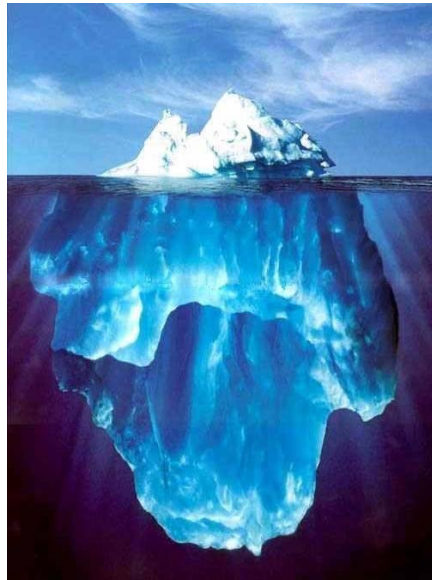


Formale Begriffsanalyse



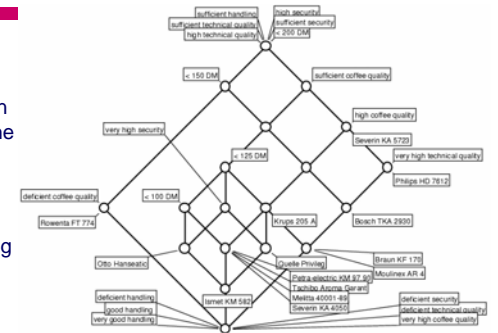
Formale Begriffsanalyse ist um 1980 in Darmstadt entstanden als mathematische Theorie, die eine Formalisierung des Begriffs vom „Begriff“ liefert.

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.

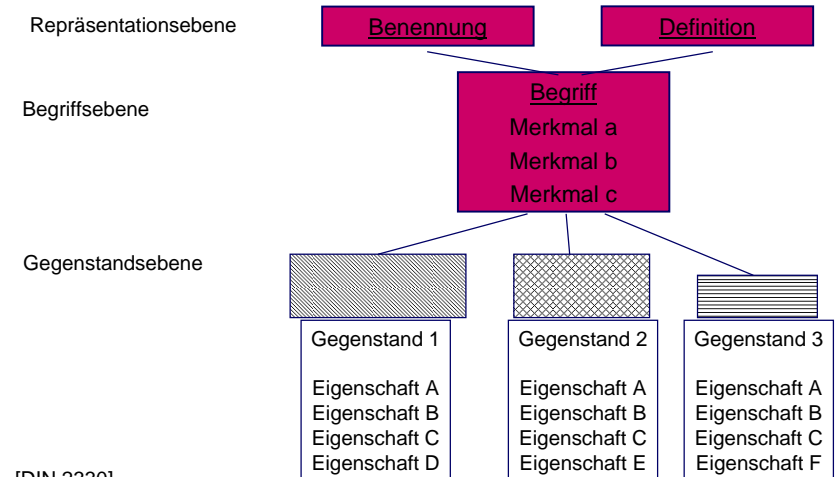


STIFTUNG WARENTEST		KAFFEEESCHNITTEN MIT WÄRMHALTEKANNE (8 bis 10 Tassen)					
test		test, Ausgabe: 1993					
	Mittlerer Preis in DM ca.	Preis für Ersatzkaffe/Gewinnanteil in DM ca.	technische Qualität	Sicherheit	Handhabung	test. Qualitätsurteil	
Gewichtung	35 %	30 %	10 %	10 %	25 %		
Neckermann Bez.-Nr. 8628489	40,-	35,-/7/□	bezug mit Otto Hansa Bez.-Nr. 4227357	○	+	+	zufriedenst.
Otto Hansa Bez.-Nr. 4227357	40,-	30,-/7/□	bezug mit Otto Hansa Bez.-Nr. 4227357	○	+	+	zufriedenst.
Quelle Privileg Bez.-Nr. 7090729	40,-	24,50/7,50	bezug mit Otto Hansa Bez.-Nr. 4227357	○	+	+	zufriedenst.
Severin KA 4650	50,-	35,-/23,-	bezug mit Otto Hansa Bez.-Nr. 4227357	○	+	+	zufriedenst.
Severin KA 4050	80,-	50,-/30,-		+	+	+	gut
Tchibo Aroma Garant An.-Nr. 48489	80,-	27,50/19,50		+	+	+	gut
Imet KM 582 starlight	94,-	47,-/14,-		+	+	+	gut

- Die Formale Begriffsanalyse modelliert Begriffe als Einheiten des Denkens, die aus zwei Teilen bestehen:
 - Der **Begriffsumfang** besteht aus allen Gegenständen, die unter den Begriff fallen.
 - Der **Begriffsinhalt** enthält alle Merkmale, die all diesen Gegenständen zukommen
- FBA wird unter anderem in der Datenanalyse, dem Information Retrieval, Data Mining, und Software Engineering angewandt.

DIN 2330: Begriffe und Ihre Benennungen

Die Begriffsanalyse bewegt sich auf der Begriffsebene, die Repräsentationsebene spielt keine (bzw. nur eine geringe) Rolle.



[DIN 2330]

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

5

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

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

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

6

Für $A, A_1, A_2 \subseteq G$ gilt:

- $A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1'$
- $A \subseteq A''$
- $A' = A'''$

Für $B, B_1, B_2 \subseteq M$ gilt:

- $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$
- $B \subseteq B''$
- $B' = B'''$

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

7

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.

Umfang

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

8

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

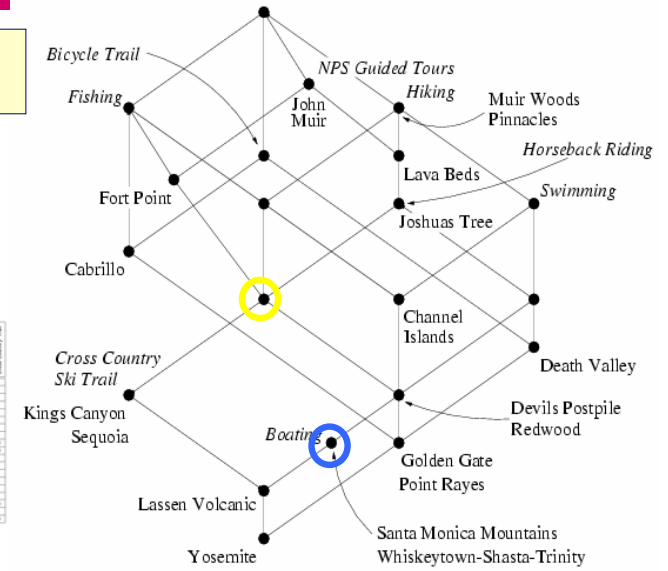
der blaue Umfang ist im gelben Umfang enthalten.

(\Leftrightarrow der gelbe Inhalt ist im blauen Inhalt enthalten.)

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

Der **Begriffsverband** zu dem Nationalpark-Kontext

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



Implikationen

Def.: Eine **Implikation**

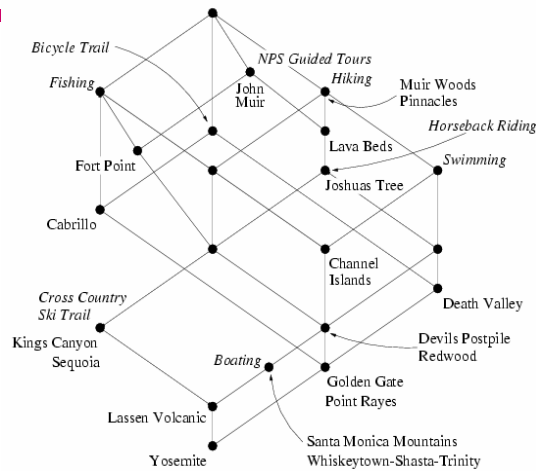
$X \rightarrow Y$ gilt in einem Kontext, wenn jeder Gegenstand, der alle Merkmale aus X hat, auch alle Merkmale aus Y hat.

• **Beispiele:**

{Swimming} \rightarrow {Hiking}

{Boating} \rightarrow {Swimming, Hiking, NPS Guided Tours, Fishing}

{Bicycle Trail, NPS Guided Tours} \rightarrow {Swimming, Hiking}



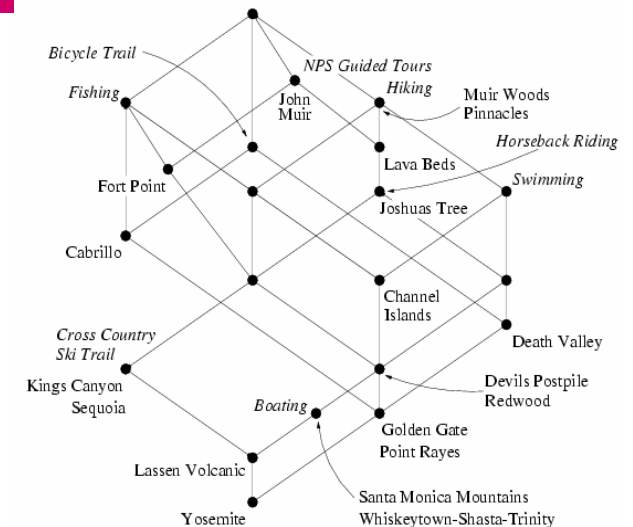
Unabhängigkeit

Def.: Sei $X \subseteq M$. Die Merkmale in X sind **unabhängig voneinander**, wenn es keine nichttrivialen Implikationen zwischen ihnen gibt.

Beispiel:

- Fishing
- Bicycle Trail
- Swimming

sind voneinander unabhängige Merkmale



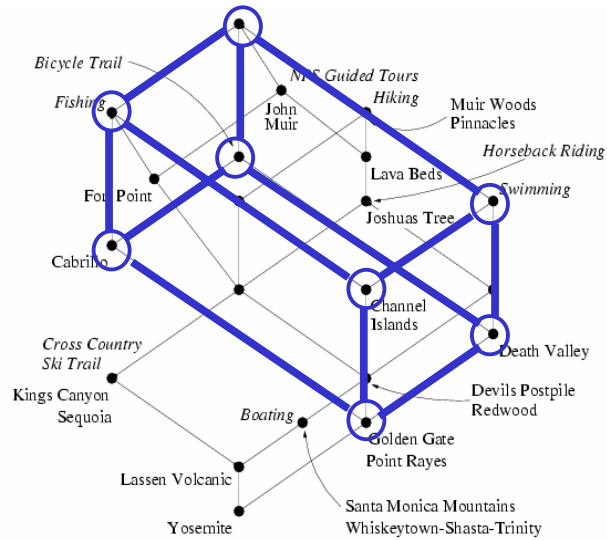
Unabhängigkeit

Lemma: Merkmale sind unabhängig, wenn sie einen (Hyper-) Würfel aufspannen.

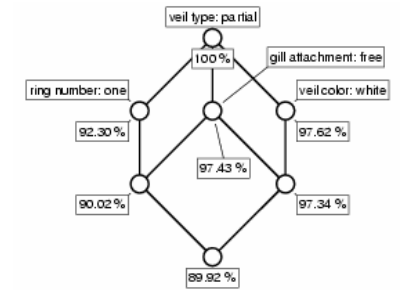
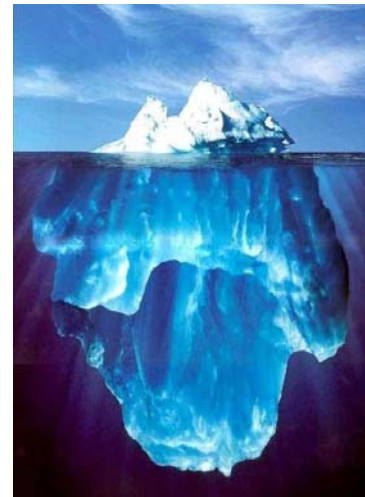
Beispiel:

- Fishing
- Bicycle Trail
- Swimming

sind voneinander unabhängige Merkmale



Iceberg Concept Lattices



minsup = 85%

For minsup = 85% the seven most general of the 32.086 concepts of the Mushrooms database <http://kdd.ics.uci.edu> are shown.

The support of a set $X \subseteq M$ of attributes is given by

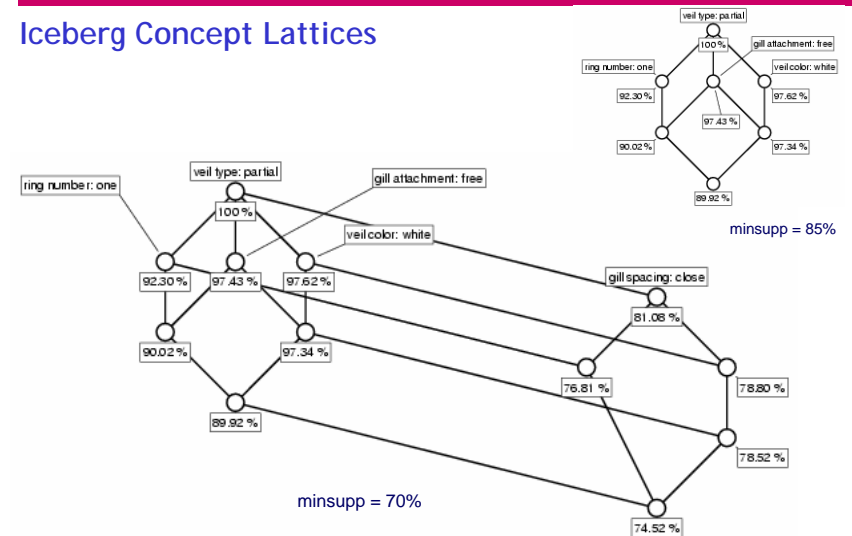
$$\text{supp}(X) = \frac{|X|}{|G|}$$

• Def.: The **iceberg concept lattice** of a formal context (G, M, I) for a given minimal support minsup is the set

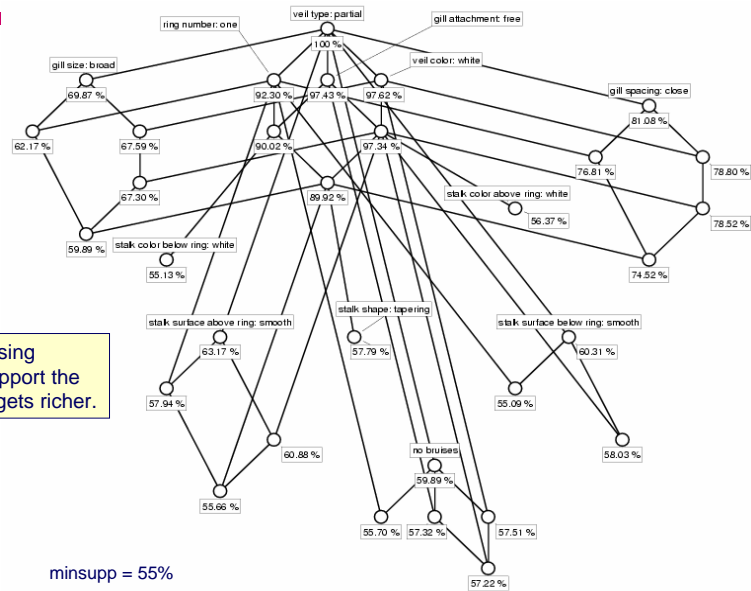
$$\{ (A, B) \in \mathbf{B}(G, M, I) \mid \text{supp}(B) \geq \text{minsup} \}$$

• It can be computed with **TITANIC**. [Stumme et al 2001]
(Will be discussed after the Association Rule Section.)

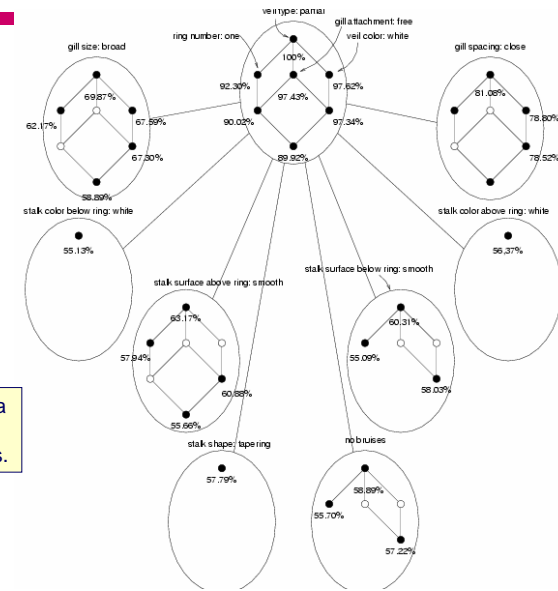
Iceberg Concept Lattices



minsup = 70%



17



18

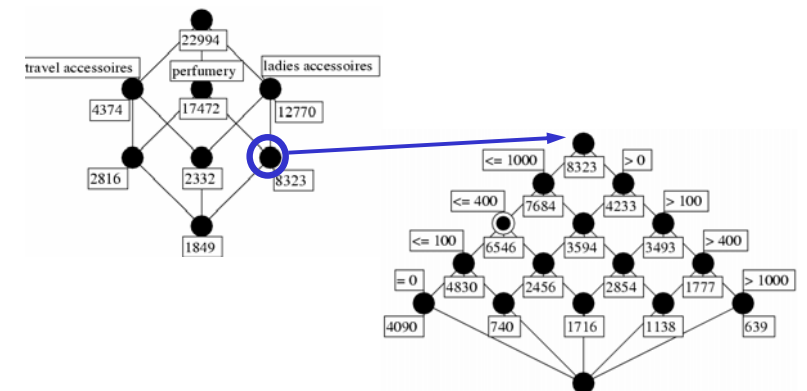
Einige Anwendungen der Formalen Begriffsanalyse

- Analyse von Diabetes-kranken Kindern
- Entwicklung qualitativer Theorien in Musik-Ästhetik
- Database Marketing in einem Schweizer Warenhaus
- Baurecht in NRW
- Analyse der Flugbewegungen am Frankfurter Flughafen
- IT Sicherheitsmanagement
- Begrifflicher Email Manager

19

2.5 Einige Anwendungen der Formalen Begriffsanalyse

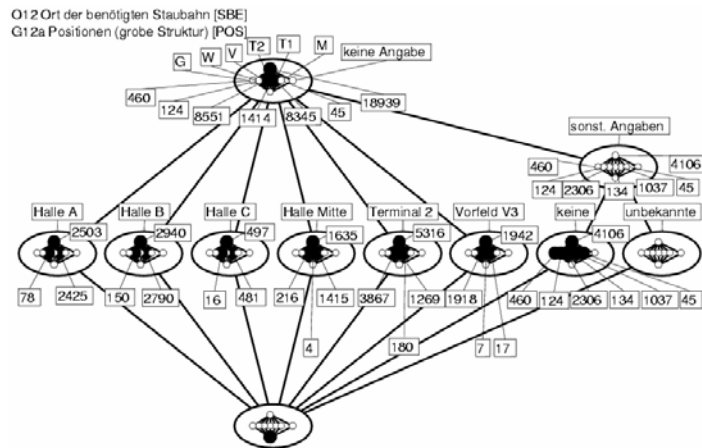
- Database Marketing bei Jelmoli AG, Zürich



20

2.5 Einige Anwendungen der Formalen Begriffsanalyse

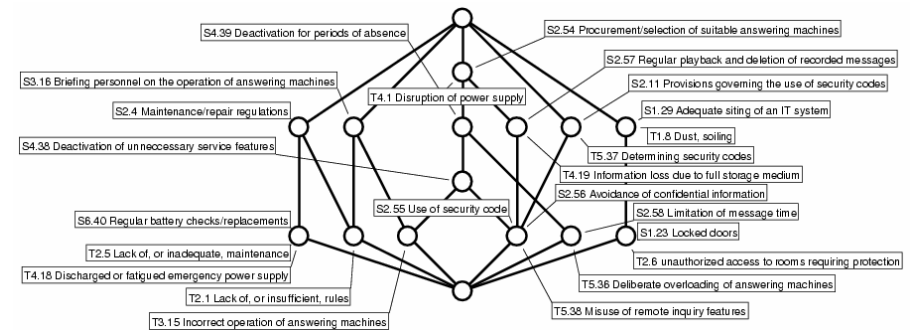
• Analyse der Flugbewegungen am Flughafen Frankfurt



21

2.5 Einige Anwendungen der Formalen Begriffsanalyse

• IT-Grundschutzhandbuch



22

list of references: ICCS-2000 - Netscape-Ordner

Name | Betreff | Datum | Priorität

final notification | Guy Mineau | 25.05.2000 16:37

ICCS2000 | Janos Sarbo | 26.05.2000 15:20

Re: Returned mail: Host u... | Alex Borgida | 26.05.2000 17:40

status of all papers | Guy Mineau | 23.05.2000 16:59

expenses covered to go t... | Guy Mineau | 29.05.2000 20:22

Re: Confirmation ICCS2000 | Gala Angelova | 30.05.2000 09:29

additional reviewer for ICC... | Harry Delugach | 30.05.2000 21:30

list of referes: ICCS-2000 | Guy Mineau | 30.05.2000 21:32

other referes: reminder | Guy Mineau | 30.05.2000 21:59

list of referes: ICCS-2000 | Peter Eklund | 31.05.2000 11:11

Additional reviewers | Ulrike Sattler | 31.05.2000 11:46

Re: List of Refereres | Pavel Kocura | 31.05.2000 12:40

Re: ICCS 2000 | Deborah L. McGuinness | 31.05.2000 20:38

Please help with accomo... | Guy Mineau | 31.05.2000 21:20

Betreff: list of referes: ICCS-2000
Datum: Wed, 31 May 2000 11:01:11 +0200 (MEST)
Von: Peter Eklund <Peter.Eklund@sophia.inra.fr>
An: stumme@mathematik.tu-darmstadt.de
CC: ganter@math.tu-dresden.de

Referenzen: 1

Hi Bernhard/Gerd. .

The referes I u

Richard Cole
Bernd Groh

Nachrichten insgesamt: 187 | Ungelesene Nachrichten: 1

Conferences/ICCS2000 vs. AUSTRALIA/eklund.peter

In konventionellen Email-Managern erfolgt Abspeicherung der Mails in Baumstruktur
→ nur ein möglicher Suchpfad, der bereits bei Abspeicherung festgelegt werden muss

23

Concept Email Manager

File Lattice View

From Friends 185

From Organisation 1878

From Griffith Uni 1431

From KVO Members 937

From Darmstadt Group 308

From Rudolf Wille 0

From Jo Hereth 10

From Gerd Stumme 296

from Gerd 296

from stumme@ 286

From Darmstadt 12

From Mailing List 48

CG Mailing List 2817

To Hermes 329

To Hermes Elec 2117

To Hermes Chat 427

To Hermes Joke 893

Text Retrieval List 736

Conferences 171

ICCS 143

ICCS 00 114

ICCS Paper with Stumme 26

ICCS 99 1

185

1878

1431

937

308

0

10

296

296

286

12

48

2817

329

2117

427

893

736

171

143

114

26

1

7

Blank Navigation View Email

From	Subject
Gerd Stumme	Paper
Gerd Stumme	lincs.cls
Gerd Stumme	Paper
Gerd Stumme	Re: [Fwd: Umschlagsent...

to: "r.cole@gu.edu.au" <r.cole@gu.edu.au>
<stumme@mathematik.tu-darmstadt.de>
from: "Gerd Stumme" <g.stumme@gu.edu.au>
Subject: Paper

Hi Richard.

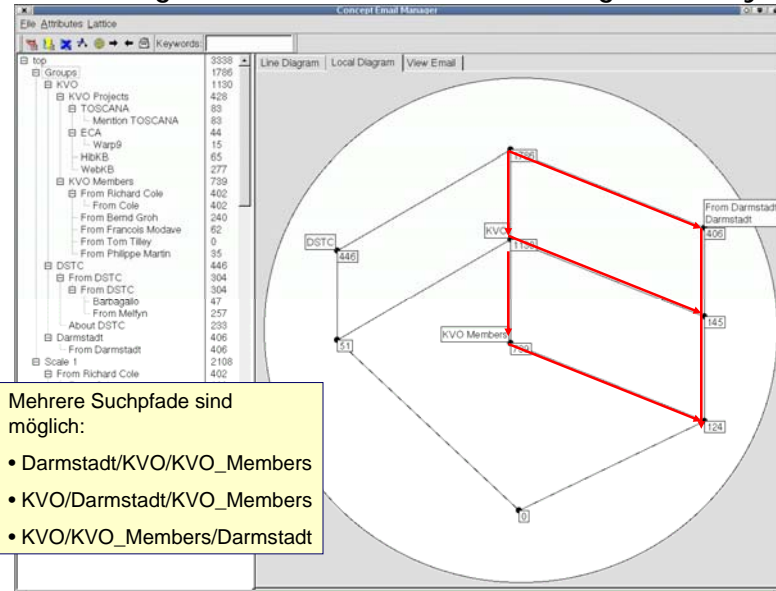
here's the Tex-File of our paper. :
lincs.cls, please have a look at t
follow the links to the Springer A

See you at the Sushi place
Gerd

Im CEM kann eine Email mehreren Schlagworten zugeordnet werden.

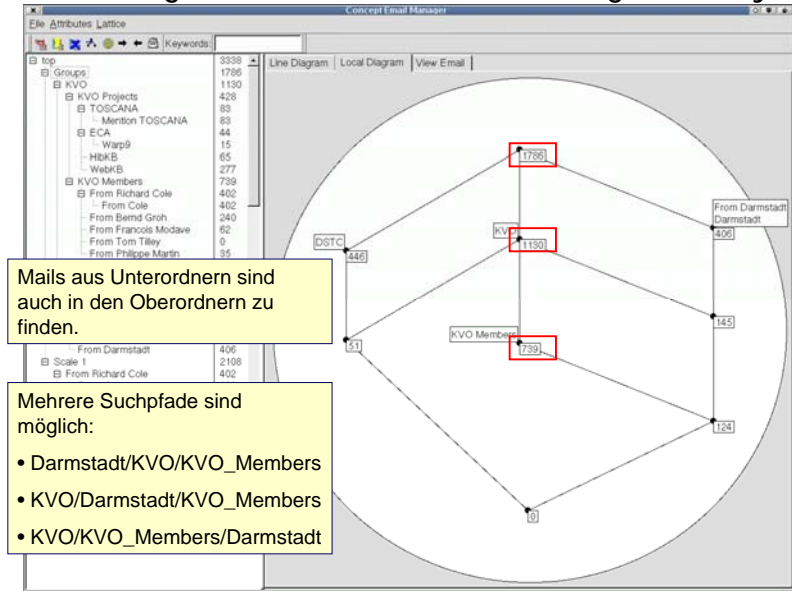
24

Browsing basierend auf Formaler Begriffsanalyse



25

Browsing basierend auf Formaler Begriffsanalyse



26

