# 3.3 Das Relationenkalkül

zugrundeliegende Idee beim Relationenkalkül:

- Ergebnis einer Anfrage wird als Menge von Tupeln betrachtet.
- Beschreibung der Ergebnisrelation ohne dabei explizit die Konstruktion der Relation anzugeben.
- basiert auf der Sprache der Prädikatenlogik erster Stufe

zwei verschiedene Techniken:

- Relationales Tupelkalkül
- Bereichskalkül

# 3.3.1 Das Tupelkalkül

- Eine Anfrage im Tupelkalkül wird wie folgt formuliert:

$$\{t(RS) \mid \psi(t)\}$$

  - $\Psi$ ist hierbei eine Formel, die den Wert *true* oder *false* liefert.
  - die Ergebnismenge der Anfrage ist durch die Tupel t aus dem Schema RS gegeben mit $\psi(t) = true$
  - Häufig ergibt sich das Schema aus der verbalen Formulierung der Anfrage. Das Schema wird deshalb bei den Formeln nicht immer explizit angegeben.

## Beispiele:

- Was sind die Namen der Mitarbeiter aus Abteilung A63?
  $\{t(\text{PName}) \mid \exists s \in$ Personal mit t[PName] = s[PName] und s[abtnr] = “A63“$\}$

# Formeln des Tupelkalküls

## Formel $\Psi$ setzt sich zusammen aus Atomen der Form

- R(s): s ist Element der Relation R (s ist *Tupelvariable*)
- $s[i]\ \theta\ u[j]$ mit $\theta \in \{\ =, \neq, <, \leq, >, \geq\ \}$
- $s[i]\ \theta\ a$

## Beispiele:

- Personal(t), t[Fähigkeit] > 4, t[abtnr] = u[abtnr]

## Eine Formel ist gegeben durch

- ein Atom
- $\Psi_1 \wedge \Psi_2$, $\Psi_1 \vee \Psi_2$, $\neg\Psi_1$, $(\Psi_1)$
- $\forall s(RS)\Psi$, $\exists s(RS)\Psi$, wobei s eine Tupelvariable in $\Psi$ und RS eine Menge von Attributen ist, auf der das Tupel definiert ist.
  - RS wird als Schema der Variable s bezeichnet

# Bemerkungen:

- Reihenfolge der Ausführung: $\forall$ und $\exists$; $\neg$; $\wedge$ ; $\vee$
- Klammern setzten die Reihenfolge außer Kraft

# Beispiele:

Annahme: Tupelvariable x ist aus dem Schema {pnr,mnr,Fähigkeit}

- $\neg x[\text{Fähigkeit}] > 4$
- $(x[pnr] = y[pnr]) \vee \neg x[\text{Fähigkeit}] > 4$
- $\text{P-M-Zuteilung}(x) \wedge (x[pnr] = y[pnr]) \vee \neg x[\text{Fähigkeit}] > 4$
- $\exists x(\text{P-M-Zuteilung}(x) \wedge (x[pnr] = y[pnr]) \vee \neg x[\text{Fähigkeit}] > 4)$

# Freie und gebundene Tupelvariablen

- entspricht dem Prinzip globaler und lokaler Variablen in einem Programm
- falls ein Quantor vor einer Variablen steht, wird diese zu einer gebundenen Variablen
- folgende Bedingungen gelten:
  - das Auftreten einer Tupelvariablen in einem Atom ist stets frei
  - für $f = \neg g$ und $f = (g)$ sind alle freien Variablen von g auch frei in f
  - für $f = g \wedge h$ und $f = g \vee h$ sind die Variablen in f frei, falls sie es in g und h sind
  - für $f = \exists x\ (RS)(g)$ und $f = \forall x\ (\mathrm{RS})(g)$ muß x eine freie Variable in g sein, die in f gebunden ist. Dabei ist RS das Schema der Variable x.

## Beispiel:

$\forall x(\{pnr, mnr, Fhigkeit\})(\neg \text{P-M-Zuteilung}(x) \vee x[\text{Fähigkeit}] > 4)$

# Berechnung der Formeln des Tupelkalküls

- ein Ausdruck des Tupelkalküls hat die Form

$$\{t(RS) \mid \Psi(t)\}$$

  wobei t (aus dem Schema *RS*) die einzig freie Variable in $\Psi$ ist.

## Substitution:

- Sei $\Psi(s)$ eine Formel. Dann ist $\Psi(t/s)$ die Substitution von der Tupelvariablen s in $\Psi$ durch das Tupel t, falls in jedem Atom, das ein freies Auftreten von s enthält, wie folgt verfahren wird:
  - R(s) wird durch “wahr” ersetzt, falls $t \in R$. Andernfalls durch “falsch”.
  - s[i] $\theta$ u[j] wird ersetzt durch $c$ $\theta$ u[j] mit $c = t$[i] (Ann.: $\mathrm{u} \neq \mathrm{s}$)
  - s[i] $\theta$ $c$ wird ersetzt durch “wahr”, falls $t[i]\ \theta\ c$ gilt. Andernfalls durch “falsch”.

## Bemerkung:

- Durch Substitution gewinnt man eine Formel die nur noch Konstanten der Form “wahr” und “falsch” und Atome mit gebundenen Variablen enthält.

# Beispiele

- u und t seien aus dem Schema {pnr,mnr, Fähigkeit}
- gegeben sei die Formel.

  $\forall u(\neg\text{P-M-Zuteilung(u)} \vee \neg u[pnr] = t[pnr] \vee u[\text{Fähigkeit}] < t[\text{Fähigkeit}])$

  für das Tupel t = (73,84,5) gilt:

  $\forall u(\neg\text{P-M-Zuteilung(u)} \vee \neg u([pnr]) = 73 \vee u([\text{Fähigkeit}]) < 5)$

- gegeben sei die Formel:

  $\neg\text{P-M-Zuteilung(u)} \vee \neg u[pnr] = 73 \vee u[\text{Fähigkeit}] < 5$

  für das Tupel u = (51,93,2) gilt:

  $\neg wahr \vee \neg falsch \vee wahr$

# Interpretation der Formel

- Sei f eine Formel ohne freie Variablen. Die Interpretation I(f) ist wie folgt definiert:
  - Falls f “wahr”, dann ist I(f) := true. Andernfalls, I(f) := false.
  - …
  - …
  - Sei f gleich $\exists x(RS)\ (g(x))$. Dann ist I(f) := true, falls es mindestens ein Tupel t aus dem Schema *RS* gibt, so daß I(g(t/x)) = true ist. Andernfalls, I(f) := false.
  - Sei f gleich $\forall x(RS)(g(x))$. Dann ist I(f) := true, falls für alle t aus dem Schema *RS* I(g(t/x)) = true gilt. Andernfalls, I(f) := false.
- Sei $E = \{x(RS) \mid \Psi(x)\}$ ein Ausdruck des Tupelkalküls und sei *RS* = $\{A_1,\ldots,A_n\}$ das Schema von x und $D_1 \times D_2 \times \ldots \times D_n$ der zugehörige Wertebereich. Der Wert von E zu einer gegebenen Datenbank besteht aus allen Tupel $t \in D_1 \times D_2 \times \ldots \times D_n$, die

$$\mathrm{I}(\Psi(t/x)) = true$$

erfüllen.

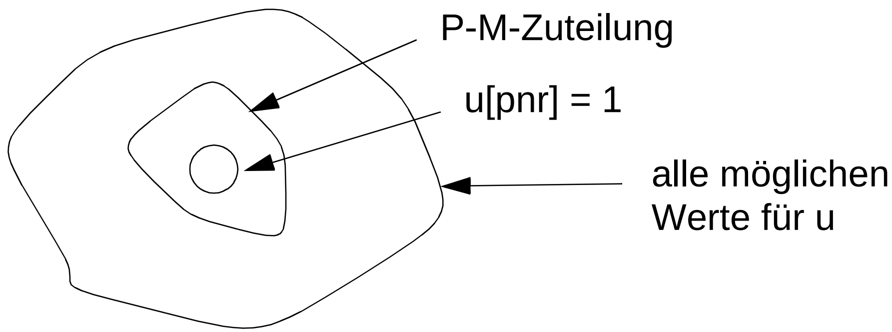
# Beispiele

- Gib alle Personalnummern von Personen, die an einer Maschine ausgebildet sind.

  $\{t(\{pnr\}) \mid \exists u(\text{P-M-Zuteilung}(u) \wedge u[pnr] = t[pnr])\}$

- Gib alle Personalnummern der Personen, die an keiner Maschine genügend gut ausgebildet sind.

  $\{t(\{pnr\}) \mid \forall u(\neg\text{P-M-Zuteilung}(u) \vee u[\text{Fähigkeit}] \geq 5 \vee \neg u[pnr] = t[pnr])\}$



# Einführung von Kurzschreibweisen

- $\exists u \in R(\psi(u)) := \exists u(R(u) \wedge \psi(u))$
- $\forall u \in R(\psi(u)) := \forall u(\neg R(u) \vee \psi(u))$
- $(\Psi_1 \Rightarrow \Psi_2) := (\neg\Psi_1 \vee \Psi_2)$

- Finde die Namen der Personen, die an keiner Maschine genügend gut ausgebildet sind (Schema der Tupelvariablen t ist {PName}).
  $\{t \mid \exists x \in Personal(\Psi_1(x, t)) \wedge \exists y(\{pnr\})(y[pnr] = x[pnr] \wedge \Psi_2(y)))\}$

  – $\Psi_1(x, t) = (x[PName] = t[PName])$

  – $\Psi_2(y) = \forall u \in \text{P-M-Zuteilung}(u[\text{Fähigkeit}] \geq 5 \vee \neg u[pnr] = y[pnr])$

  Vereinfachen der Formel $\Psi_2$:

  $\Psi_2(y) = \forall u \in \text{P-M-Zuteilung}((u[pnr] = y[pnr]) \Rightarrow (u[\text{Fähigkeit}] \geq 5))$

# Sichere Ausdrücke

- Probleme des Tupelkalküls:
  - Beschreibung unendlich großer Relationen als Ergebnis
  - keine effektive Berechnung möglich (d.h. nur durch Testen jedes Elements aus dem Wertebereich)

- Idee:
  - Beschränkung der Wertebereiche nur auf die Werte, die tatsächlich in vorhandenen Relationen vorkommen.

    DOM($\psi$) :
  - Menge aller Werte, die explizit in $\Psi$ vorkommen, oder in Relationen, die in $\Psi$ erwähnt werden (zur Erinnerung: Relationen sind hier stets endlich).

    Beispiel: $\psi$ = "$t[2] = 7 \vee R(t)$" mit

| R | | |
|---|---|---|
| | a | 5 |
| | b | 3 |

Dann gilt: $\mathrm{DOM}(\psi) = \{a, b, 3, 5, 7\}$

## Ein Ausdruck $\{t(\mathrm{RS}) \mid \psi(t)\}$ ist sicher, falls gilt:

- $u \notin \mathrm{DOM}^k(\psi) \Rightarrow \mathrm{I}(\psi(u/t)) = \text{false}$
- für jede Teilformel $\exists u(\omega(u))$ gilt: $v \notin \mathrm{DOM}^k(\omega) \Rightarrow \mathrm{I}(\omega(v/u)) = \text{false}$
- für jede Teilformel $\forall u(\omega(u))$ gilt: $v \notin \mathrm{DOM}^k(\omega) \Rightarrow \mathrm{I}(\omega(v/u)) = \text{true}$

Hierbei bezeichnet k stets die Anzahl der Attribute im Schema der Tupelvariablen und $\mathrm{DOM}^k(\psi)$ das k-malige kartesische Produkt über $\mathrm{DOM}(\psi)$.

## Beispiele: Sicher oder nicht sicher?

- $\{t\{\text{pnr, PName, Vorname, abtnr, Lohn}\} \mid Personal(t) \wedge t[Vorname] = \text{“Willi“}\}$
- $\{t\{\text{pnr, PName, Vorname, abtnr, Lohn}\} \mid (Personal(t) \vee t[Vorname] = \text{“Willi“})\}$
- für eine Relation $R$ und eine Formel $\Psi$ sind die folgenden Ausdrücke sicher:
  - $\exists u \in R(\psi(u))$
  - $\forall u \in R(\psi(u))$

# 3.3.2 Das Bereichskalkül

- ähnliche Idee wie beim Tupelkalkül mit dem Unterschied, daß Variablen sich auf die Komponenten der Tupel beziehen.
- Ausdrücke haben die Form:

$$\{x_1 \ldots x_k \mid \psi(x_1, \ldots, x_k)\}$$

wobei die Bereichsvariable $x_i$ einem Attribut $A_i$ zugeordnet ist, $1 \leq i \leq k$. Die Bereichsvariablen $x_1,\ldots,x_k$ sind die einzigen freien Variablen.

## Beispiele:

- $\{v, w, x, y, z \mid Personal(v, w, x, y, z)$ mit $x =$ “Willi“$\}$

oder kürzer

$\{v, w, \text{Willi}, y, z \mid Personal(v, w, \text{“Willi“}, y, z)\}$

oder wenn nur der Nachname erwünscht ist:

$\{w \mid \exists v\ \exists y\ \exists z\ Personal(v, w, \text{“Willi“}, y, z)\}$

## Atome haben die Form:

- $R(x_1 \ldots x_k)$ (oder auch $R(x_1, \ldots, x_k)$)
- $x\ \theta\ y$ mit Bereichsvariablen x und y
- Substitution wird entsprechend wie beim Tupelkalkül vorgenommen

## Beispiele:

- Gebe alle Namen von Personen, die an einer Maschine ausgebildet sind.

  $$\{t \mid \exists u, v, w, x, y, z(\text{P-M-Zuteilung}(u, v, w) \wedge Personal(u, t, x, y, z))\}$$

- Gebe alle Personalnummern der Personen, die an keiner Maschine genügend gut ausgebildet sind.

  $$\{t \mid \forall x\ \forall y\ (\neg\text{P-M-Zuteilung}(t, x, y) \vee y \geq 5)\}$$