

Textdatenbanken

Sommersemester 2020

6. Vorlesung

- Thesaurusrelationen -

Uwe Quasthoff

Universität Leipzig

Institut für Informatik

quasthoff@informatik.uni-leipzig.de

Angaben zu Wortgruppen: Synonyme

Relationen: Tabelle **rel** mit Spalten **w1_id**, **group_id** und **type**

Index auf **w1_id**, **group_id**

word	w_id	group_id	type
Abstammung	12488	37989	V
Abkunft	105235	37989	v
Geblüt	216266	37989	v
Herkunft	3223	37989	v
Herkommen	147265	37989	v
Deszendenz	1822552	37989	v
Herleitung	154039	37989	v
Provenienz	36948	37989	v
Abstammungslehre	382861	37990	V
Darwinismus	117034	37990	v

Synonymgruppen: Wörter mit gleicher **group_id** gehören zu einer Gruppe, **type** gibt den Typ an (hier: „vergleiche“, Daten aus Wörterbuch) . Ein Eintrag jeder Gruppe (mit **type** in Großbuchstaben) dient der Benennung der Gruppe.

Umfang der Synonyme

- 449248 Einträge in der Tabelle relation
- 81816 verschiedene Gruppen
- Angaben für 95311 Wörter

Größte Gruppe: 172 Einträge für *Gewebeart*:

vergleiche: Atlas, Barchent, Batist, Baumwolle, Beiderwand, Bobinet, Bouclé, Bourrette, Brokat, Buckskin, Cheviot, Chiffon, Chintz, Cord, Cotton, Covercoat, Cretonne, Damast, Dralon, Drapé, Drell, Drillich, Duchesse, Duvetine, Düffel, Ekrüseide, Elastik, Etamin, Etamine, Felbel, Filz, Filüre, Flanell, Flaus, Flausch, Flor, Foulard, Foulé, Fries, Frottee, Frotté, Gabardin, Gaze, Georgette, Gingan, Gingham, Glace, Glencheck, Gregseide, Halbseide, Homespun, Jacquard, Jersey, Jute, Kaliko, Kalmuck, Kambrik, Kamelhaar, Kamelott, Kammgarn, Kanevas, Kaschmir, Kattun, Khaki, Kord, Koton, Krepeline, Krepp, Kreppsatin, Kreton, Kretonne, Krimmer, Kräuselkrepp, Kunstseide, Köper, Lama, Lampas, Lamé, Lastex, Lasting, Lavable, Leinen, Leinwand, Levantine, Linnen, Linon, Loden, Lüster, Madapolam, Madras, Mako, Manchester, Marengo, Marocain, Marquissette, Mohair, Moiré, Mokette, Moleskin, Mollino, Molton, Mouliné, Mousseline, Mull, Mungo, Musselin, Nanking, Napolitaine, Natté, Nessel, Nylon, Opal, Organdy, Organsin, Organza, Orléans, Ottoman, Pepita, Perkal, Perlon, Pikee, Plüsch, Popelin, Popeline, Rasch, Ratiné, Raye, Renforcé, Rippensamt, Rips, Rohseide, Rupfen, Sammet, Samt, Satin, Schappe, Schirting, Schotten, Seide, Serge, Stramin, Streichgarn, Struck, Swanboy, Taft, Tarlatan, Textilien, Textur, Tibet, Trevira, Trikoline, Trikot, Tuch, Tweed, Twill, Tüll, Velours, Velvet, Velveton, Vigogne, Vigoureux, Voile, Wachstuch, Welliné, Wolle, Zanella, Zefir, Zephir, Zibeline, Zindeltaft, Zwillich

Umfang der Dornseiff-Daten

- 12.017 Gruppen,
- darin 90.366 Einträge (Im Wortschatz derzeit: 81.017)
- zu 53.903 Wortformen.

Die Gruppen haben auch Namen, und zwar die Dornseiff-Sachgruppe, also z.B. "19.23 Hieb- und Stichwaffe". Mit diesem Namen kann es mehrere Gruppen geben, da die Dornseiff-Sachgruppen noch nach Wortart und anderem sortiert sind.

Eine einstufige Hierarchie gibt es von den Sachgruppen nach oben zu den Kapitelnamen: Von "19.23 Hieb- und Stichwaffe" zu "19. Geräte, Technik".


Innerhalb einer Sachgruppe gibt es verschiedene Bedeutungsgruppen (ähnlich Synsets), die zusammengehörige Wörter umfassen. Bekannt ist die Wortart zu jeder Bedeutungsgruppe: V, A, N oder P (Personenbezeichner).



Wortschatz - Result - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://wortschatz.uni-leipzig.de/ Google



W O R T S C H A T Z Wort: **Suche!** ? Beachte Groß-/Kleinschreibung
UNIVERSITÄT LEIPZIG

Wort: Abstammung
Anzahl: 2358
Häufigkeitsklasse: 13 (d.h. *der* ist ca. 2^{13} mal häufiger als das gesuchte Wort)
Beschreibung: genealogischer Begriff
Sachgebiet: Familienrecht
Eigenschaften
Morphologie: ab|stamm|ung
Grammatikangaben: Wortart: Substantiv
Geschlecht: weiblich
Flexion: die Abstammung, der Abstammung, der Abstammung, die Abstammung

Relationen zu anderen Wörtern:

- ◆ Synonyme: [Abkunft](#), [Ableitung](#), [Geburt](#), [Geschlecht](#), [Herkommen](#), [Herkunft](#), [Herleitung](#), [Stamm](#), [Ursprung](#)
- ◆ vergleiche: [Abkunft](#), [Deszendenz](#), [Geblüt](#), [Herkommen](#), [Herkunft](#), [Herleitung](#), [Provenienz](#)
- ◆ ist Synonym von: [Abkunft](#), [Geburt](#), [Herkommen](#), [Herkunft](#)
- ◆ wird referenziert von: [Geblüt](#), [Herkunft](#), [Provenienz](#)

Dornseiff-Bedeutungsgruppen:

- ◆ 5.41 Herkunft: [Abkunft](#), [Abstammung](#), [Genesis](#), [Herkunft](#), [Mutterschaft](#), [Stammbaum](#), [Ursprung](#), [Urstoff](#), [Vaterschaft](#)
- ◆ 12.16 Bezeichnung, Wort: [Abstammung](#), [Etymologie](#)
- ◆ 18.3 Nation: [Abstammung](#), [Blut](#), [Eigenart](#), [Kultur](#), [Menschenschlag](#), [Nationalität](#), [Prägung](#), [Rasse](#)

Fertig PR:0

Umfang der Germanet-Daten

Germanet im Wortschatz:

- 41577 Synonymgruppen,
- darin 60640 Einträge (in `r_word_group`) Gruppennummern sind direkt aus Germanet übernommen und verteilen sich grob gestreut auf die Gruppennummern 3286 bis 2578224.

Die Hyponymhierarchie des Germanet ist (in `r_group_rels`) reflektiert.

Zusatzdaten: Genauso wie für WordNet wurden die Kohyponymgruppen berechnet, hier war das Limit 20. Anzahl: 219544

EN: WordNet

- 115424 Synsets
- 203157 Wörter
- 328768 Relationen zwischen Synsets (beidseitig codiert)

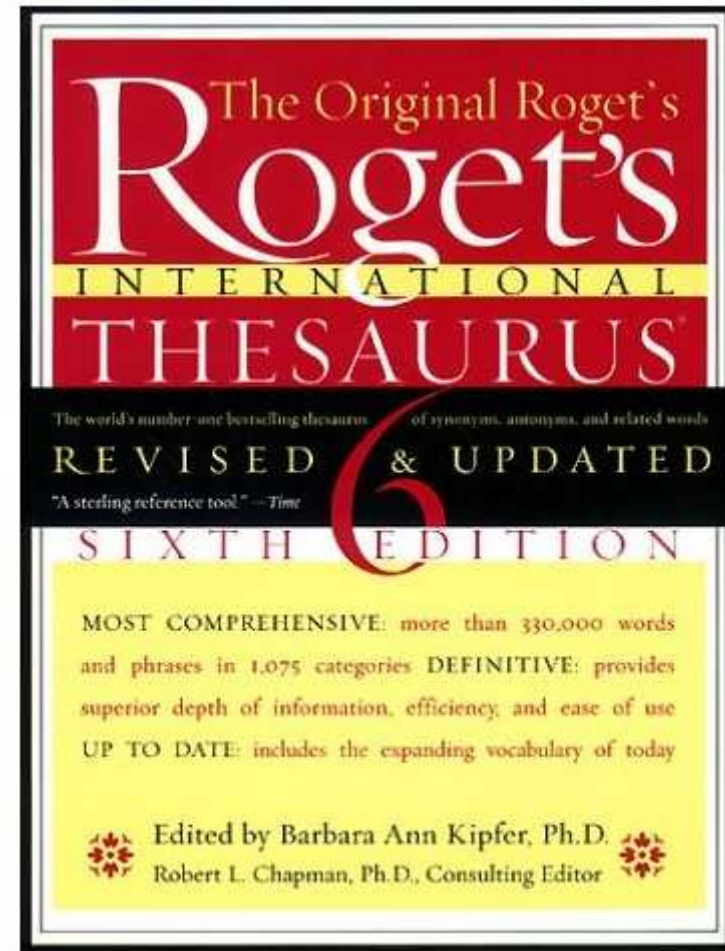
Relationen:

group_type	type_name
1	Wordnet20 Synset
2	Wordnet Antonym
3	Wordnet Domain of Synset c
4	Wordnet Similar
5	Wordnet also see
6	Wordnet Domain of Synset r
7	Wordnet Domain of Synset u
8	Wordnet Participle of verb
9	Wordnet Hyponym
10	Wordnet Hypernym
11	Wordnet Derivation
12	Wordnet Part Meronym
13	Wordnet Member of Domain c
14	Wordnet Substance Meronym
15	Wordnet Part Holonym
16	Wordnet Member Holonym
17	Wordnet Member Meronym
18	Wordnet Member of Domain u
19	Wordnet Substance Holonym
20	Wordnet Member of Domain r
21	Wordnet Entailment
22	Wordnet Verb Group
23	Wordnet Cause
24	Wordnet Cohyponym

EN: Rogets Thesaurus

Umfang:

- 3142 Gruppen
- 57618 Wörter zu Gruppen



Einheitliche Datenstruktur für Dornseiff, Germanet, WordNet und Roget's

TABLE r_word_group (wort_nr, group_nr, sig) Zuordnung wort_nr zur GruppenID.
Signifikanz vorgesehen, aus sicheren Quellen=1000

TABLE r_group_type (group_nr, group_type, group_name, wortart) Zuordnung
GruppenID zur Art der Gruppe. wortart= {N,A,V,S}

TABLE r_type_names (group_type, type_name, wa_1, wa_2) Zuordnung
Relationstyp (group_type oder rel_type) zu einem Namen. wa_1,2} Wortart des
ersten und zweiten in Relation stehenden Wortes

TABLE r_group_rels (group_nr1, group_nr2, rel_type) Speicherung von Relationen
zwischen GruppenIDs.

Mehr Thesaurusdaten erzeugen

Erster Schritt:

Man besorge sich eine Grundmenge von Thesaurusdaten. Notfalls von Hand selber erzeugen.

Zweiter Schritt: Man verwende automatische Verfahren, um die Datenmenge zu erweitern.

ACHTUNG: Automatische Verfahren machen Fehler, die Daten müssen von Hand verifiziert werden.

Erster Schritt

Betrachtet wird der (lokale) Kookkurrenzgraph.

- Den Knoten werden semantische Primitiven vergeben
- Den Kanten werden Relationen zugeteilt.

Frage: Was sind die „richtigen“ semantischen Primitiven und Relationen?

Antwort: Diejenigen, welche man häufig in den Graphen findet.

Primitiven von Hand annotiert

Bis 1.1.05 waren folgende Primitiven annotiert (Primitiven mit weniger als 1000 Instanzen wurden weggelassen).

```
select count(r.wort_nr) as anz,  
name from r_prim_types n,  
r_word_prim r where  
r.prim_type=n.prim_type group by  
r.prim_type order by anz desc ;
```

409106	Personenname
30016	Abstrakt
11193	Konkret
10444	Unbelebt
8522	Zustand konkret
7895	künstlich
7825	Beruf/Personenbezeichner
6137	Zustand abstrakt
6071	Belebt
4571	sonstige Organisation
4436	allgemeiner Ort
3728	V Bewegung
3297	kein Wort
3132	Tätigkeit
2898	Abkürzung
2751	sonstiger Name
2663	menschliche Aktivität
2592	natürlich
1853	Vergrößerungswirkung
1851	Stadt
1575	Ort/Region/Land
1561	mentale Aktion
1449	Unternehmen/Firma
1217	Äußerung
1033	Zustand von Lebewesen

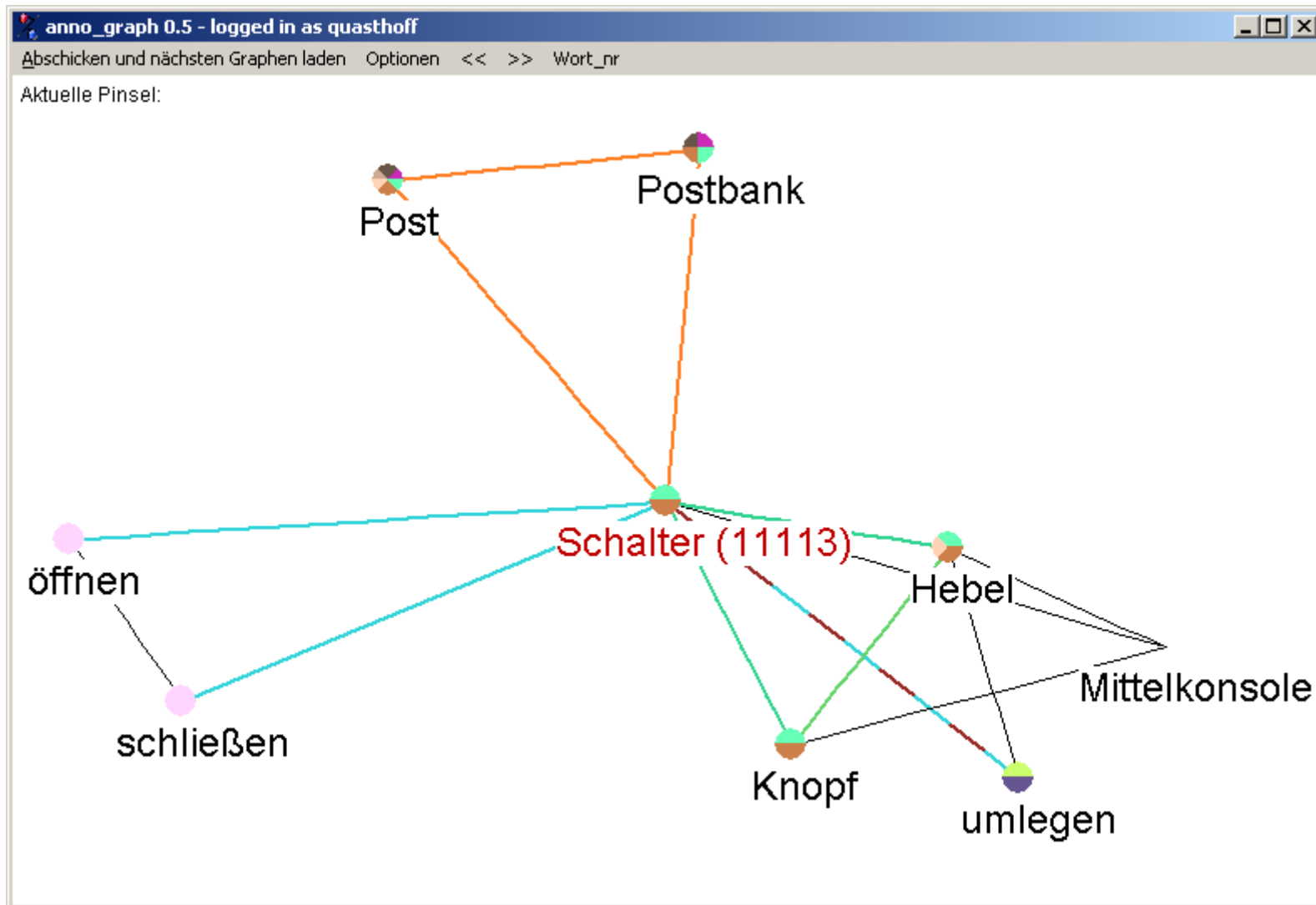
Relationen von Hand annotiert

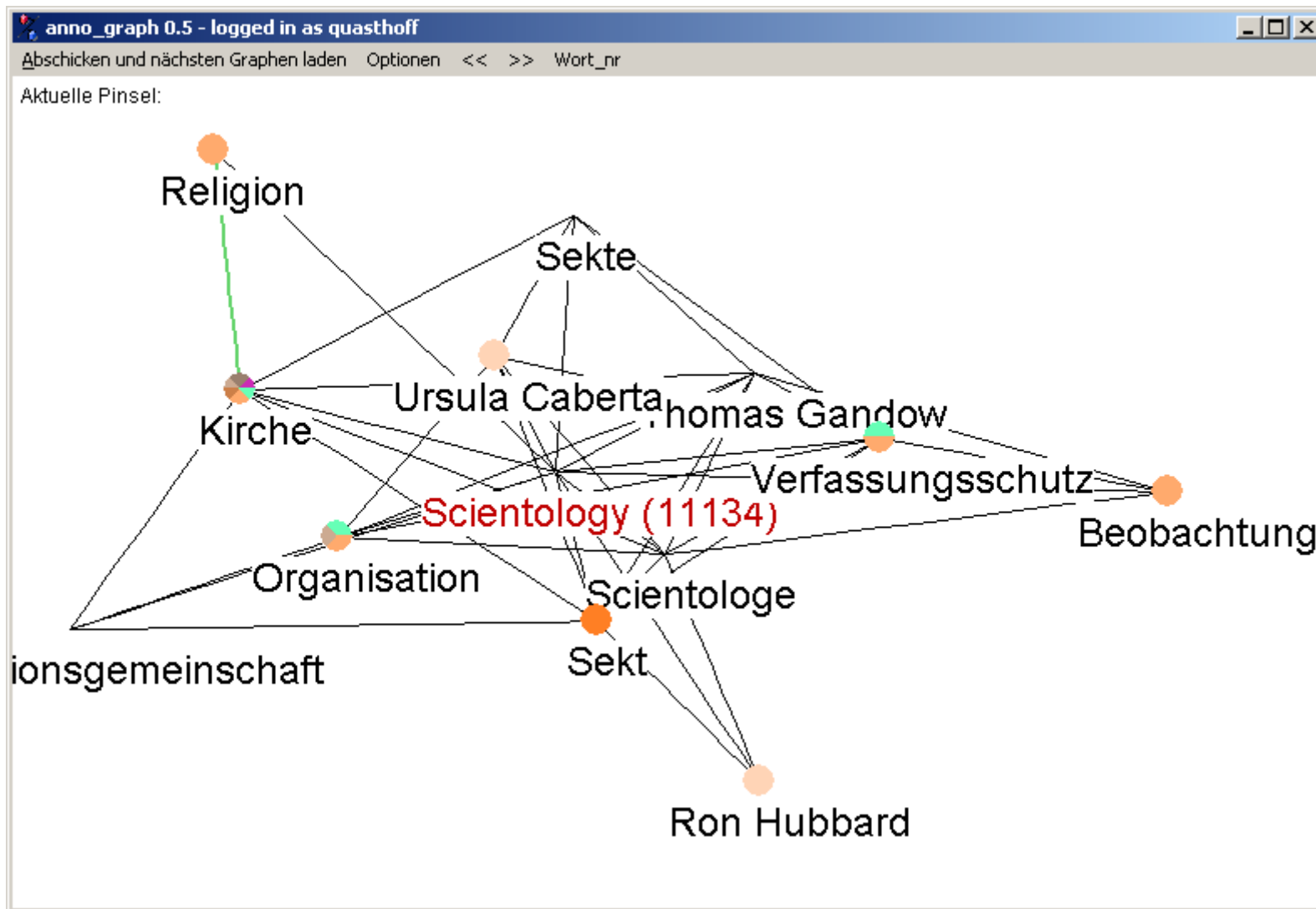
Bis 1.1.05 waren folgende Relationen annotiert (Relationen mit weniger als 3000 Instanzen wurden weggelassen).

```
select count(r.word_nr1) as  
anz, type_name from r_word_rels  
r, r_type_names n where  
r.rel_type=n.group_type group  
by rel_type order by anz desc;
```

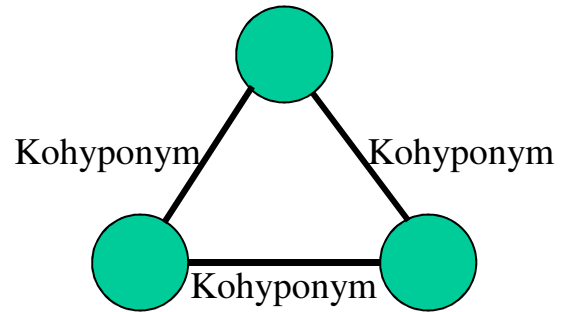
98048	N N	Kohyponym	
43567	N N	Synonym	
42462	N N	Oberbegriff	
42450	N N	Unterbegriff	
16032	A N	typische Eigenschaft	
15902	N A	hat typische Eigenschaft	
12413	A A	Kohyponym	
12079	V N	hat typisches Objekt/Instrument	
11435	N V	typisches Objekt/Instrument von	
10270	V V	Synonym	
9475	V V	Oberbegriff	
9473	V V	Unterbegriff	
9269	A A	Synonym	
8519	N N	Teil von/Material von	
8499	N N	hat Teil/hat Material	
6620	N N	hat typischen Ort	
6596	N N	typischer Ort für	
5705	V V	Kohyponym	
5442	A A	Oberbegriff	
5440	A A	Unterbegriff	
4874	V N	typische Tätigkeit	
4812	N V	typisches Subjekt von	
3528	N N	Eigename zu	
3526	N N	hat Eigename	

... ..



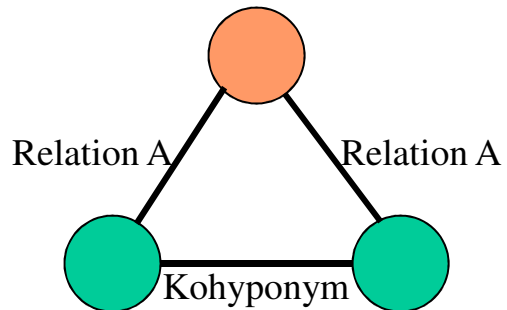


Zweiter Schritt: Schlussregeln



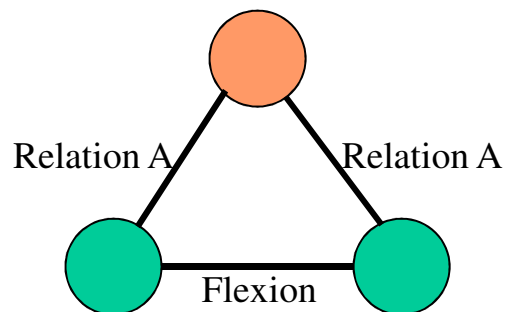
Beispiele:

- Namen mit Sachgebiet
- Stoffe wie Metalle, Drogen, ...



Beispiele:

- Lungenentzündung, Aids, erkrankt
- Rügen, Usedom, Insel



Beispiele:

- weißes, weiße, Hemd
- Fahren, gefahren, schnell
- Stich, lassen, gelassen

Schließen mit Kohyponymen

- Zu X und Y gibt es ein gemeinsames Kohyponym (oder einseitig auch Synonym) Z. Auch Oberbegriff oder beide sind Eigenname für Z. Dann sind X und Y Kohyponym. Bsp.: VW, Rover, Automobilhersteller.
- Allgemeiner: Ist R eine der Relationen Kohyp, Unter, Eigenname, isttypSubj, isttypObj, IsttypOrt, istChef. Und Q eine der Relationen Kohyp, Syn, Anto. Dann gilt:
 - Aus XRZ und YRZ folgt XQY
 - Aus XRZ und XKohypY folgt YRZ
- Für R= hattypOrt funktioniert die zweite Regel, nicht die erste. Hier sind zusätzliche Primitive für X und Y nötig, um Gleichartigkeit der Objekte am Ort zu sichern. Bsp.: X=Gottesdienst / Y=Pfarrer / Z=Kirche
- Beispiel für Q=Anto: Notarzt, Verunglückter, Rettungswagen

Lokaler Kanten-Algorithmus

Betrachtet werden Dreiecke mit zwei bewerteten Kanten. Die Schlussregel legt die Bewertung der dritten Kante mit einer bestimmten Plausibilität nahe.

Wird die gleiche Bewertung einer bestimmten Kante durch mehrere verschiedene Schlüsse nahegelegt, werden die entsprechenden Plausibilitäten addiert.

Wird bei der Addition ein gewisser Schwellwert überschritten, so

- wird die Kante tatsächlich bewertet
- werden die Prioritäten der Endknoten der Kante zur Bearbeitung um 2 erhöht.

Die Auswahl der zu bearbeitenden Dreiecke erfolgt über die Scheitelpunkte gegenüber der zu bewertenden Kante. Ausgewählt werden die Knoten nach fallender Priorität.

Initialisiert wird die Prioritätenliste durch die Anzahl der bereits bekannten Relationen für dieses Wort.

Für jeden Knoten werden aus der Relationenliste alle Kanten herausgesucht. Dann wird für aus je zwei davon eine V-Form gebildet und es wird versucht, Regeln anzuwenden.

Nach fertiger Bearbeitung wird die Priorität des bearbeiteten Knoten auf null gesetzt.

Globaler Kanten-Algorithmus

Annotationsschritt:

Betrachtet werden alle Dreiecke mit zwei bewerteten Kanten. Die Schlussregel legt die Bewertung der dritten Kante mit einer bestimmten Plausibilität nahe. Diese Bewertungen werden für jede Kante gesammelt.

Die Auswahl der zu bearbeitenden Dreiecke erfolgt über die Scheitelpunkte gegenüber der zu bewertenden Kante.

Bewertungsschritt:

Für jede Kante wird an Hand der vorläufigen Bewertungen entschieden, ob daraus eine endgültige Bewertung gemacht wird. Kriterien sind Anzahl der Bewertungen, Sicherheit, Widersprüchlichkeit.

Danach wird das ganze Vorgehen iteriert, bis keine neuen Kanten mehr annotiert werden.

Annotationstool - Mozilla

File Edit View Go Bookmarks Tools Window Help

9 rule N-Kohyponymdreieck ohne Eigennamen **aktiv** [\[Regel testen\]](#) [\[Regel bearbeiten\]](#) [\[Regel klonen\]](#) [\[Regel deaktivieren\]](#)
[\[Regel vertrauen\]](#)

"Wenn A-B Kohyponym und A-C auch, dann auch B-C."

$(rel_{ab} = "N N Kohyponym") \wedge (rel_{ac} = "N N Kohyponym") \wedge (prim_c \notin \{Unternehmen/Firma, sonstige Organisation, Personennamen, Stadt, Ort/Region/Land\}) \wedge (sig_{bc} \geq "10")$
 $\Rightarrow rel_{bc} = "N N Kohyponym"$

correct: 4104, wrong: 1421, coverage: 6368, sure_flag: 0, dont_use_flag: 0

10 Kohyponyme sind beides typische Objekte **aktiv** [\[Regel testen\]](#) [\[Regel bearbeiten\]](#) [\[Regel klonen\]](#) [\[Regel deaktivieren\]](#)
[\[Regel vertrauen\]](#)

"Falls A und C Kohyponym und A typisches Objekt von B, dann auch C"

$(rel_{ab} = "N V typisches Objekt/Instrument von") \wedge (rel_{ac} = "N N Kohyponym") \wedge (sig_{bc} \geq "4")$
 $\Rightarrow rel_{bc} = "V N hat typisches Objekt/Instrument"$

correct: 1649, wrong: 201, coverage: 1993, sure_flag: 0, dont_use_flag: 0

11 Kohyponyme sind beides typische Subjekte-11 **aktiv** [\[Regel testen\]](#) [\[Regel bearbeiten\]](#) [\[Regel klonen\]](#) [\[Regel deaktivieren\]](#)
[\[Regel vertrauen\]](#)

"Falls A und C Kohyponym und A typisches Subjekt von B, dann auch C-11"

$(rel_{ab} = "N V typisches Subjekt von") \wedge (rel_{ac} = "N N Kohyponym") \wedge (sig_{bc} \geq "4")$
 $\Rightarrow rel_{bc} = "V N typische Tätigkeit"$

correct: 564, wrong: 209, coverage: 809, sure_flag: 0, dont_use_flag: 0

12 Unterbegriffe sind auch typische Objekte-12 **aktiv** [\[Regel testen\]](#) [\[Regel bearbeiten\]](#) [\[Regel klonen\]](#) [\[Regel deaktivieren\]](#)
[\[Regel vertrauen\]](#)

"Falls A Oberbegriff zu C und A typisches Objekt von B, dann auch C-12"

$(rel_{ab} = "N V typisches Objekt/Instrument von") \wedge (rel_{ac} = "N N Oberbegriff") \wedge (sig_{bc} \geq "4")$
 $\Rightarrow rel_{bc} = "V N hat typisches Objekt/Instrument"$

PageRank

Annotationstool: Projektleiter - Mozilla

File Edit View Go Bookmarks

[angemeldet als quasthoff: [abmelden](#)]
Spalten: 2

<p>Projektleiter (N) <input type="checkbox"/> Wo</p> <p>med (A) <input type="checkbox"/> als</p> <p><keine Beschreibung> A Bewegung Abkürzung Bitte Überprüfen!</p>	<p><kein Typ> ?? Bitte Überprüfen! ?? selbe Wortgruppe N N Aggregat von N N Beruf zu N N Chef von N N Eigenname zu N N Gegenteil N N Grund von N N hat Aggregat N N hat Beruf N N hat Chef N N hat Eigenname N N hat Maßeinheit N N hat Teil/hat Material N N hat typischen Ort N N Kohyponym N N Maßeinheit für N N Oberbegriff N N Synonym <kein Typ></p>	<p>phil (A) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p><keine Beschreibung> A Bewegung Abkürzung Bitte Überprüfen!</p> <p><kein Typ></p>
<p>Abteilungsleiter (N) <input type="checkbox"/></p> <p><keine Beschreibung> Abkürzung Abstrakt allgemeiner Ort</p>		<p>Gerhard Haag (N) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p><keine Beschreibung> Abkürzung Abstrakt allgemeiner Ort</p> <p><kein Typ></p>
<p>Geschäftsführer (N) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p>Abstrakt allgemeiner Ort Belebt Beruf/Personenbezeichner</p> <p><kein Typ></p>		<p>Hernichel (N) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p><keine Beschreibung> Abkürzung Abstrakt allgemeiner Ort</p> <p><kein Typ></p>
<p>Motte (N, N) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p>natürlich non-profit-Organisation</p>		<p>München-Ost (N) <input type="checkbox"/> als Ref.-Wort <input type="checkbox"/></p> <p>Naturereignis natürlich</p>

Referenz: Projektleiter

Done PageRank

wort1	wort2	Relation(wort1, wort2)		Verbesserung (nur falls falsch)
senden	RTL	V N typische Tätigkeit	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (Kohyponyme sind beides typische Subjekte-11)	richtig wäre ...
abweisen (abgewiesen)	Verwaltungsgerichtshof	V N typische Tätigkeit	<input type="radio"/> richtig <input checked="" type="radio"/> falsch <input type="radio"/> ungeprüft (Kohyponyme sind beides typische Subjekte-11)	richtig wäre ...
bedrohen (bedrohte)	Spezies	V N hat typisches Objekt/Instrument	<input type="radio"/> richtig <input type="radio"/> falsch <input checked="" type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...
einlegen	Protest	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (Kohyponyme sind beides typische Objekte)	richtig wäre ...
einlegen	Revision	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...
einlegen	Veto	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (Kohyponyme sind beides typische Objekte)	richtig wäre ...
erfordern	Geschick	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...
begehen	Verbreche (Verbrechen)	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...
obduziert	Leichnam	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...
obduziert	Tote (Toten)	V N hat typisches Objekt/Instrument	<input checked="" type="radio"/> richtig <input type="radio"/> falsch <input type="radio"/> ungeprüft (V-Instr-Instr-Dreieck 1)	richtig wäre ...

Abschicken

Regeln lernen

Betrachte vollständig bewertete Dreiecke.

Wenn wir eine Bewertung weglassen: Kann diese wieder erschlossen werden?

Entscheidung ist (wie beim Pendel) wieder mit den bisher annotierten Daten möglich.

Bsp. für annotierte Relationen

```
select w1.wort_bin, w2.wort_bin, rt12.type_name,w3.wort_bin, rt13.type_name from wortliste w1, wortliste w2, wortliste w3, r_word_rels rw12, r_type_names rt12, r_word_rels rw13, r_type_names rt13, kollok_sig where w2.wort_nr=wort_nr1 and w3.wort_nr=wort_nr2 and w2.wort_nr<w3.wort_nr and rt12.group_type=rw12.rel_type and rt13.group_type=rw13.rel_type and w1.wort_nr=rw12.word_nr1 and w2.wort_nr=rw12.word_nr2 and w1.wort_nr=rw13.word_nr1 and w3.wort_nr=rw13.word_nr2 limit 100;
```

Deutschland	Welt	N N Unterbegriff	Land	N N Unterbegriff
Deutschland	Bayern	N N hat Teil/hat Material	Saarland	N N hat Teil/hat Material
Deutschland	Bayern	N N hat Teil/hat Material	Sachsen	N N hat Teil/hat Material
Deutschland	DDR	N N hat Teil/hat Material	BRD	N N Synonym
Deutschland	DDR	N N hat Teil/hat Material	BRD	N N hat Teil/hat Material
Deutschland	Spanien	N N Kohyponym	Irland	N N Kohyponym
DM	Euro	N N Kohyponym	Aktie	N N Maßeinheit für
DM	Bank	N N hat typischen Ort	Aktie	N N Maßeinheit für
DM	Bank	N N hat typischen Ort	Kredit	N N Maßeinheit für
Jahr	Januar	N N hat Teil/hat Material	November	N N hat Teil/hat Material
Jahr	Januar	N N hat Teil/hat Material	Juni	N N hat Teil/hat Material
Jahr	Januar	N N hat Teil/hat Material	März	N N hat Teil/hat Material
Jahr	Juni	N N hat Teil/hat Material	Geschäftsjahr	N N Oberbegriff
Jahr	Vorjahr	N N Oberbegriff	Geschäftsjahr	N N Oberbegriff
Jahr	Herbst	N N hat Teil/hat Material	Frühjahr	N N hat Teil/hat Material
Mark	DM	N N Synonym	Aktie	N N Maßeinheit für
Uhr	Uhrzeit	N N Unterbegriff	Armbanduhr	N N Oberbegriff
Uhr	Armbanduhr	N N Oberbegriff	Taschenuhr	N N Oberbegriff
Uhr	Taschenuhr	N N Oberbegriff	Sonnenuhr	N N Oberbegriff
Uhr	Wanduhr	N N Oberbegriff	Pendeluhr	N N Oberbegriff

Annotierte Dreiecke nach Anzahl

```
select concat(rt12.type_name," ", rt23.type_name," ", rt31.type_name) as rela, count(*) as anz from r_word_rels rw12, r_type_names rt12, r_word_rels rw31,
r_type_names rt31, r_word_rels rw23, r_type_names rt23 where rt12.group_type=rw12.rel_type and rt31.group_type=rw31.rel_type and
rt23.group_type=rw23.rel_type and rw12.word_nr1<>rw23.word_nr1 and rw12.word_nr1<>rw31.word_nr1 and rw23.word_nr1<>rw31.word_nr1 and
rw12.word_nr2=rw23.word_nr1 and rw23.word_nr2=rw31.word_nr1 and rw31.word_nr2=rw12.word_nr1 group by rela order by anz desc limit 50;
```

rela	anz
N N Kohyponym N N Kohyponym N N Kohyponym	367521
N N Synonym N N Synonym N N Synonym	57057
A A Kohyponym A A Kohyponym A A Kohyponym	22596
N N Kohyponym N V typisches Objekt/Instrument von V N hat typisches Objekt/Instrument	18577
V N hat typisches Objekt/Instrument N N Kohyponym N V typisches Objekt/Instrument von	18577
N V typisches Objekt/Instrument von V N hat typisches Objekt/Instrument N N Kohyponym	18577
N N Unterbegriff N N Oberbegriff N N Kohyponym	17645
N N Oberbegriff N N Kohyponym N N Unterbegriff	17645
N N Kohyponym N N Unterbegriff N N Oberbegriff	17645
V V Synonym V V Synonym V V Synonym	13398
N N Kohyponym N N Kohyponym N N Synonym	9237
N N Kohyponym N N Synonym N N Kohyponym	9237
N N Synonym N N Kohyponym N N Kohyponym	9237
N N Oberbegriff N N Kohyponym N N Kohyponym	8685
N N Kohyponym N N Oberbegriff N N Kohyponym	8685
N N Kohyponym N N Kohyponym N N Oberbegriff	8685
N N Kohyponym N N Unterbegriff N N Kohyponym	8586
N N Unterbegriff N N Kohyponym N N Kohyponym	8586
N N Kohyponym N N Kohyponym N N Unterbegriff	8586
A A Synonym A A Synonym A A Synonym	8208
N V typisches Subjekt von V N typische Tätigkeit N N Kohyponym	7565
V N typische Tätigkeit N N Kohyponym N V typisches Subjekt von	7565
N N Kohyponym N V typisches Subjekt von V N typische Tätigkeit	7565
A N typische Eigenschaft N N Kohyponym N A hat typische Eigenschaft	6488
N N Kohyponym N A hat typische Eigenschaft A N typische Eigenschaft	6488
N A hat typische Eigenschaft A N typische Eigenschaft N N Kohyponym	6488
V V Kohyponym V V Kohyponym V V Kohyponym	5244
N N hat Eigennamen N N Kohyponym N N Eigennamen zu	4101
N N Kohyponym N N Eigennamen zu N N hat Eigennamen	4101
N N Eigennamen zu N N hat Eigennamen N N Kohyponym	4101

Annotierte Dreiecke nach Art

```
select concat(rt12.type_name," ", rt23.type_name," ", rt31.type_name) as rela, count(*) as anz from r_word_rels
  rw12, r_type_names rt12, r_word_rels rw31, r_type_names rt31, r_word_rels rw23, r_type_names rt23 where
  rt12.group_type=rw12.rel_type and rt31.group_type=rw31.rel_type and rt23.group_type=rw23.rel_type and
  rw12.word_nr1<>rw23.word_nr1 and rw12.word_nr1<>rw31.word_nr1 and rw23.word_nr1<>rw31.word_nr1 and
  rw12.word_nr2=rw23.word_nr1 and rw23.word_nr2=rw31.word_nr1 and rw31.word_nr2=rw12.word_nr1 group by rela
having anz>500 order by rela limit 150;
```

rela	anz
A A Gegenteil A N typische Eigenschaft N A hat typische Eigenschaft	1228
A A Kohyponym A A Kohyponym A A Kohyponym	22596
A A Kohyponym A A Kohyponym A A Synonym	723
A A Kohyponym A A Synonym A A Kohyponym	723
A A Kohyponym A A Unterbegriff A A Oberbegriff	909
A A Kohyponym A N typische Eigenschaft N A hat typische Eigenschaft	3399
N N Kohyponym N N Eigenname zu N N hat Eigenname	4101
N N Kohyponym N N Eigenname zu N N Kohyponym	1279
N N Kohyponym N N Eigenname zu N N Oberbegriff	855
N N Kohyponym N N Oberbegriff N N Kohyponym	8685
N N Kohyponym N N Oberbegriff N N Unterbegriff	1148
N N Kohyponym N N Synonym N N Kohyponym	9237
N N Kohyponym N N Synonym N N Oberbegriff	597
N N Kohyponym N N Synonym N N Synonym	1016
N N Maßeinheit für N N Kohyponym N N hat Maßeinheit	1185
N N Oberbegriff N A hat typische Eigenschaft A N typische Eigenschaft	1269
N N Oberbegriff N N Gegenteil N N Unterbegriff	836
N N Synonym N N Kohyponym N N Kohyponym	9237
N N Synonym N N Kohyponym N N Synonym	1016
N N Synonym N N Kohyponym N N Unterbegriff	608
N N Synonym N N Synonym N N Kohyponym	1016
N N Synonym N N Synonym N N Synonym	57057
N N Synonym N N typischer Ort für N N hat typischen Ort	571
N N Synonym N N Unterbegriff N N Oberbegriff	1408

Kohyponyme lernen

Wort1 ist ähnlich zu Wort2 wegen Wort3. Dazu Relationen 1-3 und 3-2, jeweils 10 oder 16. Weiter muss sig(wort1,wort2)>10 sein. Das liefert 7018 neue Kandidaten, davon viele mehrfach.

```
select w1.wort_nr, w3.wort_nr, w1.wort_bin, w3.wort_bin, w2.wort_bin as Grund, rw12.rel_type, rw23.rel_type, rw31.rel_type from wortliste w1, wortliste w2, wortliste w3, r_word_rels rw12, r_word_rels rw23, kollok_sig k left outer join r_word_rels rw31 on (rw31.word_nr1=rw23.word_nr2 and rw31.word_nr2=rw12.word_nr1) where w1.wort_nr=rw12.word_nr1 and w2.wort_nr=rw23.word_nr1 and w3.wort_nr=rw23.word_nr2 and w1.wort_nr=k.word_nr1 and w3.wort_nr=k.word_nr2 and signifikanz>10 and rw12.rel_type in (16, 10) and rw23.rel_type in (16, 10) and rw12.word_nr1<>rw23.word_nr2 and rw12.word_nr2=rw23.word_nr1 and isnull(rw31.rel_type) limit 50;
```

8730	9800	Grafik	Zeichnung	Graphik	10	16	NULL
8731	5332	Anzug	Hut	Mantel	16	16	NULL
8731	8425	Anzug	Jeans	Pullover	16	16	NULL
8731	8533	Anzug	Hemd	Pullover	16	16	NULL
8731	11850	Anzug	Kleid	Pullover	16	16	NULL
8754	914	Utopie	Idee	Vision	10	10	NULL
8754	2262	Utopie	Traum	Illusion	16	10	NULL
8803	843	Großmutter	Sohn	Großvater	16	16	NULL
8803	843	Großmutter	Sohn	Vater	16	16	NULL
8803	853	Großmutter	Kind	Enkelkind	16	16	NULL
8803	853	Großmutter	Kind	Vater	16	16	NULL
8803	929	Großmutter	Tochter	Vater	16	16	NULL
8803	1987	Großmutter	Bruder	Vater	16	16	NULL
8803	3021	Großmutter	Schwester	Vater	16	16	NULL
8803	7669	Großmutter	Enkel	Enkelkind	16	10	NULL
8803	7669	Großmutter	Enkel	Vater	16	16	NULL
8809	25310	Kokain	Cannabis	Koks	10	16	NULL
8809	84353	Kokain	Amphetamin	Koks	10	16	NULL
8824	474	Geschäftsbereich	Bereich	Zweig	16	16	NULL
8824	6172	Geschäftsbereich	Sparte	Zweig	16	10	NULL
8825	628	Theologie	Kirche	Religion	16	16	NULL
8825	9135	Theologie	Ethik	Philosophie	16	16	NULL
8825	9135	Theologie	Ethik	Religion	16	16	NULL
8825	9135	Theologie	Ethik	Religionsunterricht	16	16	NULL
8825	12394	Theologie	Theologen	Religionsunterricht	16	16	NULL
8825	14357	Theologie	Jura	Philosophie	16	16	NULL
8825	22936	Theologie	Germanistik	Philosophie	16	16	NULL
8895	8856	Getreide	Reis	Erbse	16	16	NULL
8895	8856	Getreide	Reis	Erbsen	16	16	NULL

Allg. Relationen lernen

```
mysql> select rw23.rel_type, w1.wort_nr, w3.wort_nr,w1.wort_bin, w3.wort_bin, w2.wort_bin as Grund, rw12.rel_type, signifikanz
from wortliste w1, wortliste w2, wortliste w3, r_word_rels rw12, r_word_rels rw23, kollok_sig_alle k left outer join
r_word_rels rw31 on (rw31.word_nr1 =rw23.word_nr2 and rw31.word_nr2=rw12.word_nr1) where w1.wort_nr=rw12.word_nr1 and
w2.wort_nr=rw23.word_nr1 and w3.wort_nr=rw23.word_nr2 and w1.wort_nr=k.wort_nr2 and w3.wort_nr=k.wort_nr1 and signifikanz>9
and rw12.rel_type in (10,16,20,26,30,36,13,23,33,56, 57,58,59,60,61) and rw12.word_nr1<>rw23.word_nr2 and
rw12.word_nr2=rw23.word_nr1 and isnull(rw31.rel_type) and rw12.word_nr1>10000 order by rw12.word_nr1 limit 100;
```

rel_type	wort_nr	wort_nr	wort_bin	wort_bin	Grund	rel_type	signifikanz
11	10001	10362	Überwindung	Bewältigung	Durchsetzung	10	12
16	10008	8493	Extremisten	Fundamentalisten	Entführer	16	26
16	10008	14633	Extremisten	Attentäter	Entführer	16	28
16	10008	5602	Extremisten	Terroristen	Entführer	16	84
15	10013	3520	Hütte	Eingang	Häuser	16	13
15	10013	3229	Hütte	Garten	Häuser	16	14
74	10013	27220	Hütte	abgebrannt	Häuser	16	16
16	10013	14306	Hütte	Hütten	Häuser	16	34
15	10013	1082	Hütte	Wohnung	Häuser	16	36
63	10013	1984	Hütte	Dorf	Häuser	16	37
74	10013	1707	Hütte	bauen	Häuser	16	38
16	10018	22591	Kaufhaus	Saturn	Kaufhof	10	10
16	10021	9940	Hertha	Hamburger SV	Hertha BSC	10	62
77	10021	7817	Hertha	Champions League	Hertha BSC	10	374
11	10030	19316	Medikament	Impfstoff	Arzneimittel	10	12
16	10030	18562	Medikament	Präparate	Arzneimittel	10	12
10	10030	8221	Medikament	Medikamenten	Arzneimittel	10	15
11	10030	14567	Medikament	Antibiotika	Arzneimittel	10	17
11	10030	66097	Medikament	Tablette	Arzneimittel	10	18
11	10030	49498	Medikament	Antibiotikum	Arzneimittel	10	19
12	10030	863	Medikament	Mittel	Arzneimittel	10	35
16	10038	11528	Gattin	Hausfrau	Frau	10	13
11	10038	11528	Gattin	Hausfrau	Frau	10	13
16	10038	853	Gattin	Kind	Frau	10	15
11	10038	3318	Gattin	Dame	Frau	10	26
16	10038	253	Gattin	Kinder	Frau	10	30