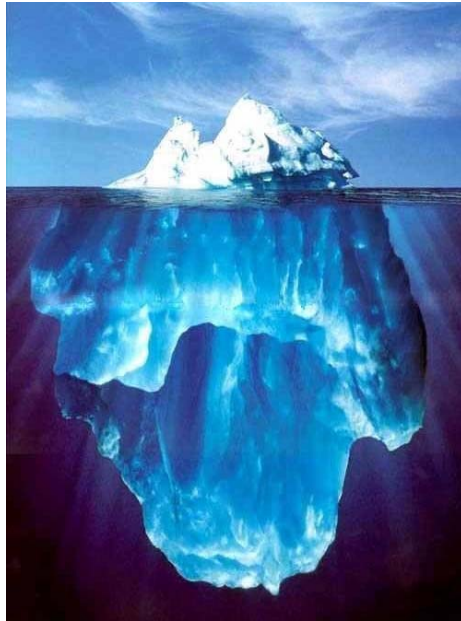


Formale Begriffsanalyse

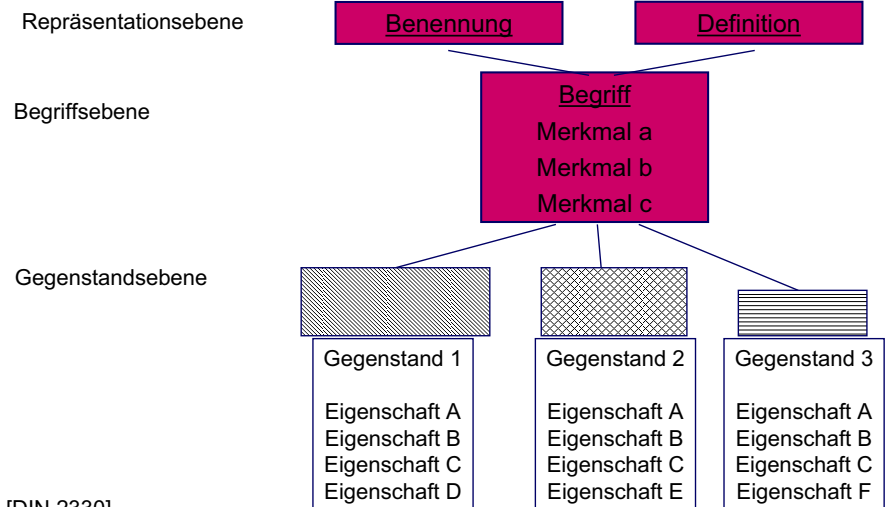


- Die Formale Begriffsanalyse modelliert Begriffe als Einheiten des Denkens, die aus zwei Teilen bestehen:
 - Der **Begriffsumfang** besteht aus allen Gegenständen, die unter den Begriff fallen.
 - Der **Begriffsinhalt** enthält alle Merkmale, die all diesen Gegenständen zukommen
- FBA wird unter anderem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

3

DIN 2330: Begriffe und Ihre Benennungen

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



[DIN 2330]

4

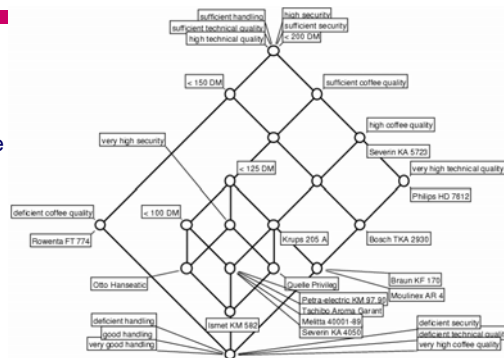
Formale Begriffsanalyse ist um 1980 in Darmstadt entstanden als mathematische Theorie, die eine Formalisierung des Begriffs vom „Begriff“ liefert.

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.



Gewichtung	Mittlerer Preis in DM ca.	Preis für Ersatzkaffein- / Gesamtsatz in DM ca.	Kaffeemaschinen mit WARMHALTEKANNE (8 bis 10 Tassen)				test. Qualitätsurteil
			35 %	30 %	10 %	25 %	
Neckermann Best.-Nr. 8628/409	40,-	35,- / 12	beugt mit Otto Hansatic Best.-Nr. 4327357	+	+	+	zufriedenst.
Otto Hansatic Best.-Nr. 4327357	40,-	30,- / 7	+	+	+	+	zufriedenst.
Quelle Privileg Best.-Nr. 7939720	40,-	24,50 / 17,50	beugt mit Otto Hansatic Best.-Nr. 4327357	+	+	+	zufriedenst.
Severin KA 9650	50,-	35,- / 23,-	beugt mit Otto Hansatic Best.-Nr. 4327357	+	+	+	zufriedenst.
Severin KA 4050	80,-	50,- / 2	+	+	+	+	gut
Tchibo Aroma Garant An.-Nr. 48469	80,-	27,50 / 19,50	+	+	+	+	gut
Ismet KM 582 starlight	84,-	47,- / 14,-	+	+	+	+	gut

2

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.							x	x
Channel Islands Natl. Park		x					x	
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.	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	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	x
Redwood Natl. Park	x	x	x	x			x	
Santa Monica Mts. Natl. Recr. Area	x	x	x	x	x	x		
Sequoia Natl. Park	x	x	x				x	x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x	x		
Yosemite Natl. Park	x	x	x	x	x	x	x	x

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	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	
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.	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	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	x
Redwood Natl. Park	x	x	x	x			x	
Santa Monica Mts. Natl. Recr. Area	x	x	x	x	x	x		
Sequoia Natl. Park	x	x	x				x	x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x	x		
Yosemite Natl. Park	x	x	x	x	x	x	x	x

A

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.							x	x
Channel Islands Natl. Park		x					x	
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.	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	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	x
Redwood Natl. Park	x	x	x	x			x	
Santa Monica Mts. Natl. Recr. Area	x	x	x	x	x	x		
Sequoia Natl. Park	x	x	x				x	x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x	x		
Yosemite Natl. Park	x	x	x	x	x	x	x	x

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	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	
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.	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	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	x
Redwood Natl. Park	x	x	x	x			x	
Santa Monica Mts. Natl. Recr. Area	x	x	x	x	x	x		
Sequoia Natl. Park	x	x	x				x	x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x	x		
Yosemite Natl. Park	x	x	x	x	x	x	x	x

Implikationen

Def.: Eine Implikation

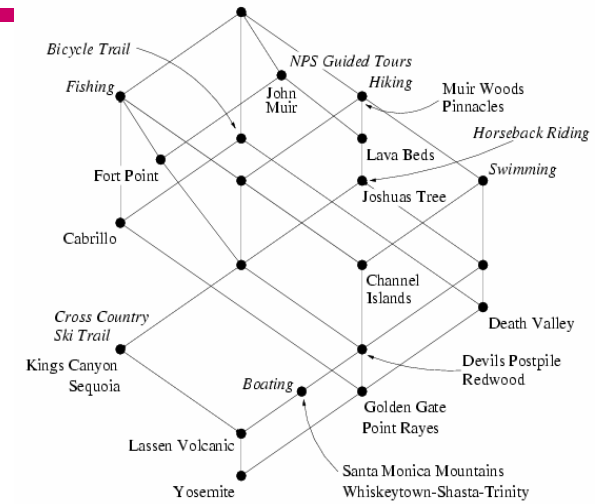
$X \rightarrow Y$ gilt in einem Kontext, wenn jeder Gegenstand, der alle Merkmale aus X hat, auch alle Merkmale aus Y hat.

• Beispiele:

{Swimming} \rightarrow {Hiking}

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

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



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		x		x				
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.		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	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	x	x	x			
Sequoia Natl. Park	x	x	x			x		x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x			
Yosemite Natl. Park	x	x	x	x	x			x

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.)

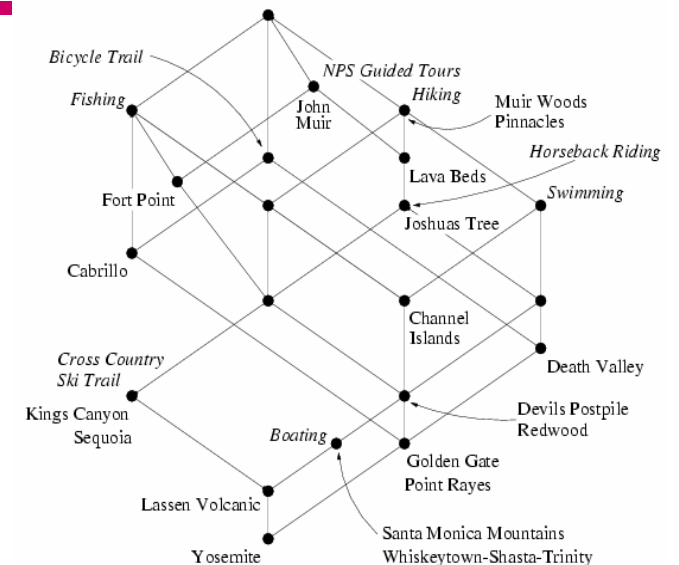
Unabhängigkeit

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

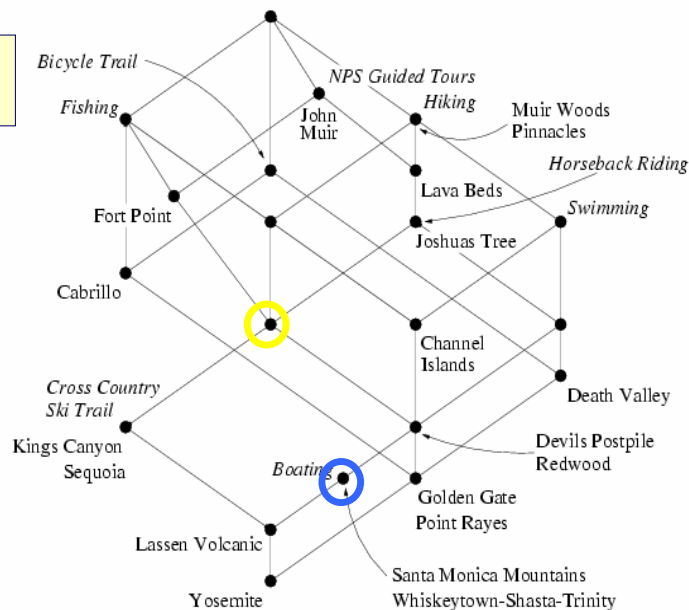
Beispiel:

- Fishing
- Bicycle Trail
- Swimming

sind voneinander unabhängige Merkmale



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.								
Channel Islands Natl. Park		x		x				
Death Valley Natl. Mon.	x	x	x	x				x
Devils Postpile Natl. Mon.		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	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	x	x	x			
Sequoia Natl. Park	x	x	x			x		x
Whiskeytown-Shasta-Trinity Natl. Recr. Area	x	x	x	x	x			
Yosemite Natl. Park	x	x	x	x	x			x

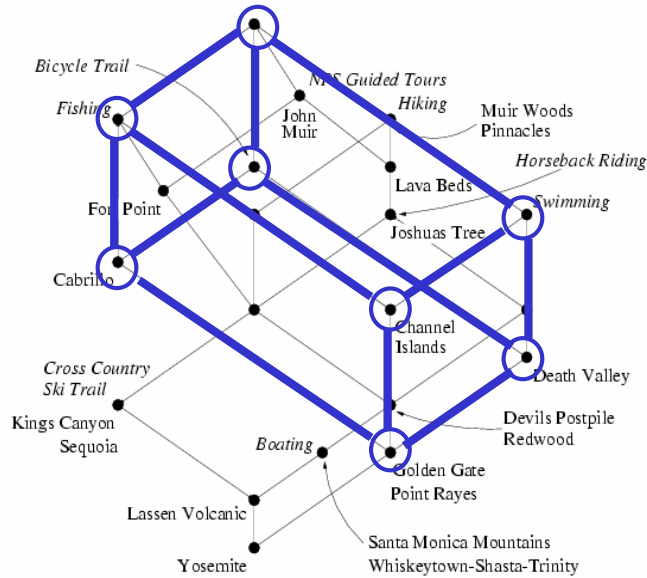
Unabhängigkeit

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

Beispiel:

- Fishing
- Bicycle Trail
- Swimming

sind voneinander unabhängige Merkmale



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

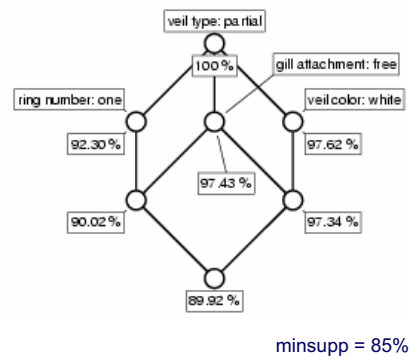
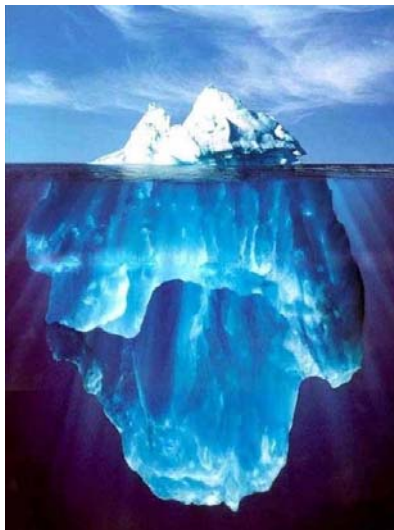
$$\text{supp}(X) = \frac{|X'|}{|G|}$$

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

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

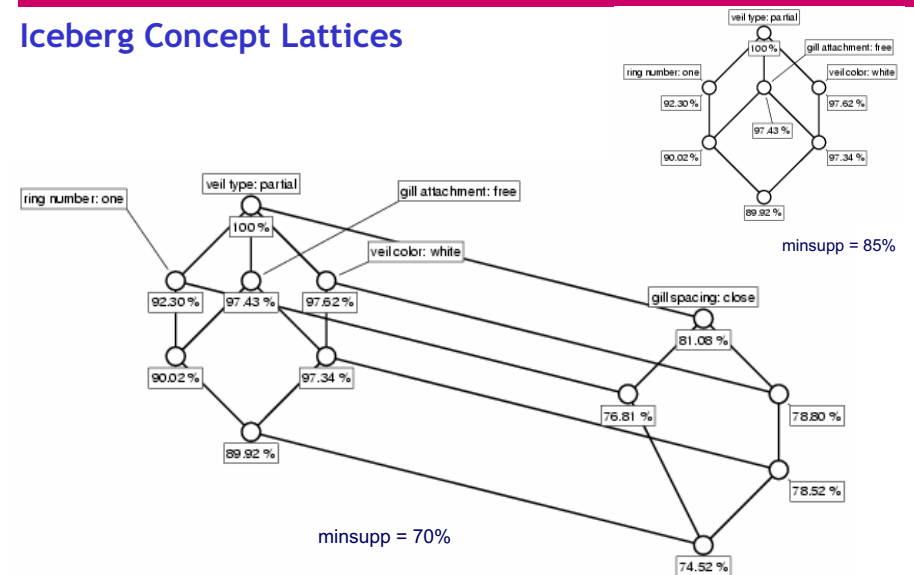
- It can be computed with **TITANIC**. [Stumme et al 2001]
(Will be discussed after the Association Rule Section.)

Iceberg Concept Lattices



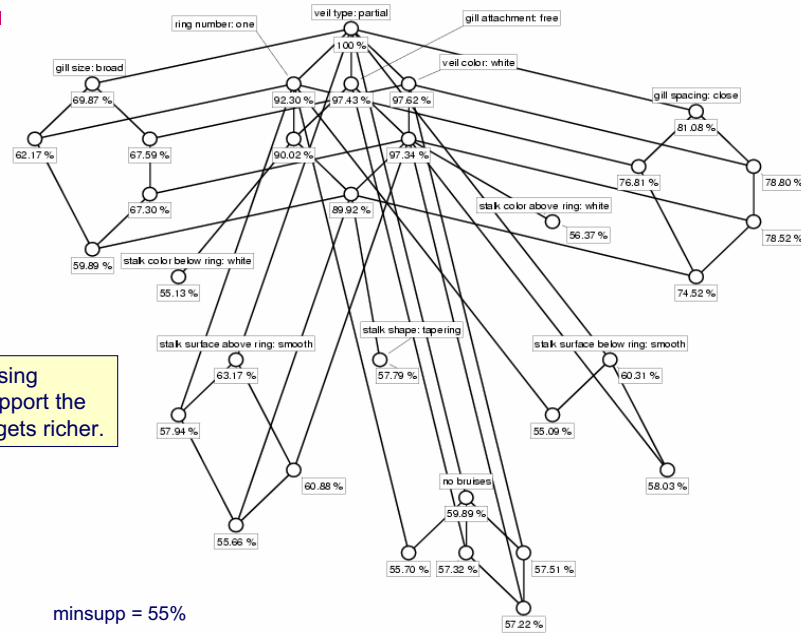
For $\text{minsupp} = 85\%$ the seven most general of the 32.086 concepts of the Mushrooms database <http://kdd.ics.uci.edu> are shown.

Iceberg Concept Lattices



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

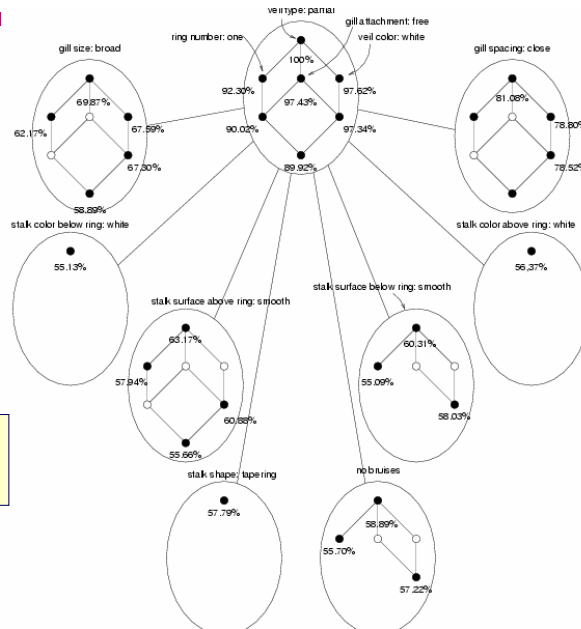


17

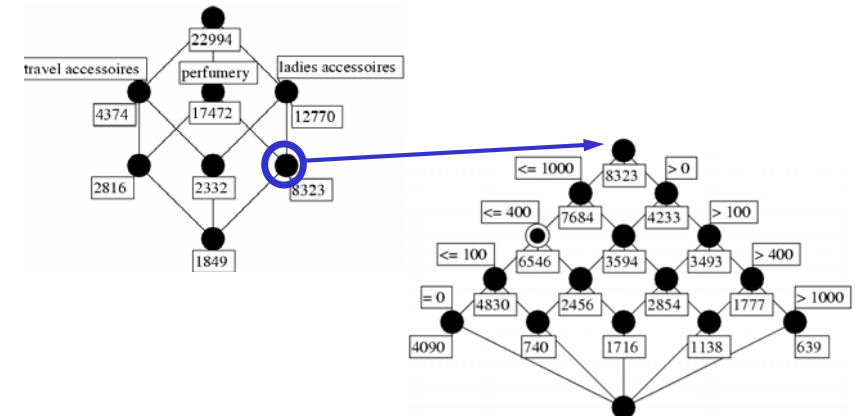
19

2.5 Einige Anwendungen der Formalen Begriffsanalyse

- Database Marketing bei Jelmoli AG, Zürich



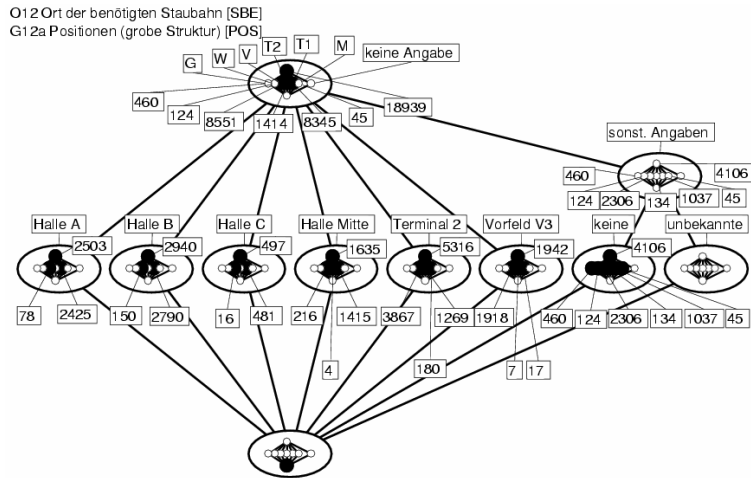
18



20

2.5 Einige Anwendungen der Formalen Begriffsanalyse

• Analyse der Flugbewegungen am Flughafen Frankfurt



21

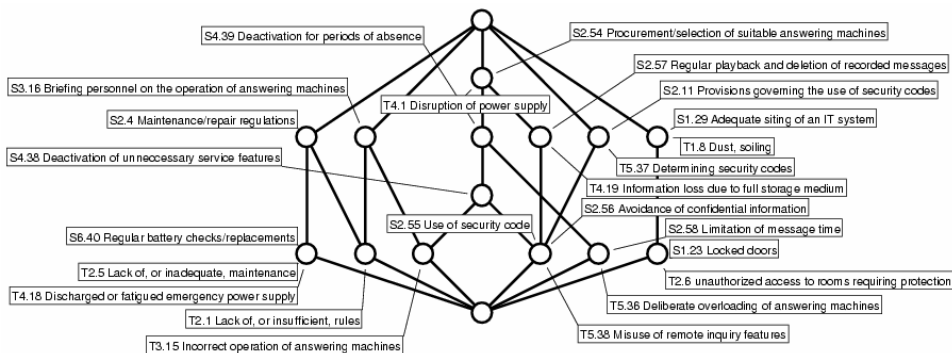
Conferences/ICCS2000
vs.
AUSTRALIA/eklund.peter

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

23

2.5 Einige Anwendungen der Formalen Begriffsanalyse

• IT-Grundschutzhandbuch



22

Im CEM kann eine Email mehreren Schlagworten zugeordnet werden.

24

Browsing basierend auf Formaler Begriffsanalyse

Mehrere Suchpfade sind möglich:

- Darmstadt/KVO/KVO_Members
- KVO/Darmstadt/KVO_Members
- KVO/KVO_Members/Darmstadt

Verschiedene Sichten können kombiniert werden.

25

Browsing basierend auf Formaler Begriffsanalyse

Mails aus Unterordnern sind auch in den Oberordnern zu finden.

Mehrere Suchpfade sind möglich:

- Darmstadt/KVO/KVO_Members
- KVO/Darmstadt/KVO_Members
- KVO/KVO_Members/Darmstadt

26