

LEIBNIZ-GEMEINSCHAFT,
INSTITUT FÜR DEUTSCHE SPRACHE, MANNHEIM
COSMAS II – DIGITALE RECHERCHE
IN FÜNF MILLIONEN BUCHSEITEN

AUF DER SUCHE NACH DEM ROTEN FADEN IN DER DEUTSCHEN SPRACHE?

Am Institut für deutsche Sprache kann man die weltweit größte Sammlung elektronischer deutschsprachiger Texte aus der Gegenwart und der neueren Vergangenheit finden. Würde man sie ausdrucken, so würde sie fast fünf Millionen Buchseiten umfassen.

Für das Durchsuchen und die Analyse dieser Textsammlung wurde am IDS die Software COSMAS entwickelt. Mit COSMAS kann man u.a. das erste Auftauchen der Redensart vom **roten Faden** in Goethes „Wahlverwandtschaften“ aufspüren und dessen spätere Verbreitung verfolgen, die Verdrängung des deutschen Genitivs durch den Dativ in der geschriebenen Sprache beobachten und die unerschöpfliche *Wortzusammensetzungskreativität* unserer Sprache sowie den Ausdruck gesellschaftlicher Trends anhand von Wörtern wie „Faden“ bestaunen (siehe mittleres Plakat).

COSMAS stellt ein wichtiges Werkzeug für die empirische linguistische Forschung dar. Aber auch wer sich als Nichtwissenschaftler für die deutsche Sprache interessiert, kann mit COSMAS den Wandel der geschriebenen deutschen Sprache miterleben.

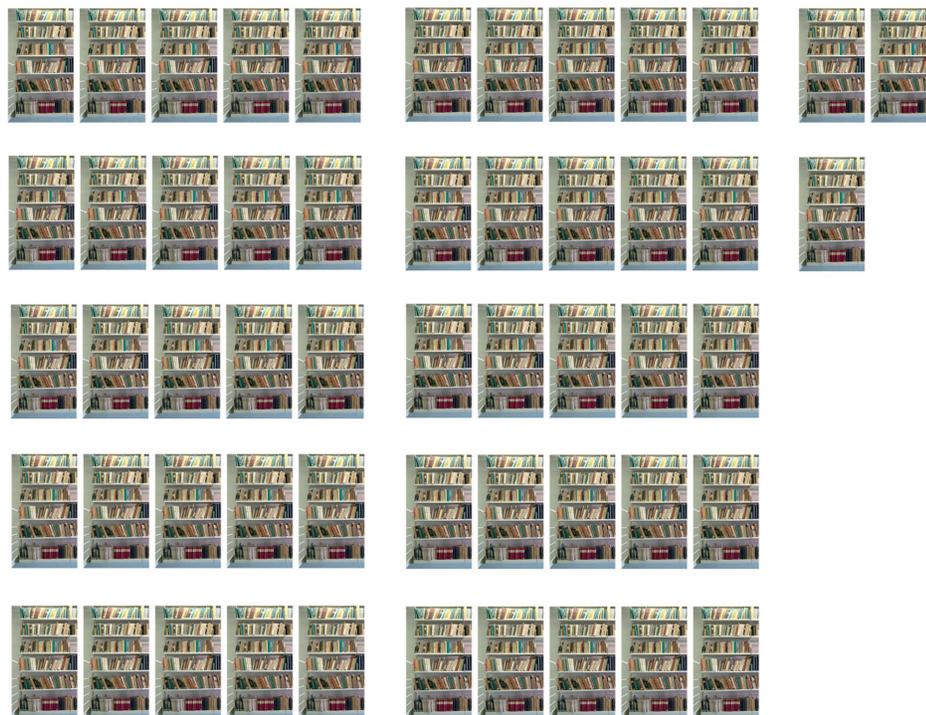
Die Textsammlung des IDS umfaßt Zeitungen, Sach-, Fach- sowie schöngestige Literatur aus Deutschland, Österreich und der Schweiz von 1772 bis heute.

Der Wissenschaftssommer bietet die Gelegenheit, COSMAS kennen zu lernen und dessen Potential durch eigene Recherchen selbst zu entdecken.



Die über COSMAS recherchierbare elektronische Textsammlung des Instituts für Deutsche Sprache umfasst zurzeit umgerechnet ca. 5 Mio Buchseiten.

Nebeneinander wären es über 300m
oder 53 dieser Bücherwände.



VOM WORT ZUR VIELFALT MIT COSMAS II

automatische Erzeugung der Flexionsformen und Wortzusammensetzungen von "Faden"

IDS INSTITUT FÜR DEUTSCHE SPRACHE

Suchanfrage: &rot /+w1 &Faden

Standard-Verknüpfungoperator: Wortabstand /+w1 logisches "ODER"

Abfrage vorbereiten | Syntzhilfe

Beispiele: direkt nacheinander

Bereits formulierte Anfragen: Treffer:

© 2003 - 2006 IDS Mannheim, Version 1.1

COSMAS-Web: Eingabefenster mit Suchanfrage für die Grundformen von "rot" und "Faden" im Abstand von 1 Wort

chronologische Auswertung der Treffer

Treffer	Jahrzehnt	→ KWIC-Ansicht
1	1800-1809	
1	1950-1959	
9	1960-1969	
2	1970-1979	
50	1980-1989	
2.306	1990-1999	
1.079	2000-2009	

Darstellung der Treffer

statistische Auswertung der Texte um "Faden" durch die Kookkurrenzanalyse

Auswahl aktiv. Alle aktiv. Alle deaktiv. KWIC

Seite 1 von 1

GGV/AGV.00000 Goethe: "Die Wahlverwandtschaften", Hamburger Ausgabe, Band 6, S. 383

gesellige Spiele, ja sogar noch Tänze, zuletzt ein freudloses Herumsitzen und Wiederauflagen einer schon gesunkenen Lust dauerten diesmal, wie sonst auch, weit über Mitternacht, denn schon hatte sich Luciane gewöhnt, morgens nicht aus dem Bette und abends nicht ins Bette gelangen zu können... um diese Zeit finden sich in Otiliens Tagebuch Ergüsse solcher Art: "doppelt häufiger auf das Leben bezogene und von Leben abgezogene Maximen und Sentenzen... weil aber die meisten derselben wohl nicht durch ihre eigene Reflexion entstanden sein können, so ist es wahrscheinlich, daß man sie irgendwo her entlehrt hat, aus dem sie sich, wie sie gewöhnlich war, ausgesprochen, manches Eigene von innerem Bezug wird an dem (rotten) Faden wohl zu erkennen sein... aus Otiliens Tagebuch wie bilden so gern in die Zukunft, weil wir das Ungelähre, was sich in ihr hin und her bewegt, durch alle Wünsche so gern zu unserm Gunsten herantreten möchten; wir beifügen uns nicht leicht in große Gesellschaft, ohne zu denken, der Zufall, der so viele zusammenbringt, solle uns auch unsere Freunde herbeiführen; man mag noch so eingezogen leben, so wird man, ehe man sich versieht, ein Schuldner oder ein Gläubiger... begegnet uns jemand, der uns dank schuldig ist, gleich fällt es uns ein...

Auswahl aktiv. Alle aktiv. Alle deaktiv. KWIC

Häufig	Kookkurrenzen	syntagmatische Muster
1810	roter	994 wie ein roter (...) Faden durch die
406	seidenen	964 an einem seidenen (...) Faden
761	sieht	804 die Fäden (...) sieht
571	Band	245 detaillierte Fäden in der Band
445	laufen	629 laufen (detaillierte) Fäden zusammen...
1810	roter	994 wie ein roter (...) Faden durch die
224	rd	844 rd (...) der) Faden
70	hängt	454 hängt an einem dünnen Faden
245	Strich	994 nach Strich und Faden
88	Hintergrund	804 im Hintergrund (...) die) Faden
91	aplanat	285 Trinne (...) den) Faden wester
255	verloren	724 den Fäden (...) verloren
87	dünnen	604 an einem dünnen Faden
49	unsichtbaren	824 diesen unsichtbaren Fäden
2962	statistisch unspesifisch	

Tabelle: Wörter mit starker Bindung zu "Faden" (Auszüge)

HABEN SIE LUST, IN DER TEXTSAMMLUNG DES IDS ZU STÖBERN?

Welche Wortverbindungen fallen Ihnen zu *Faden* ein, welche Assoziationen?

Wann ungefähr nach Ihrem Gefühl könnte *Bundeskanzlerin* in der Medienlandschaft zum ersten Mal aufgetreten sein und in welchem Zusammenhang?

Welche könnten die beiden großen Herren der deutschen Literatur sein, die Wörter wie *presumptuos* verwendet haben und was könnte dieses Wort heißen?

Fällt Ihnen vielleicht ein deutsches Wort ein, das in der Textsammlung des IDS nicht vorkommt?

COSMAS STEHT FÜR: CORPUS SEARCH, MANAGEMENT AND ANALYSIS SYSTEM.

HANDHABUNG DER KORPORA

Wer nicht in der Gesamtheit der Textsammlung (= Textkorpora) recherchieren will, kann sich spezielle Teilsammlungen definieren: z.B. alle Zeitungsartikel nach dem 11. Sept. 2001 bis zum Ende des Jahres, Texte von Schriftstellern von 1900-1930, Texte mit überwiegender neuer Rechtschreibung oder Texte aus dem deutschsprachigen Ausland.

KORPUSRECHERCHE

Lassen Sie sich überraschen auf der Suche nach dem ersten Auftreten einer Redewendung oder der morphologischen Produktivität eines Wortes wie „Globalisierung“. COSMAS lässt Sie Suchanfragen formulieren unter Verwendung von Teilwörtern, Grundformen, Wort- und Satzabständen und Wortklassen.

KORPUSANALYSE

Die Größe der IDS-Korpora erlaubt einerseits, seltene Wortkompositionen und Fälle eines Wortgebrauchs aufzuspüren, doch müssen andererseits die teils unüberschaubaren Treffermengen automatisch aufbereitet werden. Nach den chronologischen und häufigkeitsbasierten Sortierungen hilft insbesondere die am IDS entwickelte Kookkurrenzanalyse (siehe mittleres Plakat), relevante grammatikalische und semantische Muster zu entdecken.

ANWENDUNGSGEBIETE

Über 80% der in COSMAS ausgeführten Recherchen stammen von externen Nutzern. Dazu gehören sowohl Sprachwissenschaftler als auch Wissenschaftler aus interdisziplinären Wissenschaften der Psychologie, Neurologie, Kognitionswissenschaft, Sprachtherapie, Kommunikations- und Medienwissenschaft und Statistik.

Zudem finden Übersetzer und Sprachlehrer fernab vom deutschsprachigen Kernland in den IDS-Korpora anschauliche Beispiele lebendigen Sprachgebrauchs.



Mitglied der Leibniz Gemeinschaft



Deutschland
Land der Ideen



Ausgewählter Ort 2006



Einige Beispiele von beliebten Recherchen in COSMAS

Eingabe:	was gesucht wird:
&Globalisierung	Deklinationen und Wortzusammensetzungen der Grundform "Globalisierung"
&global /+w1 &Dorf	"globales Dorf", "globale Dörfer", etc.
&anzweifeln oder (&zweifeln /s0 MORPH (Vrb pref))	"anoezweifelt" oder "zweifelt ... an" (wo "an" der Ver Präfix ist)
, /+w1:1 dabei	Nebensatz, der nach einem Komma mit "dabei" eingeleitet wird
(daß oder dass) %-w1:1 ,	"dass" oder "daß" ohne vorangehendes Komma
(, /+w1:1 die) /+s0 (verlieren /w0 .)	Relativsatz, der mit "die" eingeleitet und "verlieren" beendet wird
Blut "und" Wasser schwitzen	die Redensart "Blut und Wasser schwitzen"
&auftragen /+w3 MORPH (dativ)	das Verb "auftragen", gefolgt in einem Abstand von maximal 3 Wörtern von einem Dativ
(Flinte /+w1 ins /+w1 Korn) /w3 &werfen	die Redensart "die Flinte ins Korn werfen", wobei "werfen" konjugiert und vor Flinte oder nach Korn stehen kann
(&bewegen %-w6 ,) /+w7:7 , /+w1:1 zu	eine konjugierte Form von "bewegen", gefolgt von 6 beliebigen Wörtern ohne Komma, gefolgt von einem Komma, das einen Nebensatz mit "zu" einleitet
Jamaika-Koalition	"Jamaika-Koalition"
Schwampel	"Schwampel"