

Carolin Odebrecht

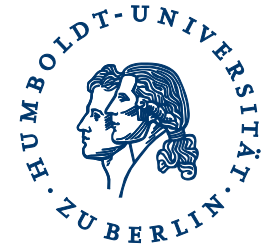
Korpuslinguistik | Humboldt-Universität zu Berlin

Vortragsreihe „Digital Humanities - Theorie und Methodik“

Universität Leipzig

ZUGANG ZU UND ZUGRIFF AUF DIGITALE FORSCHUNGSDATEN LAUDATIO UND ANNIS

Zugang und Zugriff



I. Zugang zu Korpora

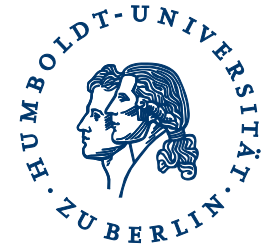
- Suche nach Korpora
- Forschungsdatenrepositorium LAUDATIO
- Funktionen von Metadaten
- Wiederverwendung von Korpora

II. Zugriff auf Korpora

- Suche in Korpora
- Such- und Visualisierungstool ANNIS
- Anfragesprache & Annotationsvisualisierung
- Analyse von Korpora

Teil I

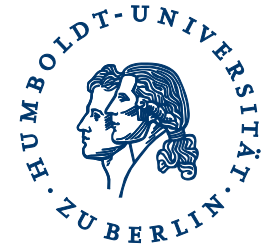
Zugang zu Korpora



- Wie kann ein Forschungsdatenrepositorium eine Menge von unterschiedlichen (textbasierten historischen) Korpusdaten beschreiben und zur Verfügung stellen,
 - um das Verständnis von **Konzepten Dritter**,
 - und eine **Auswahl für die eigene Forschung**
 - für den Zweck der **Wiederverwendung** von Korpora zu ermöglichen?
- Beispiele für Repositorien für Forschungsdaten
 - spezielle Korpora z.B. FREEbank (Salmon-Alt et al. 2006), Textgrid (Neuroth et al. 2015), Deutsches Textarchiv (Geyken & Gloning 2015)
 - für unterschiedliche Tools, Forschungsdaten und Metadaten z.B. Hamburger Zentrum für Sprachkorpora Repository <https://corpora.uni-hamburg.de/drupal/de/islandora/object/islandora%3Aroot>, Virtual Language Observatory <https://vlo.clarin.eu/;jsessionid=648DD491C3ED98FF53660F26D6D4C639?0>

Teil I

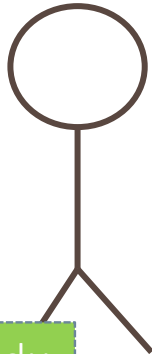
Zugang zu Korpora



- Welches Wissen über ein Korpus ist notwendig, um es wiederverwenden zu können,
 - um Korpora zu **analysieren**
 - um Korpora neu zu **kompilieren**
 - um Korpora mit **weiteren Annotationen** zu versehen?
- Funktionen von Metadaten
 - vgl. u.a. Haynes (2004), Miller (2011), Odebrecht & Krause (2013)
 - u.a. zu Lebenszyklus von Forschungsdaten (vgl. z.B. Einführung dazu Rümpel 2011)

Zugang

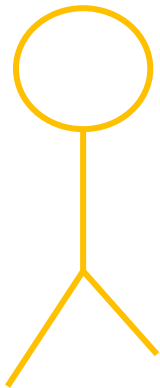
Ich suche historische Briefe aus dem Hochdeutschen!



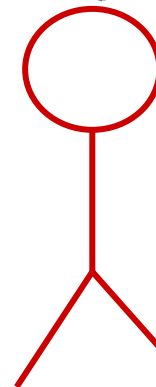
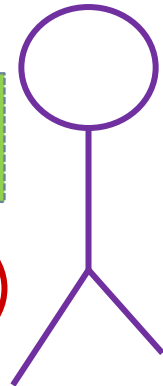
Wenn **Bücher** gemeint sind, dann...

Ich suche normalisierte Texte aus dem 17. Jahrhundert!

Was verstehst du unter Normalisierung?



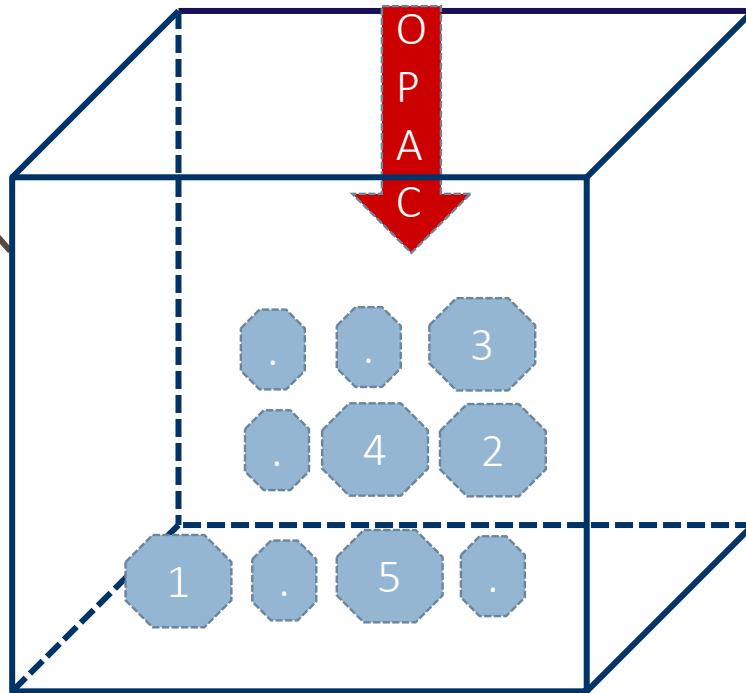
Ähm, ich kenne...



Zugang zu Büchern

Ich suche historische Briefe aus dem Hochdeutschen!

Gebäude + Regale + Bücher + OPAC



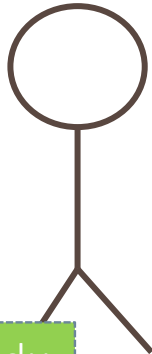
Ich suche normalisierte Texte aus dem 17. Jahrhundert!

Ähm, ich kenne...

Was verstehst du unter Normalisierung?

Zugang

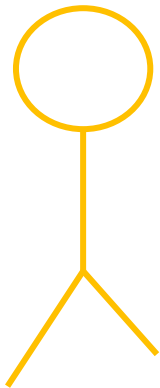
Ich suche historische Briefe aus dem Hochdeutschen!



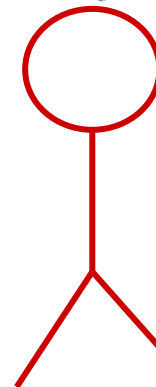
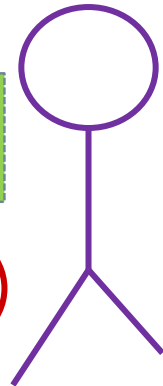
Wenn **Korpora** gemeint sind, dann...

Ich suche normalisierte Texte aus dem 17. Jahrhundert!

Was verstehst du unter Normalisierung?



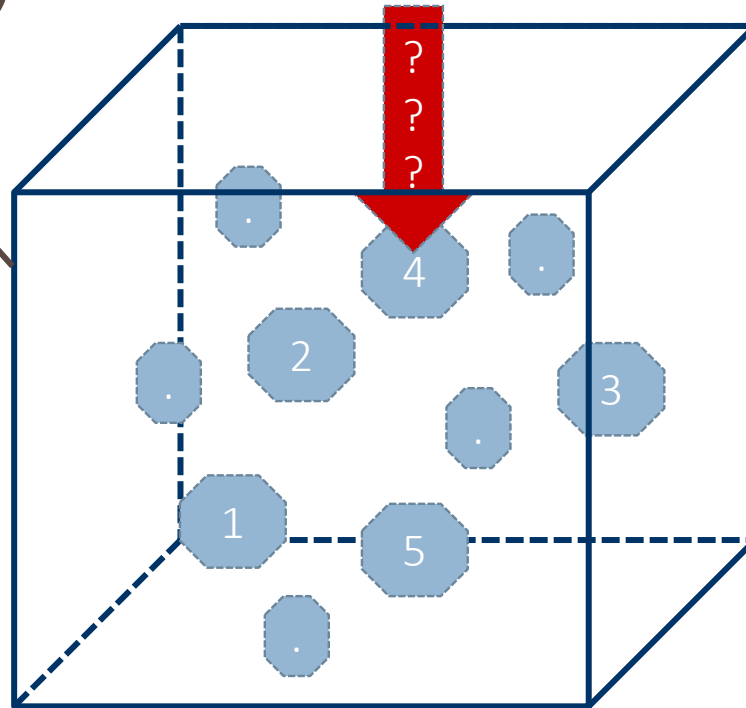
Ähm, ich kenne...



Zugang zu Korpora

Ich suche historische Briefe aus dem Hochdeutschen!

Server ? + Korpora ? + Suche?

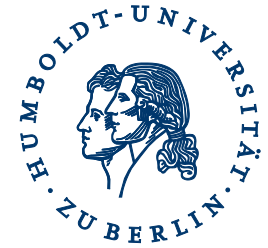


Ich suche normalisierte Texte aus dem 17. Jahrhundert!

Ähm, ich kenne...

Was verstehst du unter Normalisierung?

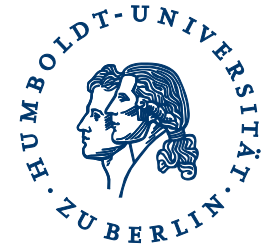
Kurzer Einstieg



- drei Beispiele historischer Korpora
 - aus historischer Linguistik und Musikwissenschaft
 - komplexe Korpora
 - unterschiedliche Aufbereitungen und Architekturen
 - aufgrund unterschiedlicher Forschungsfragen
 - Gemeinsamkeit: Arbeit mit historischen Texten
- Korpus basierte Forschung in den Geisteswissenschaften

Beispiel

Fürstinnenkorrespondenz



Datei extern: http://dwee.eu/Rosemarie_Luehr/userfiles/downloads/Projekte/Dokumentation.pdf

tok	Tokenisiert er Originaltext, der sich so nah wie möglich an die Vorlage hält; vorgefundene Zusammenschreibungen werden hier getrennt aufgeführt, durchgestrichene Textpassagen durch „#“ ersetzt und Silbentrennung am Zeilenende nicht durchgeführt.
orig	Originalspur (original): hier erscheinen originale Zusammenschreibungen, durchgestrichene Textpassagen und Silbentrennung am Zeilenende sowie Abbriviaturen, die mit % dargestellt werden.
norm	Normierte (neuhochdeutsche) Textspur, die den Einsatz des Annotationswerkzeuges TreeTagger für die Lemmatisierung und die Part -of-Speech -Zuweisung erlaubt.

	0	1	2	3	4	5	6	7	8	9	10	11
[tok]	Quetlinburgk	,	den	14,	Augustÿ	1677	Durchlechtigster	fürst	,	Gnädiger	sehr	hochgeEhrter
[orig]					Augustÿ1677							
[norm]	Quedlinburg	,	den	14.	August	1677	Durchlauchtigster	Fürst	,	gnädiger	sehr	hochgeehrter
[lemma]	Quedlinburg	,	d	14.	August	1677	durchlauchtig	Fürst	,	gnädig	sehr	hochgeehrt
[pos]	NE	\$,	ART	ADJA	FM	CARD	ADJA	NN	\$,	ADJA	ADV	ADJA
[morph]	*.Nom.Sg		Def.Masc.Acc.Sg	*.*.*			Sup.Masc.Nom.Sg	Masc.Nom.Sg		Pos.Masc.Nom.Sg		Pos.Masc.Nom.Sg
[grfunct]							voc			voc		
[s_grfunct]												

AD_JE2_1677_08_14, Lühr, Rosemarie; Faßhauer, Vera; Prutscher, Daniela; Seidel, Henry; Fuerstinnenkorrespondenz (Version 1.1), Universität Jena, DFG. <http://www.indogermanistik.uni-jena.de/Web/Projekte/Fuerstinnenkorr.htm>.

<http://hdl.handle.net/11022/0000-0000-82A0-7>

Beispiel: Verein

Musikalische Privataufführung



Datei extern

Text	Text
------	------

<p>Wien, den 17. Jänner 1921.</p>

<p>Sehr geehrter <persName role="IAV" type="Präsident" ref="Personenliste.xml#P00001">Herr Schönberg</persName>!</p>

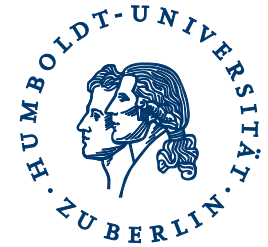
<p>

<persName role="IAV" type="Sonstige" ref="Personenliste.xml#P0007">Herr Berg</persName>

sagte mir, dass Sie einen Bericht über den Verkauf der Mitteilungen wünschen. Infolge des von Mittwoch bis gestern dauernden Poststreiks war ich bisher nicht in der Lage, Ihnen die Aufstellung zu übersenden. Jetzt, da er beendet ist, beeile ich mich, Ihrem Wunsche nachzukommen.</p>

Klarfeld_AS_1921001^7_5672, Bicher, Katrin; VereinSchoenfeld (Version 0.5), Humboldt-Universität zu Berlin. NA.
<http://hdl.handle.net/11022/0000-0000-84DD-2>

Beispiel RIDGES



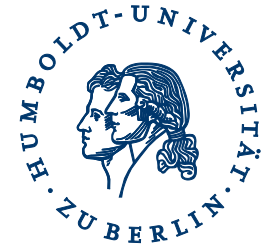
Datei extern: http://korpling.german.hu-berlin.de/ridges/documentation_v4.1_de.html

dipl	Die Transkription von Faksimiles stellt für die korpuslinguistische Aufbereitung zumeist die grundlegende, diplomatische Ebene (dipl). [...]
clean	Die clean-Ebene enthält erste vollautomatisch erstellte Normalisierungen hinsichtlich Sonderzeichen und grafischer Strukturierungen. [...].
norm	Die norm-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier die Tokenisierung und die Orthografie einheitlich nach modernen Orthografieregeln (vgl. Duden) angepasst werden, wobei die Flexion, wie z.B. Kasuszuweisungen, nicht berücksichtigt wird.[...].

dipl	von	Geiß	fen	vnnd	Hafen	zuverfthen	/
clean	von	Geissen		vnnd	Hasen	zuverstehen	/
norm	von	Geißen		und	Hasen	zu verstehen	/

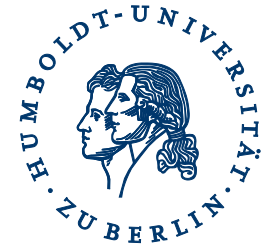
PflanzGart_1639, Lüdeling, Anke; Odebrecht, Carolin; Zeldes, Amir; RIDGES-Herbology (Version 4.1), Humboldt-Universität zu Berlin.
<http://korpling.german.hu-berlin.de/ridges/>. <http://hdl.handle.net/11022/0000-0000-8253-F>

Herausforderung der Erschließung



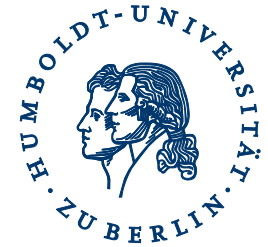
- Erschließung von Korpora direkt über deren Daten
 - Tools
 - wie z.B. Oxygen (<http://www.oxygenxml.com>), ELAN (Slotjes & Wittenburg 2008), EXMARaLDA (Schmidt et al. 2011)
 - Analyse-Software
 - wie z.B. ANNIS (Krause & Zeldes 2014) , CQP (Hardie 2012)
 - Tagsets abgebildet in dem jeweiligen Format
 - Übersicht zu linguistischen Formaten (vgl. bspw. Zipser 2014)
- Erschließung über Projekt-Homepage (Datei extern)
 - eigenständig aufgebaut
 - konzentriert auf das Forschungsziel
 - nicht immer aussagekräftig

Erschließung von Interpretationen



- notwendige Diversität in der Forschung
 - Forschungsfrage erzeugt u.U. neue Kategorien
 - theoretische Kategorien mit unterschiedliche Kategorisierungen nach Feinkörnigkeit, Ausprägung, Semantik
 - Abbildung in bestimmten Format
 - keine einheitlichen Annotationen und Formate, kaum exhaustive Auflistung möglich (wenige/keine „Standards“)
 - keine genaue Vorhersage über deren Kombinationen
 - keine genaue Vorhersage von Innovationen
 - **mehr Komplexität und Diversität durch überfachlichen Austausch von Korpora**

Erschließung durch Metadaten



- Kriterien für die Erschließung mit dem Ziel der Wiederverwendung
 - nicht basierend auf Interpretationen, Theorien
 - Diversität zwischen Fächern und innerhalb eines Faches
 - Konzentration auf die technisch-abstrakten Eigenschaften von Korpora
 - u.a. Annotationen, Art + Struktur, Erstellung
 - einheitlich über die verschiedenen Arten von historischen Textkorpora
 - wichtig für Vergleichbarkeit und einheitliche Suche
 - Gemeinsamkeiten
 - umfangreich
 - u.a. bezogen auf den Lebenszyklus von Forschungsdaten

Erschließung durch Metadaten

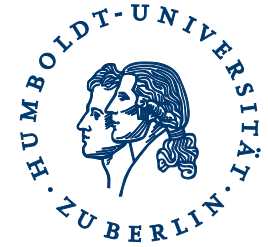


Suche nach dem jeweils „richtigen“ Korpus durch
(vgl. Odebrecht 2015)

- Metadaten der Primärquelle (bibliographisch)
 - Datum, Veröffentlichungsort, Autor
 - Register, Sprache (linguistische Eigenschaften)
- Metadaten des Korpus (administrativ)
 - Projekt, Annotatoren
 - Prüfmethode
 - Veröffentlichungsgeschichte
- Metadaten der Annotation (resultativ, deskriptiv)
 - Annotationsrichtlinien, -verfahren, -tools und -schritte

→ Kriterien für die Suche und die Auswahl eines Korpus aus einer Menge an Korpora

LAUDATIO-Repository



- Long-term **Access and Usage of Deeply Annotated Information**

Historische Korpora überfachlich

suchen

finden

speichern

veröffentlichen

dokumentieren

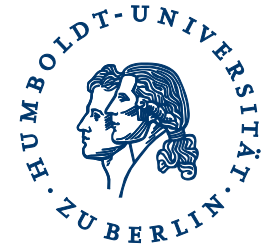
referenzieren!

- unabhängig von Format, Größe, Annotationen, Fach und Forschung!

Metadaten

- einheitliche Metadaten für die verschiedenen geisteswissenschaftlichen Korpora im LAUDATIO-Repository
 - für die Metadatenanzeige
 - Dokumentation
 - Referenzierung
 - als Grundlage für die Metadatensuche
 - Metadatenfreitextsuche
 - Metadatenfacettensuche

Metadatenanzeige für ein Korpus



LAUDATIO-Repository Register new account | Login

Home » Browse » RIDGES-Herbology

Home
Documentation
Browse
Search

RIDGES Herbology Version 4.0

2014-06-24 08:37:19

RIDGES Herbology Version 4.0, Humboldt-Universität zu Berlin, 4.0, 153732 Tokens, Fourth version. Extension of the corpus.
Licence (for corpus and related documents):

Formats: [TEI-Header](#), [EXCEL](#), [EXMARaLDA](#), [reIANNIS](#), [PDF](#)

Always quote when using this data!
Lüdeling, Anke; Odebrecht, Carolin; Zeldes, Amir; RIDGES-Herbology (Fourth version. Extension of the corpus.) Version: 4.0. Humboldt-Universität zu Berlin.
<http://korpling.german.hu-berlin.de/ridges/>
<http://hdl.handle.net/11022/0000-0000-24EC-E>

▶ Corpus RIDGES Herbology Version 4.0 ?

▶ Documents ?

▶ Annotation ?

▶ PreparationStep ?

Annotations: Wenn vorhanden: Link zum Korpus in ANNIS

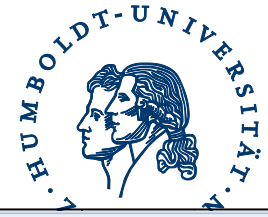
Files: Dateien zum Download

Licence: Korpus-Lizenz

Reference: Referenz und PID

Metadata: Metadaten

Metadatenuche



LAUDATIO-Repository Register new account | Login

Home » Search

Home
Documentation
Browse
Search

Freitextmetadatenuche

Full-Text Search

Filter by

Facetten Metadaten *Korpus*

Corpus

- + Corpora
- + Projects
- + Formats
- + Date - Corpus
- + Size - Corpus

Facetten Metadaten *Dokument + Annotation*

Document

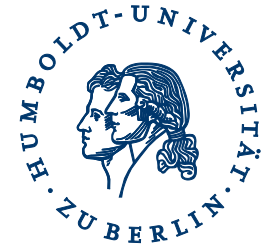
- + Annotation - Graphical
- + Annotation - Lexical
- + Annotation -

Object URL: [Direct Link to Corpus](#)
Homepage: <http://korpling.german.hu-berlin.de/ddb-doku/index.htm>
Project Description: Deutsche Diachrone Baumbank. Das durch den Berliner Senat geförderte Projekt "Interdisziplinärer Forschungsverbund Linguistik - Bioinformatik zur Berechnung von Verwandtschaft und Abstammung" hat angestrebt, Wege zu finden, wie bioinformatische Methoden dazu verwendet werden können, die Verwandtschaft zwischen (schriftlichen) Sprachdaten automatisch messbar zu ... [\(more\)](#)
Documents:
[Ein kurze und einfeltige unterweisung zum sterben nutzlich und heilsam den krancken furzuhalten an irem letzten/aus der heiligen schriften zusammen gelesen](#)
[Ein kurze und einfeltige unterweisung zum sterben nutzlich und heilsam den krancken furzuhalten an irem letzten/aus der heiligen schriften zusammen gelesen](#)
[The Monsee Fragments](#)
[\(more\)](#)

Title: GerManC_2007-04

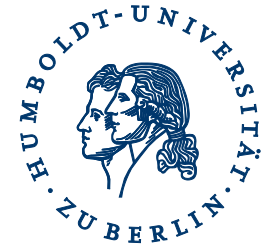
Zugang zu Korpora

Aufgabe



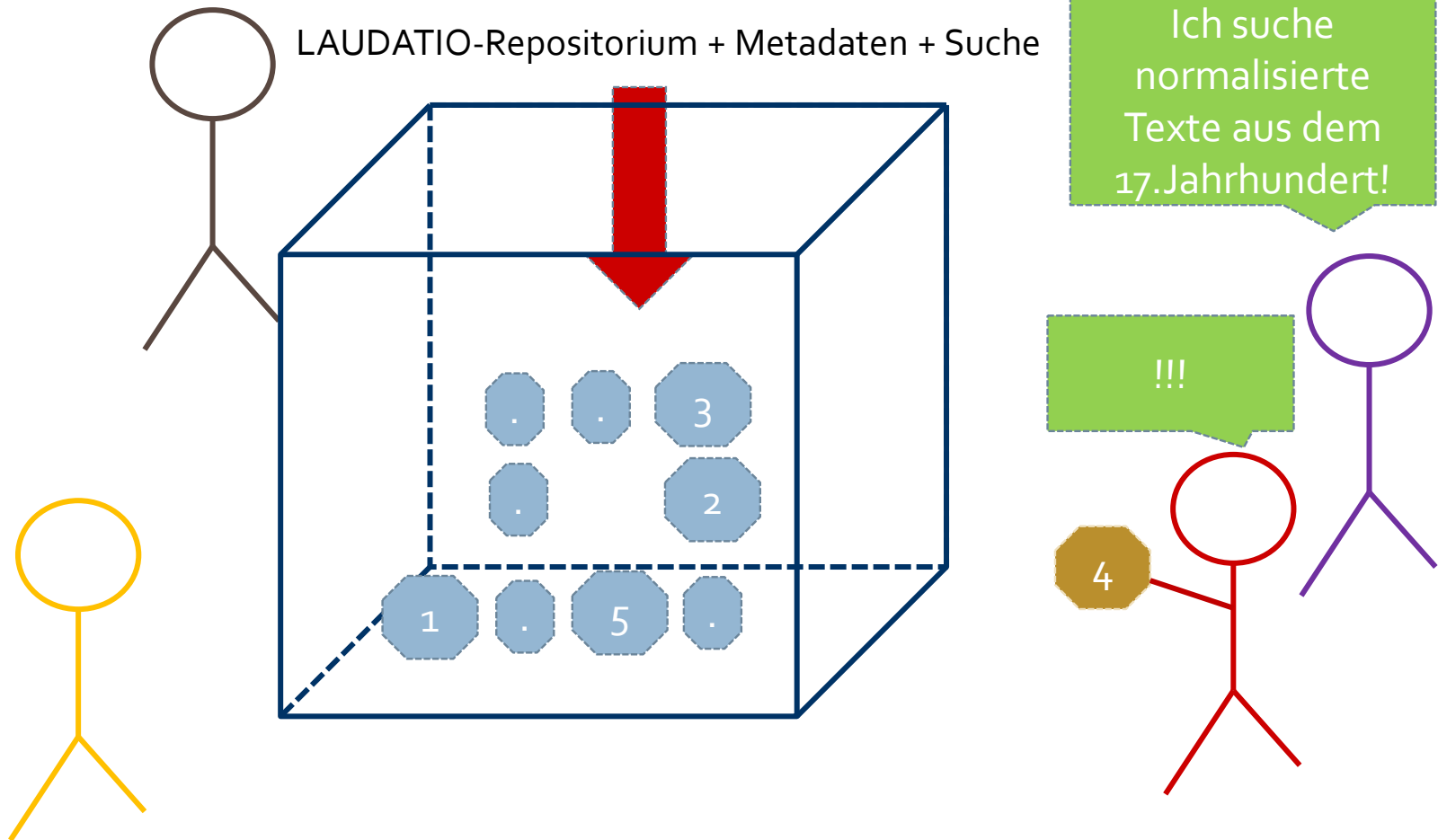
- Gesucht werden Texte aus dem 17. Jahrhundert mit Normalisierung, dazu Wissen über
 - Eigenschaften von Texten (*Dokument*)
 - Datum des Verfassens, des Veröffentlichens
 - Eigenschaften des „Text“ (*Annotation/Ebene*)
 - konkrete Werte der Annotation (Strings)
 - Wortformen, Buchstaben, Morpheme etc.
 - andere Annotationen basieren darauf
 - (alle) anderen Annotationen sind auf dieser Ebene (un-)mittelbar annotiert
 - Eigenschaften des Text-Korpus (*Korpus*)
 - Texte werden in einer Sammlungen veröffentlicht
 - Tagsets

Open Access & Data Sharing



- Unterstützung der Korpus basierten Forschung in den Geisteswissenschaften durch
 - Veröffentlichung der Korpora
 - nach, während Projekten/Forschung
 - Lizenzen
 - z.B. CreativeCommons Lizenzen für Forschungsdaten
 - creativecommons.org/
 - z.B. Apache License für Software
 - apache.org/licenses/
 - Förderung der (über-)fachlichen Zusammenarbeit
 - Austausch von Forschungsdaten
 - freier Zugang zu Infrastrukturen

Zugriff auf Korpora



Zugang und Zugriff

I. Zugang zu Korpora

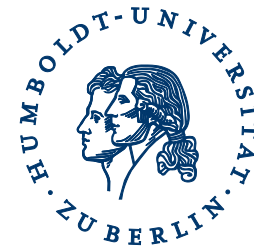
- Suche nach Korpora
- Forschungsdatenrepositorium LAUDATIO
- Funktionen von Metadaten
- Wiederverwendung von Korpora

II. Zugriff auf Korpora

- Suche in Korpora
- Such- und Visualisierungstool ANNIS
- Anfragesprache & Annotationsvisualisierung
- Analyse von Korpora

Teil II

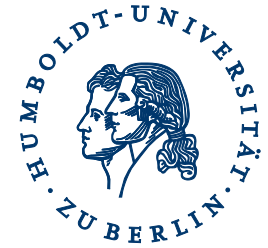
Zugriff auf Korpora



- Wie kann der Zugriff auf Korpora und deren Annotationen einheitlich und fächerübergreifend ermöglicht werden,
 - um in eigenen oder Korpora Dritter **nach Annotationen zu suchen**.
 - um Belege zu **analysieren und auszuwerten**.
 - um die **Nachvollziehbarkeit** von Analysen zu fördern.
- Korpusuchtools
 - spezialisiert auf bestimmte Korpusarten z.B. CQP (Hardie 2012), TIGERSearch (Lezius 2002), Cosmas II (Bodmer 2005)

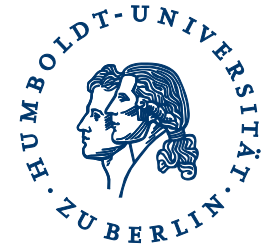


ANNIS

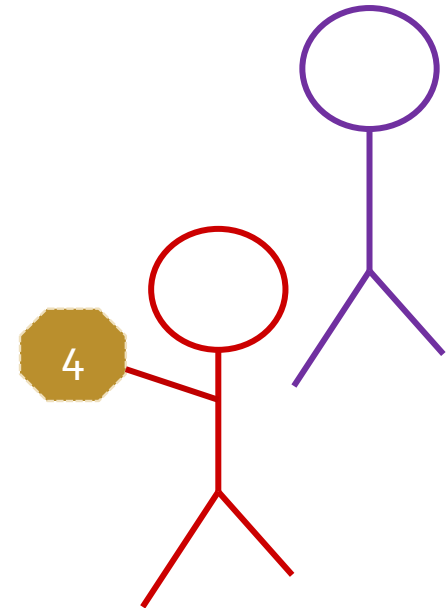


- Browser basiertes Such- und Visualisierungstool für Mehrebenen-Korpora
 - Serverinstallation, lokale Installation
- generisches Datenmodell Salt (Zipser & Romary 2010)
 - Graph basiert (Krause & Zeldes 2014)
- generischen Anfragesprache ANNIS Query Language (AQL)
 - grundsätzlich gilt:
Gefunden werden kann nur das, was auch annotiert ist!
 - kein NLP-Tool
- anpassbaren Annotationsvisualisierungen

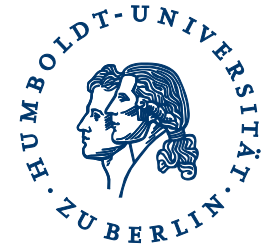
Korpora Annotationen



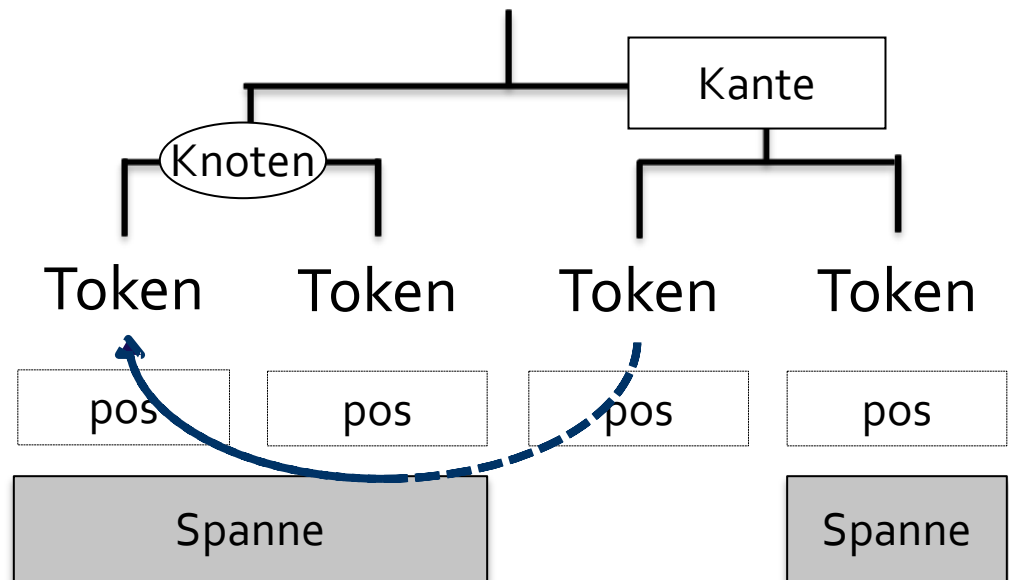
- Suche und Visualisierung diverser Annotationsarten
 - Tokenannotationen
 - Spannenannotation
 - Abhängigkeitsannotationen
 - Baumannotationen
 - Filterung nach Metadaten
 - alle Kombinationen daraus



Annotationen Arten

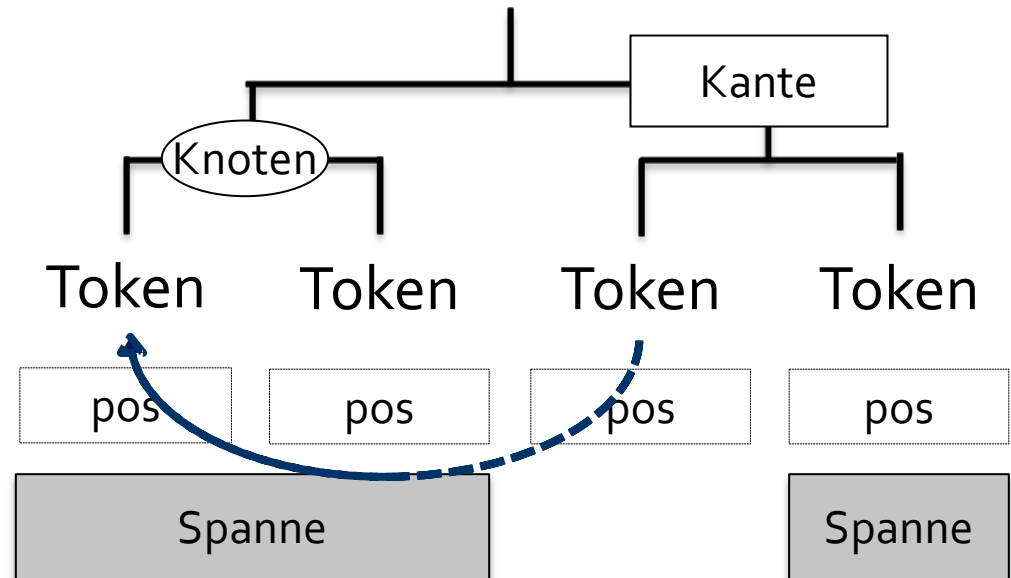


- Zugriff auf Korpora via AQL
 - für viele Arten von Annotationen
 - unabhängig von der Bedeutung von Annotationen (z.B. Tagsets)

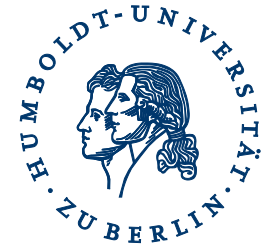


Formate

- Konverter Framework Pepper (Zipser & Romary 2010)
 - gemeinsames Datenmodell mit ANNIS → Salt
 - Unterstützung von u.a.
 - TEI XML,
 - MMAx,
 - EXMARaLDA,
 - ANNIS,
 - TIGER XML,
 - TCF,
 - PAULA

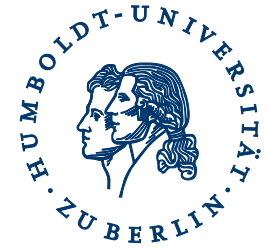


Ressourcen



- HU - Instanz
 - <https://korpling.german.hu-berlin.de/annis3/>
- Suche mit
 - Korpus-Referenz RIDGES https://korpling.german.hu-berlin.de/annis3/#_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x
 - Korpus-Referenz Fürstinnen (Fürstinnenkorrespondenzkorpus) https://korpling.german.hu-berlin.de/annis3/#_c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x
- für alle Referenz-Links
 - letzter Zugriff 20.10.2015

https://korpling.german.hu-berlin.de/annis3/



Help us to make ANNIS better! not logged in Login

Please enter AQL query

Query Builder

Search More History

Welcome to ANNIS! A tutorial is available on the right side.

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
a5.hausa.news	4	2.017	ⓘ	📄
a5.hausa.umarnin.uwa	47	10.194	ⓘ	📄
abraham.our.father	7	7.671	ⓘ	📄
Alibaba.6.all	1	202	ⓘ	📄
apophthegmata.patrum.	11	1.500	ⓘ	📄
arabic.tree.test	1	11	ⓘ	📄
b1.aja	19	257	ⓘ	📄
b1.fon	19	264	ⓘ	📄
b1.foodo	19	215	ⓘ	📄

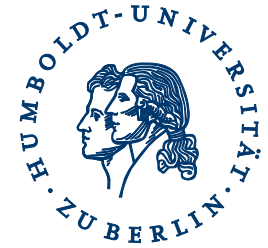
Help/Examples

Tutorial

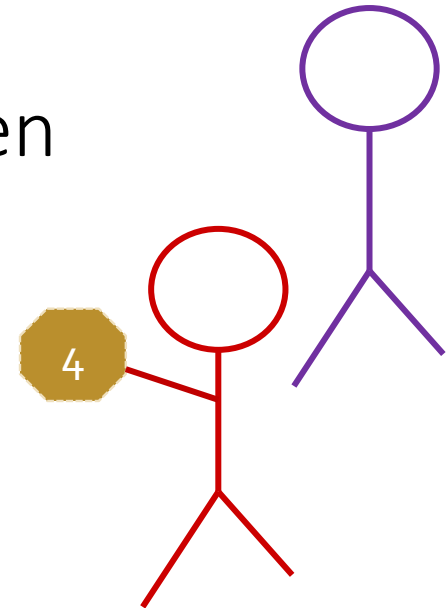
Example Queries

Example Query	Description	open corpus browser
Q "e"	search for the word "e"	Besa.letters
Q /fFfriedrich/	Search for the "friedrich" with upper or lower-case "f" (regular expression)	a5.hausa.news
Q "."	search for the word "."	a5.hausa.umarnin.uwa
Q norm="ἰοϣῖε"	search for the normalized word ἰοϣῖε	abraham.our.father
Q pos="NPRO"	search for proper names	abraham.our.father
Q "den"	search for the word "den"	fuerstinnenkorrespondenz
Q /fSslehr/	Search for the "sehr" with upper or lower-case "s" (regular expression)	fuerstinnenkorrespondenz
Q "sehr"	search for the word "sehr"	Fuerstinnenkorrespondenz1.1
Q /fAaIugusty/	Search for the "augusty" with upper or lower-case "a" (regular expression)	Fuerstinnenkorrespondenz1.1
Q sic	Search for errors and disfluencies	GUM
Q "of"	search for the word "of"	GUM
Q /fIiInternational/	Search for the "international" with upper or lower-case "i" (regular expression)	GUM
Q POS="NA"	Findet ein Nomen.	HIPKON
Q "heiligin"	search for the word "heiligin"	HIPKON
Q /fOoIratio/	Search for the "oratio" with upper or lower-case "o" (regular expression)	HIPKON
Q "ain"	search for the word "ain"	HSJ-Baseler Maesebuch
Q /fAaIin/	Search for the "ain" with upper or lower-case "a" (regular expression)	HSJ-Baseler Maesebuch
Q "un"	search for the word "un"	HSJ-Briefe
Q "/f/ with upper or lower-case "/	Search for the "/f/ with upper or lower-case "/	

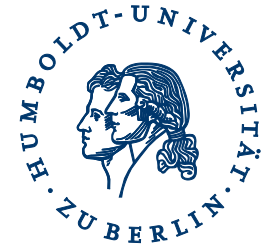
Zugriff auf Korpora durch AQL



- anhand zweier Korpora
 - RIDGES und Fürstinnenkorrespondenz
- Token- und Spannenannotation
- Relationen zwischen Annotationen
- Metadatenfilterung



Zugriff auf Korpora durch AQL

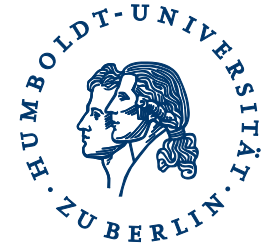


ANNIS Query Language

- Prinzip I
 - Variable-Wert-Paare
- Prinzip II
 - Relationen

➤ für alle Annotationsarten

Token



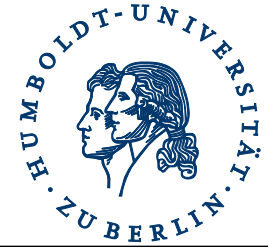
- 1) Als Token bezeichnet man häufig die **kleinste (technische) Einheit** in einem Korpus.
- 2) Ein Token entspricht oft (aber nicht immer) einem **orthographischen Wort** oder **Satzzeichen**.
- 3) Nach diesen Einheiten kann man in ANNIS **suchen**.

Token Token Token Token Token Token ...

dipl	wider	auff	ein	frisches	zerstoffenes	kraut	/	thue	es	wider	in
clean	wider	auff	ein	frisches	zerstossenes	kraut	/	thue	es	wider	in
norm	wieder	auf	ein	frisches	zerstoßenes	Kraut	/	tue	es	wieder	in

Prinzip I

Variable-Wert-Paar



- 1) Voraussetzung ist das **Vorhandensein einer Ebene namens „dipl“**. (Metadaten!)
- 2) Erwartetes Ergebnis ist es, **exakt alle Vorkommen** dieser Zeichenkette in „dipl“ im ausgewählten Korpus zu finden.

dipl= /kraut/

Variable

Wert

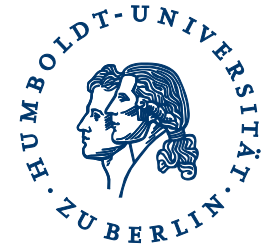
(Layer, Tier, Ebene ...)

(Wort, Lemma, Satz, Wortart ...)

dipl	wider	auff	ein	frisches	zerstoffenes	kraut	/	thue	es	wider	in
clean	wider	auff	ein	frisches	zerstossenes	kraut	/	thue	es	wider	in
norm	wieder	auf	ein	frisches	zerstoßenes	Kraut	/	tue	es	wieder	in

Variable-Wert-Paar

Beispiel RIDGES



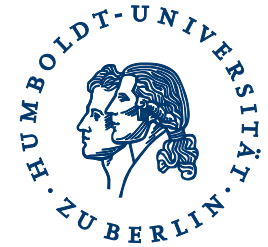
- Suche nach historischer Form von *und*
 - dipl
 - Annotationsebene mit Transkriptionen
 - eine historische Form von *und*
 - *vnnd*

dipl=/vnnd/

Such-Referenz https://korpling.german.humboldt-berlin.de/annis3/#_q=ZGlwbD0vdm5uZC8&c=UklER0VTX0hlcmJvbG9neV9WZlJzaW9uNC4x&cl=5&cr=5&s=0&l=10&seg=ZGlwbA

Variable-Wert-Paar

Beispiel Fürstinnen



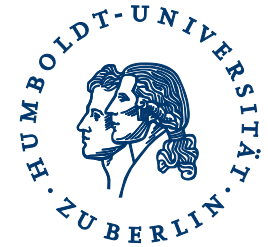
- Suche nach historischer Form von *und*
 - tok
 - Annotationsebene mit Transkriptionen
 - eine historische Form von *und*
 - *vnnd*

tok=/vnnd/

Such-Referenz https://korpling.german.humboldt-berlin.de/annis3/#_q=dG9rPS92bm5kLw&c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x&cl=5&cr=5&s=0&l=10

Variable-Wert-Paar

Beispiel RIDGES



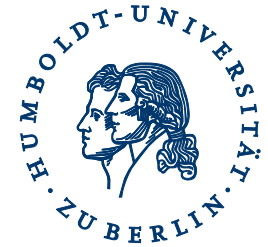
- Suche nach allen historischen Formen von *und*
 - norm
 - orthographische Normalisierung
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*

norm=/und/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=bm9ybT0vdW5kLw&c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&seg=ZGlwbA

Variable-Wert-Paar

Beispiel Fürstinnen



- Suche nach allen historischen Formen von *und*
 - norm
 - orthographische Normalisierung
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*

norm=/und/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=bm9ybT0vdW5kLw&c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x&cl=5&cr=5&s=0&l=10

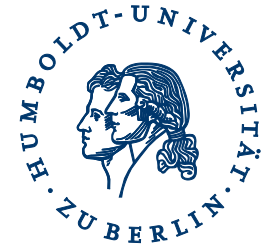
Operatoren für die Mustersuche



- . Ein beliebiges Zeichen
- ? 0 oder 1 Zeichen (des vorherigen Elementes)
- * 0 bis unendlich viele Zeichen (d. vorh. E.)
- + 1 bis unendlich viele Zeichen (d. vorh. E.)
- \\ wörtlich (folgendes Zeichen)
- ! nicht
- (a | b) a oder b (auch: [ab])

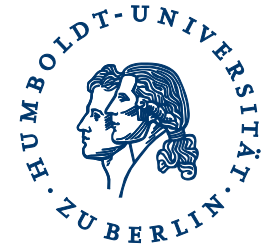
Operatoren

Beispiel RIDGES



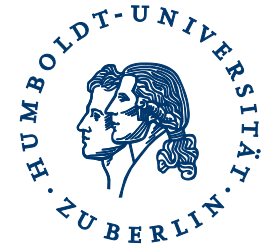
- Welche Ergebnisse erwarten Sie für folgende Anfragen, sogenannte Mustersuchen?
 - norm=/g.b./
 - *gebe, gibt* (für RIDGES, theoretisch noch andere möglich)
 - dipl=/r(a|o)t/
 - *rot, rat* (für RIDGES)
 - dipl=/meint?/
 - *mein, meint* (für RIDGES)

Such-Referenzen



- Mustersuchen in RIDGES, Folie 50
- https://korpling.german.hu-berlin.de/annis3/#_q=bm9ybT0vZy5iLi8K&_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&_seg=ZGlwbA
- https://korpling.german.hu-berlin.de/annis3/#_q=ZGlwbD0vcihhfG8pdC8K&_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&_seg=ZGlwbA
- https://korpling.german.hu-berlin.de/annis3/#_q=ZGlwbD0vbWVpbmQLwoK&_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&_seg=ZGlwbA

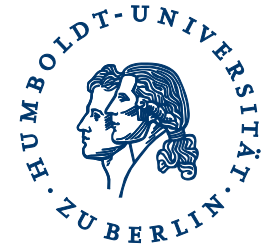
Referenzierung via Links



- Korpus-Referenzen
 - z.B. RIDGES in ANNIS
 - https://korpling.german.humboldt-berlin.de/annis3/#_c=UklERoVTXohlcmJvbGgneVgWZXJzaW9uNC4x
- Such-Referenzen in ANNIS
 - z.B. dipl=/vnnd/ in RIDGES in ANNIS
 - https://korpling.german.humboldt-berlin.de/annis3/#_q=ZGlwbDovdm5uZC8KCg&_c=UklERoVTXohlcmJvbGgneVgWZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&_seg=ZGlwbA
- geplant: Beleg-Referenzen in ANNIS (in der Testphase)
 - einen konkreten Treffer der Anfrage dipl=/vnnd/ in RIDGES

Prinzip II

Relationen RIDGES

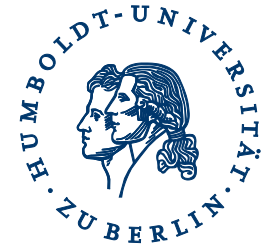


head=/head/ **VW-Paar1**
 o **Relation**
 dipl=/Wermut/ **VW-Paar2**

- 1) Es gibt **mehrere Arten** von Annotationen!
- 2) Wissen, wie diese Annotationen in **Beziehung** zu einander stehen können!
- 3) **Spannen** (head) überlappen **Tokenannotationen** (dipl)!

dipl	an	ftatt	des	Zittwer-Samens	.	Abfinthium	Vulgare	,	Wermut	.
lb	lb					lb				
head						head				

Syntax-Highlighting



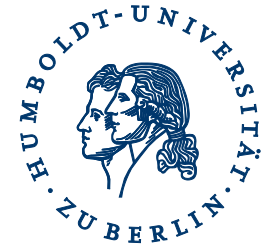
- pro VW-Paar eine Farbe:
 - Farbe **Rot** für alle Werte der Variable **head**
 - Farbe **Lila** für alle Werte der Variable **dipl**
- Treffer in genau diesen Farben:
 - mehrere Token **rot**, zusammen in einer Überschrift
 - ein Token **lila**, der gesuchte dipl-Wert

```
head=/head/  
o  
dipl=/Wermut/
```

Abfinthium **Vulgare** , **Wermut** .

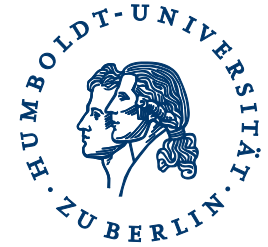
dipl	an	ftatt	des	Zittwer>Samens	.	Abfinthium	Vulgare	,	Wermut	.
lb	lb					lb				
head						head				

Such-Referenz



- RIDGES Folie 53
- https://korpling.german.hu-berlin.de/annis3/#_q=aGVhZD0vaGVhZC8KX29fCmRpcGw9L1dlcm11dC8&c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&seg=ZGlwbA

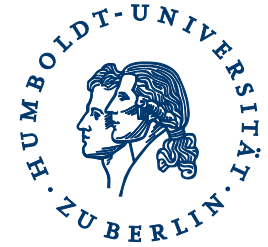
wichtige Relationen zwischen VW-Paaren



- . Direkte Präzedenz
- .* Indirekte Präzedenz
- _= Identische Abdeckung
- _i_ Inklusion
- _o_ Überlappung
- _l_ linksseitige Überlappung
- _r_ rechtsseitige Überlappung

Relationen

Beispiel RIDGES



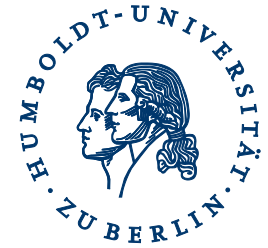
- Suche nach Phrasen in Form von Wortartenabfolgen
 - pos
 - Annotation der Wortarten
 - Artikel Adjektiv Nomen = NP

pos=/ART/ . pos=/ADJA/ . pos=/NN/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=cG9zPS9BUlQvIC4gcG9zPS9BREpBLyAuIHBvcz0vTk4vCQo&_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&_seg=ZG|w_bA

Relationen

Beispiel RIDGES



Help us to make ANNIS better | not logged in | Login

Help/Examples | Export | Frequency Analysis | Query Result

Base text | 1 / 234 | Displaying Results 1 - 10 of 2331 | Result for: pos=/ART/ . pos=/ADJA/ . pos=/NN/

1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 190 - 202) left context: 5 right context: 5

publication zuwider fein / dafz der gemeine Deutsche nicht verfehe wie gelehrt fie

- Transcription
- Lexical Annotation
- Syntactical Annotation
- Content Annotation
- Graphical Annotation
- All

2 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 541 - 553) left context: 5 right context: 5

/ weit vnd enge / eine zimliche anzahl haben . Difz nennet man

- Transcription
- Lexical Annotation
- Syntactical Annotation
- Content Annotation
- Graphical Annotation
- All

3 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 595 - 608) left context: 5 right context: 5

Kugel / vber fich mit einem lan gen Halz / werden Vio len genennet

- Transcription
- Lexical Annotation
- Syntactical Annotation
- Content Annotation
- Graphical Annotation
- All

pos=/ART/ . pos=/ADJA/ .
pos=/NN/

Query Builder

2331 matches in 29 documents

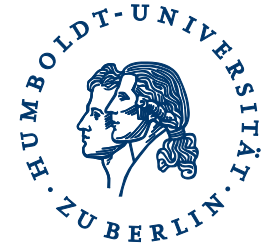
Search Options

Visible: All

Name	Texts	Tokens		
RIDGES				
RIDGES_Herbology_Version	22	122.698		
RIDGES_Herbology_Version	29	154.266		
RIDGES_Herbology_Version	29	154.267		
Ridges_Herbology_Version	13	60.811		

Relationen

Beispiel Fürstinnen



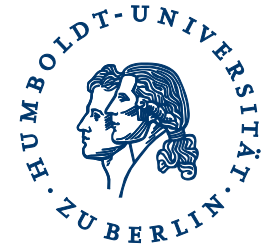
- Suche nach Phrasen in Form von Wortartenabfolgen
 - pos
 - Annotation der Wortarten
 - Artikel Adjektiv Nomen = NP

pos=/ART/ . pos=/ADJA/ . pos=/NN/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=cG9zPS9BUIQvIC4gcG9zPS9BREpBLyAuIHBvcz0vTk4v&_c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x&cl=5&cr=5&s=0&l=10

Relationen

Beispiel Fürstinnen



[Help us to make ANNIS better!](#)
not logged in [Login](#)

[Help/Examples](#)
[Export](#)
[Frequency Analysis](#)
[Query Result](#)

Base text

/ 260
 Displaying Results 1 - 10 of 2599
Result for: pos=/ART/ . pos=/ADJA/ . pos=/NN/

- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_08_14 (tokens 223 - 235)
 left context: 5 right context: 5
 , so hätte Ich wohl **einen guten Peltz** , fon nöhten , AlB
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_08_14 (tokens 436 - 448)
 left context: 5 right context: 5
 mirh zu troste , und **den gantzen lande** , zu grossen Auffnehmnen ,
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_11_07 (tokens 73 - 85)
 left context: 5 right context: 5
 Grund meiner Sellen , das **der Allmächtige Gott** , Ew Gnd, ihn gleigen
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_11_07 (tokens 82 - 94)
 left context: 5 right context: 5
 Ew Gnd, ihn gleigen Auch **die Libe Frau** Mutter Gd mit seyner himmel
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_11_07 (tokens 341 - 353)
 left context: 5 right context: 5
 Princesin von Bernburg Id, sagen **Einen demütigen Gruß** , beÿ Ihr Gnd der
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_11_07 (tokens 365 - 377)
 left context: 5 right context: 5
 Permission die H Brudter , **einen freundsvesterligen gruß** , werde E Gnd nicht
 general (grid)
 general (grid)
- Path: Fuerstinnenkorrespondenz1.1 > AD > AD_JE2_1677_11_07 (tokens 442 - 454)
 left context: 5 right context: 5
 mirh understehen darff Ew gd **eine geringe Messe** zu senden , Ervarde ich
 general (grid)
 general (grid)

2599 matches in 537 documents

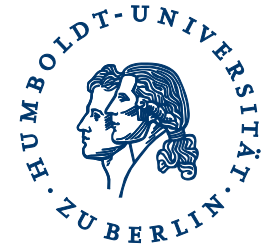
Visible:

Name	Texts	Tokens		
fuerstinnenkorrespondenz	600	262.468	<input type="checkbox"/>	<input type="checkbox"/>
Fuerstinnenkorresponden: 600	262.465		<input type="checkbox"/>	<input type="checkbox"/>

Query Builder

Relationen

Beispiel RIDGES



- Suche nach allen historischen Varianten von *und*
 - norm
 - orthographische Normalisierung
 - dipl
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*

dipl=/.*/

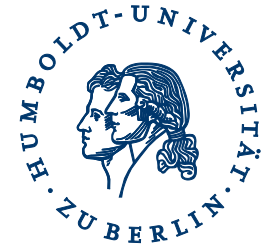
o

norm=/und/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=CQlkaXBsPS8uKi8KCQlfb18KCQlub3JtPS91bmQvCg&c=UklER0VTX0hlc mJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&seg=ZGlwbA

Relationen

Beispiel RIDGES

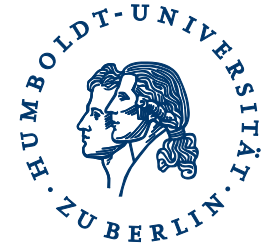


- Export des Ergebnisses in csv zur weiteren Analyse
 - zum Beispiel:

1	1_id	1_span	1_anno_default_ns:dipl	2_id	2_span	2_anno_default_ns:norm
2	35845084		vnnd 35845087	und		
3	35845137		vnd 35845140	und		
4	35845785		und 35845788	und		
5	35846030		vnd 35846034	und		
6	35846329		vnd 35846332	und		
7	35846490		vnnd 35846493	und		
8	35846939		vnnd 35846942	und		
9	35847160		vnd 35847164	und		
10	35848362		vnd 35848366	und		
11	35848723		vnd 35848726	und		
12	35848775		vnd 35848778	und		
13	35848933		vnd 35848936	und		
14	35849029		vnd 35849032	und		
15	35849061		vnd 35849064	und		

Relationen

Beispiel Fürstinnen



- Suche nach allen historischen Varianten von *und*
 - norm
 - orthographische Normalisierung
 - tok
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*

tok=/.*/
o
norm=/und/

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=CQl0b2s9Ly4qLwoJCV9vXwoJCV5vcm09L3VuZC8K&_c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x&cl=5&cr=5&s=0&l=10

Export von Treffern

Text-Exporter

```
0. of the International Brotherhood of [Magicians] Wednesday , October 9
1. Magic Month in the United [States] . Wikinews spoke with William
2. of the International Brotherhood of [Magicians] , about the current st
3. - " Scarne on Card [Tricks] " and " Scarne on
4. and " Scarne on Magic [Tricks] " . That started me
```

Grid-Exporter

```
0.      tok      of the International Brotherhood of Magicians Wednesday
      pos      IN[1-1] DT[2-2] NP[3-3] NP[4-4] IN[5-5] NPS[6-6] NP[7-7]
      cat      S[1-6] VP[1-6] NP[1-6] PP[1-6] NP[2-4] PP[5-6] NP[6-6] NP[7-12]
```

CVS-Exporter

```
'11318611','the current state','NP','11318616','current','AJ0','current','JJ'
'11318686','magic','NP','11318688','magic','AJ0','magic','JJ'
'11318757','some basic tricks','NP','11318760','basic','AJ0','basic','JJ'
```

Export von Treffern



Wekaexporter

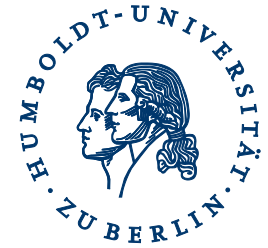
```
@relation name

@attribute #1_id string
@attribute #1_span string
@attribute #1_anno_const:cat string
@attribute #2_id string
@attribute #2_span string
@attribute #2_anno_GUM:claws5 string
@attribute #2_anno_GUM:lemma string
@attribute #2_anno_GUM:pos string

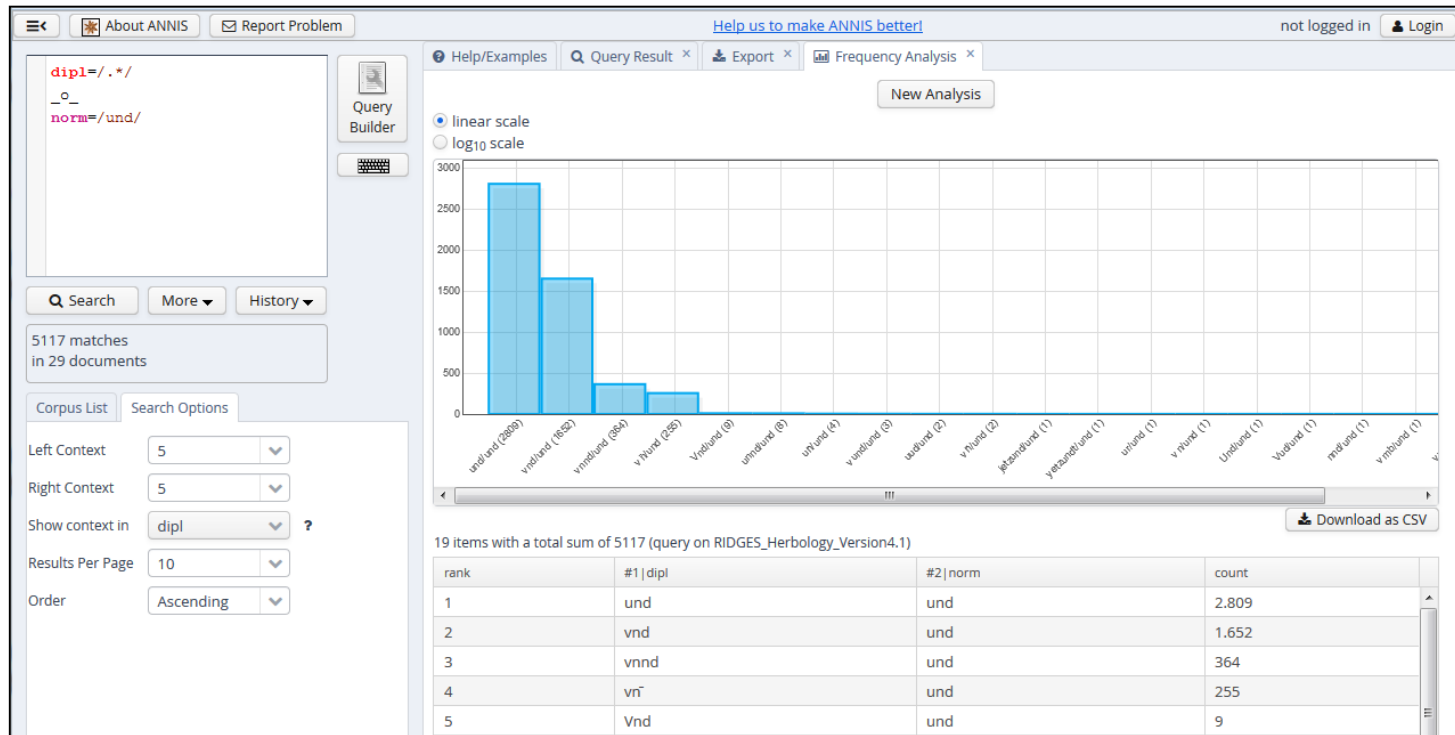
@data

'11318611','the current state','NP','11318616','current','AJ0','current','JJ'
'11318686','magic','NP','11318688','magic','AJ0','magic','JJ'
'11318757','some basic tricks','NP','11318760','basic','AJ0','basic','JJ'
```


Relationen Frequenzanalyse



- Suche mit Frequenzanalyse kombinieren
 - Vorkommen der einzelnen Varianten von *und*



Frequenzanalyse RIDGES



The screenshot shows the ANNIS RIDGES interface. At the top, there are navigation links like 'About ANNIS' and 'Report Problem'. The main area is divided into several sections:

- Query Builder:** Contains the query `dipl=/.*/`, `_o_`, and `norm=/und/`.
- Search Results:** Shows '5117 matches in 29 documents'. A red arrow points to the 'More' button.
- Search Options:** Includes 'Left Context' (5), 'Right Context' (5), 'Show context in' (dipl), 'Results Per Page' (10), and 'Order' (Ascending).
- Frequency Analysis Table:** A table with columns 'Node number/name', 'Selected annotation of node', and 'Comment'. It contains two rows:

Node number/name	Selected annotation of node	Comment
1	dipl	automatically created from <code>dipl=/.*/</code>
2	norm	automatically created from <code>norm=/und/</code>
- Metadata:** A 'Select' dropdown menu with 'Add' and 'Delete selected row(s)' buttons.
- Buttons:** 'Perform frequency analysis' and 'Automatic mode' (checked).

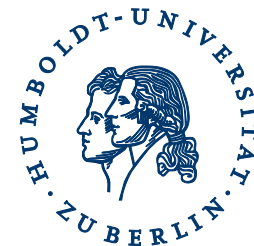
Anfrage bereits gestellt

unter „More“

Beispiel: Finden aller Wortformen (dipl), die als *und* (norm) normalisiert worden sind

Start der Frequency Analysis

Frequenzanalyse RIDGES



ANNIS interface showing a search query and frequency analysis results.

Anfrage bereits gestellt

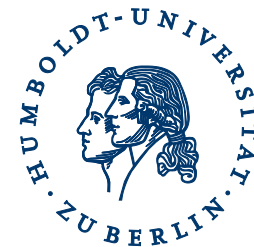
Beispiel: Finden aller Wortformen (dipl), die als und (norm) normalisiert worden sind, + Angabe über die Häufigkeit (Vorkommen) der Kombination

Vorkommen

Liste der Frequenzanalyse Download als cvs möglich

rank	#1 dipl		
1	und	und	2.809
2	vnd	und	1.652
3	vnnd	und	364
4	vn̄	und	255
5	Vnd	und	9

Frequenzanalyse Fürstinnen



[About ANNIS](#) | [Report Problem](#) | [Help us to make ANNIS better!](#) | not logged in | [Login](#)

[Help/Examples](#) | [Export](#) | [Frequency Analysis](#) | [Query Result](#)

linear scale
 log₁₀ scale

Token	Count
vnd	7218
vndt	1409
und	752
vnnd	315
Vnd	98
vndtund	88
Vndtund	69
Vndtund	67
Vndtund	19
vndtund	19
Vndtund	12
#und	10
v und	6
V und	6
v und	2
und und	2
und und	2
V und und	1
vnd und	1

23 items with a total sum of 9969 (query on Fuerstinnenkorrespondenz1.1)

rank	#1 spanned text	#2 norm	count
1	vnd	und	7.218
2	vndt	und	1.409
3	und	und	752
4	vnnd	und	315
5	Vnd	und	98

tok=/.*/
 o
 norm=/und/

| |

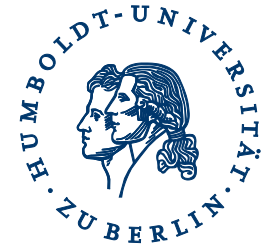
9969 matches in 599 documents

|

Visible: All

Name	Texts	Tokens
Fuerstinnen	600	262.468
Fuerstinnenkorresponden; 600	262.465	

Filterung nach Metadaten RIDGES

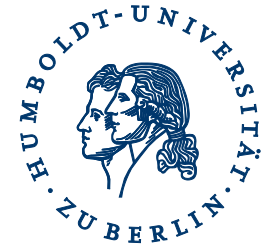


- Suche nach allen historischen Varianten von *und* im 16. Jh.
 - norm
 - orthographische Normalisierung
 - dipl
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*
 - Datumsangabe

```
dipl=/.*/  
_o_  
norm=/und/  
& meta::date=/15.*/
```

Such-Referenz https://korpling.german.hu-berlin.de/annis3/#_q=CQlkaXBsPS8uKi8KCQlfb18KCQlub3JtPS91bmQvCgkJJiBtZXRhOjpkYXRIPS8xNS4qLwo&c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x&cl=5&cr=5&s=0&l=10&seg=ZGlwbA

Filterung nach Metadaten Fürstinnen

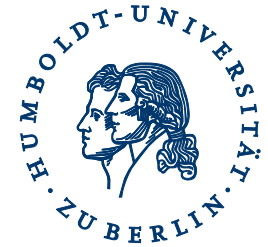


- Suche nach allen historischen Varianten von *und* im 16. Jh.
 - norm
 - orthographische Normalisierung
 - tok
 - historische Formen von *und*
 - *vnnd, vnd, vn, und ...*

```
tok=/.*/  
_o_  
norm=/und/  
& meta::date=/15.*/
```

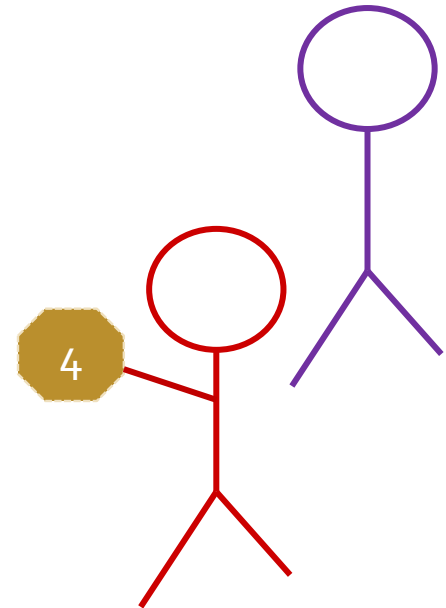
Such-Referenz https://korpling.german.humboldt-berlin.de/annis3/#_q=dG9rPS8uKi8KX29fCm5vcv09L3VuZC8KJiBtZXRhOjpkYXRIPS8xNS4qLwo&c=RnVlc nN0aW5uZW5rb3JyZXNwb25kZW56MS4x&cl=5&cr=5&s=0&l=10

Korpora Annotationen

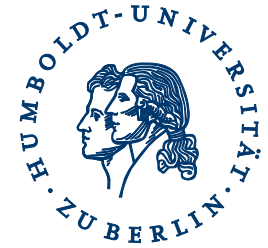


- Suche und Visualisierung diverser Annotationsarten
 - Tokenannotationen
 - Spannenannotation
 - Dependenzannotationen
 - Baumannotationen
 - **Filterung nach Metadaten**
 - alle Kombinationen daraus

➤ für die Analyse verschiedene
Visualisierungen

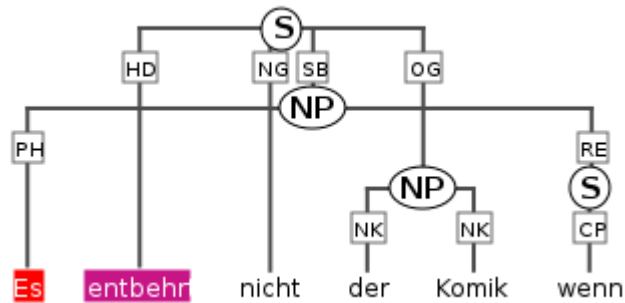


Visualisierungen Annotationen

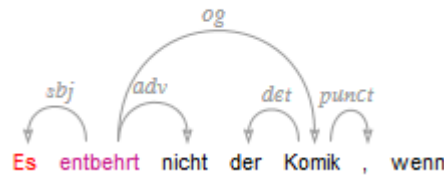


- z.B. ppc2, 11299

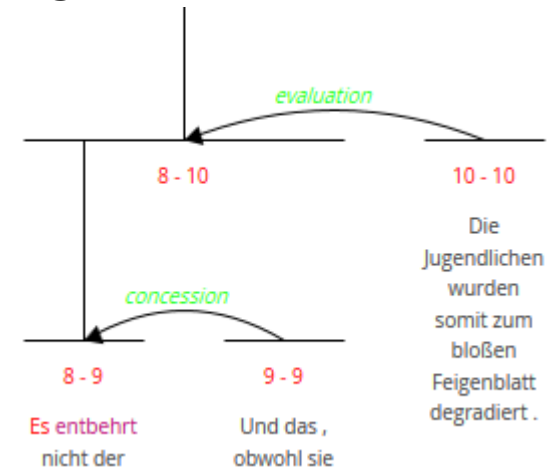
Tree



Dependencies



RST



Discourse

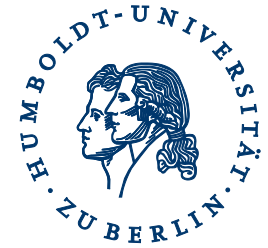
Und aberwitzig dazu . **Es entbehrt** nicht der Komik , wenn sich drei erbündnis/FDP) , Susanne Michler (CDU) und Joachim Zanow (SPD) die wollen und brauchen , ohne auf die Idee zu kommen , sie selbst zu

Grid

Inf-Stat	idiom				idiom		
NP	NP				NP		
Sent	s						
tok	Es	entbehrt	nicht	der	Komik	,	wenn

Such-Referenz: https://korpling.german.hu-berlin.de/annis3/#_q=cG9zPS9QLiiovICYgcG9zPS9WLkZJTl8gJiAjMiAtPmRlcFtmdW5jPSJzYmoiXSAjMQ&c=cGNjMg&cl=5&cr=5&s=0&l=10

Visualisierungen Dokumente



- z.B. Coptic Scriptorium (Zeldes & Schroeder to appear), Shenoute.a22
- <http://copticSCRIPTORIUM.org> [letzter Zugriff 20.10.2015]

Diplomatische Ansicht

Analytische Ansicht

ρῖηεζραῖεγ	περιτογερη
βαγκαλιον`	τεῖνογα`εἰν
ἡεζραῖεγφω	σομῖμοϛ`ε
φωγεζραῖ	μοοϛε`εει
5 εσελααγῖζνα	5 εβολ`εγε.ενε
αγ`εϛτῖτων	πζλλο αγω
εναῖ εμπογ	Χωρισναρην
ογερσαρνε	φηρεφνη
ναφριτῖπζλ	ἡσελααγῖρω
10 λω ειμητι	10 μεεγφανερ
πιαῖῖρωμε	ρατογεζραῖζῖν
ετφωνεῖ	λααγῖζνααγ
ματεῖῖνῖζλ	ἡτεῖῖνεεγ
λο`αγρζαζῖ	εφωπε`εγ

Help us to make ANNIS better! not logged in [Login](#)

Help/Examples

Base text ▾

1 2 3 4 5 / 4 Displaying Results 1 - 10 of 32 Result for: norm="νουτε"

1 Path: shenoute.a22 > a22.YA421-428 (norm_group 123 - 133) left context: 5 right context: 5

ζραι ρησγναγωγη τηρογ μπχοεις · ετρεπνουτε σωντ εγνοσ ιμνηθε ηκκτο μπεφρο

annotations (grid)

analytic view (document)

V ADV PREF ART N CONJ ADV PREF ART N ADV PREF ART N

ρῖη εζραι ε | γ | βαγκαλιον η εζραι ε | γ | φωφωγ εζραι ε | σε | λααγ

FREP N CCIRC PFERS VSTAT PREP PDEM CREL ANEGPST PFERS V N PREP PFERO

η | ζνααγ ε | φ | τῖτων ε | ναῖ ε | μπ | ογ | ογερ | σαρνε να | φ

FREP ART N CONJ ART N PREP ART N CREL V PREP V PREP ART N

ζιτι | π | ζλλω ειμητι π | μα | η | η | ρωμε ετ | φωνε μ | ματε μῖ | η | ζλλο

APST PFERS V PREP N CONJ PDEM N PREP ACAUS PFERS V PREP N PFERO

α | γ | ρζαζ η | ρουπε αγω νεῖ | κοογε ε | τρε | γ | φῖνε η | τωστ | φ

FREP ART N

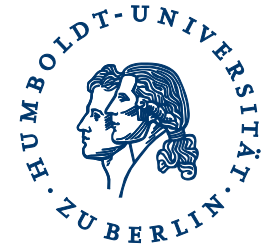
μ | π | ζλλο

... if one urinates into a vessel with a narrow neck or jar, into any other vessel like these, without having been ordered by the elder, except for the place of those who are very ill and the elders of advanced years and these others, he is to ask the elder.

Such-Referenz

https://corpling.uis.georgetown.edu/annis/scriptorium#_q=bm9ybT0i4rKb4rKf4rKp4rKn4rKJlg&_c=c2hlbm91dGUuYTYy&cl=5&cr=5&s=0&l=10&_seg=bm9ybV9ncm91cA

Ressourcen



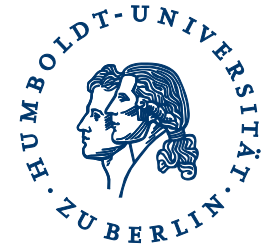
LAUDATIO

- Technische Dokumentation: <http://www.laudatio-repository.org/repository/technical-documentation/>
- Metadaten Spezifikationen: <https://github.com/korpling/LAUDATIO-Metadata>
- Repository : <http://www.laudatio-repository.org>

ANNIS

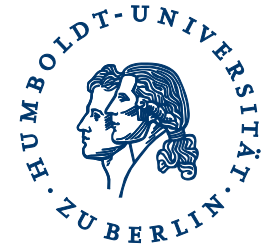
- Homepage <http://corpus-tools.org>
- Suche <https://korpling.german.hu-berlin.de/annis3/>

Ressourcen



- HU - Instanz
 - <https://korpling.german.hu-berlin.de/annis3/>
- Suche mit
 - Korpus-Referenz RIDGES https://korpling.german.hu-berlin.de/annis3/#_c=UklER0VTX0hlcmJvbG9neV9WZXJzaW9uNC4x
 - Korpus-Referenz Fürstinnen (Fürstinnenkorrespondenzkorpus) https://korpling.german.hu-berlin.de/annis3/#_c=RnVlcnN0aW5uZW5rb3JyZXNwb25kZW56MS4x
- für alle Referenz-Links
 - letzter Zugriff 20.10.2015

Kooperation



- Finanzierung
 - DFG-Projekt LAUDATIO
 - laudatio-repository.org
 - DFG-SFB 632 Informationsstruktur
 - ANNIS + SaltNPepper (corpus-tools.org)

- Kooperation zwischen

Humboldt-Universität
Berlin

Friedrich Schiller
Universität Jena

Georgetown University
Washington

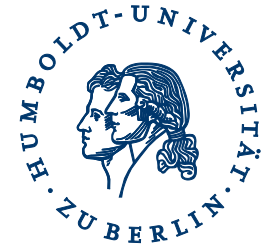


seit 1558



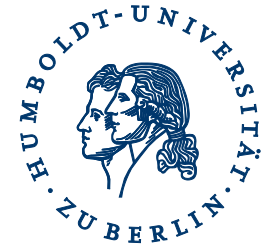
und viele weitere!

Referenzen



- **Bodmer, F.** (2005). COSMAS II. Recherchieren in den Korpora des IDS. Sprachreport, 21(3):2–5.
- **Geyken, A., Gloning, T.** (2015) A living text archive of 15th-19th-century German. Corpus strategies, technology, organization. In: Gippert, J., Gehrke, R. (Hg.): Historical Corpora. Challenges and Perspectives. Tübingen 2015, S. 165–180.
- **Hardie A.** (2012). CQPweb - Combining power, flexibility and usability in a corpus analysis tool. International Journal of Corpus Linguistics 2012;17(3):380-409.
- **Haynes, D.** (2004). Metadata for information management and retrieval. London: Facet publishing.
- **Krause, T., Zeldes, A.** (2014): ANNIS3: A new architecture for generic corpus query and visualization. in: Digital Scholarship in the Humanities 2014 <http://dsh.oxfordjournals.org/cgi/content/abstract/fqu057?ikey=GJBr0LhNfKW1g8i&keytype=ref> [letzter Zugriff: 20.10.15]
- **Lezius, W.** (2002) Ein Suchwerkzeug für syntaktisch annotierte Textkorpora Ph.D. thesis IMS, University of Stuttgart Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), volume 8, number 4.
- **Miller, S. J.** (2011). Metadata for Digital Collections. A How-To-Do-It Manual. New York; London: Neal-Schuman Publishers (How-To-Do-It Manuals, 179).
- **Neuroth, H., Rapp, A., Söring, S.** (Hg.) (2005) TextGrid: Von der Community — für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Universitätsverlag Göttingen, Verlag Werner Hülsbusch, Glückstadt.
- **Odebrecht, C., Belz, M., Zeldes, A., Lüdeling, A.** (eingereicht) RIDGES Herbology - Designing a Diachronic Multi-Layer Corpus. <https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiter-innen/anke/pdf/odebrechtetalridges-submitted.pdf> [letzter Zugriff 20.10.15]
- **Odebrecht, C.** (2015) Interdisziplinäre Nutzung von Forschungsdaten mithilfe einer technisch-abstrakten Modellierung. Vortrag. Von Daten zu Erkenntnissen. 2. Jahrestagung des Verbandes der Digital Humanities im deutschsprachigen Raum. 25.02.-27.02.2015, Graz.
- **Odebrecht, C., Krause, T.** (2013). Metadata in an Infrastructure for Historical Corpora. SFB 732 Incremental Specification in Context. Kolloquium. 20.06.2013, Stuttgart.

Referenzen



- **Rümpel**, S. (2011). Der Lebenszyklus von Forschungsdaten. In Büttner, St., Hobohm, H. & Müller, L. (Hg.) Handbuch Forschungsdatenmanagement. Bock und Herchen Verlag. Bad Honnef. 25-31.
- **Salmon-Alt**, S., Romary, L., Pierrel, J. (2006). Un modèle générique d'organisation de corpus en ligne : application à la FReeBank. *Traitement Automatique des Langues, ATALA*, 2006, 45, 145-169. <hal-00110970>
- **Sloetjes**, H., Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).
- **Schmidt**, T., Wörner, K., Hedeland, H., Lehmborg T. (2011). New and future developments in EXMARaLDA In Thomas Schmidt, Kai Wörner, eds., *Multilingual Resources and Multilingual Applications*. Proceedings of GSCL Conference 2011 Hamburg.
- **Zeldes**, A., Schroeder, C. T. (to appear). Computational Methods for Coptic: Developing and Using Part-of-Speech Tagging for Digital Scholarship in the Humanities. *Digital Scholarship in the Humanities*.
https://corpling.uis.georgetown.edu/amir/pdf/Computational_Methods_for_Coptic_prepub.pdf [letzter Zugriff 20.10.2015]
- **Zipser**, F. (2014). SaltNPepper und das Formatpluriversum. LAUDATIO Workshop 2014. Berlin, 07.- 08.10.2014.
<http://dx.doi.org/10.5281/zenodo.17557>
- **Zipser**, F., Romary, L. (2010). A Model Oriented Approach to the Mapping of Annotation Formats using Standards. Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC-2010 . Valletta, Malta, pp. 7–18.