Begriffsanalyse (Concept Analysis)

Begriffsanalyse

Lernziele:

- Einführung von Begriffsanalyse
- Anwendung der Begriffsanalyse zur Erkennung logischer Module

Kontext:

- Begriffsanalyse ist Technik zur Analyse binärer Relationen
- Begriffsanalyse wird vielseitig eingesetzt (z.B. zur Restrukturierung von Vererbungshierarchien im nächsten Kapitel)

Variablenreferenz

(Restriktive) Forderung für ein abstraktes Datenobjekt D:

- *jede Funktion in D referenziert jede Variable in D*
- *Maximale Kohäsion mit gleichzeitigem Versuch zur*
- *Minimierung von Kopplung*
- *korrespondiert mit Rechtecken in der Variablenreferenztabelle (mit geeigneten Spalten− und Zeilenpermutationen):*

	V1	V2	V3	V4	V5	V6	V7	V8
F1	X	X						
F2			X	X	X			
F3			X	X		X	X	X
F4			X	X	X	X	X	X

Variablenreferenz

Weitere Forderung für ein abstraktes Datenobjekt D:

- *ein abstraktes Datenobjekt soll kein weiteres Datenobjekt enthalten*
  - ➢ *die Rechtecke in der Variablenreferenztabelle müssen maximal groß sein*

	V1	V2	V3	V4	V5	V6	V7	V8
F1	X	X						
F2			X	X	X			
F3			X	X		X	X	X
F4			X	X	X	X	X	X

Begriff (Concept)

- Menge $\mathcal{O}$ von **Objekten** (hier Funktionen)
- Menge $\mathcal{A}$ von **Attributen** (hier Variablen)
- **Relation** $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{A}$
  (hier: Funktion *referenziert* Variable)
- das Tripel $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ wird **Kontext** genannt
- für eine Menge von Objekten $O \subseteq \mathcal{O}$ ist σ (O) die Menge gemeinsamer Attribute:
  $\sigma(O) = \{ a \in \mathcal{A} \mid \forall(o \in O)\ (o, a) \in \mathcal{R} \}$
- für eine Menge von Attributen $A \subseteq \mathcal{A}$ ist τ (A) die Menge gemeinsamer Objekte:
  $\tau(A) = \{ o \in \mathcal{O} \mid \forall(a \in A)\ (o, a) \in \mathcal{R} \}$

Begriff (Concept)

- (O, A) ist ein **Begriff** (**Concept**), wenn
  $A = \sigma(O)$ und $O = \tau(A)$
- ein Begriff entspricht einem maximalen Rechteck in der Tabelle (modulo Zeilen− und Spaltenpermutationen)
- für einen Begriff c = (O, A):
  - O = extent (c)
  - A = intent (c)

Begriff (Concept)

	V1	V2	V3	V4	V5	V6	V7	V8
F1	X	X						
F2			X	X	X			
F3			X	X		X	X	X
F4			X	X	X	X	X	X

C1 = ({F1, F2, F3, F4}, ∅)

C2 = ({F2, F3, F4}, {V3, V4})

C3 = ({F1}, {V1, V2})

C4 = ({F2, F4}, {V3, V4, V5})

C5 = ({F3, F4}, {V3, V4, V6, V7, V8})

C6 = ({F4}, {V3, V4, V5, V6, V7, V8})

C7 = (∅, {V1, V2, V3, V4, V5, V6, V7, V8})

({F3, F4}, {V6, V7, V8}) ist kein Begriff wegen:

σ ({F3, F4})

= {V3, V4, V6, V7, V8}

≠ {V6, V7, V8}

Partielle Ordnung der Begriffe

Die Menge der Begriffe bildet eine partielle Ordnung mit:

$$(O1, A1) \leq (O2, A2) \Leftrightarrow O1 \subseteq O2$$

oder alternativ:

$$(O1, A1) \leq (O2, A2) \Leftrightarrow A1 \supseteq A2$$

Beispiele:

C6 = ({F4}, {V3, V4, V5, V6, V7, V8}) ≤ C4 = ({F2, F4}, {V3, V4, V5})

aber weder:

C6 = ({F4}, {V3, V4, V5, V6, V7, V8}) ≤ C3 = ({F1}, {V1, V2})

noch: C3 = ({F1}, {V1, V2}) ≤ C6 = ({F4}, {V3, V4, V5, V6, V7, V8})

Partielle Ordnung der Begriffe

Beispiel: ({F4}, {V3, V4, V5, V6, V7, V8}) $\leq$ ({F2, F4}, {V3, V4, V5})

Falls C1 $\leq$ C2:

- C2 hat mehr Objekte wegen
  - extent (C1) $\subseteq$ extent (C2)

  und weniger Attribute wegen
  - intent (C2) $\subseteq$ intent (C1)

  - C1 ist ein **Unterbegriff** (**Subconcept**) von C2
  - C2 ist ein **Oberbegriff** (**Superconcept**) von C1

Supremum

Gemeinsamer Oberbegriff (**Supremum**) zweier Begriffe A und B umfasst alle gemeinsamen Attribute von A und B

C4 = ({F2, F4}, {V3, V4, V5})

C5 = ({F3, F4}, {V3, V4, V6, V7, V8})

- Schnitt von *intent* (A) und *intent* (B)
- daraus ergibt sich ein neuer *extent:*

(O1, A1) $\vee$ (O2, A2) = (t (A1 $\cap$ A2), A1 $\cap$ A2)

C4 $\vee$ C5 = ({F2, F3, F4}, {V3, V4})

- es gilt allgemein: O1 $\cup$ O2 $\subseteq$ τ (A1 $\cap$ A2), d.h. es können Objekte hinzukommen, die in den ursprünglichen Konzepten nicht vorhanden waren.

Infimum

Gemeinsamer Unterbegriff (**Infimum**) zweier Begriffe A und B umfasst alle gemeinsamen Objekte von A und B

C4 = ({F2, F4}, {V3, V4, V5})

C5 = ({F3, F4}, {V3, V4, V6, V7, V8})

- Schnitt von *extent* (A) und *extent* (B)
- daraus ergibt sich ein neuer *intent*
- $(O1, A1) \wedge (O2, A2) = (O1 \cap O2, \sigma (O1 \cap O2))$

$C4 \wedge C5$ = ({F4}, {V3, V4, V5, V6, V7, V8})

- es gilt allgemein: $A1 \cup A2 \subseteq \sigma (O1 \cap O2)$, d.h. es können Attribute hinzukommen, die in den ursprünglichen Konzepten nicht vorhanden waren.

Begriffsverband (Concept Lattice)

Die Menge $\mathcal{L}$ der Begriffe für einen Kontext $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$:

$\mathcal{L}(\mathcal{K}) = \{(O, A) \in \wp(\mathcal{O}) \times \wp(\mathcal{A}) \mid A = \sigma(O) \wedge O = \tau(A)\}$

mit den Operatoren $\vee$ und $\wedge$ bilden einen vollständigen Verband, d.h. es gilt:

- Kommutativität:

  $x \vee y = y \vee y$ $\qquad$ $x \wedge y = y \wedge x$

- Assoziativität:

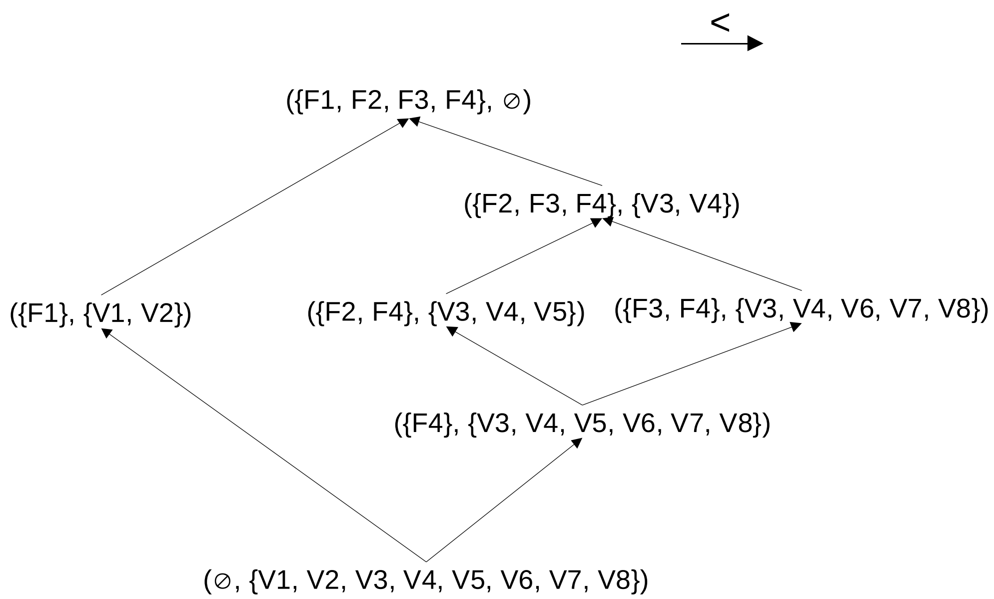
  $(x \vee y) \vee z = x \vee (y \vee z)$ $\qquad$ $(x \wedge y) \wedge z = x \wedge (y \wedge z)$

- Absorbtionsgesetze (Verschmelzungsgesetze):

  $x \vee (x \vee y) = x$ $\qquad$ $x \wedge (x \wedge y) = x$

- Top− und Bottomelement existieren

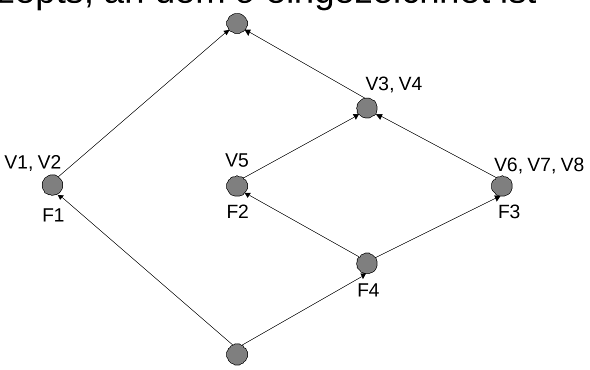
Beispielsbegriffverband



Begriffsverband

Äquivalente, aber leichter lesbare Darstellung:

- Attribut *a* ist enthalten in allen Konzepten ≤ des Konzepts, an dem *a* eingezeichnet ist
- Objekt *o* ist enthalten in allen Konzepten ≥ des Konzepts, an dem *o* eingezeichnet ist

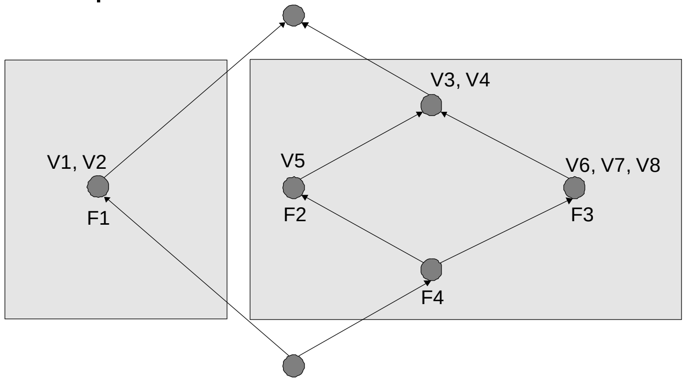


Erkennung von abstrakten Datenobjekten (ADO)

- Erstelle Begriffsverband für Variablenreferenz.
- Jedes Konzept ist prinzipiell ein ADO−Kandidat, allerdings überlappen die Konzepte in ihren *intents* und *extents*.
- Das ist unproblematisch, solange der Begriffsverband horizontal zerlegbar ist (was eine schwächere und realistischere Forderung als die eingangs aufgestellte ist).

Horizontal zerlegbare Verbände

- ein Begriffsverband ist **horizontal zerlegbar**, wenn er in Unterverbände zerlegt werden kann, die nur über Top− und Bottomelement verbunden sind



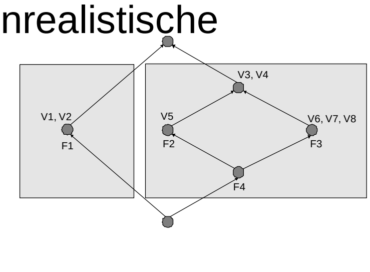
Horizontal zerlegbare Verbände

- ein Begriffsverband ist **horizontal zerlegbar**, wenn die ursprüngliche Tabelle in Quadranten eingeteilt werden kann

	V1	*V2*	*V3*	*V4*	*V5*	*V6*	*V7*	*V8*
F1	*X*	*X*						
F2			*X*	*X*	*X*			
F3			*X*	*X*		*X*	*X*	*X*
F4			*X*	*X*	*X*	*X*	*X*	*X*

Erkennung von ADOs

- horizontal zerlegbare Unterverbände enthalten Unterprogramme, die auf eine Menge von Variablen des Unterverbandes zugreifen
  - die Unterprogramme greifen auf keine Variablen außerhalb des Unterverbandes zu
  - kein Unterprogramm außerhalb des Unterverbandes greift auf Variablen des Unterverbandes zu
- folglich ist Information Hiding gewährleistet
- nicht jedes Unterprogramm muss auf alle Variablen zugreifen (wäre unrealistische Anforderung)



Interferenzen

Was geschieht, wenn Information Hiding nicht umgesetzt ist?

	V1	*V2*	*V3*	*V4*	*V5*	*V6*	*V7*	*V8*
F1	*X*	*X*						
F2		*X*	*X*	*X*	*X*			
F3			*X*	*X*		*X*	*X*	*X*
F4			*X*	*X*	*X*	*X*	*X*	*X*

Neue Konzepte: ({F1, F2}, {V2}) und ({F2}, {V2, V3, V4, V5}).

Interferenzen