

COSMAS II Hilfe – Übersicht

Siehe auch die Onlinehilfe <http://www.ids-mannheim.de/cosmas2/web-app/hilfe/suchanfrage/>.

Lemmasuche: Die Default-Interpretation einer Worteingabe ist die individuelle Wortform, d.h. **Arzt** findet nur *Arzt* aber nicht *Arztes*. Man muss aber das Verhalten bzgl. Groß-/Kleinschreibung beachten (**Arzt** findet auch *arzt*, *ARZT*, *Ärzt* usw.)! Für die Lemmasuche muss der **Grundformoperator** & dem Wort unmittelbar vorangestellt werden, z.B. **&Arzt** (Suche nach literalem ‘&’ mit ***&**). Spezielle Optionen für die Lemmasuche können in den Suche-Optionen (siehe Randmenüpunkt **Lemmatisierung**) eingestellt werden.

Wortartensuche ist nur möglich in den getaggten Archiven. Die Texte in TAGGED-T sind mit dem STTS-Tagset **Schiller et al. (1999)** annotiert. Der MORPH-Assistent hilft, die Anfrage zusammenzustellen, z.B. **ein /w0 MORPH(PTK vz)** sucht nach der trennbaren Verbpartikel *ein* (PTKVZ).

Groß-/Kleinschreibung: In der Default-Einstellung werden Groß-/Kleinschreibung und diakritische Zeichen ignoriert, d.h. *Nähe* sucht nicht nur nach *Nähe*, sondern auch nach *NAHE*, *nahe* usw. Sie können die Grundeinstellung über die allgemeinen Einstellungen in den **Optionen (zur Suche)** ändern. Optional wird die Groß-/Kleinschreibung am Wortanfang, für den Rest des Wortes oder für das ganze Wort beachtet. Wenn aktiviert, kann ein einzelner Suchausdruck mit vorangestelltem **\$** davon ausgenommen werden, z.B. **\$ein** findet auch bei aktiviertem Groß-/Kleinschreibungfilter sowohl *ein* als auch *Ein*.

Satzzeichen und andere Sonderzeichen müssen in Anführungszeichen eingegeben werden (Ausnahme: Für die Suche nach Fragezeichen wird davor ein “Backslash” gesetzt: `\?`, ebenso vor den Backslash selbst `\\`, sowie vor `*` und `+`). Die Wörter *und*, *oder*, *nicht*, *and*, *or*, *not* werden als Operatoren interpretiert, es sei denn sie stehen in Anführungszeichen

Beispielanfragen mit COSMAS II

Suchanfrage	Ergebnis
&Arzt	Arzt, Arztes, Ärzte, ... (alle flektierten Formen von “Arzt”) Wenn Lemmatisierungs-Optionen gesetzt, auch: “Ärzte”-Song, Allgemeinärztes, ...
Arzt	nur die Wortform “Arzt” – weitere Formen in Abhängigkeit von den Options-Einstellungen
Arzt*	Arzt, Arztbesuch, Arztberuf, Arzt/Patient, ... Platzhalter für einzelnes Zeichen: <code>?</code> , z.B. Ra? = Rad, Rat, Rau, ...
*arzt	Hausarzt, Zahnarzt, Zoo-Arzt, ...
&gut /+w1 &Arzt	guter Arzt, besten Ärzten, ... Lemmatisierung: nur Flexionsformen aktiviert Abstandsoperatoren: Zeit /+w2 Geist (“Geist” steht 0-2 Wörter nach “Zeit”) Zeit /+w2:2 Geist /“Geist” steht genau 2 Wörter nach “Zeit”) Zeit /-w2:5 Geist (“Geist” steht 2-5 Wörter vor “Zeit”) Analog: satzbezogen: <code>/s</code> , paragraphenbezogen: <code>/p</code> . Siehe auch die ausschließenden Abstandsoperatoren: z.B. <code>%w2</code> .
(&d- /+w1 &gut) /+w1 &Beispiel	Das gute Beispiel, die besten Beispiele, ... Lemmatisierung: nur Flexionsformen aktiviert Obligatorische Klammern markieren den Bezugsbereich bei mehr als einem Abstandsoperator.
das gute Beispiel	das gute Beispiel, Das gute Beispiel

Suchanfrage

Kanzlerin /+w2 Merkel

Kanzlerin oder Merkel

(Kanzlerin %s0 Merkel) oder
(Merkel %s0 Kanzlerin)

Kanzlerin %s0 Merkel

MORPH(N ne) /w0 Herzog

Ägide /+w3 MORPH(N ne)

MORPH(N nn) /w0 *zeit

&üben /+w10 (aus /w0
MORPH(PTK vz))

&üben /+s0 (aus /w0 MORPH(PTK
vz))

(&üben /+w10 (aus /w0
MORPH(PTK vz))) #IN <s>

(&reden /w0 MORPH(VRB fin v))
/+s0 (aus /w0 MORPH(PTK vz))

(&sein /w0 MORPH(VRB a)) /+s0
(MORPH(VRB pp) /w1:1 worden)

","

(MORPH(N ne) /+w1 folgend)
/w0", "

"\$"

Ergebnis

Kanzlerin Merkel, Kanzlerin Angela Merkel ("Kanzlerin" und "Merkel" im Abstand von höchstens einem Wort)

Entweder "Kanzlerin" oder "Merkel" (oder beides) **im Text**.

Entweder nur "Kanzlerin" oder nur "Merkel" im Satz

"Kanzlerin" ohne "Merkel" im selben Satz

Roman Herzog, Werner Herzog, Ed Herzog, ...

Ägide Huub Stevens, Ägide von Premierministerin Magaret "Maggie" Thatcher, ...

Weihnachtszeit, ... (Substantive, die auf "-zeit" enden)

... übt ein Wahlamt aus ... (ermittelt Präfixverben) **Problem:** Ergebnis ist nicht zwingend auf einen Satz beschränkt.

Lösung 1 für das Problem oben.

Lösung 2 für das Problem oben mit dem **Kombinationsoperator** #IN und der Satzannotation < s >.

"reden" als finites Vollverb gefolgt von einem trennbaren Präfix "aus" im selben Satz.

Phrasensuche mit drei Wortformen: Auxiliar "sein", gefolgt einem Partizip und "worden" im selben Satz.

Suche nach Komma (siehe Stichwort "Satzzeichen" oben).
Zeigt als Treffer das Komma mit dem vorangehenden Token.

Phrasensuche "NE folgend, ".

Suche nach \$.

Die Cosmas II -Anfragen sind nach dem Vorbild der Anfragen auf der DWDS-DDC-Hilfe formuliert, vgl. <http://www.dwds.de/help/ddc/>.

Literatur

Anne Schiller, Simone Teufel, Christine Stöckert, and Christine Thielen. Guidelines für das Tagging deutscher Textcorpora mit STTS. Technical report, Institut für maschinelle Sprachverarbeitung, Stuttgart, 1999. URL <http://www.ims.unistuttgart.de/projekte/corplex/TagSets/stts-1999.pdf>.