

Datenbanken

05 Relationale Algebra

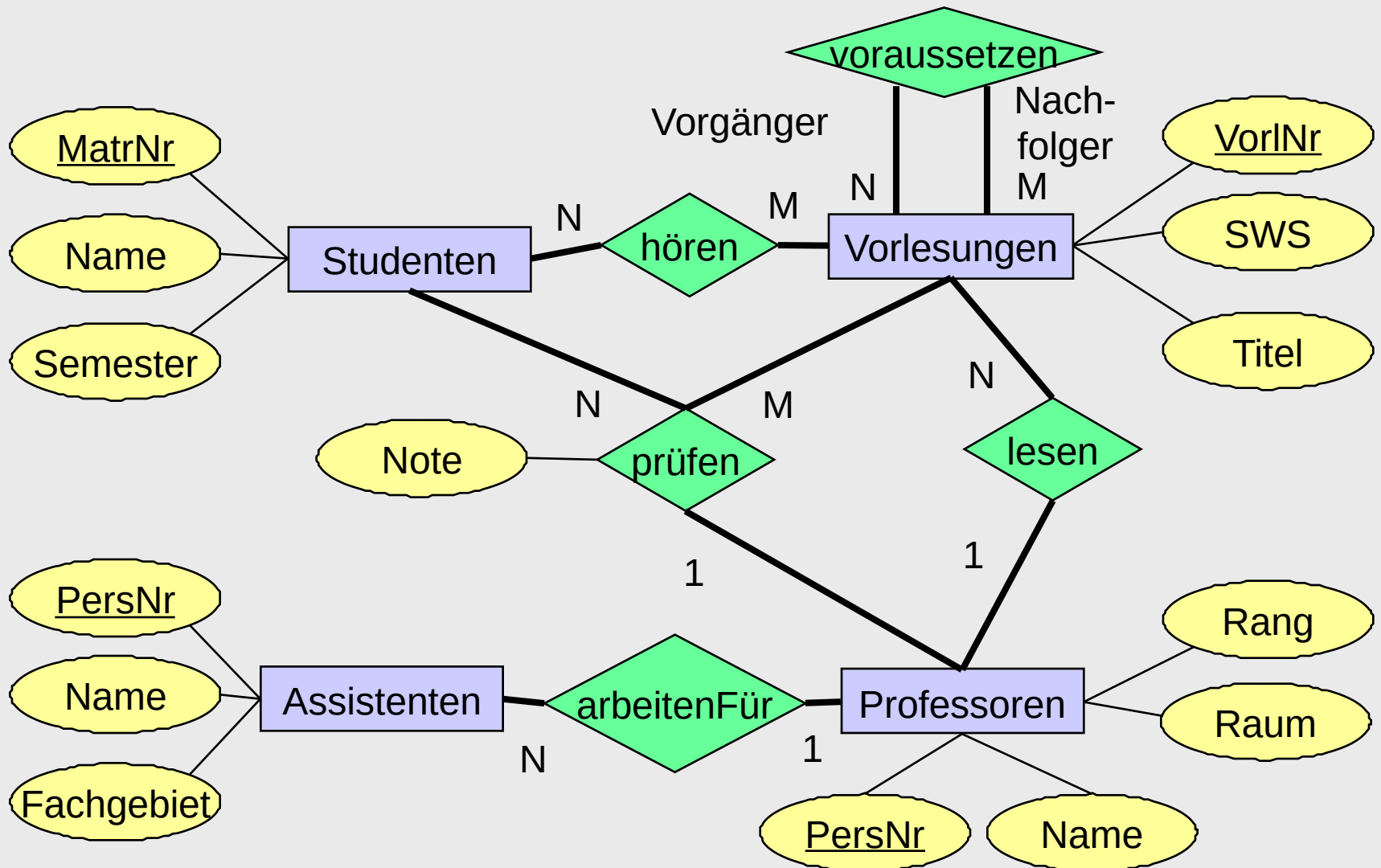
Seminaristischer Unterricht

Prof. Dr.-Ing. Hendrik Gärtner

Gliederung

- Einführung:
 - Was ist Relationale Algebra?
 - Basisoperatoren der Relationalen Algebra
 - Operatorbäume
- Fortgeschrittene Operatoren
 - Vereinigungsmenge
 - Differenz
 - Schnittmenge
- Beispiele aus der Praxis

Universitätsschema



Schema der Beispieldatenbank

Studenten: {[MatrNr:integer, Name: string, Semester: integer]}

Vorlesungen: {[VorlNr:integer, Titel: string, SWS: integer, gelesenVon:integer]}

Professoren: {[PersNr:integer, Name: string, Rang: string, Raum: integer]}

Assistenten: {[PersNr:integer, Name: string, Fachgebiet: string, Boss:integer]}

voraussetzen: {[Vorgänger:integer, Nachfolger:integer]}

hören: {[MatrNr:integer, VorlNr:integer]}

prüfen: {[MatrNr:integer, VorlNr:integer, PersNr:integer, Note:integer]}

Professoren			
PersNr	Name	Rang	Raum
2125	Sokrates	C4	226
2126	Russel	C4	232
2127	Kopernikus	C3	310
2133	Popper	C3	52
2134	Augustinus	C3	309
2136	Curie	C4	36
2137	Kant	C4	7

Studenten		
MatrNr	Name	Semester
24002	Xenokrates	18
25403	Jonas	12
26120	Fichte	10
26830	Aristoxenos	8
27550	Schopenhauer	6
28106	Carnap	3
29120	Theophrastos	2
29555	Feuerbach	2

Vorlesungen			
VorlNr	Titel	SWS	gelesen Von
5001	Grundzüge	4	2137
5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125
4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137

voraussetzen	
Vorgänger	Nachfolger
5001	5041
5001	5043
5001	5049
5041	5216
5043	5052
5041	5052
5052	5259

hören	
MatrNr	VorlNr
26120	5001
27550	5001
27550	4052
28106	5041
28106	5052
28106	5216
28106	5259
29120	5001
29120	5041
29120	5049
29555	5022
25403	5022

Assistenten			
PerslNr	Name	Fachgebiet	Boss
3002	Platon	Ideenlehre	2125
3003	Aristoteles	Syllogistik	2125
3004	Wittgenstein	Sprachtheorie	2126
3005	Rhetikus	Planetenbewegung	2127
3006	Newton	Keplersche Gesetze	2127
3007	Spinoza	Gott und Natur	2126

prüfen			
MatrNr	VorlNr	PersNr	Note
28106	5001	2126	1
25403	5041	2125	2
27550	4630	2137	2

Relationale Algebra

- Formale Sprache für die Formulierung von Abfragen innerhalb eines relationalen Schemas
- Ermöglicht Relationen miteinander zu verknüpfen oder zu reduzieren und komplexere Informationen daraus herzuleiten
- Definiert Operationen, die sich auf einer Menge von Relationen anwenden lassen (z.B. verknüpfen, filtern, etc.)
- Ergebnisse aller Operationen sind ebenfalls **Relationen (ist abgeschlossen)**
- Relationale Algebra ist die Basis für die Datenbanksprache **SQL (Structured Query Language)**
(Übersetzung intern SQL \rightarrow RA)

Operatoren der Relationalen Algebra

- σ Selektion \Leftarrow
- π Projektion \Leftarrow
- \times Kreuzprodukt \Leftarrow
- \bowtie Join (Verbund)
- ρ Umbenennung \Leftarrow
- - Mengendifferenz
- \div Division
- \cup Vereinigung
- \cap Mengendurchschnitt
- \bowtie Semi-Join (linker)
- \bowtie Semi-Join (rechter)
- \bowtie linker äußerer Join
- \bowtie rechter äußerer Join

Operatoren stammen aus der Mengenlehre

SQL: SELECT Statement

Π - Projektion

SELECT \downarrow \downarrow
Attribut1, Attribut2, ... oder * (für alle)

Π - Projektion (Herausgreifen einzelner Attribute)

\downarrow
FROM Tabelle1, Tabelle2, ...

X- Kartesisches Produkt
(Elemente verknüpfen)

WHERE Bedingung1 \downarrow **AND/OR** Bedingung2 **AND/OR** ...

σ Selektion (Auswählen von einzelnen Tupeln, die den Bedingungen entsprechen)

Operator Projektion

Definition Projektion

Bei der Projektion werden die Attribute/Spalten einer (Argument-)Relation R extrahiert. D.h. es sind nur die Attribute vorhanden die ausgewählt wurden.

- Projektions-Symbol: Π
- Menge der Attributnamen im Subskript:

$\Pi_{\text{Attributnamen}}(\text{Relationenname}) \leftarrow$

- z.B.: $\Pi_{\text{Rang}}(\text{Professoren}) \leftarrow$ $\begin{matrix} c_3 \\ c_4 \end{matrix} \leftarrow$

- eventuell auftretende Duplikate werden entfernt

- Wenn mehrere Spalten extrahiert werden sollen, dann werden diese durch Komma getrennt:

$\Pi_{\text{Name, Rang, Raum}}(\text{Professoren})$

$\text{SELECT } \text{DISTINCT Rang FROM Professoren} \leftarrow$

Π

$\Pi_{\text{Rang, Name}}(\text{Professoren})$

Beispiel Projektion

$\Pi_{\text{MatrNr, Name}}(\text{Studenten})$

Herausgreifen von
Attributen – dabei
werden alle
Doppelten eliminiert
(nur bei RA/nicht
SQL)

$\Pi_{\text{Rang}}(\text{Professoren})$

$\Pi_{\text{MatrNr, Name}}(\text{Studenten})$	
MatrNr	Name
24002	Xenokrates
25403	Jonas
...	...

$\Pi_{\text{Rang}}(\text{Professoren})$	
Rang	
C4	
C3	

Duplikate in SQL versus mathematische Relationen

Duplikate

Mathematische Relationen kennen keine Duplikate. In SQL Tabellen sind Duplikate aber erlaubt bzw. diese werden aus Effizienzgründen nicht automatisch beseitigt. Falls gewünscht müssen diese explizit beseitigt werden.

DISTINCT

Operator Selektion

Definition Selektion

Bei der Selektion werden die Tupel einer Relation R mittels eines Selektionsprädikat gefiltert, d.h. das Ergebnis einer Selektion sind die Tupel der Relation R, die das Selektionsprädikat erfüllen.

- Selektions-Symbol: σ
- Selektionsprädikat als Subskript:
- $\sigma_{\text{Attributnamen} \text{ bed. (Relationenname)}}$
z.B.: $\sigma_{\text{Semester} > 10}(\text{Studenten})$
- Selektionsprädikat ist ein sogenannter Boolescher Ausdruck (kann nur wahr oder falsch sein und wird mit arithmetischen Vergleichsoperatoren und den logischen Operatoren errechnet)

SELECT * FROM

Studenten where
Semester > 10

$\pi_{\text{titel}} (\sigma_{\text{sws}=4} (\text{Vorlesungen}))$ \neq $\sigma_{\text{sws}=4} (\pi_{\text{titel}} (\text{Vorlesungen}))$ \neq $(\pi_{\text{titel}} (\text{Vorlesungen}))$

Boolescher Ausdruck / Selektionsprädikat

- Im Allgemeinen ist das Selektionsprädikat eine Formel F mit:
 - Attributnamen der Argumentrelation R oder Konstanten als Operanden
 - den arithmetischen Vergleichsoperatoren $=, <, >, \leq, \geq, \neq$
 - den logischen Operatoren; \vee, \wedge, \neg
- Das Ergebnis der Selektion sind alle Tupel $t \in R$, für die die Formel F erfüllt ist

Was ist aus der Prädikatenlogik bereits bekannt?

Beispiel Selektion

$\sigma_{\text{Semester} > 10}$ (Studenten)

Selektion

Selektionsprädikat
bezieht sich immer
nur auf eine Zeile

$\sigma_{\text{Name}='Sokrates'}$ (Professoren)

$\sigma_{\text{Semester} > 10}$ (Studenten)		
MatrNr	Name	Semester
24002	Xenokrates	18
25403	Jonas	12

Professoren			
PersNr	Name	Rang	Raum
2125	Sokrates	C4	226

SQL: SELECT Statement



SELECT Attribut1, Attribut2,... oder * (für alle)

Π - Projektion (Herausgreifen einzelner Attribute)

FROM Tabelle1, Tabelle2,...

X- Kartesisches Produkt (Elemente verknüpfen)

WHERE Bedingung1 **AND/OR** Bedingung2 **AND/OR** ...

σ - Selektion (Auswählen von einzelnen Tupeln, die den Bedingungen entsprechen)



Beispiel Projektion und Selektion

Projektion: Wie heißen die Professoren der Beispieluniversität?

$\Pi_{\text{Name}}(\text{Professoren})$

SELECT name FROM Professoren

Wie heißt der Student mit der Matrikelnummer 25403?

$\Pi_{\text{Name}}(\sigma_{\text{MatrNr}=25403}(\text{Studenten}))$

SELECT \uparrow Name FROM Studenten WHERE MatrNr = 25403

Wie sind Name und Matrikelnummer der Studenten, die bereits mehr als 6 Semester studiert haben?

$\Pi_{\text{Name}, \text{MatrNr}}(\sigma_{\text{Semester} > 6}(\text{Studenten}))$

SELECT Name, MatrNr FROM Studenten WHERE semester > 6

Beispiel Projektion und Selektion

Projektion: Wie heißen die Professoren der Beispieluniversität?

$\Pi_{\text{Name}}(\text{Professoren})$ 

SELECT Name **FROM** Professoren 

Selektion:

Wie heißt der Student mit der Matrikelnummer 25403?

$\Pi_{\text{Name}}(\sigma_{\text{MatrNr}=25403}(\text{Studenten}))$

SELECT Name **FROM** Studenten **WHERE** MatrNr=25403

Wie sind Name und Matrikelnummer der Studenten, die bereits mehr als 6 Semester studiert haben?

$\Pi_{\text{Name, MatrNr}}(\sigma_{\text{Semester}>6}(\text{Studenten}))$

SELECT name, MatrNr **FROM** Studenten **WHERE** Semester>6

Operator Kartesisches Produkt

Definition Kartesisches Produkt

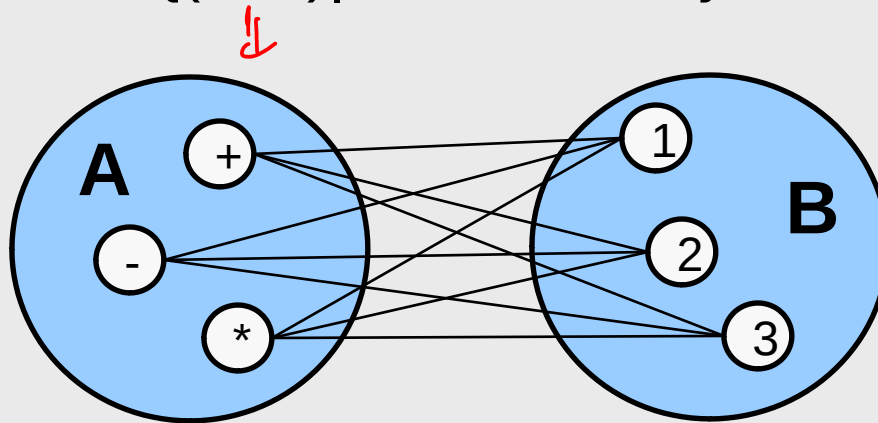
Das Kartesische Produkt zwischen zwei Relationen R und S enthält alle $|R| * |S|$ möglichen Paare von Tupeln aus R und S . Das Schema der Ergebnisrelation $\text{sch}(R \times S)$ ist die Vereinigung der Attribute aus $\text{sch}(R)$ und $\text{sch}(S)$.

- Symbol: \times
- Verbindet zwei Relationen R und S : $R \times S$
- Ergebnisschema:
 $\text{sch}(R \times S) = \text{sch}(R) \cup \text{sch}(S)$

Operator Kartesisches Produkt

Das kartesische Produkt zwischen zwei Mengen A und B ist die Kombination aller Elemente zwischen den Elementen aus A und B:

$$A \times B = \{(a; b) | a \in A \wedge b \in B\}$$



$A \times B = \{(+,1), (+,2), (+,3), (-,1), (-,2), (-,3), (*,1), (*,2), (*,3)\}$
Durch das kartesische Produkt zwischen A und B wird eine Paar-Menge definiert.

Beispiel Kartesisches Produkt

Kartesisches Produkt (Kreuzprodukt)

Professoren x Vorlesungen

Professoren				Vorlesungen			
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS	gel.v.
2125	Sokrates	C4	226	5001	Grundzüge	4	2137
...		
2125	Sokrates	C4	226	5041	Ethik	4	2125
...		
2137	Kant	C4	7	5041	Ethik	4	2125

- Problem: riesige Zwischenergebnisse
- Beispiel: (Professoren x Vorlesungen)
- "bessere" Operation: Join

Operatoren der Relationalen Algebra

Kartesisches Produkt (Kreuzprodukt)

Professoren x Vorlesungen

Professoren				Vorlesungen			
PersNr	Name	Rang	Raum	VorlNr	Titel	SWS	gel.v.
2125	Sokrates	C4	226	5001	Grundzüge	4	2137
...
2125	Sokrates	C4	226	5041	Ethik	4	2125
...
2137	Kant	C4	7	5041	Ethik	4	2125

Welcher Professor liest Ethik?

$\Pi_{p.Name} (\sigma_{p.PersNr=v.gelesenVon \wedge v.Titel='Ethik'} (Professoren p \times Vorlesungen v))$

SELECT Name FROM Professoren, Vorlesungen WHERE
PersNr=gelesenVon AND Titel='Ethik'

Beispiel Projektion, Selektion und Kartesisches Produkt

Welche Vorlesungen (Titel) hält Professor Sokrates?

$\Pi_{\text{titel}} (\sigma_{\text{Vorlesung} = \text{ProfNr} \wedge \text{name} = \text{'Sokrates'}} (\text{Professoren} \times \text{Vorlesungen}))$

SELECT titel FROM Professoren, Vorlesungen WHERE Vorlesung = ProfNr AND name = 'Sokrates'

Welche Assistenten arbeiten für Kopernikus?

$\Pi_{\text{Assistenten}} (\sigma_{\text{boss} = \text{ProfNr} \wedge \text{Professoren.Name} = \text{'Kopernikus'}} (\text{Professoren} \times \text{Assistenten}))$

SELECT Assistenten.Name FROM Professoren, Assistenten WHERE Professoren.Name = 'Kopernikus' AND boss = Professoren.ProfNr

Welche Studenten (Name und MatrNr) hören die Vorlesung Grundzüge?

$\Pi_{\text{Name}, \text{Studenten.MatrNr}} (\sigma_{\text{Studenten.MatrNr} = \text{ hoeren.MatrNr} \wedge \text{ hoeren.VorNr} = \text{Vorlesungen.VorNr} \wedge \text{titel} = \text{'Grundzüge'}} (\text{S} \times \text{h} \times \text{V}))$

SELECT Name, S.MatrNr

FROM Studenten as S, hoeren as h, Vorlesungen as V

WHERE S.MatrNr = h.MatrNr and h.VorNr = V.VorNr and titel = 'Grundzüge'

Beispiel Projektion, Selektion und Kartesisches Produkt

Welche Vorlesungen (Titel) hält Professor Sokrates?

$\Pi_{\text{Titel}}(\sigma_{\text{Name}='Sokrates' \wedge \text{gelesenVon}=\text{PersNr}}(\text{Vorlesungen} \times \text{Professoren}))$

SELECT Titel **FROM** Vorlesungen, Professoren **where** Name='Sokrates' and gelesenVon=PersNr

Welche Assistenten arbeiten für Kopernikus?

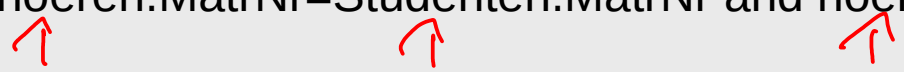
$\Pi_{\text{Assistenten.Name}}(\sigma_{\text{Name}='Kopernikus' \wedge \text{Boss}=\text{PersNr}}(\text{Assistenten} \times \text{Professoren}))$

SELECT Assistenten.Name **FROM** Assistenten, Professoren **WHERE** Name='Kopernikus' and Boss=PersNr

Welche Studenten (Name und MatrNr) hören die Vorlesung Grundzüge?

$\Pi_{\text{Studenten.Name, Studenten.MatrNr}}(\sigma_{\text{Titel}='Grundzuege' \wedge \text{ hoeren.MatrNr}=\text{Studenten.MatrNr} \wedge \text{ hoeren.VorlNr}=\text{Vorlesungen.VorlNr}}(\text{Studenten} \times \text{ hoeren} \times \text{ Vorlesungen}))$

SELECT Name, Studenten.MatrNr from Studenten, hoeren, Vorlesungen **where** Titel='Grundzuege' and hoeren.MatrNr=Studenten.MatrNr and hoeren.VorlNr=Vorlesungen.VorlNr



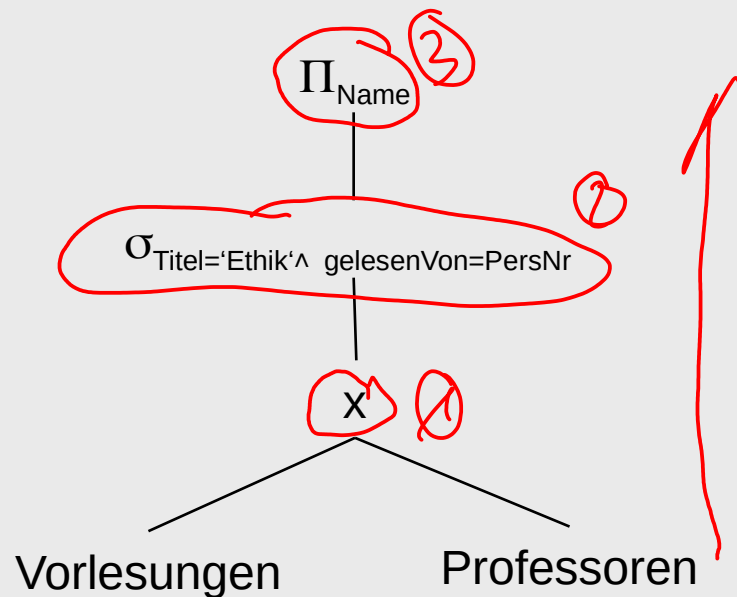
Inline versus Operatorbaum

Anfrage: Welcher Professor liest Ethik?

Inline:

$\Pi_{\text{Name}}^{(3)} (\sigma_{\text{Titel}='Ethik' \wedge \text{gelesenVon}=\text{PersNr}}^{(2)} (\text{Vorlesungen} \times \text{Professoren}))$

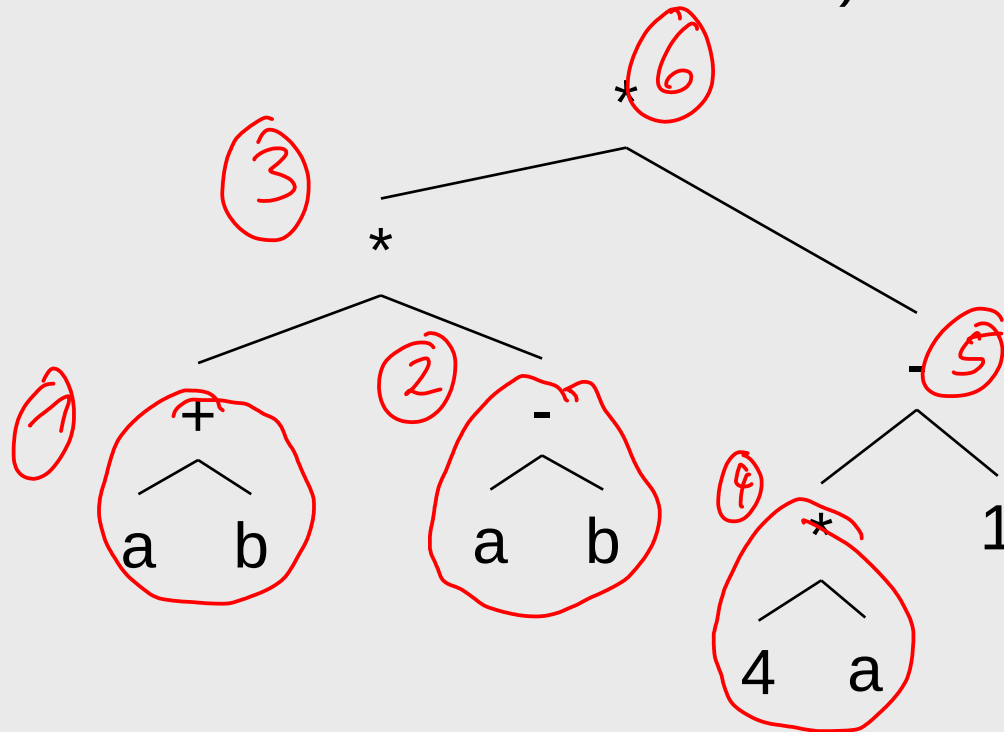
Operatorbaum:



Vergleich Mathematische Formel

Zum Beispiel $(a+b)*(a-b)*(4*a-1)$ Punkt vor Strichrechnung

Auswertungsbaum des Mathematischen Ausdrucks
(wird im Rechner so intern verwendet!)



Übersetzung in die relationale Algebra

Allgemein hat eine (ungeschachtelte) SQL-Anfrage die Form:

select A_1, \dots, A_n
from R_1, \dots, R_k
where P ;

Übersetzung in die relationale Algebra:

$$\Pi_{A_1, \dots, A_n}(\sigma_P(R_1 \times \dots \times R_k))$$

$$\Pi_{A_1, \dots, A_n}$$

$$\sigma_P$$

Binärer Operator

\times

\times

R_k

R_3

R_1

R_2

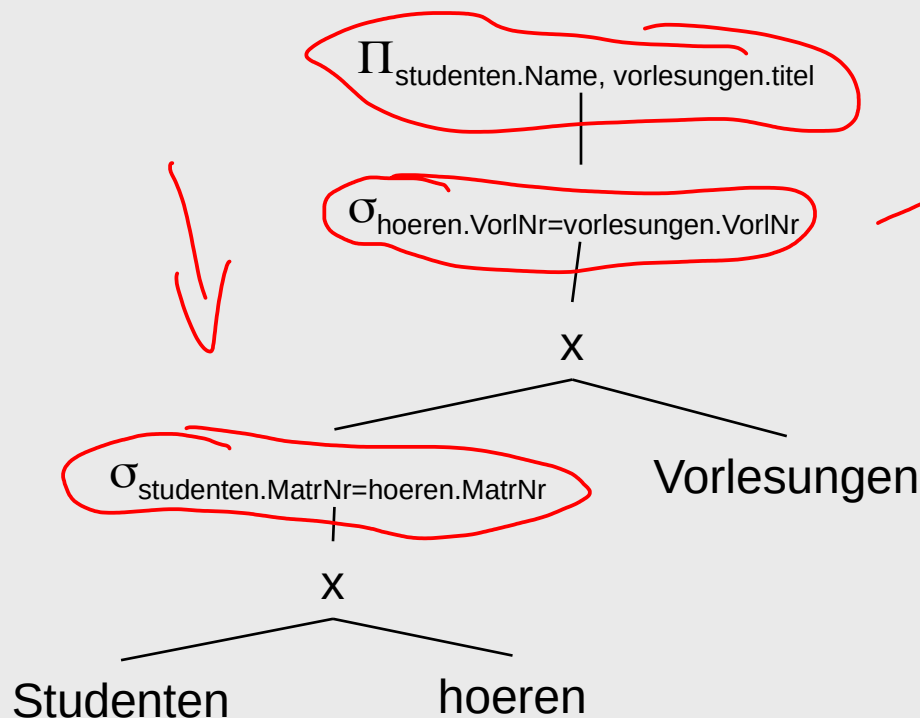
Bäume starten unten mit
Relation

~~$R_1 R_2 R_3$~~

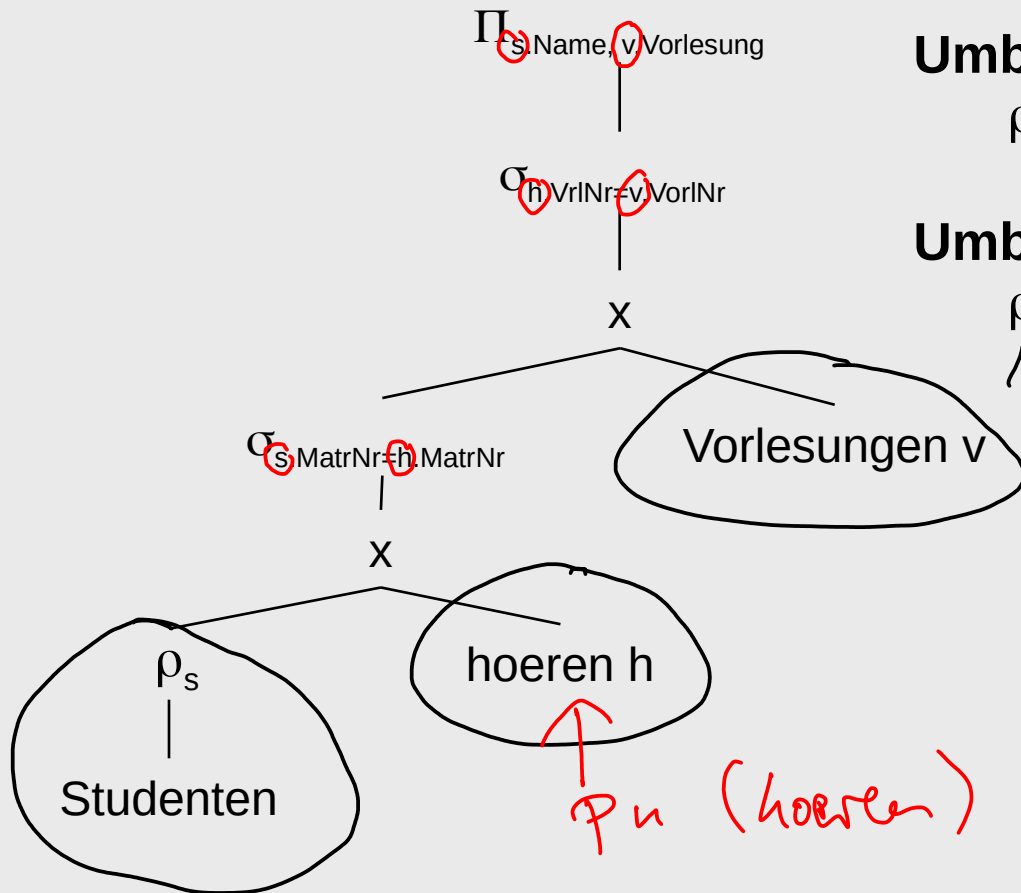
Beispiel Operatorbaum

Welche Studenten hören welche Vorlesungen?

$\Pi_{\text{studenten.Name, vorlesungen.titel}} (\sigma_{\text{ hoeren.VorlNr=vorlesungen.VorlNr}} ($
 $\text{Vorlesungen} \times \sigma_{\text{ studenten.MatrNr= hoeren.MatrNr}} (\text{Studenten} \times \text{ hoeren})))$



Umbenennungsoperator



Umbenennung von Attributen:

$\rho_{\text{neuerName}} \leftarrow \text{alterName (Relation)}$

$PID \leftarrow MatrNr \text{ (Studenten)}$

Umbenennung von Relationen:

$\rho_{\text{neuerName}} \text{ (Relation)}$

\uparrow
 Rho

$p_s(\text{Studenten})$

count....

\times
 π
 σ
 ρ

Welcher Professor hält mindestens 2 Veranstaltungen?

Welcher Professor liest mindestens 2 Vorlesungen? (Self Join)

↙
V x V

Vorlesungen v1				Vorlesungen v2			
VorlNr	Titel	SWS	gelesenVon	VorlNr	Titel	SWS	gelesenVon
5001	Grundzüge	4	2137	5001	Grundzüge	4	2137
5041	Ethik	4	2125	5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126	5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125	5049	Mäeutik	2	2125
4052	Logik	4	2125	4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126	5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126	5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133	5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134	5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137	4630	Die 3 Kritiken	4	2137

V1.VorlNr <> v2.VorlNr

V1.gelesenVon = v2.gelesenVon

SELECT name FROM

Professoren, Vorlesungen (as) V1,
Vorlesungen (as) V2

$\sigma_{\text{PersNr} = \text{V1.f gelesen von}}$

Where V1.f gelesen von = V2.f gelesen von
And V1.VorNr != V2.VorNr and
PersNr = gelesen von

Professoren

Vorlesungen V1

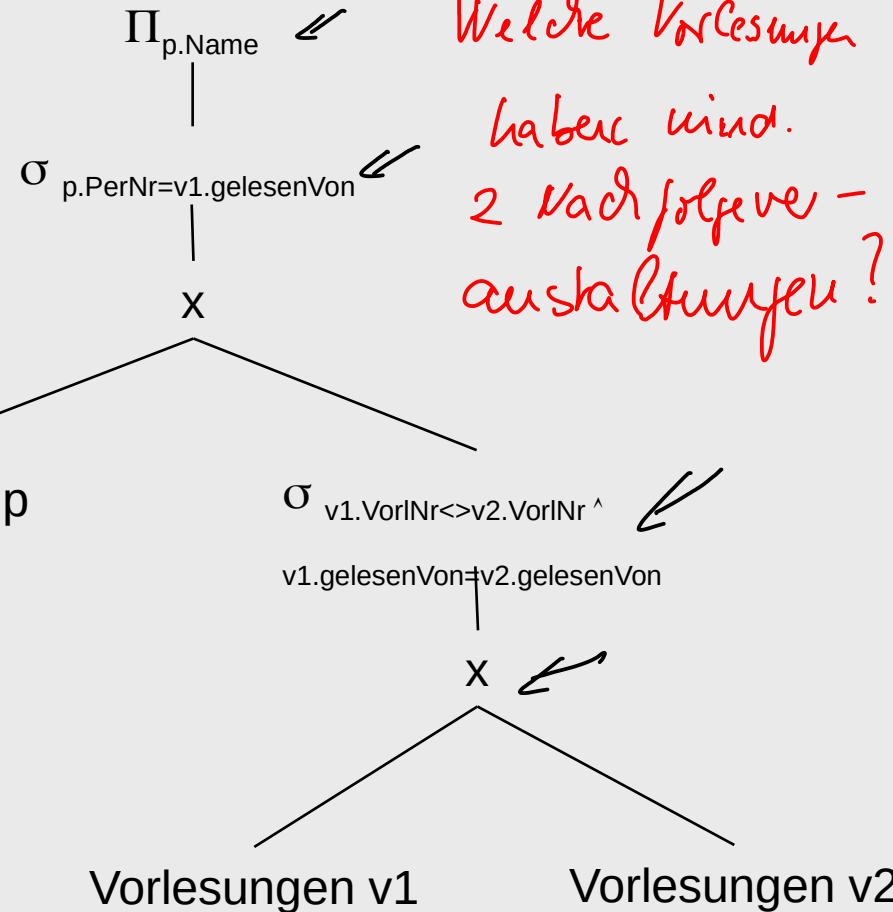
Vorlesungen V2

Operatorbaum zur Anfrage

Wie sieht die Abfrage in SQL aus?

select p.Name from Professoren p,
Vorlesungen v1, Vorlesungen v2
where v1.VorlNr<>v2.VorlNr AND
v1.gelesenVon=v2.gelesenVon AND
p.PersNr=v1.gelesenVon

Proj 1 → Proj 2 → Proj 3



Wie sieht die Inline-Darstellung des Operatorbaums aus?

$\Pi_{p.Name}(\sigma_{p.PersNr=v1.gelesenVon}(\text{Professoren } x(\sigma_{v1.VorlNr \neq v2.VorlNr \wedge v1.gelesenVon=v2.gelesenVon}(\text{Vorlesungen } v_1 \times \text{Vorlesungen } v_2))))$