

Begriffsverbände

"Begriff": in Kapitel 2 (Generalisierungs-Halbordnung, Entscheidungsbäume) bisher nur beschrieben durch Menge von Gegenständen (Instanzen), die dazugehören (*Begriffsumfang*), bzw. durch entsprechende Boolesche Funktion

– zu einem Begriff gehört aber auch eine Menge von Merkmalen (Attributen), die alle Gegenstände des Begriffs erfüllen müssen:

Begriffsinhalt

Dualität und Antagonismus zwischen Begriffsumfang und -inhalt:

viele Merkmale \Rightarrow nur wenige Gegenstände können alle erfüllen
viele Gegenstände \Rightarrow nur wenige Merkmale treffen auf alle zu

formal könnten die Rollen von Gegenständen und Merkmalen vertauscht werden: "dualer Begriff"

Um mit Gegenständen und Merkmalen "gleichwertig" umgehen zu können, müssen wir das Datenmodell ändern

zunächst nur binärwertige Merkmale: trifft zu oder trifft nicht zu

Erweiterung auf mehrwertige Merkmale möglich, später

Definition eines "formalen Kontexts":

- $K := (G, M, I)$: formaler Kontext
- G : Menge der Gegenstände
- M : Menge der Merkmale
- $I \subseteq G \times M$: Inzidenzrelation $((g, m) \in I$
 \Leftrightarrow „m trifft auf g zu“)

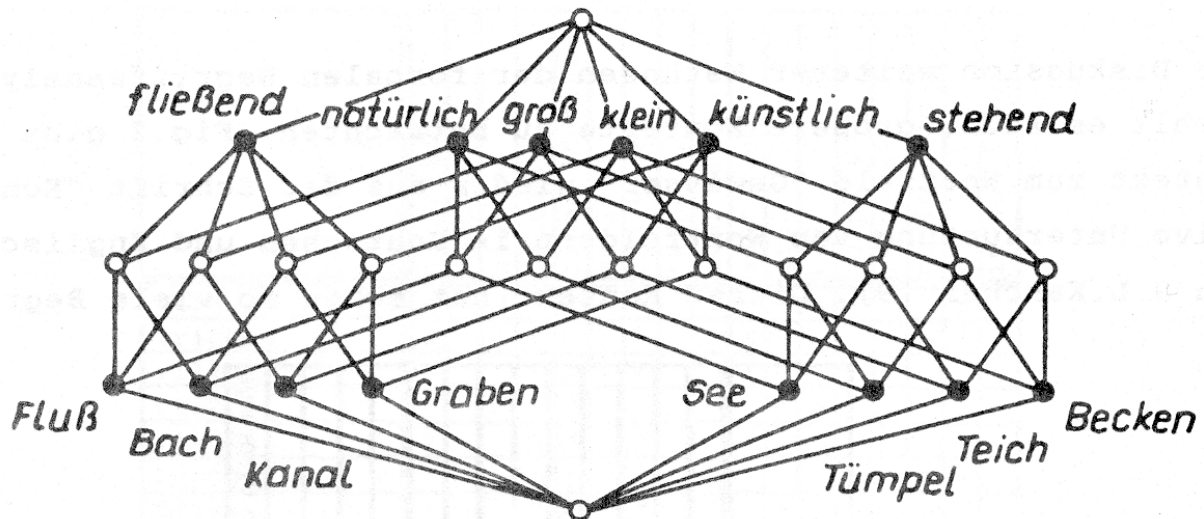
(hier u. im Folgenden nach Ganter und Wille (o.J.),
 vgl. a. Ganter & Wille 1996)

Darstellung eines formalen Kontexts als Kreuzchentabelle
 Zeilen: Gegenstände, Spalten: Merkmale

Beispiel:

	natürlich	künstlich	stehend	fließend	binnenländisch	maritim	konstant	temporär	sehr groß	sehr klein	groß	klein	sumpfig	vulkanischen Ursprungs
Rinnsal	x			x	x		x			x				
Bach	x			x	x		x					x		
Fluß	x			x	x		x				x			
Strom	x			x	x		x		x					
Kanal		x		x	x		x							
Haff	x		x			x	x				x			
Meer	x		x			x	x		x					
Lache	x		x		x			x		x			x	
Pfütze	x		x		x			x				x	x	
Pfuhl	x		x		x		x					x	x	
Tümpel	x		x		x		x				x		x	
Teich	x		x		x		x				x			
Weiher		x	x		x		x				x			
See m.	x		x		x		x		x					
Maar	x		x		x		x							x

Ein Ergebnis der "formalen Begriffsanalyse" wird eine übersichtlichere Darstellung sein, das *Linien-
diagramm*:



(Beisp. aus Kipke & Wille 1987)

Definition "*formaler Begriff*":

$$A' := \{m \in M \mid g \text{ I } m \text{ für alle } g \in A\}$$

$$B' := \{g \in G \mid g \text{ I } m \text{ für alle } m \in B\}$$

(A,B) heißt **formaler Begriff** des Kontextes (G,M,I) mit **Umfang** A und **Inhalt** B, wenn $A \subset G, B \subset M, A' = B$ und $B' = A$.

$B(G, M, I)$: Menge der Begriffe des Kontextes (G, M, I)

A' : Übergang von Zeilen auf die mit Kreuzchen (in allen ausgewählten Zeilen) markierten Spalten

B' : dual, Übergang von Spalten auf die mit Kreuzchen markierten Zeilen

doppelte Anwendung des "Strich-Operators": führt wieder von Gegenstandsmengen auf Gegenstandsmengen, bzw. von Merkmalsmengen auf Merkmalsmengen

Bemerkungen zu Begriffen

- (A'', A') ist stets ein Begriff
- A'' ist der kleinste Begriffsumfang, der A umfaßt
- $A \subseteq G$ ist genau dann ein Begriff, wenn $A = A''$

Definition:

Begriffsverband

(A_1, B_1) ist **Unterbegriff** von (A_2, B_2) ,
falls $A_1 \subseteq A_2$.

Dies definiert eine Ordnung \leq auf $B(G, M, I)$.

$(B(G, M, I), \leq)$ ist der **Begriffsverband**
zu (G, M, I) .

Hauptsatz über Begriffsverbände

$B(G, M, I)$ ist ein vollständiger Verband.

$$\text{Infimum: } \bigwedge_{t \in T} (A_t, B_t) = \left(\bigcap_{t \in T} A_t, \left(\bigcup_{t \in T} B_t \right)'' \right)$$

$$\text{Supremum: } \bigvee_{t \in T} (A_t, B_t) = \left(\left(\bigcup_{t \in T} A_t \right)'', \bigcap_{t \in T} B_t \right)$$

Bedeutung von Begriffsverbänden

Ein Begriffsverband kann gesehen werden als:

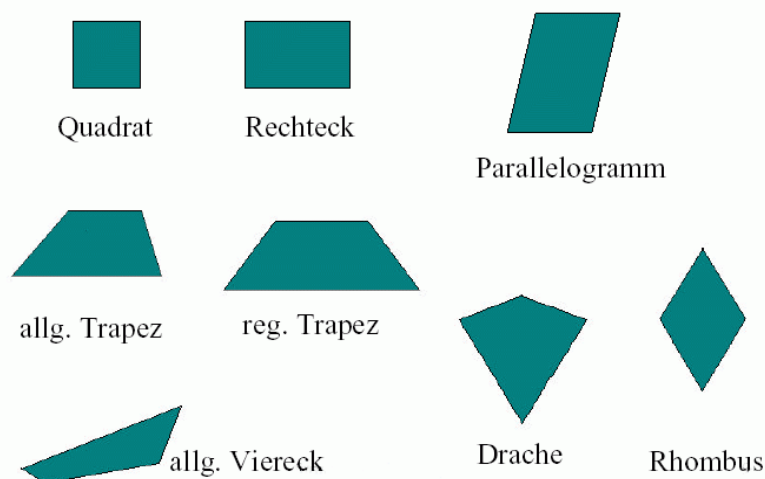
- Hierarchische Klassifikation von Gegenständen.
- System von Merkmalsimplikationen.
- Struktur zum Darstellen und Abfragen von Wissen.

Gegenstands- und Merkmalbegriffe

- **Gegenstandsinhalt:** $g' := \{m \in M \mid gIm\}$
- **Merkmalumfang:** $m' := \{g \in G \mid gIm\}$
- **Gegenstandsbegriff:** $\gamma_g := (g'', g')$
(Begriff mit kleinstem Umfang, der g umfaßt)
- **Merkmalbegriff:** $\mu_m := (m', m'')$

Ein Beispiel für die Bestimmung eines Begriffsverbandes:

Ebene Vierecke



Kontext der ebenen konvexen Vierecke

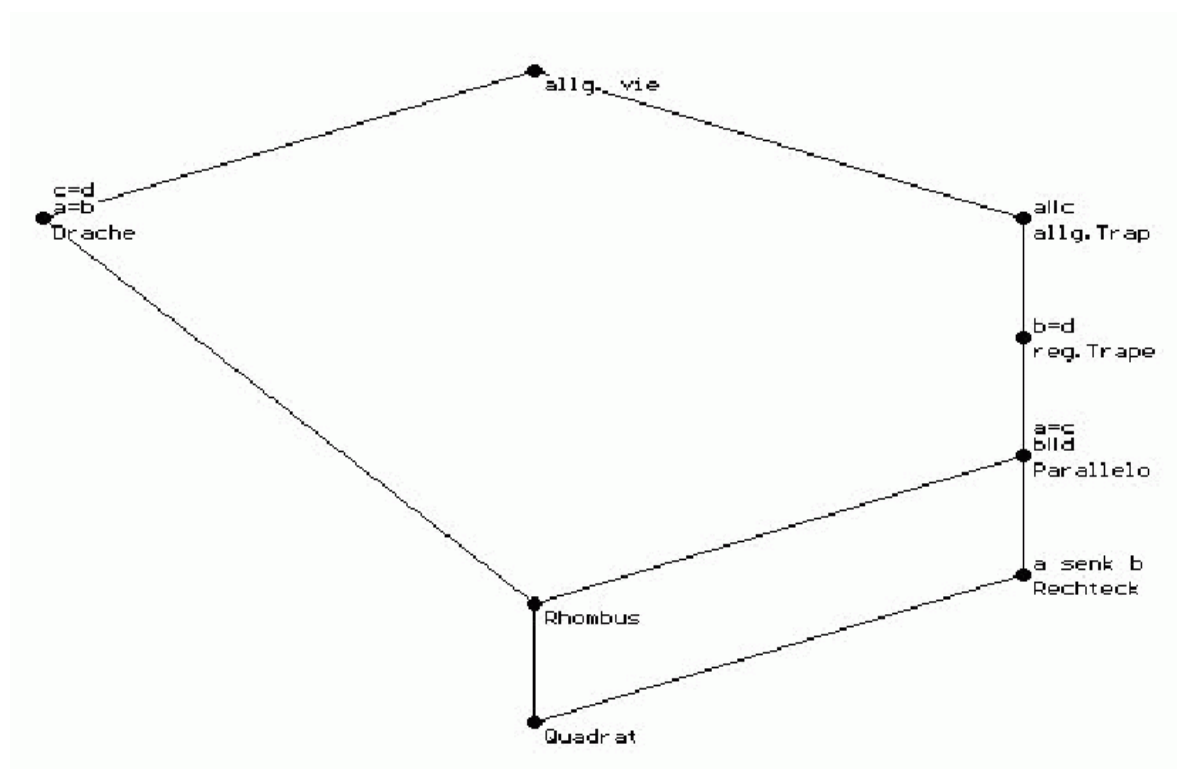
	a c	b d	a=b	c=d	a=c	b=d	a⊥b
Quadrat	X	X	X	X	X	X	X
Rechteck	X	X			X	X	X
Parallelo- gramm	X	X			X	X	
allg. Trapez	X						
reg. Trapez	X					X	
Drache			X	X			
Rhombus	X	X	X	X	X	X	
allg. Viereck							

Wie bestimme ich die Begriffsmenge?

- Jeder Begriffsumfang ist Durchschnitt von “Spaltenumfängen” m' ($m \in M$).
- Jeder Begriffsinhalt ist Durchschnitt von “Zeileninhalten” g' ($g \in G$).
- In Liniendiagrammen reicht es, die Merkmal- und Gegenstandsbegriffe zu beschriften.

Liniendiagramm: Diagramm der partiellen Ordnung " \leq " auf den Begriffen

Liniendiagramm des Begriffsverbandes "ebene, konvexe Vierecke":



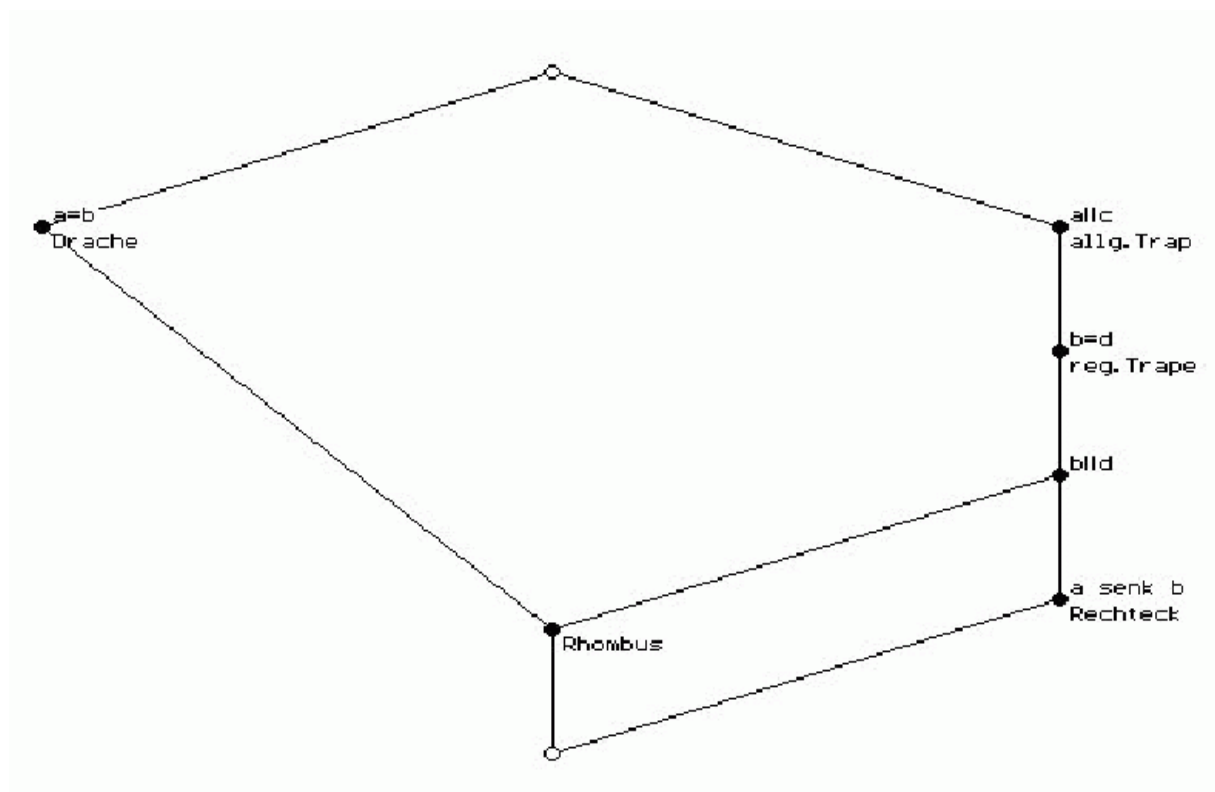
Kontextmanipulationen, die die Struktur des Verbandes erhalten

- Bereinigter Kontext (Gegenstände bzw Merkmale mit gleichem Informationsgehalt werden gestrichen)
- reduzierter Kontext (reduzible Merkmale, solche, die sich als Kombination anderer Merkmale schreiben lassen, werden gestrichen (analog: reduzible Gegenstände))

Beispiel: Reduzierter Kontext der ebenen, konvexen Vierecke

	$a \parallel c$	$b \parallel d$	$a=b$	$b=d$	$a \perp b$
Rechteck	X	X		X	X
allg. Trapez	X				
reg. Trapez	X			X	
Drache			X		
Rhombus	X	X	X	X	

Reduzierter Begriffsverband zu diesem Beispiel:



1. Algorithmus zur Bestimmung der Begriffe eines Kontextes

Satz: Jeder Begriffsumfang ist Durchschnitt von Merkmalsumfängen und jeder Begriffsinhalt ist Durchschnitt von Gegenstandsinhalten.

Schritt 1: Trage G in die Liste der Begriffsumfänge ein.

Schritt m ($m \in M$): Für jede Menge A , die bei einem früheren Schritt in die Liste eingetragen wurde, bilde die Menge $A \cap m'$ und trag sie, falls noch nicht vorhanden, ein.

Besserer Algorithmus (Implementierung: ConImp)

Es gibt einen wesentlich schnelleren und leichter zu programmierenden *Algorithmus zur Erzeugung aller Hüllen eines gegebenen Hüllenoperators*, der auf dieses Problem anwendbar ist.

Implementierung: z.B. ConImp von Peter
Burmeister

Nun: Verallgemeinerung

Mehrwertige Kontexte

- Preis, Gewicht, Genus, Kasus, ... sind typische **mehrwertige Merkmale**.

	Abbruchnach	Teilnehmer	Platzbedarf	Ballspiel	Zuschauerzahl
Handball	60 min.	2×7	mittel	direkt	mittel
Fußball	90 min.	2×11	hoch	direkt	hoch
Volleyball	Ergebnis	2×6	mittel	direkt	mittel
Basketball	40 min.	2×5	mittel	direkt	mittel
Golf	Ergebnis	bel.×1	hoch	mit Schläger	niedrig
Tennis-Einzel	Ergebnis	2×1	mittel	mit Schläger	mittel
Tennis-Doppel	Ergebnis	2×2	mittel	mit Schläger	mittel
Schach	Ergebnis	2×1	gering	nein	niedrig

formaler mehrwertiger Kontext

formal: Ein **mehrwertiger Kontext** ist ein 4-Tupel (G, M, W, I) mit Mengen G, M, W und einer Relation $I \subseteq (G \times M \times W)$. Es gilt:
aus $(g, m, w) \in I$ und $(g, m, v) \in I$ folgt $w = v$

- (G, M, W, I) heißt **n-wertig**, falls W n Elemente hat.

Wie kann man einem mehrwertigen Kontext Begriffe zuordnen?

- Der mehrwertige Kontext wird in einen einwertigen Kontext umgewandelt.
- Die Begriffe des abgeleiteten einwertigen Kontext werden als Begriffe des mehrwertigen Kontextes gedeutet (**begriffliche Skalierung**).
- Die begriffliche Skalierung ist nicht eindeutig!!!

begriffliche Skalierung

1.) Jedes Merkmal eines mehrwertigen Kontextes wird durch einen Kontext (begriffliche Skala) interpretiert.

Def.: Eine **Skala** zum Merkmal m eines mehrwertigen Kontextes ist ein einwertiger Kontext $\mathbb{S}_m := (G_m, M_m, I_m)$ mit $m(G) \subseteq G_m$. Die Gegenstände der Skalen heißen **Skalenwerte**, die Merkmale **Skalenmerkmale**.

2.) Die einzelnen Skalen werden zu einem einwertigen Kontext zusammengefaßt (**Skalierung**).

Beispiel

	Zeit	Ergebnis
60 min.	×	
90 min.	×	
Ergebnis		×
40 min.	×	

§ Abbruch nach

	Ball- spiel	Ballspiel direkt	Ballspiel mit Schläger
direkt	×	×	
mit Schläger	×		×
nein			

§ Ballspiel

	≥hoch	≥mittel	≤mittel	≤gering
hoch	×	×		
mittel		×	×	
gering			×	×

§ Platzbedarf

	≥hoch	≥mittel
hoch	×	×
mittel		×
niedrig		

§ Zuschauerzahl

	2 Parteien	bel. viele Parteien	1 Spieler je Partei	>1 Spieler je Partei
2×7	×			×
2×11	×			×
2×6	×			×
2×5	×			×
bel.×1		×	×	
2×2	×			×
2×1	×		×	

§ Teilnehmer

Schlichte Skalierung

Def: Zum mehrwertigen Kontext (G, M, W, I) mit Skalenkontexten S_m ist (G, N, J) der abgeleitete Kontext bzgl. der schlichten Skalierung, dabei gilt:

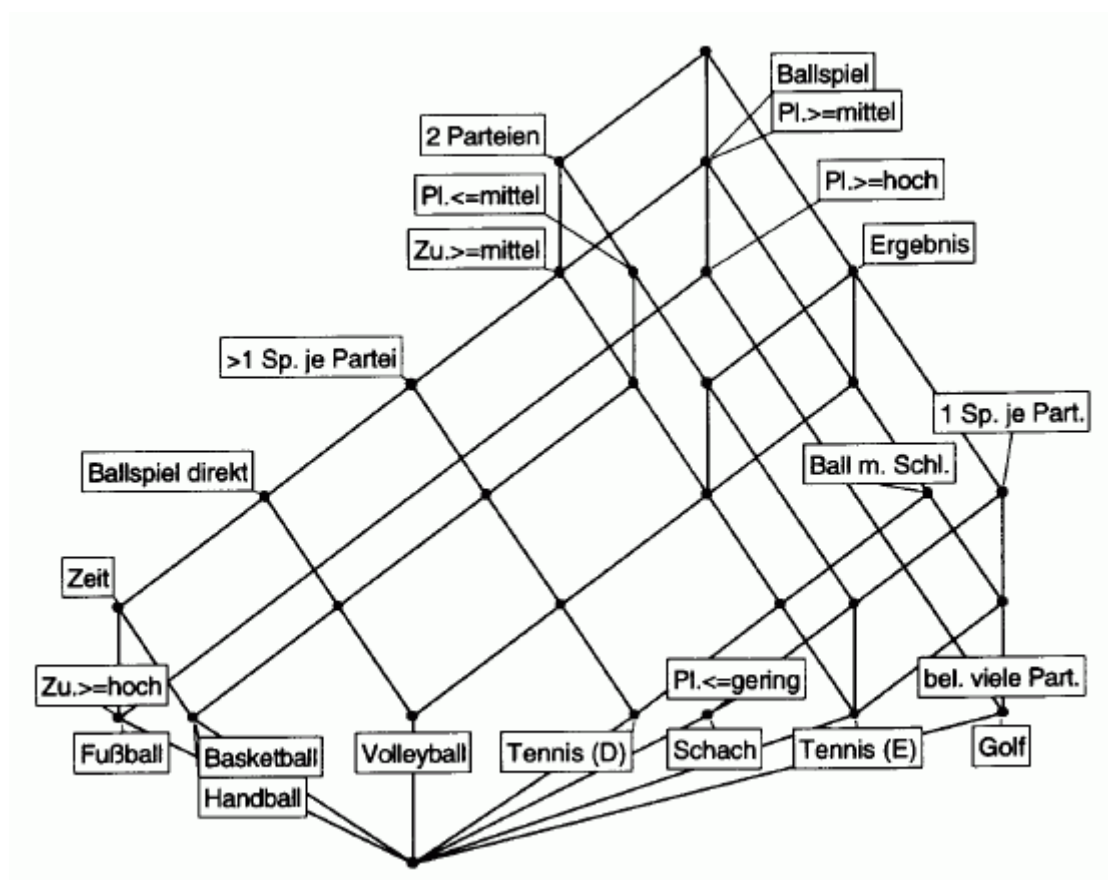
$$N := \bigcup_{m \in M} M_m \quad \text{und} \quad gJ(m, n) := \Leftrightarrow m(g) = w \quad \text{und} \quad wI_m n$$

anschaulich: In der Tabellendarstellung des Kontextes (G, M, W, I) wird jede Merkmalsausprägung $m(g)$ durch die zu $m(g)$ gehörende Zeile des Skalenkontextes S_m ersetzt.

Beispiel

	Abbruch nach		Teilnehmer			Platzbedarf				Ballspiel mit			Zu- schauer	
	Zeit	Erg.	2 bel.	1	>1	$\geq h$	$\geq m$	$\leq m$	$\leq g$	ja	dir.	Schl.	$\geq h$	$\geq m$
Handball	x		x		x		x	x		x	x			x
Fußball	x		x		x	x	x			x	x		x	x
Volleyball		x	x		x		x	x		x	x			x
Basketball	x		x		x		x	x		x	x			x
Golf		x		x	x	x	x			x		x		
Tennis-E.		x	x		x		x	x		x		x		x
Tennis-D.		x	x		x		x	x		x		x		x
Schach		x	x		x			x	x					

Liniendiagramm des resultierenden Begriffsverbandes:



Elementarskalen

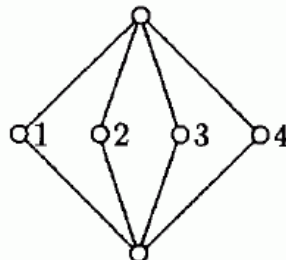
- Skalen können theoretisch beliebig frei gewählt werden, sollten sie aber nicht!
 - Die Eigenschaften des abgeleiteten einwertigen Kontextes hängen von den gewählten Skalen ab.
- ⇒ wähle bedeutungstragende Skalen
(Interpretationsentscheidung)
- nützliche, häufig verwendete Skalen sind die Elementarskalen.

Nominalskala

- Skalierung von Merkmalen, deren Ausprägungen sich gegenseitig ausschließen (z.B. Kasus, Genus,...).

	1	2	3	4
1	×			
2		×		
3			×	
4				×

Die Nominalskala N_4 .

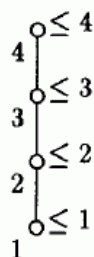


Ordinalskala

- Skalierung mehrwertiger Merkmale, deren Ausprägungen geordnet sind und bei denen jede Merkmalsausprägung die jeweils schwächere impliziert.

$O_4 =$

	1	2	3	4
1	×	×	×	×
2		×	×	×
3			×	×
4				×

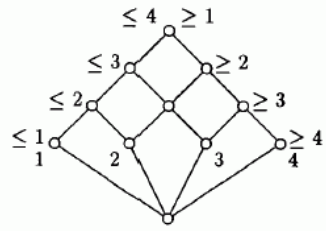


Interordinalskala

- Skalierung von Intervallzugehörigkeiten (kann gut bei der Auswertung von skalierten „trifft zu“-„trifft nicht zu“-Fragebögen verwendet werden).

$$\mathbb{I}_4 =$$

	≤ 1	≤ 2	≤ 3	≤ 4	≥ 1	≥ 2	≥ 3	≥ 4
1	×	×	×	×	×			
2		×	×	×	×	×		
3			×	×	×	×	×	
4				×	×	×	×	×

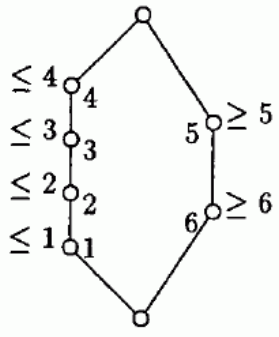


Biordinalskala

- Skalierung von gewerteten Gegensatzpaaren (laut \supset sehr laut leise \supset sehr leise).

$$\mathbb{M}_{4,2} =$$

	≤ 1	≤ 2	≤ 3	≤ 4	≥ 5	≥ 6
1	×	×	×	×		
2		×	×	×		
3			×	×		
4				×		
5					×	
6					×	×



Dichotome Skala

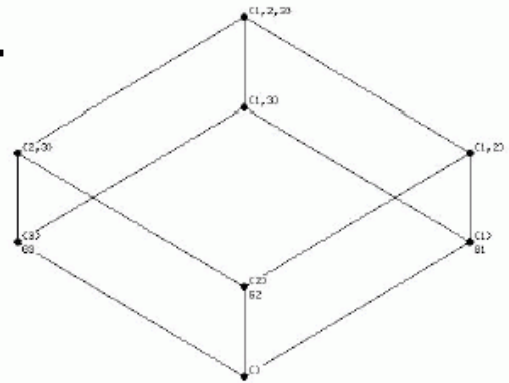
- In einwertigen Kontexten sind „leere Zellen“ nicht begriffsbildend. Mit Hilfe der dichotomen Skala wird die Negation eines Merkmals zur Begriffsbildung hinzugezogen.

	0	1
0	×	
1		×

Boolsche Skala

- In Fällen in denen die Merkmalsausprägung auch eine Menge von Werten sein kann, bietet sich eine Boolesche Skala an.

	{}	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
1		X			X	X		X
2			X		X		X	X
3				X		X	X	X



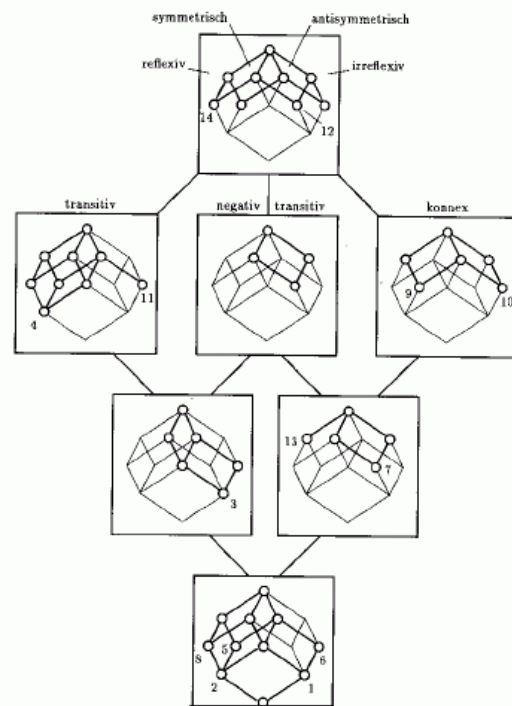
Was macht Diagramme so interessant?

- Diagramme beinhalten die volle Information des Kontextes.
- Visualisierung von Information.
- Strukturelle Einsichten werden sichtbar; Möglichkeit zur Hypothesenbildung.
 - Implikationen sind ablesbar.
 - Hierarchie ist sichtbar.

Probleme bei großen Datensätzen (viele Gegenstände, viele Merkmale): Liniendiagramm wird unübersichtlich

Abhilfe: "**gestuftes Liniendiagramm**", hierarchisch

Bessere Lesbarkeit bei größeren Diagrammen



Gestuftes Liniendiagramm:

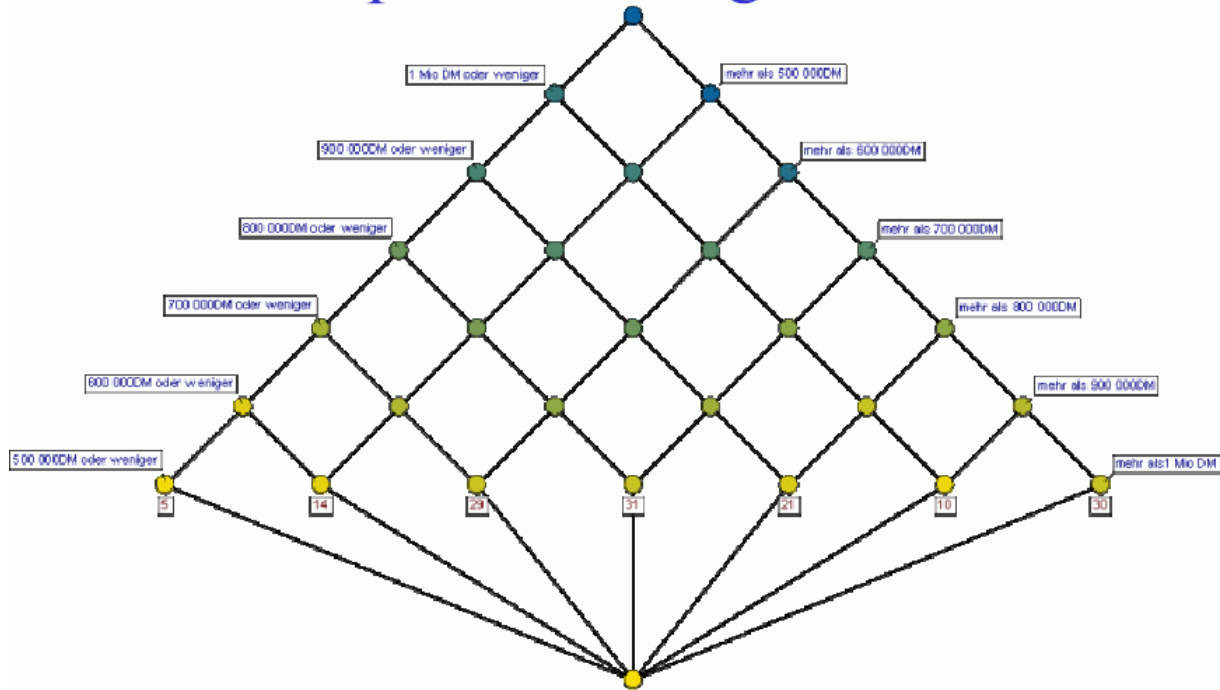
Wie wird ein gestuftes Liniendiagramm gelesen?

- Ein gestuftes Liniendiagramm besteht aus abgegrenzten Feldern.
- Eine einfache Linie zwischen zwei Feldern zeigt an, dass alle Knotenpunkte, die bei der Verschiebung des einen Feldes auf das andere zusammenfallen, im gewöhnlichen Liniendiagramm verbunden sind.
- Enthalten die verbundenen Felder nur Teile einer kongruenten Figur, wird die kongruente „Hintergrundstruktur“ zusätzlich eingezeichnet.

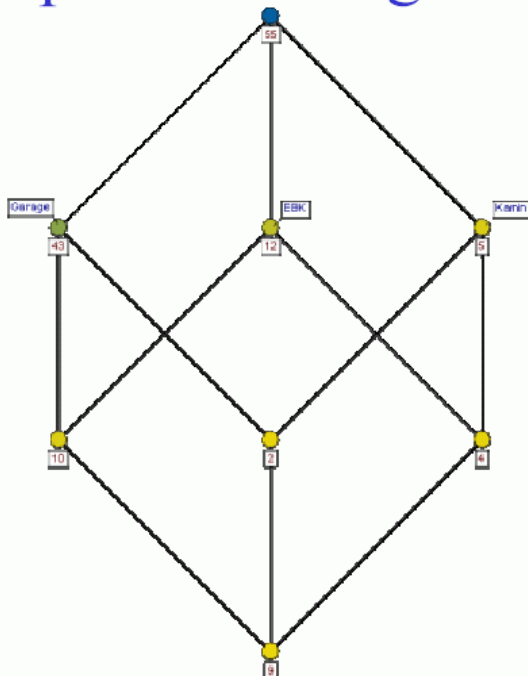
Wie entsteht ein gestuftes Liniendiagramm?

- Teile die Merkmalsmenge in zwei nicht notwendigerweise disjunkte Teilmengen (möglichst bedeutungstragend).
- Zeichne zu beiden Teilmengen die zugehörigen Liniendiagramme.
- Ersetze in einem der Diagramme die Knotenpunkte durch Kästen, in die du jeweils eine Kopie des zweiten Diagramms zeichnest, achte darauf, welche Knotenpunkte tatsächlich besetzt sind.

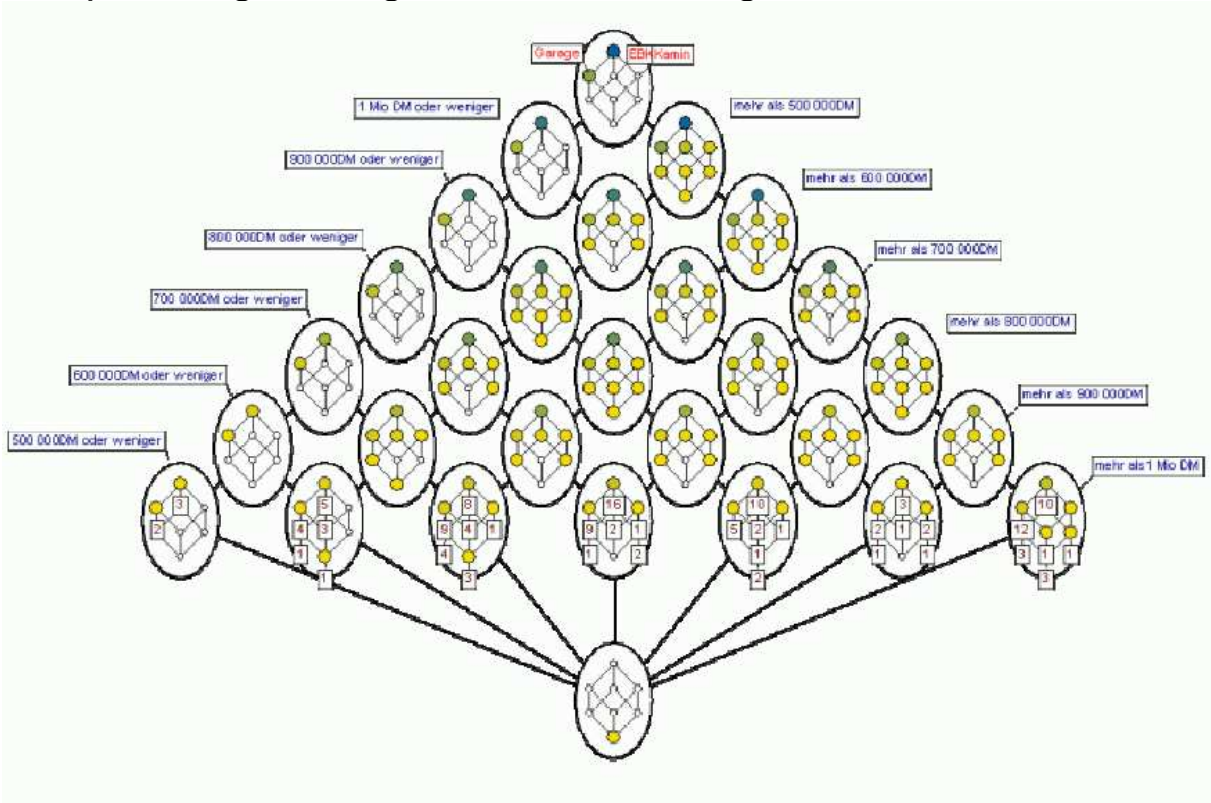
Beispiel: Teildiagramm 1



Beispiel: Teildiagramm 2



Beispiel / Ergebnis: gestuftes Liniendiagramm



(aus Scheich et al. 1992)

Anwendungsbeispiele der formalen Begriffsanalyse:

- Bestimmungsschlüssel

z.B. Pflanzen

Tiere

Bodentypen

Vegetationsgemeinschaften

Fehlertypen in technischen Systemen

Vorteile gegenüber herkömmlichen, regelbasierten top-down-Schlüsseln:

- Merkmale können in beliebiger Reihenfolge erhoben werden
- Einbeziehung von Vorwissen möglich
- jederzeit Überblick über alle möglichen Lösungen
- grafische Navigation im Liniendiagramm

Begriffsverband dieses mehrwertigen Kontexts:

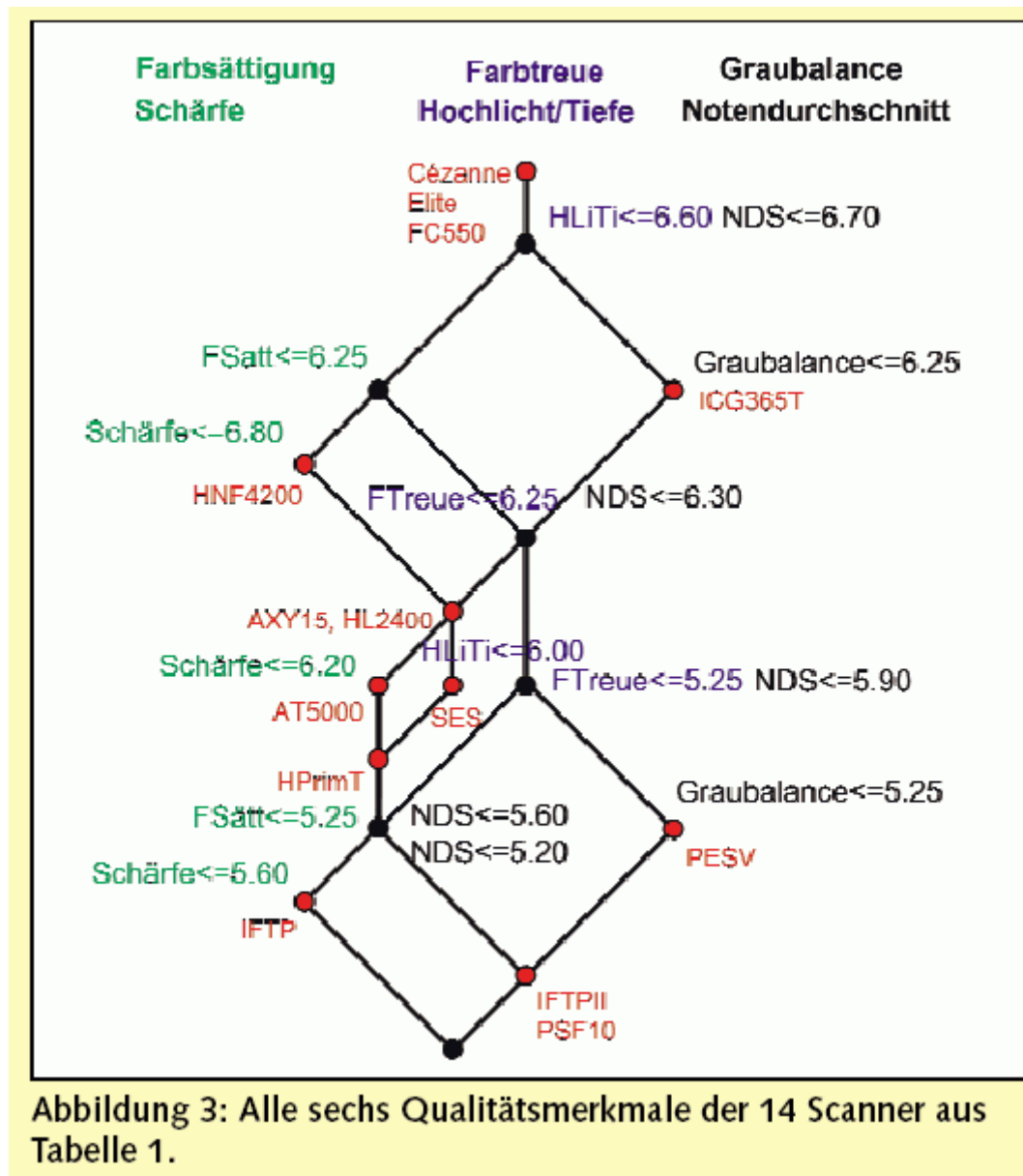


Abbildung 3: Alle sechs Qualitätsmerkmale der 14 Scanner aus Tabelle 1.

(K.-E. Wolff, Internetpublikation)

Vorteile:

- Übersichtlichkeit
- Rangordnungen inhärent in der Tabelle
- bei genügend flexiblem Begriffsanalysesystem auch Wechsel der Skalen und Ausblenden einzelner Merkmale möglich

- Diagnose

Beispiel: Psychologische Tests an Magersucht-Patientinnen (K.-E. Wolff)

Datenerhebungsblatt aus einem Test:

Grid O11E2V1	SE	ID	FA	MO	SI	BR	EL	JA	MA	JE	DI
rational – emotional	2	3	2	5	4	3	2	5	3	4	4
sincere – insincere	1	2	2	3	3	3	2	3	3	4	2
optimistic – pessimistic	5	1	3	5	3	3	4	4	3	2	2
interested – not interested	5	2	2	3	3	3	2	4	4	3	3
flexible – timid	5	1	3	3	2	3	4	4	3	2	2
materialistic – idealistic	5	5	4	2	4	2	3	1	4	3	4
not fashion-conscious – fashion-conscious	2	2	2	5	2	4	4	5	3	4	4
light-hearted – depressive	6	1	3	4	3	3	3	3	4	2	2
resolute – insecure	5	2	3	4	4	3	4	4	5	2	3
unconstrained – constrained	4	2	2	3	2	3	4	5	4	3	2

Table 1: A repertory grid of a patient (example O11E2V1)

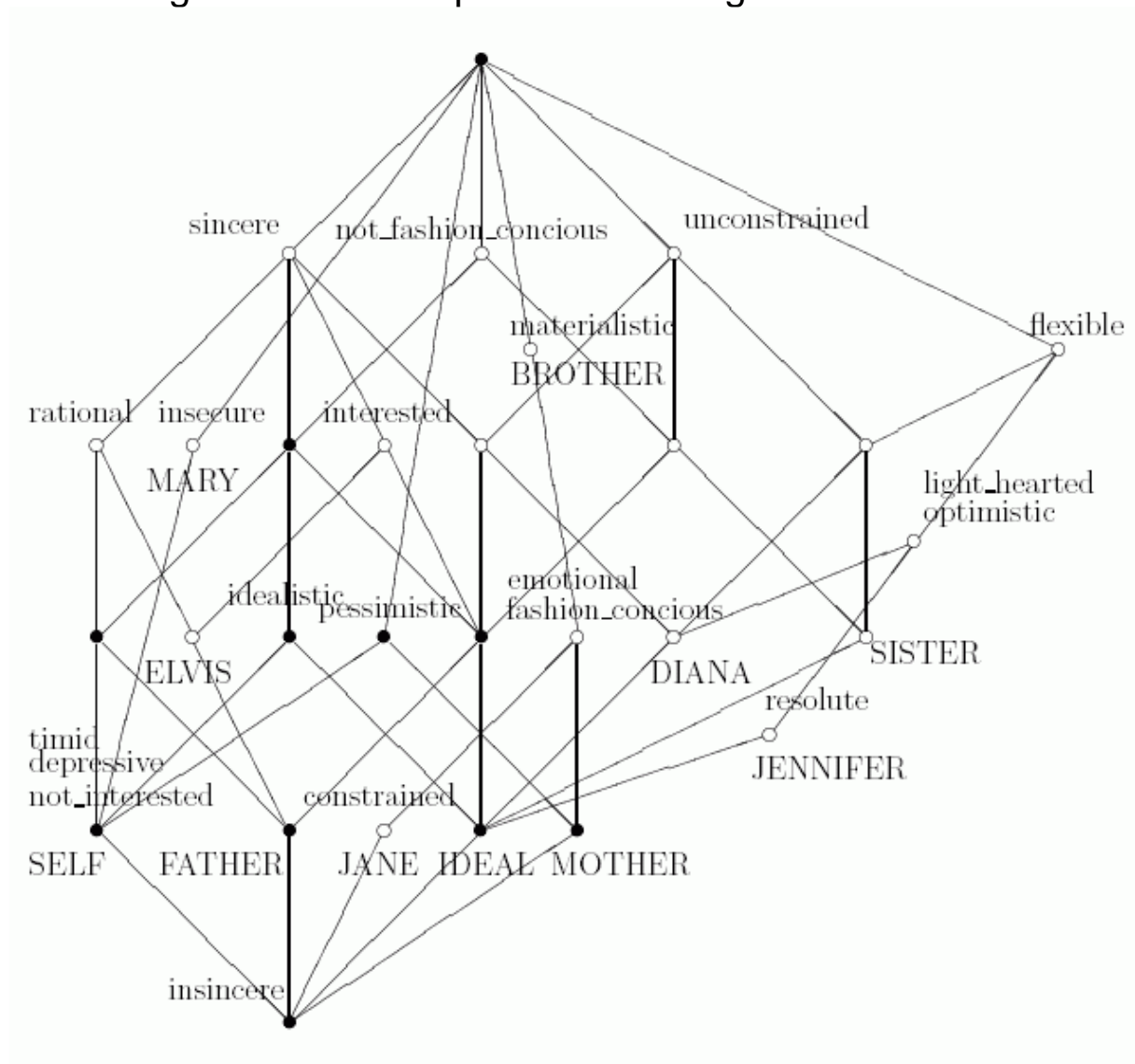
(SE = Selbsteinschätzung, ID = Selbstideal, FA = father usw.)

Verwandlung in einen einwertigen, skalierten Kontext:

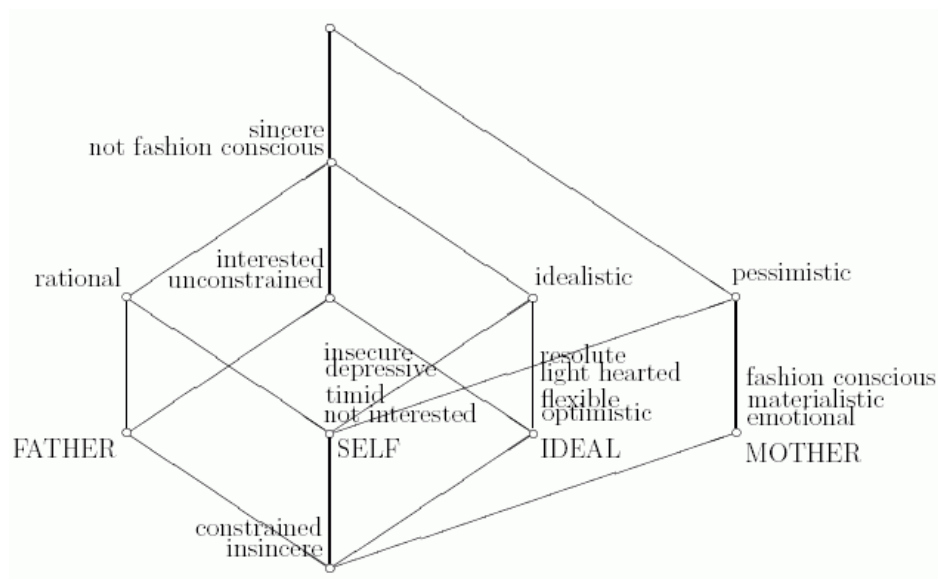
Context	SE	ID	FA	MO	SI	BR	EL	JA	MA	JE	DI
rational	×	.	×	.	.	.	×
emotional	.	.	.	×	.	.	.	×	.	.	.
sincere	×	×	×	.	.	.	×	.	.	.	×
insincere
optimistic	.	×	×	×
pessimistic	×	.	.	×
interested	.	×	×	.	.	.	×
not interested	×
flexible	.	×	.	.	×	×	×
timid	×
materialistic	.	.	.	×	.	×	.	×	.	.	.
idealistic	×	×
not fashion-conscious	×	×	×	.	×
fashion-conscious	.	.	.	×	.	.	.	×	.	.	.
light-hearted	.	×	×	×
depressive	×
resolute	.	×	×	.
insecure	×	×	.	.
unconstrained	.	×	×	.	×	×
constrained	×	.	.	.

Table 2: The one-valued scaled context for the above grid for example O11E2V1:
“1” or “2”: a cross for the first, “5” or “6”: a cross for the second property

Liniendiagramm des entsprechenden Begriffsverbandes:



Einschränkung auf die ersten 4 Personen (Spalten) der Tabelle:

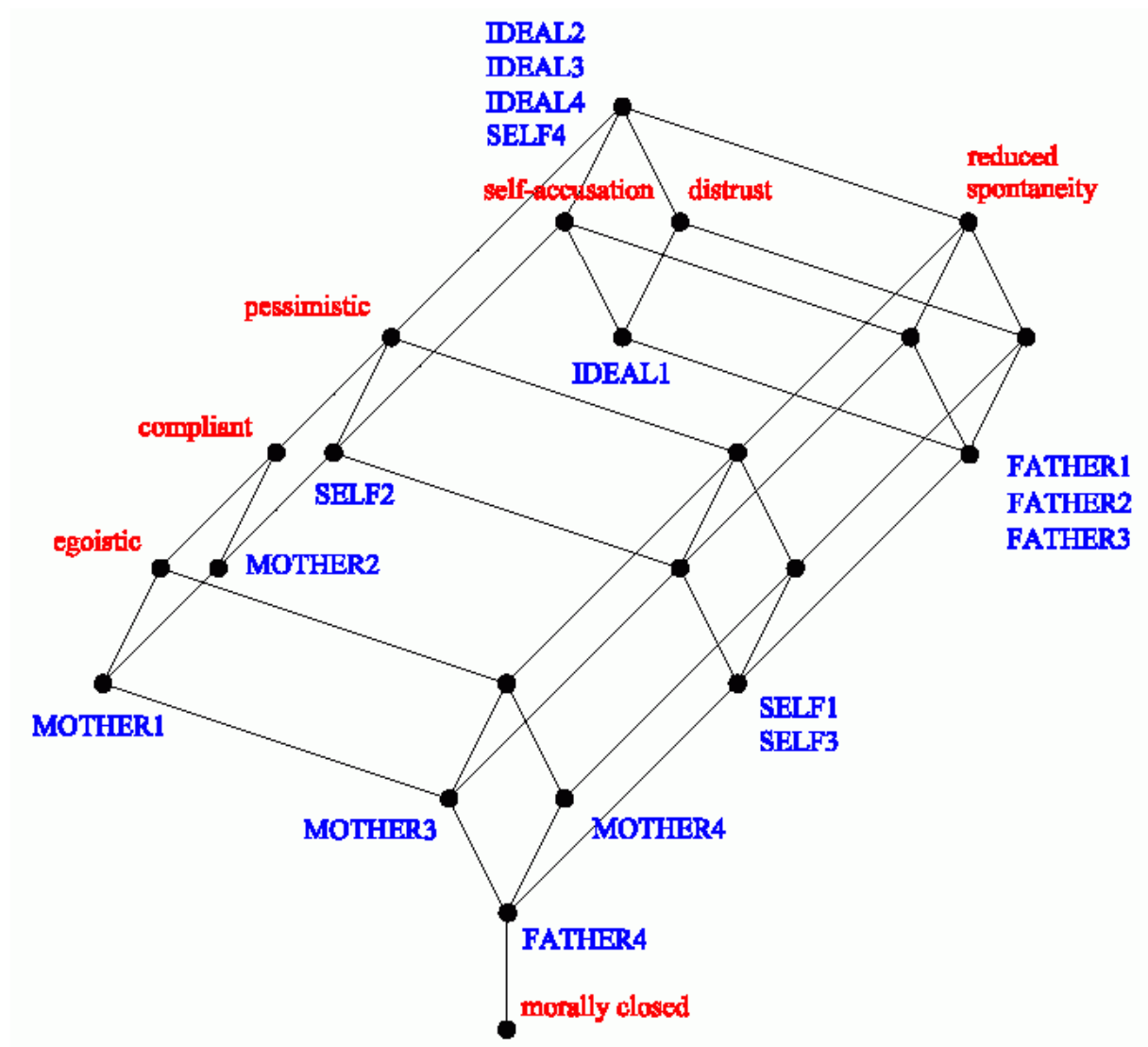


(Peter Burmeister)

Vergleich der Liniendiagramme verschiedener Patientinnen oder derselben Patientin zu verschiedenen Zeitpunkten der Therapie unterstützt "Hineinsehen" in die Daten (besser als Tabelle)

"Timed context analysis":

Merkmalsbelegungen zu verschiedenen Zeitpunkten werden ins selbe Liniendiagramm aufgenommen (mit entspr. Labels für die Zeitpunkte)



(K.-E. Wolff)

Ausschnitte entnommen aus
 Ganter & Wille (o.J.), Ganter & Wille (1996),
 Kipke & Wille (1987), Röder (1997)
 (genaue Quellenangaben siehe http://www.uni-forst.gwdg.de/~wkurth/fs10_lit.htm)