Seminar: Formale Begriffsanalyse
Überblick und Anwendungsbeispiele

mit Folien von Gerd Stumme und Marteen Janssen (siehe Fußzeile)

Dozentin: Wiebke Petersen
petersew@uni-duesseldorf.de

18. April 2006

Seminarplan

1. Einige Anwendungsbeispiele zur Motivation
2. Formale Kontexte und formale Begriffe
   - Grundlagen der Mengenlehre, Eigenschaften von Relationen
3. Begriffsverbände
   - Grundlagen der Ordnungs- und Verbandstheorie
   - Hüllensystem
4. Zeichnen von Begriffsverbänden
   - bereinigte und reduzierte Kontexte
   - additive Liniendiagramme
   - gestufte Liniendiagramme

5. Algorithmen zur Berechnung der formalen Begriffe
   - naive Methode
   - next-concept-Algorithums
6. Mehrwertige Kontexte und Skalierung
   - schlichte Skalierung
   - Nominalskalen
   - Ordinalskalen
   - Interordinalskalen
   - Biordinalskalen
   - dichotome Skala
7. Implikationen
   - einige Begriffe aus der Aussagenlogik
   - Berechung der Stammbasis

8. Wenn die Resultate der FBA zu genau sind:
   - Pruning der Begriffsverbände — Eisbergverbände
   - Assoziationsregeln
9. Wissensakquisition mit FBA
   - Merkmalsexploration

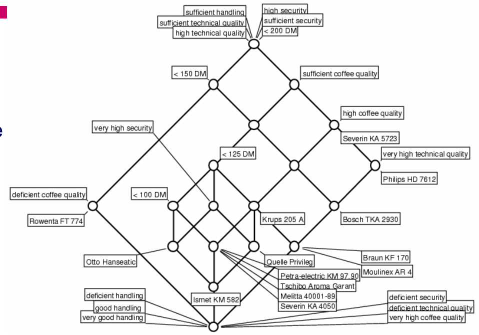
Formale Begriffsanalyse ist um 1980 in Darmstadt entstanden als mathematische Theorie, die eine Formalisierung des Begriffs vom „Begriff“ liefert.

FBA hat seitdem zunehmend Verbreitung in der Informatik gefunden, u.a. in

- der Datenanalyse,
- der Wissensentdeckung,
- dem Software Engineering.

Ausgehend von Datensätzen leitet FBA Begriffshierarchien ab.

FBA bietet die Erzeugung und Visualisierung der Begriffshierarchien auf einer mathematisch fundierten Basis.



STIFTUNG WARENTEST test KOMPASS — KAFFEEMASCHINEN MIT WARMHALTEKANNE (8 bis 10 Tassen) — test Ausgabe 12/98

	Mittlerer Preis in DM ca.	Preis für Ersatzkanne/ Glaseinsatz in DM ca.	Kaffee-qualität	Technische Prüfung	Sicherheit	Handhabung	test-Qualitätsurteil
Gewichtung			35 %	30 %	10 %	25 %	
Neckermann Best.-Nr. 8628/409	40,–	35,–[1] / ❑	baugl. mit Otto Hanseatic Best.-Nr. 4327357				zufriedenst.
Otto Hanseatic Best.-Nr. 4327357	40,–	30,–[2] / ❑	O	+	++	O	zufriedenst.
Quelle Privileg Best.-Nr. 7030720	40,–	24,50 / 17,50	baugl. mit Otto Hanseatic Best.-Nr. 4327357				zufriedenst.
Severin KA 9660	50,–	35,– / 23,–	baugl. mit Otto Hanseatic Best.-Nr. 4327357				zufriedenst.
Severin KA 4050	80,–	50,– / ❑	+	+	+	O	gut
Tchibo Aroma Garant Art.-Nr. 48469	80,–	27,50 / 19,50	+	+	+	O	gut
Ismet KM 582 starlight	84,–	47,– / 14,–	+	+	++	O	gut

Formale Begriffsanalyse

Def.: Ein **formaler Kontext** ist ein Tripel (G,M,I), wobei

- G eine Menge von Gegenständen,
- M eine Menge von Merkmalen
- und I eine Relation zwischen G und M ist.

- $(g,m) \in I$ wird gelesen als „Gegenstand g hat Merkmal m“.

National Parks in California	NPS Guided Tours	Hiking	Horseback Riding	Swimming	Boating	Fishing	Bicycle Trail	Cross Country Trail
Cabrillo Natl. Mon.						×	×	
Channel Islands Natl. Park		×		×		×		
Death Valley Natl. Mon.	×	×	×	×			×	
Devils Postpile Natl. Mon.	×	×	×	×		×		
Fort Point Natl. Historic Site	×					×		
Golden Gate Natl. Recreation Area	×	×	×	×		×	×	
John Muir Natl. Historic Site	×							
Joshua Tree Natl. Mon.	×	×	×					
Kings Canyon Natl. Park	×	×	×			×		×
Lassen Volcanic Natl. Park	×	×	×	×	×	×		×
Lava Beds Natl. Mon.	×	×						
Muir Woods Natl. Mon.		×						
Pinnacles Natl. Mon.		×						
Point Reyes Natl. Seashore	×	×	×	×		×	×	
Redwood Natl. Park	×	×	×	×		×		
Santa Monica Mts. Natl. Recr. Area	×	×	×	×	×	×		
Sequoia Natl. Park	×	×	×			×		×
Whiskeytown-Shasta-Trinity Natl. Recr. Area	×	×	×	×	×	×		
Yosemite Natl. Park	×	×	×	×	×	×	×	×

Für $A \subseteq G$ definieren wir

$A' := \{ m \in M \mid \forall g \in A: (g,m) \in I \}$.

Für $B \subseteq M$ definieren wir dual

$B' := \{ g \in G \mid \forall m \in B: (g,m) \in I \}$.

A´

National Parks in California	NPS Guided Tours	Hiking	Horseback Riding	Swimming	Boating	Fishing	Bicycle Trail	Cross Country Trail
Cabrillo Natl. Mon.						×	×	
Channel Islands Natl. Park		×		×		×		
Death Valley Natl. Mon.	×	×	×	×			×	
Devils Postpile Natl. Mon.	×	×	×	×		×		
Fort Point Natl. Historic Site	×					×		
Golden Gate Natl. Recreation Area	×	×	×	×		×	×	
John Muir Natl. Historic Site	×							
Joshua Tree Natl. Mon.	×	×	×					
Kings Canyon Natl. Park	×	×	×			×		×
Lassen Volcanic Natl. Park	×	×	×	×	×	×		×
Lava Beds Natl. Mon.	×	×						
Muir Woods Natl. Mon.		×						
Pinnacles Natl. Mon.		×						
Point Reyes Natl. Seashore	×	×	×	×		×	×	
Redwood Natl. Park	×	×	×	×		×		
Santa Monica Mts. Natl. Recr. Area	×	×	×	×	×	×		
Sequoia Natl. Park	×	×	×			×		×
Whiskeytown-Shasta-Trinity Natl. Recr. Area	×	×	×	×	×	×		
Yosemite Natl. Park	×	×	×	×	×	×	×	×

A

Def.: Ein **formaler Begriff** ist ein Paar (A,B) mit

- $A \subseteq G$ und $B \subseteq M$,
- $A' = B$,
- $B' = A$.

A ist der **Umfang** und B der **Inhalt** des Begriffs.

Inhalt

National Parks in California	NPS Guided Tours	Hiking	Horseback Riding	Swimming	Boating	Fishing	Bicycle Trail	Cross Country Trail
Cabrillo Natl. Mon.						×	×	
Channel Islands Natl. Park		×		×		×		
Death Valley Natl. Mon.	×	×	×	×			×	
Devils Postpile Natl. Mon.	×	×	×	×		×		
Fort Point Natl. Historic Site	×					×		
Golden Gate Natl. Recreation Area	×	×	×	×		×	×	
John Muir Natl. Historic Site	×							
Joshua Tree Natl. Mon.	×	×	×					
Kings Canyon Natl. Park	×	×	×			×		×
Lassen Volcanic Natl. Park	×	×	×	×	×	×		×
Lava Beds Natl. Mon.	×	×						
Muir Woods Natl. Mon.		×						
Pinnacles Natl. Mon.		×						
Point Reyes Natl. Seashore	×	×	×	×		×	×	
Redwood Natl. Park	×	×	×	×		×		
Santa Monica Mts. Natl. Recr. Area	×	×	×	×	×	×		
Sequoia Natl. Park	×	×	×			×		×
Whiskeytown-Shasta-Trinity Natl. Recr. Area	×	×	×	×	×	×		
Yosemite Natl. Park	×	×	×	×	×	×	×	×

Umfang

2.1 Basic Notions

Der blaue Begriff ist ein **Unterbegriff** des gelben Begriffs, denn:

der blaue Umfang ist im gelben Umfang enthalten.

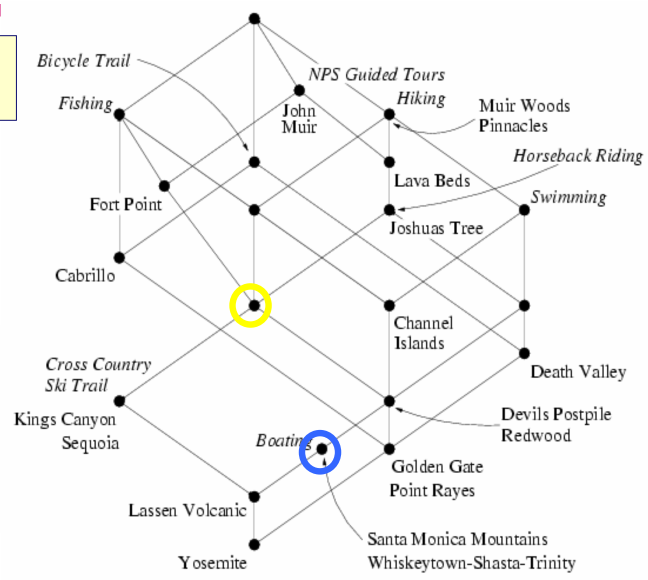
(⇔ der gelbe Inhalt ist im blauen Inhalt enthalten.)

Def.: $(A_1,B_1) \leq (A_2,B_2)$
$:\Leftrightarrow A_1 \subseteq A_2$
$(\Leftrightarrow B_1 \supseteq B_2)$

National Parks in California	NPS Guided Tours	Hiking	Horseback Riding	Swimming	Boating	Fishing	Bicycle Trail	Cross Country Trail
Cabrillo Natl. Mon.						×	×	
Channel Islands Natl. Park		×		×		×		
Death Valley Natl. Mon.	×	×	×	×			×	
Devils Postpile Natl. Mon.	×	×	×	×		×		
Fort Point Natl. Historic Site	×					×		
Golden Gate Natl. Recreation Area	×	×	×	×		×	×	
John Muir Natl. Historic Site	×							
Joshua Tree Natl. Mon.	×	×	×					
Kings Canyon Natl. Park	×	×	×			×		×
Lassen Volcanic Natl. Park	×	×	×	×	×	×		×
Lava Beds Natl. Mon.	×	×						
Muir Woods Natl. Mon.		×						
Pinnacles Natl. Mon.		×						
Point Reyes Natl. Seashore	×	×	×	×		×	×	
Redwood Natl. Park	×	×	×	×		×		
Santa Monica Mts. Natl. Recr. Area	×	×	×	×	×	×		
Sequoia Natl. Park	×	×	×			×		×
Whiskeytown-Shasta-Trinity Natl. Recr. Area	×	×	×	×	×	×		
Yosemite Natl. Park	×	×	×	×	×	×	×	×

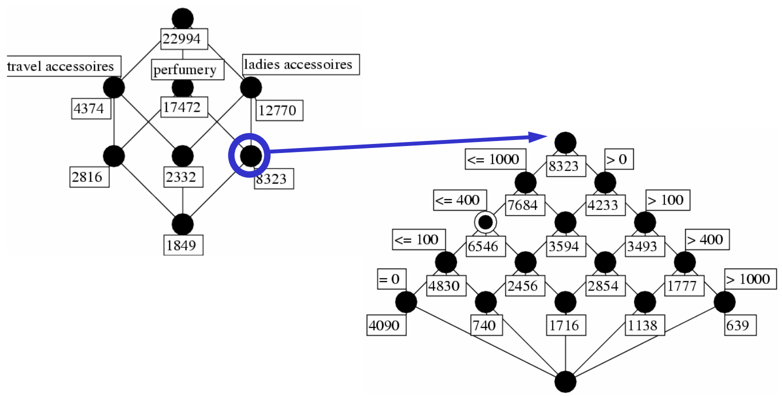
2.1 Basic Notions

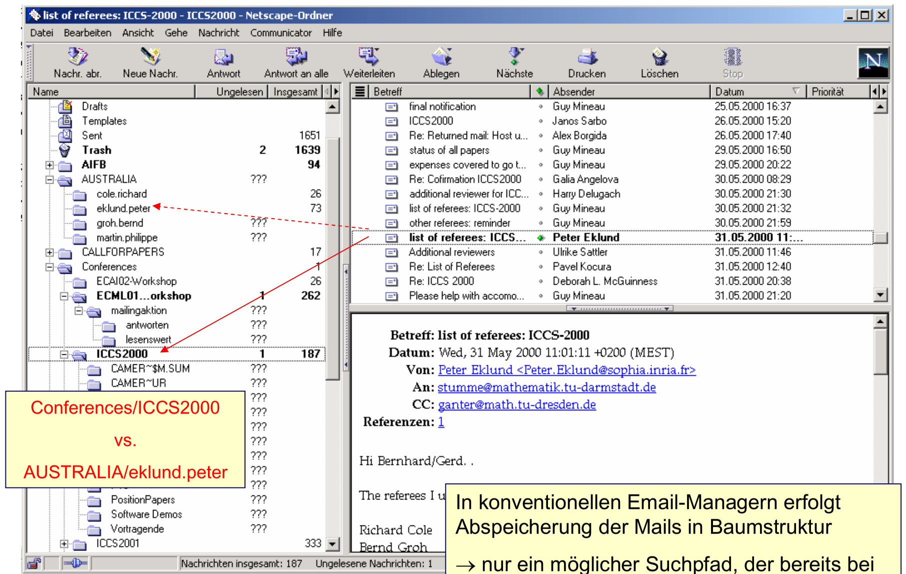
Der **Begriffsverband** zu dem Nationalpark-Kontext



Einige Anwendungen der Formalen Begriffsanalyse

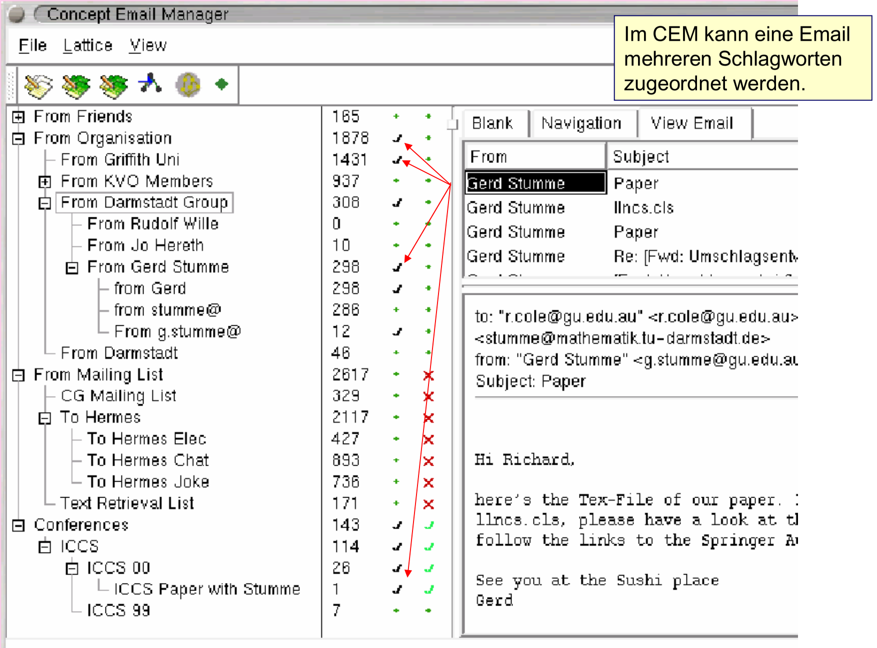
- **Database Marketing bei Jelmoli AG, Zürich**



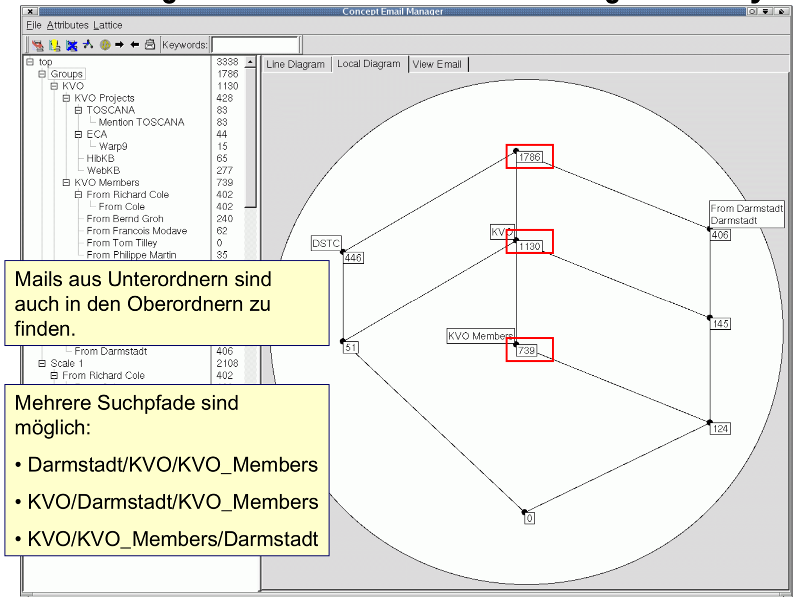


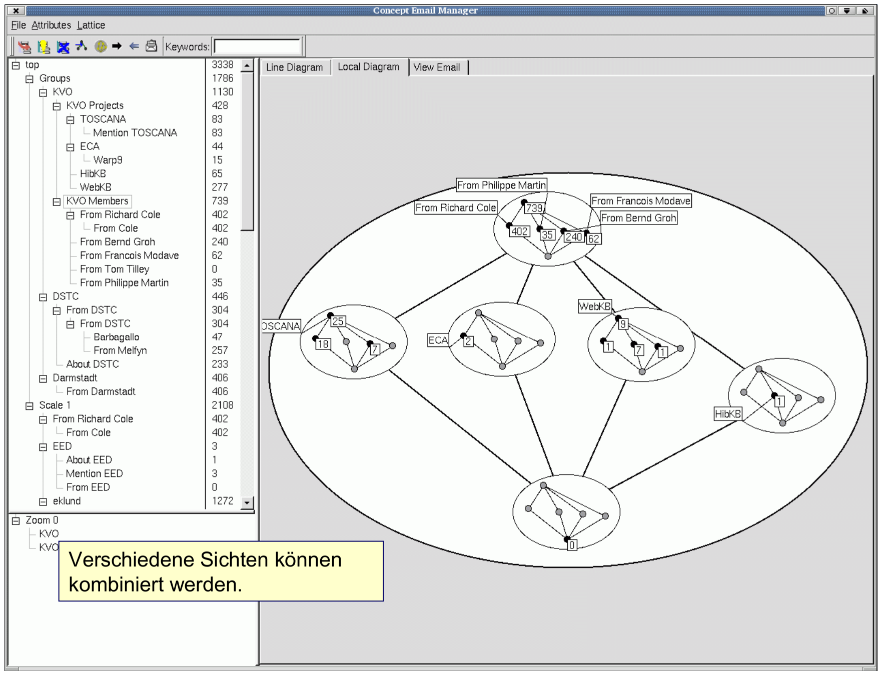
In konventionellen Email-Managern erfolgt Abspeicherung der Mails in Baumstruktur

→ nur ein möglicher Suchpfad, der bereits bei Abspeicherung festgelegt werden muss



Browsing basierend auf Formaler Begriffsanalyse

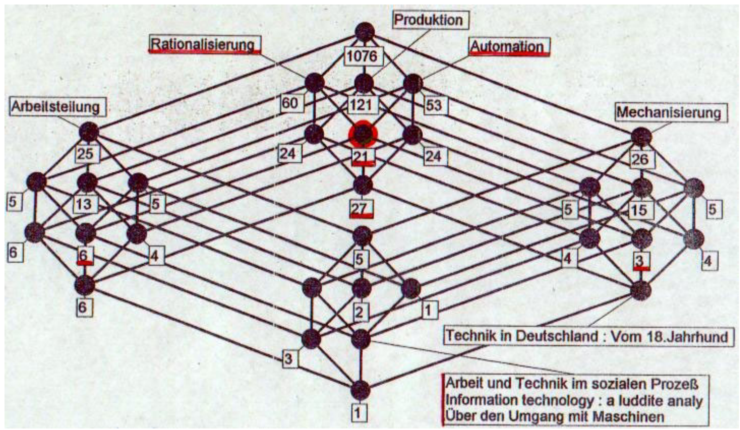




Information Retrieval in der Bibliothek des Zentrums für Interdisziplinäre Technikforschung, TU Darmstadt

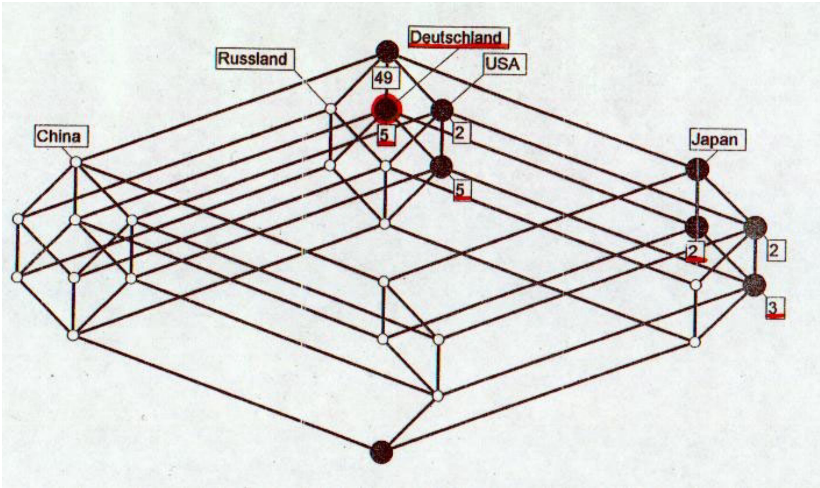
AIFB

Beispiel: Suche nach älterer Literatur über Automatisierung in den wichtigsten Industrienationen



Skala Produktionswandel

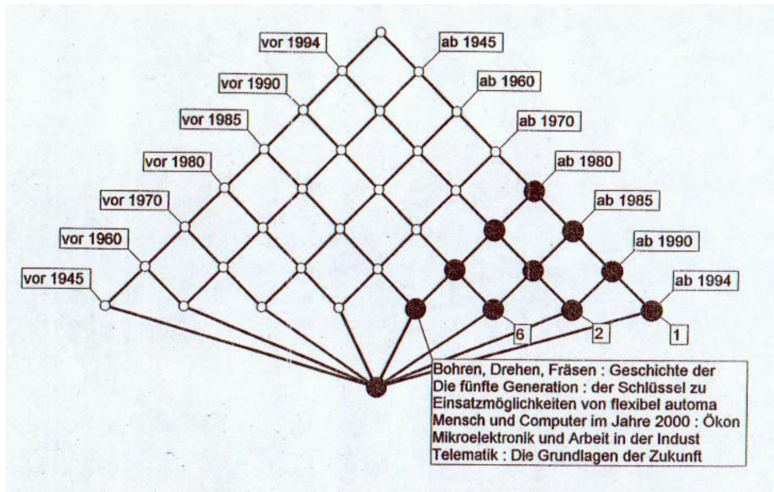
III Formale Begriffsanalyse



Skala Wichtige Industriestaaten

(eingeschränkt auf Bücher mit den Schlagwörtern Automation und Rationalisierung)

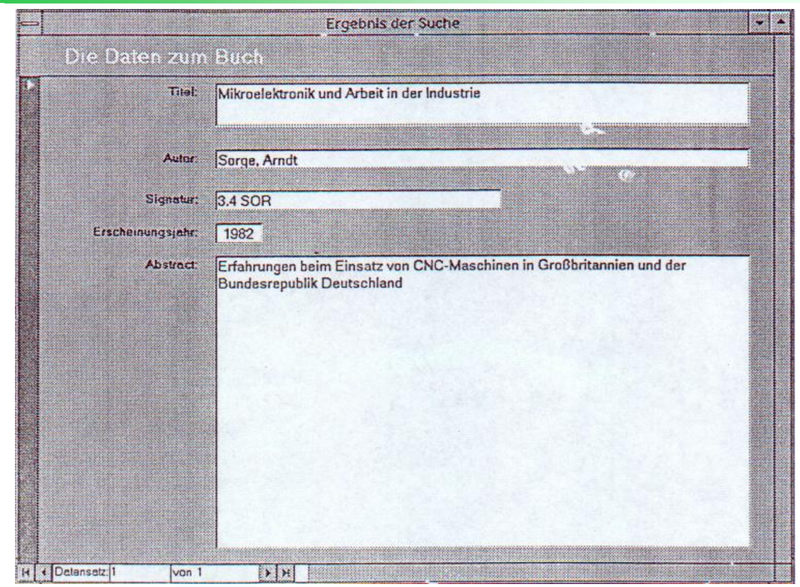
III Formale Begriffsanalyse



Skala Erscheinungsjahr

(eingeschränkt auf Bücher mit den Schlagwörtern Deutschland, Automation und Rationalisierung)

III Formale Begriffsanalyse



III Formale Begriffsanalyse

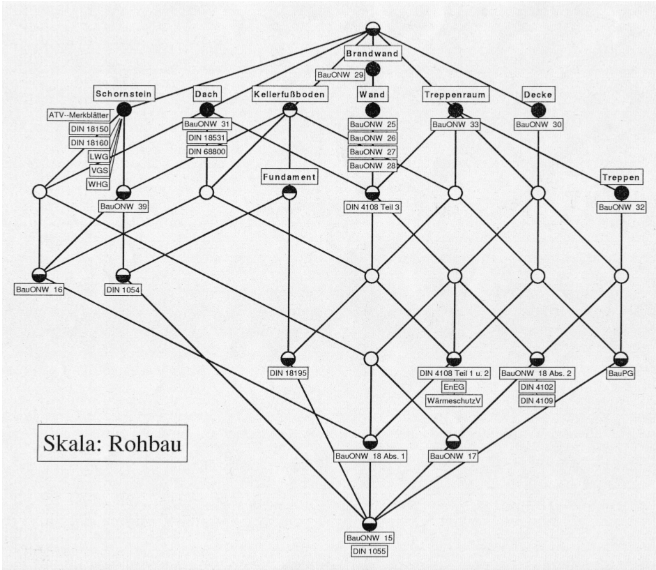
4.4 Beispiel

Baurecht
in Nordrhein-
Westfalen

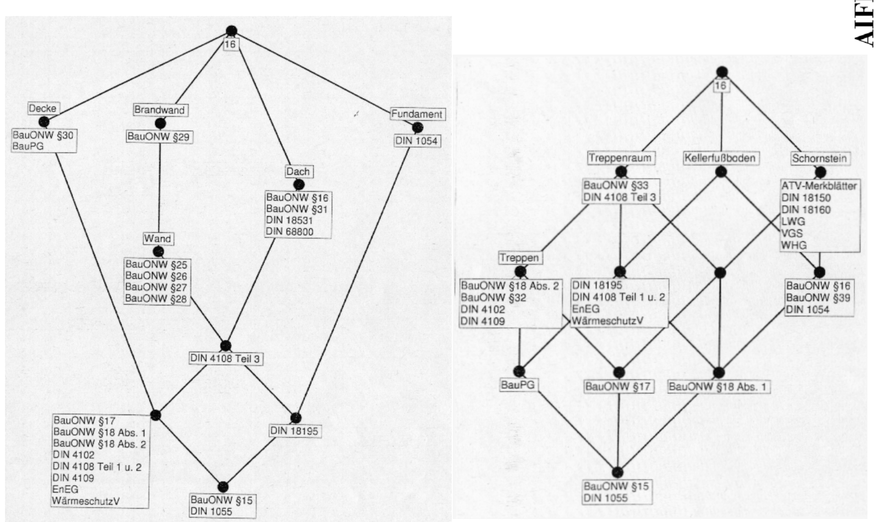
	Dach	Decke	Wand	Brandwand	Treppen	Treppenraum	Fundament	Kellerfußboden	Schornstein
BauONW 15	X	X	X	X	X	X	X	X	X
BauONW 16	X							X	X
BauONW 17	X	X	X	X	X	X			X
BauONW 18 Abs. 1	X	X	X	X		X		X	X
BauONW 18 Abs. 2	X	X	X	X	X	X			
BauONW 25			X	X					
BauONW 26			X	X					
BauONW 27			X	X					
BauONW 28			X	X					
BauONW 29				X					
BauONW 30		X							
BauONW 31	X								
BauONW 32					X	X			
BauONW 33						X			
BauONW 36									
BauONW 39								X	X
BauONW 40									
BimSchG									
BauPG		X			X	X		X	
EnEG	X	X	X	X		X		X	
WHG									X
LWG									X
WärmeschutzV	X	X	X	X		X		X	
HeizAnlV									
BImSchV									
VGS									X
DIN 1054							X	X	X
DIN 1055	X	X	X	X	X	X	X	X	X
DIN 4102	X	X	X	X	X	X			
DIN 4108 Teil 1 u. 2	X	X	X	X		X		X	
DIN 4108 Teil 3	X		X	X		X			
DIN 4109	X	X	X	X	X	X			
DIN 18150									X
DIN 18160									X
DIN 18195	X		X	X		X	X	X	
DIN 18531	X								
DIN 68800	X								
DIN-Normen für Feuerungsanlagen									
DIN-Normen für Entwässerung									
ATV-Merkblätter									X

Aus: D. Eschenfelder, W. Kollewe, M. Skorsky, R. Wille: Ein Erkundungssystem zum Baurecht: Mothoden der Entwicklung eines TOSCANA-Systems. In: G. Stumme, R. Wille (Hrsg.): Begriffliche Wissensverarbeitung - Methoden und Anwendungen. Springer 2000

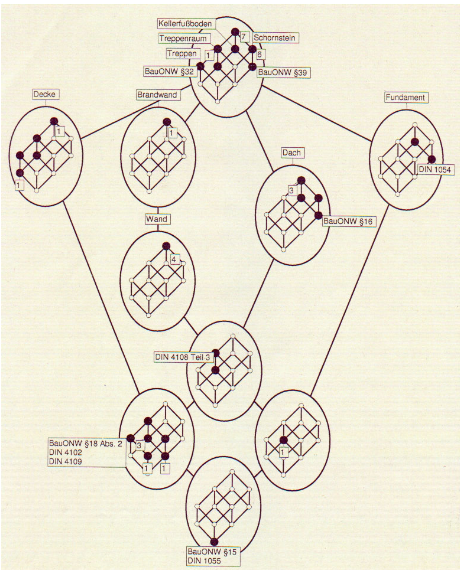
AIFB



AIFB



AIFB



SIMuLLDA

Structured **I**nterlingua **M**ulti**L**ingual
Lexical **D**atabase **A**pplication

Marteen Janssen

Alle Abbildungen aus
SIMuLLDA - a Multilingual Lexical Database Application
using a Structural Interlingua, Dissertation M. Janssen 2002

	horse	male	female	adult	young
horse	×				
stallion	×	×		×	
mare	×		×	×	
foal	×				×
filly	×		×		×
colt	×	×			×

Table 2.5: Analysis of Definitions for Horses

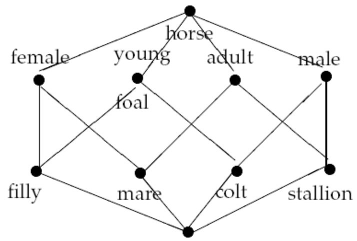


Figure 2.5: Concept Lattice for Horses

English	horse	stallion	mare	foal	filly	colt
Dutch	paard	hengst	merrie	veulen	merrieveulen	hengstveulen
German	Pferd	Hengst	Stute	Fohlen Füllen	Stutenfohlen Stutenfüllen	Hengstfohlen Hengstfüllen
French	cheval	étalon	jument	poulain	pouliche	*no spec. word*
Italian	cavallo	stallone	cavalla giumenta	puledro	puledra	*no spec. word*
Hungarian	ló	csődör	kanca	csikó	fruska	*no spec. word*
Swedish	häst	hingst	sto	föl	ungsto	unghingst
Russian	ло́шадь	жеребе́ц	кобыла	жеребёнок	кобылица	*no spec. word*
Georgian	ცხენი		ჭაკი	კვიცი	[illegible]	*no spec. word*
Malay	kuda	kuda jantan	kuda betina	anak kuda	anak kuda betina	anak kuda jantan

Table 2.6: Words for Horses in Different Languages

SIMuLLDA - M. Janssen, 2002

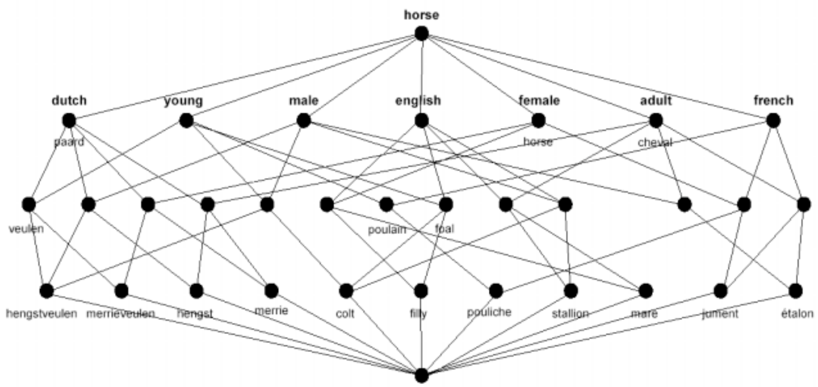


Figure 2.6: Multilingual Connotative Context

SIMuLLDA - M. Janssen, 2002

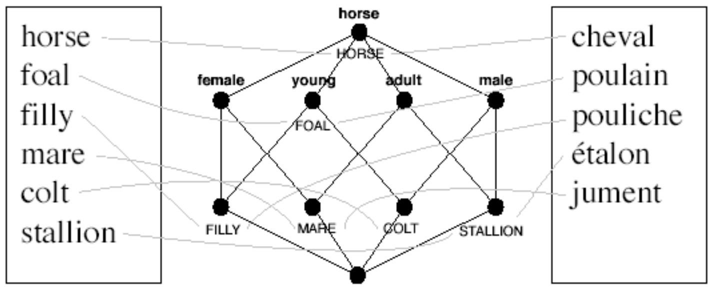


Figure 2.7: Partial Multilingual Set-up

SIMuLLDA - M. Janssen, 2002

Seminar: Formale Begriffsanalyse
Grundlagen der Mengenlehre

Dozentin: Wiebke Petersen
petersew@uni-duesseldorf.de

18. April 2006

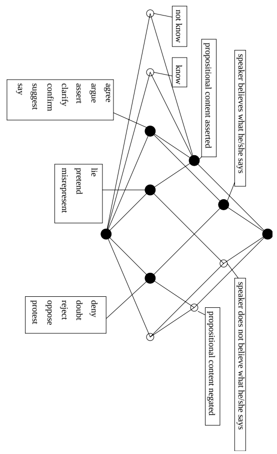


Fig. 1. Verb speech acts in analogy to Großkopf & Harras (1999)

Mengen

Definition 1. *Eine* ***Menge*** *ist eine Zusammenfassung beliebiger Objekte, genannt Elemente, zu einer Gesamtheit, wobei keines der Objekte die Menge selbst sein darf. Zwei Mengen sind* ***gleich****, g.d.w. sie die gleichen Elemente enthalten. Es gibt genau eine Menge, die keine Elemente enthält, die* ***leere Menge*** *$\emptyset$.*

Bemerkung: (Polaritätsprinzip) Für jedes Objekt a und jede Menge M gilt entweder $a \in M$ oder $a \notin M$.

Mengenbeschreibungen — explizit und implizit

explizite Mengendarstellung $\{a_1, a_2, \ldots, a_n\}$ ist die Menge, die genau die Elemente $a_1, a_2, \ldots, a_n$ enthält.
Beispiel: $\{2, 3, 4, 5, 6, 7\}$

implizite Mengendarstellung $\{x : A\}$, bzw. $\{x \mid A\}$, ist die Menge, die genau die Objekte x enthält, auf die die Aussage A zutrifft.
Beispiel: $\{x : x \in \mathbb{N}$ und $x < 8$ und $1 < x\}$

Teilmenge / Obermenge

Definition 2. *Eine Menge T ist eine **Teilmenge** der Menge M ($T \subseteq M$) g.d.w. alle Elemente von T auch Elemente von M sind. M heißt dann **Obermenge** von T. Die Relation $\subseteq$ heißt **Mengeninklusion**. T ist eine **echte Teilmenge** von M ($T \subset M$) g.d.w. $T \subseteq M$ und $T \neq M$.*

Bemerkung: Die Mengeninklusion ist reflexiv, transitiv und antisymmetrisch

Definition 3. *Die **Potenzmenge** $\wp(M)$ einer Menge M ist die Menge aller Teilmengen von M.*

$$\wp(M) \stackrel{\text{def}}{=} \text{def=}\{T \mid T \subseteq M\}$$

Bemerkung: Die Potenzmenge $\wp(M)$ einer n-elementigen Menge M enthält $\wp(M)$ 2^n Elemente.

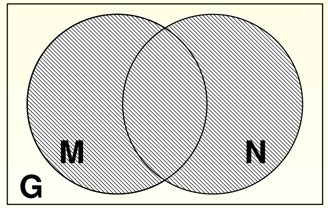
Vereinigung

Definition 4. *Die **Vereinigung** zweier Mengen M und N ($\mathbf{M \cup N}$) ist die Menge aller Objekte, die Element von M oder von N sind:*

$$M \cup N \stackrel{\text{def}}{=} \{x : x \in M \text{ oder } x \in N\}$$

Proposition 5. [Eigenschaften der Vereinigung] *Für beliebige Mengen M und N gilt:*

a) $M \cup M = M$
b) $M \cup \emptyset = M$
c) $M \cup N = N \cup M$
d) $M \cup N \supseteq M$



$M \cup N$

Bemerkung: Mengen werden häufig mit Hilfe von so genannten Venn-Diagrammen visualisiert (siehe Bild).

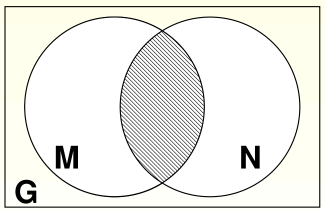
Durchschnitt

Definition 6. *Der **Durchschnitt** zweier Mengen M und N ($\mathbf{M \cap N}$) ist die Menge aller Objekte, die sowohl Element von M als auch von N sind:*

$$M \cap N \stackrel{\text{def}}{=} \{x : x \in M \text{ und } x \in N\}$$

Proposition 7. [Eigenschaften des Durchschnitts] *Für beliebige Mengen M, N und P gilt:*

a) $M \cap M = M$
b) $M \cap \emptyset = \emptyset$
c) $M \cap N = N \cap M$
d) $M \cap N \subseteq M$
e) $M \cup (N \cap P) = (M \cup N) \cap (M \cup P)$
f) $M \cap (N \cup P) = (M \cap N) \cup (M \cap P)$



$M \cap N$

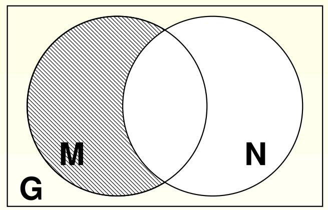
Differenz

Definition 8. *Die **Differenz** zweier Mengen M und N ($\mathbf{M \backslash N}$) ist die Menge aller Objekte, die Element von M aber nicht von N sind:*

$$M \backslash N \stackrel{\text{def}}{=} \{x : x \in M \text{ und } x \notin N\}$$

Proposition 9. [Eigenschaften der Differenz] *Für beliebige Mengen M und N gilt:*

a) $M \backslash M = \emptyset$
b) $M \backslash \emptyset = M$
c) $M \backslash N \subseteq M$
d) $M \backslash N = M$ g.d.w. $M \cap N = \emptyset$
e) $M \subseteq (M \backslash N) \cup N$
f) $M = (M \backslash N) \cup N$ g.d.w. $N \subseteq M$

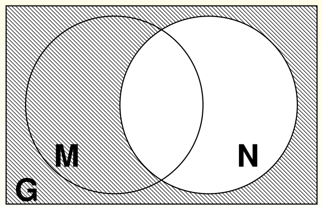


$M \backslash N$

Komplement

Definition 10. *Sei M eine Teilmenge einer gegebenen Grundmenge G, dann ist das Komplement von M bzgl. G ($\mathcal{C}_G(M)$) die Differenz von G und M:*

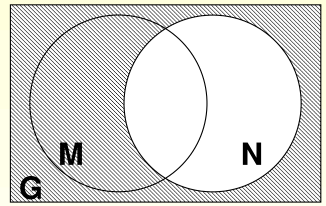
$$\mathcal{C}_G(M) \stackrel{\text{def}}{=} \{x : x \in G \text{ und } x \notin M\}$$



$\mathcal{C}_G(N)$

Eigenschaften des Komplements

Proposition 11. [Eigenschaften des Komplements] *Für beliebige Teilmengen M und N von G gilt:*
a) $\mathcal{C}_G(\emptyset) = G$
b) $\mathcal{C}_G(G) = \emptyset$
c) $\mathcal{C}_G(\mathcal{C}_G(M)) = M$
d) $M \cup \mathcal{C}_G(M) = G$
e) $M \cap \mathcal{C}_G(M) = \emptyset$
f) $M \backslash N = M \cap \mathcal{C}_G(N)$
g) $\mathcal{C}_G(M \cup N) = \mathcal{C}_G(M) \cap \mathcal{C}_G(N)$
(1. Gesetz von ***de Morgan****)*
h) $\mathcal{C}_G(M \cap N) = \mathcal{C}_G(M) \cup \mathcal{C}_G(N)$
(2. Gesetz von de Morgan)



$\mathcal{C}_G(N)$

Kreuzprodukt

Definition 12. *Seien M uns N beliebige Mengen, dann ist das* ***Kreuzprodukt*** *von M und N ($\mathbf{M} \times \mathbf{N}$) die Menge aller geordneten Paare, deren erstes Element ein Element von M ist und deren zweites Element ein Element von N ist.*

$$M \times N \stackrel{\text{def}}{=} \{(m, n) : m \in M \text{ und } n \in N\}$$

Übungsaufgaben

Aufgaben:

1. Zeichnen Sie zu den folgenden Mengen jeweils ein Venn-Diagramm:

   (a) $\mathcal{C}_G(M) \cap \mathcal{C}_G(N)$
   (b) $(M \backslash N) \cup N$
   (c) $(M \cup N) \backslash N$

2. Beweisen Sie mindestens drei Teilaussagen von Proposition 7.

3. Beweisen Sie mindestens drei Teilaussagen von Proposition 11.