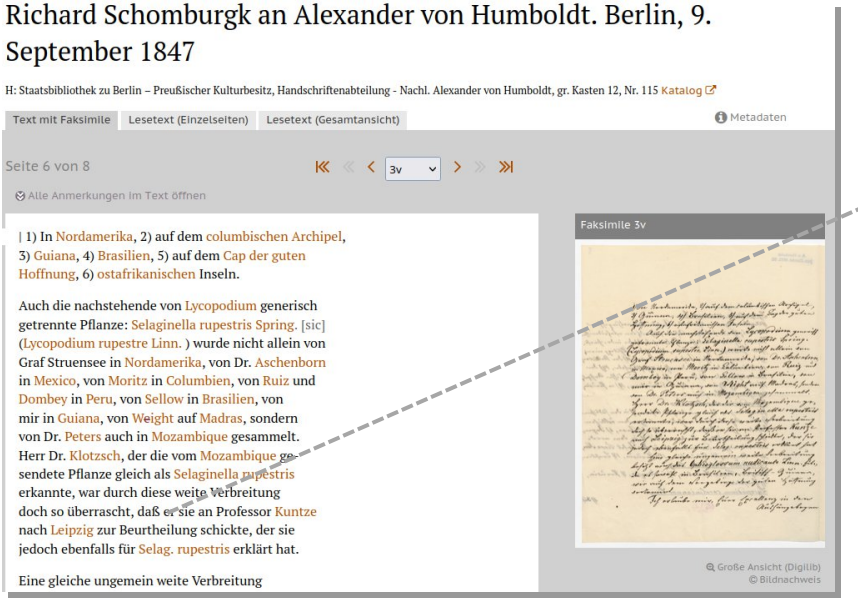
Latrine

Latrine f. „(behelfsmäßiger) Abort“ wurde im 16. Jh. aus lat. *lātrīna*, *-ae* f. „Abort; Kloake“ entlehnt. Das lateinische Wort ist eine Ableitung zum Verb lat. *lavāre* „waschen, baden“.

Kluge, Friedrich 2002: Etymologisches Wörterbuch der deutschen Sprache. Begr. Friedrich Kluge, Bearb. Elmar Seebold. 24., durchges. und erw. Auflage. Berlin u.a.: de Gruyter, s.v. *Latrine*.
Pfeifer, Wolfgang (Hg.) 1993: Etymologisches Wörterbuch des Deutschen. 2 Bde. 2., durchges. u. erg. Aufl. Berlin: Akad. Verl., s.v. *Latrine*.

Autorin: Bettina Bock

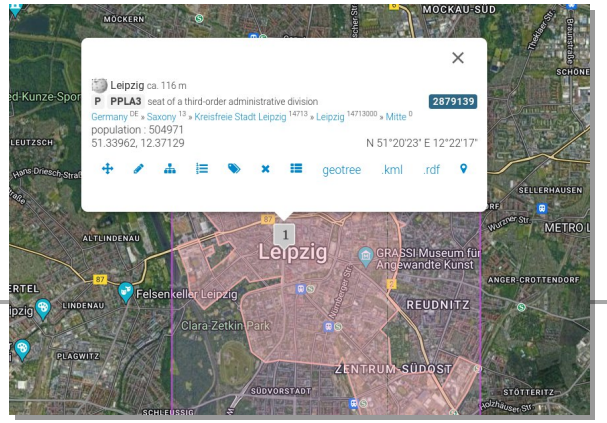
Wo? edition humboldt digital!

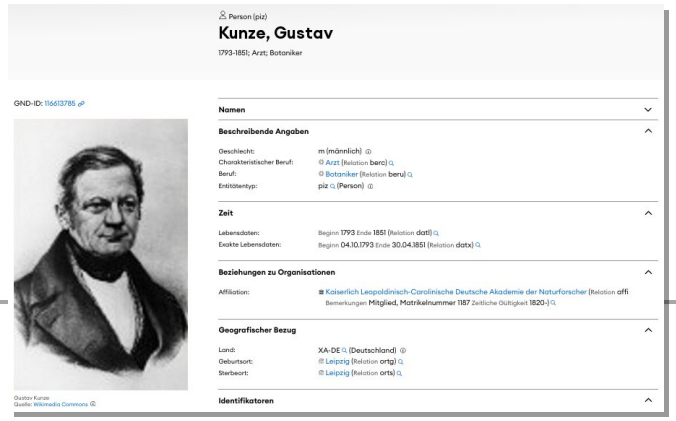



```
doch so überrascht, daß er sie an Professor
<persName ref="https://editon-humboldt.de/H0020225 https://d-nb.info/gnd/116613785">Kuntze</persName>
<lb/>
nach
<placeName ref="https://editon-humboldt.de/H0011410 https://www.geonames.org/2879139">Leipzig</placeName>
zur Beurtheilung schickte, der sie
<lb/>
jedoch ebenfalls für
```

GeoNames

GND GEMEINSAME NORMDATEI





Wie (automatisch) annotieren**?**

Was ist Entity Linking?

*[...] Auch **Ferdinand Schmidt** war 1846 Mitbegründer eines „Verein zum Wohle der arbeitenden Klassen“. [...]*

Bildquellen:Wikipedia (gemeinfrei) und Wikimedia Commons

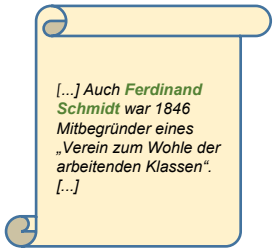
„***Named Entity Linking*** *(EL) is the task of resolving named entity mentions to entries in a structured* ***Knowledge Base*** *(KB).“*

Hachey et al. 2013: Evaluating Entity Linking with Wikipedia

Subtasks des Entity Linkings

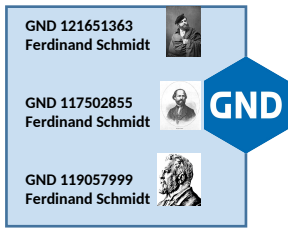
Entity Linking (EL)

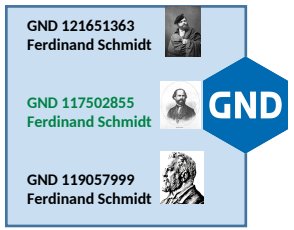
= Named Entity Recognition (NER) + Entity Disambiguation (ED)



Entity Disambiguation (ED)

= Candidate Search + Candidate Disambiguation





Aktuelle Forschung

- Modelltraining meist **überwacht**
  - Große Anzahl an Zielklassen
  - Große Mengen annotierter Trainingsdaten nötig
  - Trainingsdaten dadurch oft automatisch aus Wikipedia abgeleitet

- **Aber:** nur selten **einheitliche** Experimente, u.a. bzgl.:
  - Trainingsdaten
  - genutztem Entitätenvokabular
  - zusätzlichen Signale (Kandidatenlisten)

- Nur wenige „**Meta-Benchmarks**“ bisher, wie [*Milich and Akbik, 2023*]

Aktuelle Ansätze

- Auch hier inzwischen große Fortschritte durch **Large Language Models** (LLMs).

- Erster Schritt oft **Embedding** von Entity-Mention, Textkontext und (optional) zusätzlichen Informationen.

- Danach **Decoding**, z.B.:
  - Generative (Seq2seq) [*De Cao et al., 2021*]
  - Sequence Classification [*Ravi et al., 2021*]
  - Candidate Scoring with KB facts [*Ayoola et al., 2022*]
  - ...

EL für deutschsprachige Daten?

- Relativ **spärlich**, egal ob für Daten, Werkzeuge oder Benchmarks...
- Ausgewählte existierende **Werkzeuge**:
  - **entity-fishing**
  - **OpenTapioca**
  - **DBpedia Spotlight**

Evaluation EL-Tools für deutschsprachige Daten

Model	F1	P	R	TP	FP	FN
Spacyfishing	0.28	0.351	0.233	234	432	772
DBpedia Spotlight	0.29	0.192	0.557	675	2837	536
OpenTapioca	0.421	0.407	0.437	497	724	641
HIPE Task 2022 team2[1]	0.464	0.462	0.466	535	623	612

[1] Beste Scores für Task End-to-end EL hipe2020 German relaxed @1 (literal sense) von Team „L3i“ von La Rochelle University, La Rochelle, Frankreich, https://hipe-eval.github.io/HIPE-2022/results#hipe-2022-track-evaluation-results

Schwarz, Pia, Barth, Florian: „Classification and Linking of Named Entites“, Beitrag in:
Pollin, Christopher, et al. “Workshop generative KI, LLMs und GPT bei digitalen Editionen”. Zenodo, March 29, 2024. https://doi.org/10.5281/zenodo.10893761.

– Evaluation auf **HIPE 2020** Testdaten (historische Zeitungen!)

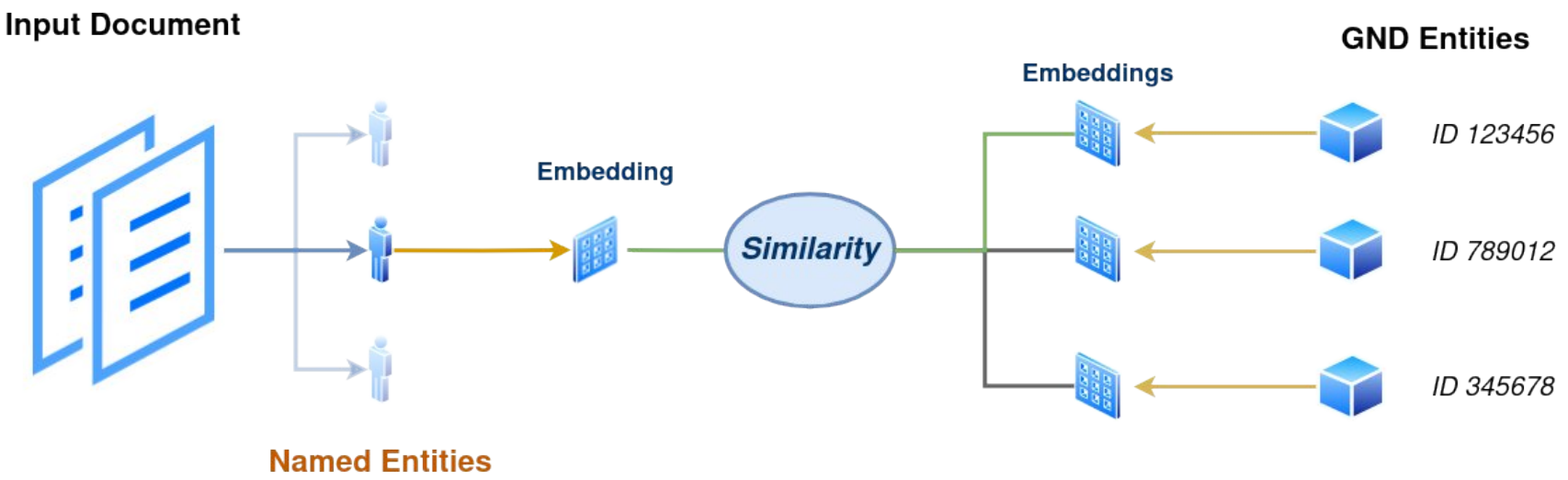
– Alle Modelle nutzen ***spaCy German NER*** (*de_core_news_lg*).

– Evaluiert nach „**fuzzy regime**“: TP auch wenn Tokengrenzen *nicht* komplett stimmen

– Für DBpedia Spotlight: **Interlinking** von Q-items und DBpedia-Entität

Experimente: GND-Entity-Embeddings

- In-House-Experimente mit **GND-Entity-Embeddings** basierend auf Wikipedia-Artikeln.



Experimente: GND-Entity-Embeddings II

- Erster **End-to-End-Prototyp** existiert.
- Aber: in aktueller Form noch **ausbaufähig**:
  - Noch begrenzte Anzahl an Embeddings
  - Weitere Embedding-Ansätze erproben (bisher: fastText)
  - Candidate Search!
- Weitere **Testdatensätze** nötig... ⟶ tlw. *selbst erstellen!*
- Und: *hardwarehungrig*! Bei großen KBs jedoch kaum vermeidbar...

Experimente: Relationsextraktion

- GND-Relationen als Vorgabe einer zu extrahierenden **Mikrostruktur**:

“Auf der anderen Seite gibt es zunehmend Solisten und Musiker in kleinen Ensembles wie Howard Alden, John Pizzarelli und sein Sohn Bucky Pizzarelli, die bei der Vorliebe für siebensaitige Archtops an George Van Eps anknüpfen.”

Bucky Pizzarelli (GND 121085007) **Sohn** (GND rel:bezf „Sohn“) → John Pizzarelli (GND 134681231)

- Initial durchgeführt als Masterarbeit von Marc Richter[1], nutzt **aufwendige Pipeline** zur Erkennung von Entitäten, Erzeugung und Filterung von Relations-Tripeln.

- Angewandt auf Textkorpora des **Wortschatz Leipzig.**

[1] *„Korpusbasierte Relationsextraktion und Normdaten-Disambiguierung zur Identifizierung von Belegstellen am Beispiel des GND-Netzwerkes“*

Experimente: Relationsextraktion II

Ausgewählte Zahlen aus der Arbeit von M. Richter:

Tabelle 6.11: Korpora-Statistiken

Korpus	Anzahl extrahierte RE-Triples	Anzahl Textstellen-Belege
News	599.435	21.600
Web	443.561	21.947
Wikipedia	1.073.149	56.325
Insgesamt	2.116.145	99.872

Tabelle 6.12: Häufigste extrahierte Relationen in den Korpora

News	Web	Wikipedia
associatedPlace (1.195)	placeOfBusiness (1.672)	placeOfBirth (3.559)
placeOfBusiness (1.013)	associatedPlace (1.647)	dateOfDeath (3.506)
placeOfBirth (1.011)	placeOfBirth (1.413)	dateOfBirth (3.409)
isPartOf (787)	dateOfBirth (1.143)	associatedPlace (3.040)
dateOfBirth (524)	isPartOf (1.109)	placeOfBusiness (2.577)
dateOfDeath (241)	dateOfDeath (462)	isPartOf (2.322)

Vorteil: höhere **Interpretierbarkeit** der Annotationen als bei rein numerischen Konfidenzwerten!

Fazit Entity Linking

- Weiterhin **schwieriges Problem** des NLP - in Durchführbarkeit wie auch Vergleichbarkeit.
- Insbesondere im Deutschen noch großer **Mangel** an verlässlichen Werkzeugen und Testdatensätzen.
- Aber: auch **lohnenswert** zu lösenden Problem…

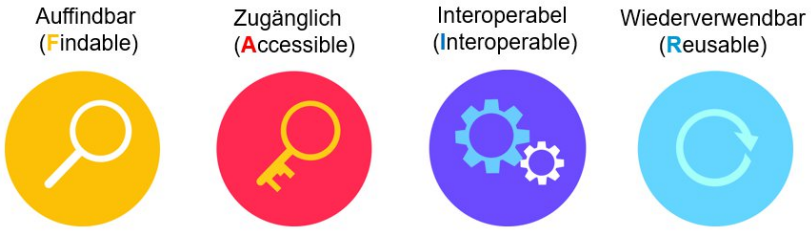
… denn der **Nutzen** von annotierten Daten ist groß!

Wofür ist das alles gut**?**

Text+ @ SAW

- **Daten** und **Verfahren** — was nun?
- Bekannte Probleme?
  - Verteilte Daten — Zugriff für und Austausch mit anderen?
  - Verschiedene Strukturen / Typen und Umfang von Annotationen
- Wichtig: **FAIR**e Nutzung

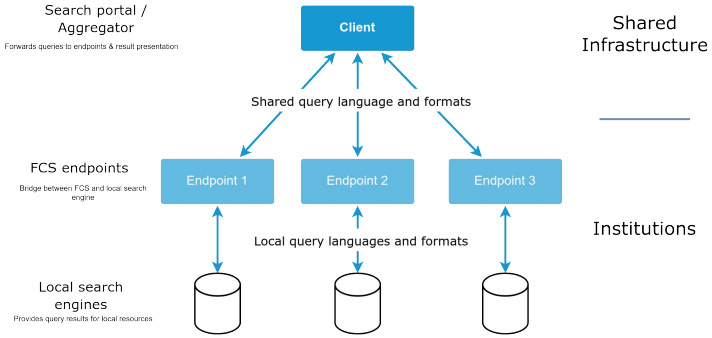
➜ **NFDI — Text+**



Quelle: forschungsdaten.info

Federated Content Search (FCS)

- Spezifikation und Suchplattform für verteilte Ressourcen
- National und EU (aus CLARIN)



Quelle: CLARIN.eu

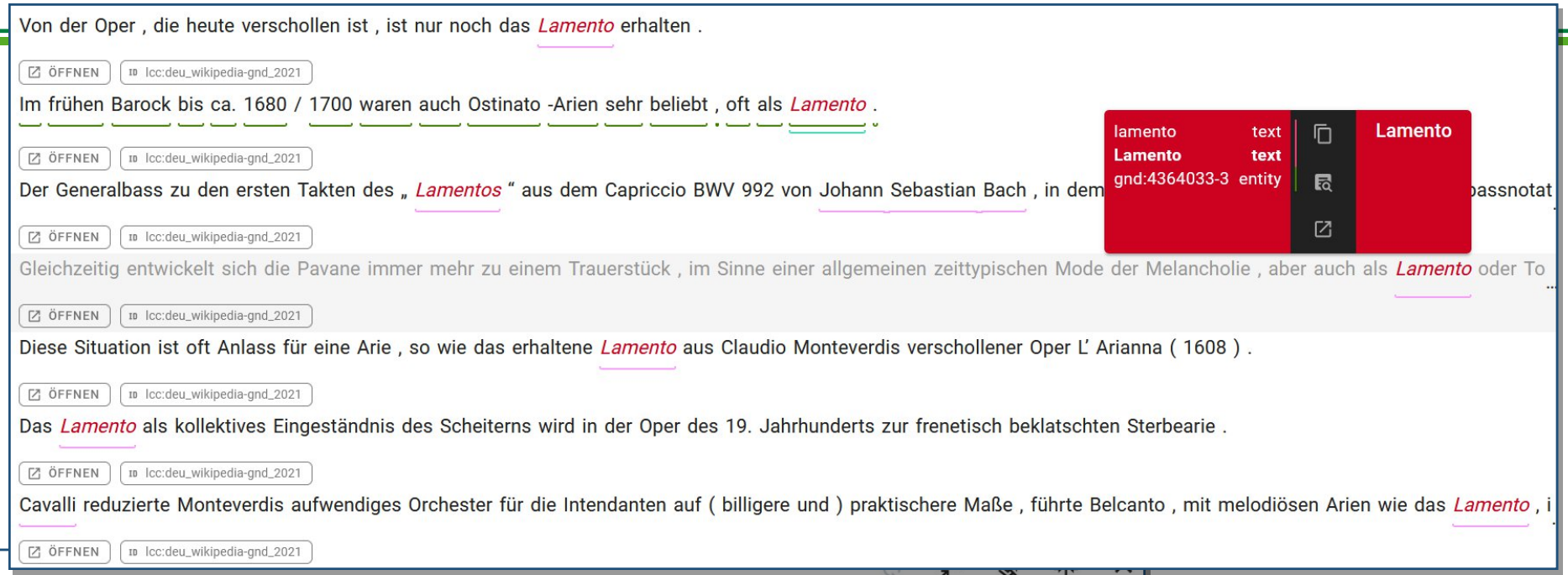
Besonderheiten

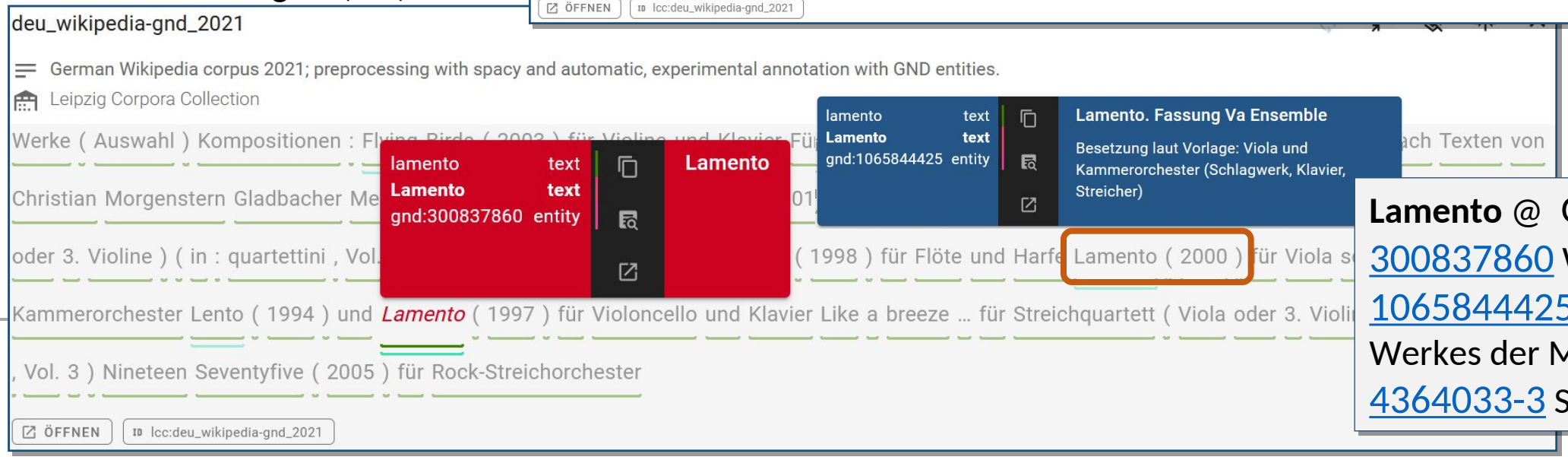
- Suche auf (räumlich) verteilte Ressourcen
- Inhaltssuche, keine Metadatensuche (vgl. OPAC, …)
- Suche auf annotierten Daten, nicht nur Volltext à la Google
- Suche mit Text, linguistischen Mustern oder Entitäten

Beispiele – Lamento

Disambiguierung

- [entity="gnd:300837860"]
  - Werk der Musik (wim)
  - „Lamento“ (1997) von Rüdiger Blömer
- [entity="gnd:1065844425"]
  - „Lamento“ (2000) R. B.
- [entity="gnd:4364033-3"]
  - Sachbegriff (saz)





Lamento @ GND-Explorer
300837860 Werk der Musik (wim)
1065844425 Fassung eines Werkes der Musik (wif)
4364033-3 Sachbegriff (saz)

Beispiele – Latrine

Suche in Lexikalischen Ressourcen, verschiedene Definitionen

- „Latrine“ in LRs
  - Ergebnisse in 8 Ressourcen
- senseRef="pwn:04453410-n"
  - Princeton-WordNet; 2 LRs
- oder Volltext „Latrine“ …

Semantic Field Etymology in German and in European Context

"Semantic Field Etymology in German and in European Context" has been sponsored by the Saxon Academy of Sciences and Humanities in Leipzig since April 2007 and thus aligns with the series of dictionary projects, which traditionally form a special focus in regard to the academy's research activities. The implementation of the long-term project is carried out by scientific employees as well as by team members on the basis of a work contract. Under the direction of Prof. Dr. Rosemarie Lühr, the project is seated at the seminar for Indo-European studies at the Friedrich Schiller University Jena.

Saxon Academy of Sciences and Humanities in Leipzig

Abort : Die Bezeichnungen für den „Ort zur Verrichtung der Notdurft“ (Pfeifer) sind oft aus Tabugründen verhüllend und werden häufig durch neue Bildungen ersetzt (siehe Kapitel 3.5). Als Erstbeleg für Abort m. in der Bedeutung „Toilette“ wird im DWb2 ein Beleg von 1755 gebucht: an eenen af-ort gahn (Richey...

ÖFFNEN Abort

Klo : Klo n. ist eine Kurzform von Klosett n. „Toilettenraum“. Das Wort wurde im 19. Jh. aus engl. water-closet *„abgeschlossener Raum mit Wasser“ entlehnt. Toiletten mit Wasserspülung gab es eigentlich schon bei Römern. Als „Vater des WCs“ gilt aber Sir John Harington, der eine Toilette mit Spülung 1596...

ÖFFNEN Klo

Klosett : Klo n. ist eine Kurzform von Klosett n. „Toilettenraum“. Das Wort wurde im 19. Jh. aus engl. water-closet *„abgeschlossener Raum mit Wasser“ entlehnt. Toiletten mit Wasserspülung gab es eigentlich schon bei Römern. Als „Vater des WCs“ gilt aber Sir John Harington, der eine Toilette mit Spülung 1596...

ÖFFNEN Klosett

Latrine : Latrine f. „(behelfsmäßiger) Abort“ wurde im 16. Jh. aus lat. lātrīna, -ae f. „Abort; Kloake“ entlehnt. Das lateinische Wort ist eine Ableitung zum Verb lat. lavāre „waschen, baden“. Kluge, Friedrich 2002: Etymologisches Wörterbuch der deutschen Sprache. Begr. Friedrich Kluge, Bearb. Elmar Seebold...

ÖFFNEN Latrine

Wortschatz Leipzig German

Wortschatz Leipzig German Resource

Saxon Academy of Sciences and Humanities in Leipzig

Toilette (*NOUN*): *Die Toilette //, auch Klosett, Abort, Latrine, Null-Null, WC oder Lokus ist eine sanitäre Vorrichtung zur Aufnahme von Körperausscheidungen. Daneben wird der Raum, in dem sich eine solche Vorrichtung befindet, ebenfalls Toilette genannt. Eine Toilette dient einer umfassenderen Nutzung als das lediglich zur Abführung von Urin errichtete Urinal.*

A. Muka: Dictionary of the Lower Sorbian Language and Its

Lemma běžadło | die Rinne [der/die/das Rinne] ; der Rinnstein [der/di
сточный камень ; сточная канавка

Digital Dictionary of the German Language – Berlin-Brandenburg

Lemma Latrine : POS NOUN . [salopp, abwertend] Def. Gerücht

Lemma Latrine : POS NOUN . Def. behelfsmäßiger, einfacher Abort, besc

Lemma Kübel : POS NOUN . [spezieller] Def. hauptsächlich in Haftanstal

Lemma Donnerbalken : POS NOUN . [veraltend, Soldatensprache, derb] Def. Sitzstange der behelfsmäßigen Latrine

Ihre Suche nach **Latrine** erzielte **16259 Ergebnisse** (206 geladen) in **35 Ressourcen**

Beispiele – Leyptzigk

Leipzig – Schreibweisen

- [entity="gnd:4035206-7" & word != "Leipzig"]
- Varianten: (Stadt) Leipzig, Lipsk, Liptzigk, Leiptzk, Leyptzigk, Leipziger, …

bakfj_vol1

24 weitere Ergebnisse

Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johanns des Beständigen 1513 bis 1532

Leipzig Corpora Collection

[1] Der *Leipziger* Dominikaner und Konventslesemeister Marcus von Weida erklärt , dass er für Kf …

ÖFFNEN bakfj:1791

… Feilitzsch und Heinrich vom Ende sollen mit Hz . Georg von Sachsen wegen des *Leipziger* Briefgewölbes verhandeln .

ÖFFNEN bakfj:2216

Zur nächsten *Leipziger* Ostermesse soll das Buch den Mönchen unversehrt zurückgegeben werden .

ÖFFNEN bakfj:1763

… r abte zu Salfeldt und Kemnitz , dergleichen auch der probist zu sant Thomas in *Liptzigk* zu conservatoren und handhabern solchs privilegii von babstlicher heiligkeit gege …

ÖFFNEN bakfj:1978

… , weren wir bdacht , uns des selbten auch zuhalten und hierauff den probste zu *Liptzigk* , deme dan , alß wir warhafftig bericht worden , solch privilegien kundigk , auch …

ÖFFNEN bakfj:1978

[4] Nachdem aber gmelter probst zu *Liptzigk* in des durchlauchtigen hochgbornen fursten und herren herzcog Georgen , eurer …

ÖFFNEN bakfj:1978

… urfft gnedige furschrifft und bevelch an offtgmelten probsten zu sant Thomas in *Liptzigk* gnediglich widerfarn lassen und ime entpfehlen , das er als conservator solchs b …

ÖFFNEN bakfj:1978

… en , so weyt umbfarn solten und sonderlich wenn die tag kurtz sein geladen gen *Leiptzk* nicht farn konten und alßdann untderwegen vorgebene unkost zcu unsers armen …

ÖFFNEN bakfj:2615

deu_wikipedia-gnd_2021

German Wikipedia corpus 2021; preprocessing with spacy and automatic, experimental annotation with GND entities.

Leipzig Corpora Collection

Bach sah sich daher gezwungen , in einer Eingabe an den Rat der *Stadt* Leipzig vom …

ÖFFNEN lcc:deu_wikipedia-gnd_2021

Am 24. Oktober 1730 schickte Zedler dem Rat der *Stadt* Leipzig den Vorabdruck des geplanten Titelblattes , diesmal jedoch ohne Privilegantr…

ÖFFNEN lcc:deu_wikipedia-gnd_2021

S. 17. sind inzwischen durch Belege aus dem Leichenbuch und dem Grabregister der *Stadt* Leipzig widerlegt .

ÖFFNEN lcc:deu_wikipedia-gnd_2021

…w.leipzig.de/de/buerger/freizeit/leipzig/baum/strasse/allg/03447.shtml Website der *Stadt* Leipzig , Leipzigs Stadtgri

ÖFFNEN lcc:deu_wikipedia-gnd_2021

…mein akzeptiert ist die Etymologie des Ortsnamens Leipzig als vom sorbischen Wort *Lipsk* kommend (gleichlautend

ÖFFNEN lcc:deu_wikipedia-gnd_2021

in : Enno Bünz (Hrsg.): Geschichte der *Stadt* Leipzig .

ÖFFNEN lcc:deu_wikipedia-gnd_2021

…-und-verwaltung/stadtrat/fraktionen / Fraktionszusammensetzung] auf der Seite der *Stadt* Leipzig

ÖFFNEN lcc:deu_wikipedia-gnd_2021

…ahl des Leipziger Oberbürgermeisters am 02.02.2020 (1. Wahlgang)] , Webseite der *Stadt* Leipzig

ÖFFNEN lcc:deu_wikipedia-gnd_2021

leipzig text
Leipzig **text**
gnd:4035206-7 entity

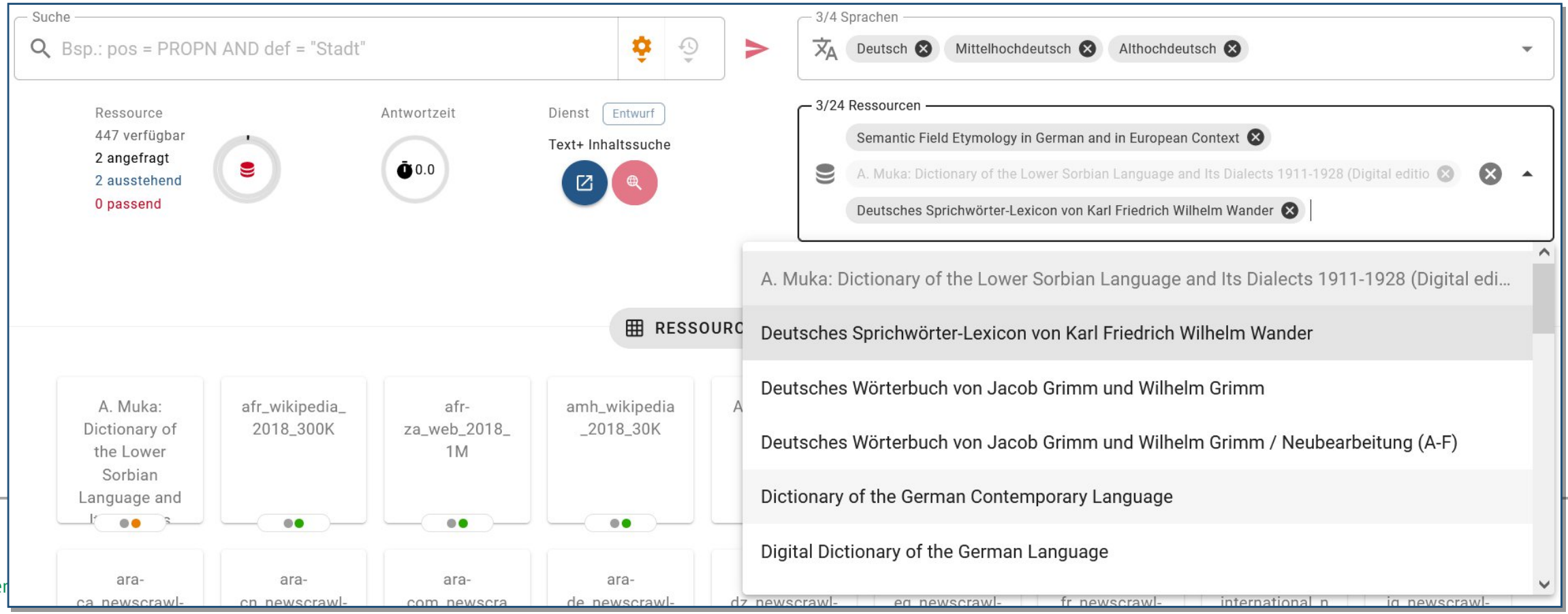
Leipzig

Stadt in der Leipziger Tieflandsbucht, aus seit 7./8. Jh. bestehender slaw. Siedlung entstanden, 1015 u. 1050 urkundl. erwähnt, um 1165 Stadtrecht; Verwaltungssitz des gleichnamigen Regierungsbezirks; 1937-1945 Ehrentitel "Reichsmessestadt"

Leipzig @ GND-Explorer 4035206-7

Demo – FCS Benutzeroberfläche

- Filterung relevanter Ressourcen über Facetten (Sprache, Name)
- Vorschlag von Beispielanfragen



Vielen Dank für Ihre Aufmerksamkeit!

www.saw-leipzig.de

Credits

Nach- und Hinweise

Das Titelbild wurde generiert mit Adobe Firefly (Firefly Image 3-Modell).

Die vorliegende Publikation wurde im Rahmen des Konsortiums Text+ im Kontext der Arbeit des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. verfasst. NFDI wird von der Bundesrepublik Deutschland und den 16 Bundesländern finanziert, und das Konsortium Text+ wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 460033370.

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

Literatur

Balog, Krisztian: „Entity-Oriented Search“, The Information Retrieval Series (INRE, volume 39), Springer Cham, 2018.

Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johanns des Beständigen 1513 bis 1532. Reformation im Kontext frühneuzeitlicher Staatswerdung. Online-Edition: https://bakfj.saw-leipzig.de

Deutsche Wortfeldetymologie in Europäischem Kontext (DWEE). https://dwee.saw-leipzig.de

Edition Humboldt digital. https://edition-humboldt.de

Hachey et al.: Evaluating Entity Linking with Wikipedia, Artificial Intelligence, Volume 194, 2013, Pages 130-150, ISSN 0004-3702, https://doi.org/10.1016/j.artint.2012.04.005.

Marcel Milich and Alan Akbik: ZELDA: A Comprehensive Benchmark for Supervised Entity Disambiguation. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 2023, pages 2061–2072, Dubrovnik, Croatia. Association for Computational Linguistics.

De Cao et al.: Editing Factual Knowledge in Language Models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pages 6491–6506, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ravi et al.: CHOLAN: A Modular Approach for Neural Entity Linking on Wikipedia and Wikidata. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021, pages 504–514, Online. Association for Computational Linguistics.

Ayoola et al.: Improving Entity Disambiguation by Reasoning over a Knowledge Base. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022, pages 2899–2912, Seattle, United States. Association for Computational Linguistics.

Schwarz, Pia, Barth, Florian: „Classification and Linking of Named Entites“, Beitrag in: Pollin, Christopher, et al. “Workshop generative KI, LLMs und GPT bei digitalen Editionen”. Zenodo, March 29, 2024. https://doi.org/10.5281/zenodo.10893761.

M. Richter, „Korpusbasierte Relationsextraktion und Normdaten-Disambiguierung zur Identifizierung von Belegstellen am Beispiel des GND-Netzwerkes“, Masterarbeit, Institut für Informatik, Universität Leipzig, 2023.