



ANNIS Einführung RIDGES Herbology 5.0-beta1

Laura Perlitz und Carolin Odebrecht
Humboldt-Universität zu Berlin

Korpus RIDGES Herbology



- Wiederholung
 - Kräuterkundekorpus
 - 15.-20. Jahrhundert
 - verschiedene Dialekte aus dem deutschsprachigen Raum
 - Arbeit von vier Seminaren aus B.A. und M.A. von ca. 60 Studenten
 - Projekthomepage
 - <http://korpling.org/ridges/>
 - Annotationsrichtlinie
 - http://korpling.org/ridges/documentation_de/

ANNIS

Such- und Visualisierungstool



- webbasiertes Suchtool für Korpora
 - Projekthomepage: <http://corpus-tools.org/annis/>
 - Zugang auf das ANNIS-Tool: <https://korpling.org/annis3>
- Was kann ANNIS?
 - findet Annotationen in einem Korpus
 - findet mehrere Annotationen in bestimmter Relation zu einander
 - gibt die gesuchten Annotationen als Treffer in vielfältigen Visualisierungen aus
 - filtert nach Metadaten
 - exportiert Treffer
 - u.v.m.
- Hier im Seminar lernen wir mit dem Korpus RIDGES Version 5.0-beta1, wie man ANNIS benutzt!

Weitere Informationen zu ANNIS

- Zum Nachlesen:
 - Dokumentation: <http://corpus-tools.org/annis/documentation.html>
 - Nutzerhandbuch: [http://corpus-tools.org/annis/resources/ANNIS User Guide 3.4.3.pdf](http://corpus-tools.org/annis/resources/ANNIS_User_Guide_3.4.3.pdf)
 - Krause, Thomas and Zeldes, Amir (2016) "ANNIS3: A New Architecture for Generic Corpus Query and Visualization". *Digital Scholarship in the Humanities* 31(1), 118-139.
<http://dsh.oxfordjournals.org/cgi/content/abstract/fqu057?ijkey=GJBr0LhNfKW1g8i&keytype=ref>
- Zum Nachschauen (Video-Tutorials):
 - <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/corpus-tools/annis-tutorials>

Historische Korpora

Variation, Segmentierung

Historische Texte

- Varianz der Orthographie bzw. Setzung
- vieles nicht wirklich „vorhersehbar“
- graphische/(text-) strukturelle Informationen
 - Zeilenumbruch, Seitenumbruch, Setzung

Kräutern	<u>Alchymistische Practic</u> (1603)
Kraut	<u>Alchymistische Practic</u> (1603)
fraut	<u>Alchymistische Practic</u> (1603)
Kreutern	<u>Alchymistische Practic</u> (1603)
Kreutter	<u>Alchymistische Practic</u> (1603)
Kreüter	New <u>Kreüterbüch</u> (1543)
Kräuteren	<u>Pflantz-Gart</u> (1639)
Kreuter	<u>Alchymistische Practic</u> (1603)
Kräuter	Deutsche Pflanzennamen (1870)

Normalisierung

- Um alle Schreibvarianten eines Wortes in historischen Korpora finden zu können, sind Normalisierungen notwendig.
- In RIDGES gibt es daher drei Textebenen:
 - **dipl**: diplomatische Transkription
 - **clean**: Ersetzung aller heute unüblichen Sonderzeichen und Zusammenziehung von Wörtern, die durch einen Zeilenumbruch getrennt wurden (mit Bindestrich)
 - **norm**: Anpassung an die moderne Orthographie

Segmentierung I

- Annotationen in RIDGES (vgl. Belz et al. 2015)
 - dipl: diplomatische Transkription
 - clean: automatische Normalisierungen
 - norm: manuelle Normalisierung
 - lb: Zeilenumbrüche
 - lemma: Lemmatisierungen


clean	die	aus	drei	Worten	zusammengesetzt		sind
dipl	die	aus	drei	Worten	zufammen ^s	gefezt	find
lb	lb					lb	
lemma	d	aus	drei	Wort	zusammensetzen		sein
norm	die	aus	drei	Worten	zusammengesetzt		sind

Trefferreferenz: <https://korpling.org/annis3/?id=cb133ed8-ae93-4132-82d3-642ef9b92f58>

Belz, Malte; Odebrecht, Carolin; Perlitz, Laura & Voigt, Vivian (2015): Annotationsrichtlinien zu Ridges Herbiology Version 4.1, Humboldt-Universität zu Berlin.

Segmentierung II

- Segmentierung
 - mehrere primäre Textebenen in RIDGES: dipl, clean, norm
 - Jede dieser Ebenen besitzt eine eigenständige Einteilung in kleinste Einheiten (Segmentierung)
 - Bei der weiteren Annotation muss IMMER festgelegt werden, auf welcher Segmentierung sie sich bezieht!



clean	die	aus	drei	Worten	zusammengesetzt		sind
dipl	die	aus	drei	Worten	zufammen s	gefetzt	find
lb	lb					lb	
lemma	d	aus	drei	Wort	zusammensetzen		sein
norm	die	aus	drei	Worten	zusammengesetzt		sind

Trefferreferenz: <https://korpling.org/annis3/?id=cb133ed8-ae93-4132-82d3-642ef9b92f58>

Segmentierung III

Position dipl	1	2	3	4	5	6	7
Position norm	1	2	3	4	5	6	

Anzahl der
Einheiten
dipl=7
norm=6

clean	die	aus	drei	Worten	zusammengesetzt		sind
dipl	die	aus	drei	Worten	zufammen	gefetzt	find
lb	lb					lb	
lemma	d	aus	drei	Wort	zusammensetzen		sein
norm	die	aus	drei	Worten	zusammengesetzt		sind

Position dipl	1	2	3	4	5	
Position norm	1	2	3	4	5	6

Anzahl der
Einheiten
dipl=5
norm=6

clean	Den	Winter-spinet	sehr	grosz	zubekommen	
dipl	Den	Winter-spinet	lehr	grofz	zubekommen	
lb	lb					
lemma	d	Winterspinat	sehr	groß	zu	bekommen
norm	Den	Winterspinat	sehr	groß	zu	bekommen

Trefferreferenz 1: <https://korpling.org/annis3/?id=cb133ed8-ae93-4132-82d3-642ef9b92f58>

Trefferreferenz 2: <https://korpling.org/annis3/?id=4c675f63-5890-420e-b3a7-fd9e637b31a3>

Suche

Bitte öffnen!

- RIDGES-Dokumentation (Version 4.1, für den Kurs gibt es eine aktualisierte Annotationsrichtlinie):
http://korpling.org/ridges/documentation_de/
- STTS: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html>
- ANNIS-Instanz: <https://korpling.org/annis3/>

<https://korpling.org/annis3/>

ANNIS interface showing a list of corpora and example queries.

Corpus List

Name	Texts	Tokens		
a5.hausa.umarnin.uwa_v	47	10.194	i	
abraham.our.father	7	7.671	i	
apophthegmata.patrum.	11	1.500	i	
arabic.tree.test	1	11	i	
b2.hausa	50	6.991	i	
b4.tatian2.0	2.031	11.295	i	
b4.tatian2.1	2.030	11.295	i	
b7.wolof.web.V2	4	14.676	i	
b7.wolof.wiki.V4	14	12.737	i	
ReMaTaC 1.1 2013-01	12	11.192	i	

Example Queries

Example Query	Description	open corpus browser
pos="NN"	Findet einen Nomen.	RIDGES Herbiology Version4.1
lang="foreign"	search for foreign language (non-Hausa) segments within the Hausa text	a5.hausa.umarnin.uwa_V2
pos="NPR"		
norm="NOY"		
pos="VSTA"		m.11
norm="am"		m.11
morph="u"		
pos="PPER"		
norm="NKC"		
edition="e"		el
document		el
edition="e"		
document	Access all texts in the corpus for reading purposes	DDD-Genesis
edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-Heliand
document	Access all texts in the corpus for reading purposes	DDD-Heliand
edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-Isidor
document	Access all texts in the corpus for reading purposes	DDD-Isidor
edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-IsidorLatein
document	Access all texts in the corpus for reading purposes	DDD-IsidorLatein

Benutzer: histkorp
Passwort: 5220049

Interface

Startseite, Hilfestellung, Query Fenster

ANNIS Query Language



- oder: Wie sage ich ANNIS, was ich suche?
- für die Suche nach Annotationen gelten zwei Prinzipien:
 - Attribut-Wert-Paar (AW-Paar)
 - Relationen zwischen AW-Paaren

Prinzip I: Attribut-Wert-Paar

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

dipl= /kraut/



Attribut

(auch „Layer“, „Tier“, „Ebene“ ...)

Wert

(auch „Value“)



dipl	wider	auff	ein	frifches	zerftoffenes	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

Attribut-Wert-Paar

dipl= /kraut/



Screenshot of the ANNIS search interface showing results for the query `dipl=/kraut/`.

The interface includes a search bar, a corpus list, and a detailed view of search results with annotations.

Corpus List:

Name	Texts	Tokens		
Ridges-SE-CompAnalyse	20	92,368		
ridges.herbology	14	63,734		
RIDGES_Herbology_Version	22	122,698		
RIDGES_Herbology_Version	29	154,266		
RIDGES_Herbology_Version	29	154,267		
RIDGES_Herbology_Version	36	183,710		
Ridges_Herbology_Version	13	60,811		

Search Results:

1 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1915 - 1925) left context: 5 right context: 5

M. b. 1. gedör rets **kraut** / zerhacket es ein wenig

2 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 2466 - 2476) left context: 5 right context: 5

difillirte waffer vff ein frisches **kraut** / vnd difillir es wider

3 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 3870 - 3880) left context: 5 right context: 5

wider auff ein frisches zerftoffenes **kraut** / thue es wider in

4 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 4005 - 4015) left context: 5 right context: 5

Suchreferenz: <https://korpling.org/annis3/?id=8ebb8273-3048-4c06-b697-81a19573eae4>

Aufgabe

Schreibvarianten



- Finden Sie alle Schreib- und Flexionsvarianten von *Kraut*!
 - Für die Suche nach allen Varianten wählen Sie die passende Annotationsebene (Attribut) aus
 - Passendes Attribut wäre hier **lemma**.
 - und setzen den Wert **/Kraut/** ein.
 - lemma=/Kraut/

Aufgabe Schreibvarianten



Screenshot of the ANNIS interface showing search results for the query "lemma=/Kraut/".

Search Interface:

- Query: lemma=/Kraut/
- Search Results: 586 matches in 31 documents
- Corpus List:

Name	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_2.0	13	60,811
SMULTRON_Banana	2	3,782

Search Results:

Base text: Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 888 - 898)

Result for: lemma=/Kraut/

1. weilt vnd weg / aufz **Kräutern** / blettern / wurz tzelz

2. weg / viel waffer aufz **Kräutern** vff einmal zu diltillren .

3. / denn nimb was für **Kräuter** du wilt / zerlotz die

Suchreferenz: <https://korpling.org/annis3/?id=61d11111-124f-42e9-8a4d-7306417f9007>

Normalisierung

- Schauen Sie sich die Treffer zu lemma=/Kraut/genau an!

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 888 - 898)

weiß vnd weg / auß kräutern / blettern / wurzeln

Text Layers

dipl	weiß	vnd	weg	/	auß	kräutern	/	bletern	/	wurz	eln
clean	weis	vnd	weg	/	aus	kräutern	/	bletern	/	wurtzeln	
norm	Weisen	und	Wege	/	aus	Kräutern	/	Blättern	/	Wurzeln	

- Es werden alle historischen Wortformen von Kraut gefunden!

Trefferreferenz: <https://korpling.org/annis3/?id=2d1e7843-0bf8-45fb-b752-29f8f9415c75>

Aufgabe

Normalisierung



- Suchen Sie das Lemma von *zusammensetzen*!
 - Finden von ganz unterschiedlichen historischen Schreibvarianten!
 - lemma=/zusammensetzen/
- Treffer 1: <https://korpling.org/annis3/?id=3c470377-a87d-464f-82ce-5e2b7310809c>
 , die aus drei Worten **zufammen** **geletzt** find , wie die oben
- Treffer 2: <https://korpling.org/annis3/?id=e46cc32b-3f3c-472d-8d2e-e77b2e326c55>
 muß aus zwei einfachen Worten **zu** **fammengefetzt** fein , von denen der
- Treffer 3: <https://korpling.org/annis3/?id=83497961-4769-4382-91cc-6af6b6839c4d>
 Wenn der lateinische Artname **zufammengefetzt** ift , kann man entweder

Mustersuche

Operatoren

Operatoren für die Mustersuche



Für die Manipulation der **Annotationswerte** zu nutzen:

- . 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, z.B. spitze Klammern, Punkte)
- (a | b) a oder b (auch: [ab])

Für die Manipulation der **gesamten Anfrage** zu nutzen:

- ! nicht (steht direkt vor dem „=“ zwischen Attribut und Wert)

Möchte man nach Unicodes suchen, muss man diese mit einem Backslash \ maskieren, z.B. dipl=/\u002F/ (Hier wird nach „/“ gesucht).

Aufgabe

Operatoren



- 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, meine* (für RIDGES)

Aufgabe

Wortarten I



- Wie finden Sie alle Appellativa in RIDGES?
 - Passendes Attribut wäre hier **pos**.
 - pos=/NN/

Aufgabe Wortarten I



ANNIS interface showing search results for the query `pos=/NN/`. The interface includes a search bar, a corpus list, and a detailed view of the search results.

Search Bar: `pos=/NN/`

Search Results: 33628 matches in 36 documents.

Corpus List:

Name	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_3.0	12	60,811

Search Results Detail:

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1 - 8)

Result for: `pos=/NN/`

1 **An den Lefer** . ES lñndt mir lieber

2 **. ES lñndt mir lieber Lefer** / zwey feine Alchy^m miltliche

3 **/ zwey feine Alchy^m miltliche Tractätlein** von guten freunden zukorhen /

Suchreferenz: <https://korpling.org/annis3/?id=11b66562-075d-420b-bae1-8062d076d9c4>

Aufgabe

Wortarten II



- Wie finden Sie alle Adjektive unabhängig von ihren Bezugswörtern in RIDGES?
 - Was sagen die Richtlinien (STTS)?
 - Passendes Attribut wäre hier **pos.**
 - STTS: Unterschied ADJA und ADJD?
 - pos=/ADJ./

Aufgabe Wortarten II



Screenshot of the ANNIS interface showing search results for the query `pos=/ADJ./`.

The interface includes a top navigation bar with links like "About ANNIS", "Report Problem", and "Help us to make ANNIS better!". The user is logged in as "laura".

The search results are displayed in a table with columns: Name, Texts, Tokens, and a checkbox. The table lists various corpora, with **RIDGES_Herbology_Version 36** highlighted.

Name	Texts	Tokens	
Ridges-SE-CompAnalyse	20	92,368	
ridges.herbology	14	63,734	
RIDGES_Herbology_Versior	22	122,698	
RIDGES_Herbology_Versior	29	154,266	
RIDGES_Herbology_Versior	29	154,267	
RIDGES_Herbology_Versior	36	183,710	
Ridges_Herbology_Version	13	60,811	

The search results are also displayed in a list view, showing the path: `RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 3 - 13)`. The results are filtered by the query `pos=/ADJ./`.

The list view shows three results, each with a path and a list of annotations. The first result is:

1 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 3 - 13)

left context: 5 right context: 5

Lefer . ES findt mir lieber Lefer / zwey feine Alchy

The second result is:

2 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 7 - 17)

left context: 5 right context: 5

mir lieber Lefer / zwey feine Alchy militiche Tractätlein von guten

The third result is:

3 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 8 - 19)

left context: 5 right context: 5

lieber Lefer / zwey feine Alchy militiche Tractätlein von guten freunden zukoffen

Suchreferenz: <https://korpling.org/annis3/?id=9c15ac7c-5053-4586-be0e-5d2587b2e18e>

Aufgabe

Suchen von Annotationsebenen



- Wie finden Sie heraus, ob es in RIDGES Herbology Annotationen für Überschriften gibt?
 - Wie sieht eine Anfrage aus, um Überschriften zu finden?
 - Passendes Attribut wäre hier **head**.
 - head=/head/

Aufgabe

Suchen von Annotationsebenen



The screenshot shows the ANNIS web interface. On the left, there is a search bar with the query 'head=/head/'. Below it, a table lists corpora with their respective text and token counts. The 'RIDGES_Herbology_Version5.0-beta1' corpus is highlighted. The main area displays search results for the query, showing the path 'RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1 - 9)'. The results are displayed in a list format, showing the text 'An den Leler . ES findt mir lieber Leler' and 'namen zum belten brauchen . TRACTATVS CHYMISTARVM DE DESTILLATIONIBVS & extractionibus . IM diltilln / ja in'. The interface also includes a sidebar with a list of annotation layers (Text Layers, Lexical Annotation, Morphological Annotation, Syntactic Annotation, Graphical Annotation, Content Annotation, Other Annotation, All Annotation, Diplomatic Text, Normalized Text) and a bottom section with a list of corpora.

Corpus	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_2.0	13	60,811
SMULTRON_Banana	2	3,782

Suchreferenz: <https://korpling.org/annis3/?id=03ad8de5-6570-443d-b9b8-39bc391ed7ed>

Aufgabe

Kombinierte Suche



- Finden Sie die diplomatische Wortform *Wermut*, die in einer Überschrift vorkommt!
 - Passende Attribute wären hier **head** und **dipl**.

Prinzip II: Relationen

head=/head/ **AW-Paar1**

o

Relation

dipl=/Wermut/ **AW-Paar2**

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

/ ABSYNTHIUM MON- TANUM . Pontifcher Wermut / ABSYNTHIUM PONTICUM . Meer Wermut / ABSYNTHIUM MARI-

☐ Text Layers
☐ Lexical Annotation
☐ Morphological Annotation
☐ Syntactic Annotation
☐ Graphical Annotation

dipl	/	ABSYNTHIUM	MON-	TANUM	.	Pontifcher	Wermut	/	ABSYNTHIUM	PONTICUM	.	Meer
lb	lb			lb			lb					lb
pb	pb											pb
pb_n	305											306
head	head					head						head

Trefferreferenz: <https://korpling.org/annis3/?id=d01793c8-457e-4f02-b1dd-6b1c96192bc1>

Aufgabe Überlappung I



ANNIS interface showing search results for the query `head=/head/` and `dipl=/Wermut/`. The interface includes a Query Builder, Search Options, and a Corpus List.

Search Results:

- 7 matches in 2 documents
- Path: RIDGES_Herbology_Version5.0-beta1 > Phythologia_1662_Becher (dipl 1 - 9)
- Wermut / ABSYNTHIUM . Bergwermut / ABSYNTHIUM MON- TANUM
- Path: RIDGES_Herbology_Version5.0-beta1 > Phythologia_1662_Becher (dipl 6 - 21)
- / ABSYNTHIUM MON- TANUM . Pontifcher Wermut / ABSYNTHIUM PONTICUM . Meer Wermut / ABSYNTHIUM MARI-

Corpus List:

Name	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034

Suchreferenz: <https://korpling.org/annis3/?id=6d1e0e1b-f7ad-45cc-b833-27eab37e03f6>

Syntax-Highlighting

- pro AW-Paar eine Farbe – hier unser Beispiel:
 - Farbe **Rot** für alle Werte des Attributs **head**
 - Farbe **Lila** für alle Werte des Attributs **dipl**

```
head=/head/  
_o_  
dipl=/Wermut/
```

- Treffer erhalten genau diese Farben:
 - Mehrere dipl-Einheiten **rot**, da sie zusammen in einer Überschrift stehen.
 - Eine dipl-Einheit **lila**, da dieses der gesuchte dipl-Wert .

```
Pontifcher Wermut / ABSYNTHIUM PONTICUM .
```

Aufgabe

Überlappung II



- Suchen Sie ein Verb, das in einer Überschrift vorkommt!
 - Passende Attribute wären **pos** und **head**.
 - Operator **_o_**
 - `pos=/V.*/` **_o_** `head=/head/`

Aufgabe Überlappung II



Screenshot of the ANNIS (Annotation Network for German) interface showing search results for the query `pos=/V.*/`.

Search Query: `pos=/V.*/`
Result: 152 matches in 14 documents

Corpus List:

Name	Texts	Tokens	Info
Ridges-SE-CompAnalyse	20	92,368	Info
ridges.herbology	14	63,734	Info
RIDGES_Herbology_Version 22	22	122,698	Info
RIDGES_Herbology_Version 29	29	154,266	Info
RIDGES_Herbology_Version 29	29	154,267	Info
RIDGES_Herbology_Version 36	36	183,710	Info
Ridges_Herbology_Version 13	13	60,811	Info

Search Results (Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 510 - 555))

Result 1: `von difti ller Gläsern . Difz nennet man einen Kolben / es fey von einem Glafz oder andern Materien / vnd derfelbē mutz man klein / grofz / h`

Result 2: `von difti ller Gläsern . Difz nennet man einen Kolben / es fey von einem Glafz oder andern Materien / vnd derfelbē mutz man klein / grofz / h`

Result 3: `von difti ller Gläsern . Difz nennet man einen Kolben / es fey von einem Glafz oder andern Materien / vnd derfelbē mutz man klein / grofz / h`

Suchreferenz: <https://korpling.org/annis3/?id=6d1e0e1b-f7ad-45cc-b833-27eab37e03f6>

Aufgabe

Identität I



- Suchen Sie alle normalisierten Formen von dem Vollverb „setzen“, die infinit sind.
- Passende Attribute wären **norm** und **pos**.
 - Operator **_=_**
 - norm=/setzen/**_=_** pos=/VVINF/

Aufgabe Identität I



ANNIS interface showing search results for the query `norm=/setzen/` and `_pos=/VVINF/`. The interface includes a search bar, a corpus list, and a detailed view of the search results.

Search Query: `norm=/setzen/`
`_pos=/VVINF/`

Search Results: 11 matches in 8 documents

Corpus List:

Name	Texts	Tokens	Info	Download
Ridges-SE-CompAnalyse	20	92,368	Info	Download
ridges.herbology	14	63,734	Info	Download
RIDGES_Herbology_Version	22	122,698	Info	
RIDGES_Herbology_Version	29	154,266	Info	
RIDGES_Herbology_Version	29	154,267	Info	Download
RIDGES_Herbology_Version 36	183,710	183,710	Info	Download
Ridges_Herbology_Version	13	60,811	Info	

Search Results:

1 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 418 - 428) left context: 5 right context: 5

nicht verdrieffen laßen etwas darzu **zufetzen** / vnd wie viel ich

2 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 2815 - 2825) left context: 5 right context: 5

feu lung eines Balnei Roris **letzen** / welches derwegen ein Balneum

3 Path: RIDGES_Herbology_Version5.0-beta1 > ArtzneyBuchleinDerKreutter_1532_Tallat (dipl 2248 - 2258) left context: 5 right context: 5

vnd widder zu dem fewer **letzen** vnd wol fchaumen / vnd

Suchreferenz: <https://korpling.org/annis3/?id=2f10fab7-70a8-4047-861e-48766a301edb>

Aufgabe Identität I



Beachte: Am ersten Treffer wird deutlich, wie wichtig mehrere Segmentierungen und Normalisierungen sind. Deshalb muss man sich für eine Suchanfrage genau überlegen, auf welcher Annotationsebene (Attribut) man was genau suchen möchte.

1 ⓘ ↶ Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 418 - 428)

nicht verdrieffen laffen etwas darzu zufetzen / vnd wie viel ich

☐ Text Layers

dipl	nicht	verdrieffen	laffen	etwas	darzu	zufetzen		/	vnd	wie	viel	ich
clean	nicht	verdriessen	lassen	etwas	darzu	zusetzen		/	vnd	wie	viel	ich
norm	nicht	verdrießen	lassen	etwas	dazu	zu	setzen	/	und	wie	viel	ich

☐ Lexical Annotation

norm	nicht	verdrießen	lassen	etwas	dazu	zu	setzen	/		und	wie	viel	ich
lemma	nicht	verdrießen	lassen	etwas	dazu	zu	setzen	/		und	wie	viel	ich
pos	PTKNEG	VVINF	VVFIN	PIS	PAV	PTKZU	VVINF	/		KON	KOUS	PIS	PPER
pos_klein	PTKNG	VINF	VFIN	N	ADV	PTKZU	VINF	ZEICHEN	KOORD	PWAV	N	N	

Trefferreferenz: <https://korpling.org/annis3/?id=a0a4ab8c-695d-4523-843c-4037a13c9775>

Aufgabe

Identität II



- Suchen Sie Partikelverben, die die Partikel *auf* beinhalten!
 - Passende Attribute wären **lemma** und **pos**.
 - Operator **_=_**
 - `pos=/VV.*/` **_=_** `lemma=/auf.*/`

Aufgabe Identität II



ANNIS interface showing search results for the query `pos=/VV.* / _= lemma=/auf.* /`.

Search Results Summary:

Corpus	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_2.0	13	60,811
SMULTRON_Banana	2	3,782

Search Results Details:

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1350 - 1361)

herüber lind / muß man auff hören / das Fewer laffen abgehen

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1469 - 1479)

herüber fein / zu ditilliren auffgehört werden . Es were dann

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1526 - 1536)

würd / ilt es zeit auffzuhören / vnd das Fewer laffen

Aufgabe

Identität II



- Suchen Sie Partikelverben, die die Partikel *auf* beinhalten!
 - passend wären **lemma** und **pos**
 - Operator **_=_**
 - `pos=/VV.*/ _=_ lemma=/auf.*/`
- Was findet man damit nicht?

Aufgabe

Direkte Präzedenz



- Suchen Sie einen Artikel, der ein Nomen direkt präzediert!
 - Passendes Attribut wäre **pos**.
 - Operator .
 - pos=/ART/ . pos=/N./
- Achtung! Hier wird der Punkt-Operator in zwei verschiedenen Funktionen verwendet – sowohl für die Manipulation der Annotationswerte als auch als Relationsoperator zwischen Attribut-Wert-Paaren.

Aufgabe

Direkte Präzedenz



ANNIS interface showing search results for the query `pos=/ART/ . pos=/N./`. The interface includes a search bar, a corpus list, and a detailed view of the search results.

Search Results:

- 11113 matches in 36 documents
- Corpus List: Search Options
- Visible: All
- Filter: Name, Texts, Tokens

Name	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_2.0	13	60,811

Query Result:

Base text: `pos=/ART/ . pos=/N./`

Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 1 - 8)

Result for: `pos=/ART/ . pos=/N./`

1. `An den Lefer . ES findt mir lieber`

2. `findet . Ich hab mir das judicium Athenienfum von Hippocrate belieben laffen`

3. `zu Gottes ehr / und deß Nechften fonderlich Deutcher Nation , nutz`

Suchreferenz: <https://korpling.org/annis3/?id=d91c7647-4bc2-4351-b91f-ca2f01df90a6>

Aufgabe

Metadaten



- Finden Sie heraus, ob die normierte Wortform *Blume* in einem Text aus dem Jahr 1543 zu finden ist!
- Wenn ja, wie oft?
 - Passende Attribute wären **norm** und **meta::datum**.
 - Verknüpfungszeichen **&**
 - `norm=/Blume/ & meta::datum=/1543/`
 - 2 Treffer

Aufgabe Metadaten



Query Builder

norm=/Blume/
&
meta::datum=/1543/

Q Search More History

2 matches in 1 document

Corpus List Search Options

Visible: All

Name	Texts	Tokens
MHD_context	4	2,760
NoSta-D-1.4-bematac	22	25,934
NoSta-D-1.4-falko	10	6,034
NoSta-D-Anselm	2	2,710
NoSta-D-Kafka	2	10,388
NoSta-D-TueBaDZ	2	10,832
NoSta-D-Unicum	2	11,312
pcc2	2	399
ridges.herbology	14	63,734
RIDGES_Herbology_Version3.0	22	122,698
RIDGES_Herbology_Version4.0	29	154,266
RIDGES_Herbology_Version4.1	29	154,267
RIDGES_Herbology_Version5.0-beta1	36	183,710
Ridges_Herbology_Version_2.0	13	60,811
SMULTRON_Banana	2	3,782

Help/Examples Query Result

Base text

Path: RIDGES_Herbology_Version5.0-beta1 > NewKreuterbuch_1543_Fuchs (dipl 1897 - 1907)

left context: 5 right context: 5

. Jn ei ner yeden blümen leind die drey erite bletter

Text Layers

dipl	Jn	ei	ner	yeden	blümen	leind	die	drey	erite	bletter
clean	Jn	einer	yeden	blumen	seind	die	drey	erste	bletter	
norm	In	einer	jeden	Blume	sind	die	drei	ersten	Blätter	

Lexical Annotation
Morphological Annotation
Syntactic Annotation
Graphical Annotation
Content Annotation
Other Annotation
All Annotation
Diplomatic Text
Normalized Text

Path: RIDGES_Herbology_Version5.0-beta1 > NewKreuterbuch_1543_Fuchs (dipl 3837 - 3847)

left context: 5 right context: 5

Iporn / darumb das feine blüm einem rittersporn nit vngleich ist

Text Layers

dipl	Iporn	/	darumb	das	feine	blüm	einem	rittersporn	nit	vngleich	ist
clean	Rittersporn	/	darumb	das	seine	blum	einem	rittersporn	nit	vngleich	ist
norm	Rittersporn	/	darum	dass	seine	Blume	einem	Rittersporn	nicht	ungleich	ist

Lexical Annotation
Morphological Annotation
Syntactic Annotation
Graphical Annotation
Content Annotation
Other Annotation
All Annotation
Diplomatic Text
Normalized Text

Suchreferenz: <https://korpling.org/annis3/?id=932395db-fa1e-44a8-b993-4aa86c292221>

Tipps



- Die Trefferliste zeigt im Kontext nur die Annotationsebenen (Attribute) an, die auch genau in diesem Trefferkontext annotiert worden sind!
 - Sie können den linken und rechten Kontext immer manuell einstellen, also vergrößern oder verkleinern!
- Schauen Sie in die Korpusmetadaten und Annotationsrichtlinien, um zu wissen, welche Annotationsebenen im Korpus vorhanden sind!
- Nicht alle Dokumente in einem Korpus müssen die gleichen (Anzahl und Typ) Annotationsebenen besitzen!

Zusammenfassung



- Suche in ANNIS basiert auf
 - **Attributen** (Annotationsebenen) und **Werten** (Kategorien in den Annotationsebenen), z.B.:
 - Suche von **exakten Werten**, z.B. pos=/ADJA/
 - Suche von **Mustern**, z.B. pos=/ADJ./
 - Attribut-Wert-Paare könne in **Relation** zu einander gesucht werden, z.B.:
 - Annotationen **überlappen** sich
 - Annotationen decken einen **identischen Bereich** ab
 - Attribut-Wert-Paare können in Verbindung mit **Metadaten** gesucht werden, z.B.:
 - pos=/ADJA/ & meta::datum=/1870/

Vielen Dank!

wichtige Operatoren für Mustersuche (AW-Paar)



- . 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 (steht direkt vor dem „=“ zwischen Attribut und Wert)
- [abc] Menge (oder $[\wedge abc]$ =alles außer abc)
- (a | b) a oder b (auch: [ab])
- a{2,3} a 2 bis 3mal

wichtige Relationen zwischen AW-Paaren



- . Direkte Präzedenz
- .* Indirekte Präzedenz mit beliebigem Abstand
- .1,2 Entweder direkte Präzedenz oder ein Token dazwischen (beliebig erhöhbar, z.B. .1,4 oder auch .2,8)
- _=_ Identische Abdeckung
- _i_ Inklusion
- _o_ Überlappung
- _l_ linksseitige Überlappung
- _r_ rechtsseitige Überlappung

Übungsaufgaben in ANNIS

- Nach ein paar Aufgaben für Sie zum Üben!

Aufgabe

Inhaltliche Annotationen



- Suchen Sie technische Begriffe und Erklärungen, die ein getrennt geschriebenes Kompositum beinhalten!

Aufgabe

Inhaltliche Annotationen



- Suchen Sie technische Begriffe und Erklärungen, die ein getrennt geschriebenes Kompositum beinhalten!
 - Passende Attribute wären **term**, **komp** und **komp_orth**.

Aufgabe

Inhaltliche Annotationen



- Suchen Sie technische Begriffe und Erklärungen, die ein getrennt geschriebenes Kompositum beinhalten!
 - Passende Attribute wären **term**, **komp** und **komp_orth**.
 - Operatoren **_o_** **_=_**
 - **term**=/t/ **_o_** **komp**=/k/ **_=_** **komp_orth**=/gtr/

Aufgabe

Inhaltliche Annotationen



ANNIS interface showing search results for the query: `term=/t/_o_ komp=/k/_=_ komp_orth=/gtr/`. The interface displays 25 matches in 6 documents.

Visible: All

Name	Texts	Tokens	
Maerchenkorpus	211	295.880	
Mercurius	2	187.423	
MHD_context	4	2.760	
NoSta-D-1.4-bematac	22	25.934	
NoSta-D-1.4-falko	10	6.034	
NoSta-D-Anselm	2	2.710	
NoSta-D-Kafka	2	10.388	
NoSta-D-TueBaDZ	2	10.832	

Search Results:

1 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 681 - 692)
left context: 5 right context: 5
ding zu Alchen brennet / Reuerberir Ofen / darin alle ding mit

2 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 2981 - 2992)
left context: 5 right context: 5
denn laffe den Buttner ein Eychen Fatzlein machen / mit einem Boden

Suchreferenz: <https://korpling.org/annis3/?id=f7e2cc64-2800-4ba2-9830-f083fc1770af>

Aufgabe Abfolgen



- Suchen Sie einen Artikel direkt gefolgt von einem attribuierenden Adjektiv direkt gefolgt von einem Nomen!

Aufgabe Abfolgen



- Suchen Sie einen Artikel direkt gefolgt von einem attribuierenden Adjektiv direkt gefolgt von einem Nomen!
 - Passendes Attribut wäre **pos.**

Aufgabe Abfolgen



- Suchen Sie einen Artikel direkt gefolgt von einem attribuierenden Adjektiv direkt gefolgt von einem Nomen!
 - Passendes Attribut wäre **pos.**
 - Operator .
 - **pos**=/ART/ . **pos**=/ADJA/ . **pos**=/N./

Aufgabe Abfolgen



Screenshot of the ANNIS interface showing search results for the query `pos=ART/ . pos=ADJA/ . pos=N./`.

Search Results Summary:

Visible:	All
ridges	
Name	Texts Tokens
Ridges-SE-CompAnalyse	20 92,368
ridges.herbology	14 63,734
RIDGES_Herbology_Version	22 122,698
RIDGES_Herbology_Version	29 154,266
RIDGES_Herbology_Version	29 154,267
RIDGES_Herbology_Version 36	183,710
Ridges_Herbology_Version	13 60,811

Search Results Details:

1 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 190 - 202) left context: 5 right context: 5

publication zuwider fein / da3 der gemeine Deutsche nicht verthehe wie gelehrt fie

2 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 541 - 553) left context: 5 right context: 5

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

3 Path: RIDGES_Herbology_Version5.0-beta1 > AlchymistischePractic_1603_Libavius (dipl 594 - 607) left context: 5 right context: 5

ein Kugel / vberlich mit einem lan gen Halz / werden Vio len genennet

Suchreferenz: <https://korpling.org/annis3/?id=e39f54cb-8a90-47d0-9a09-830c12022c32>