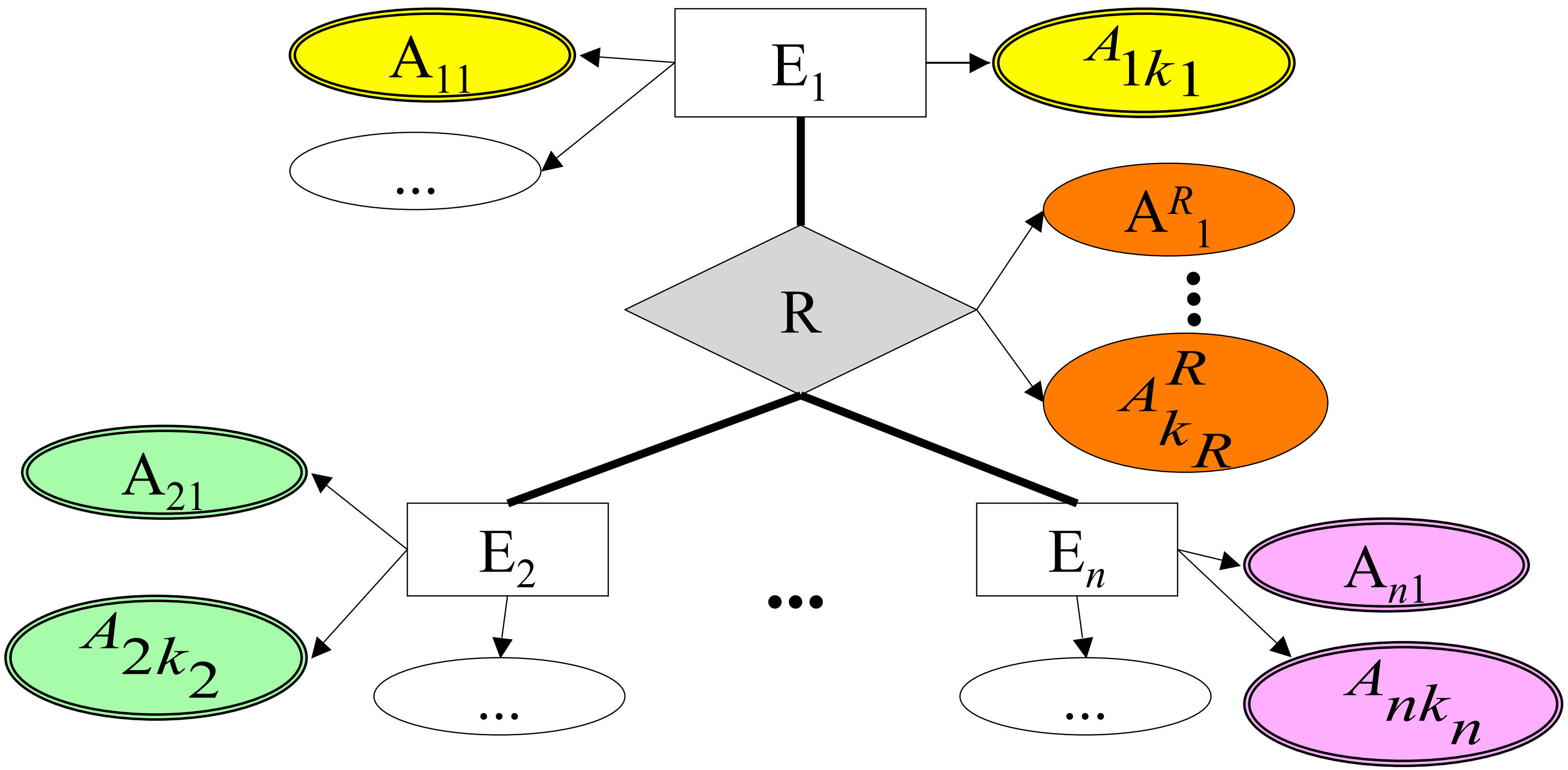
Kapitel 3: Das relationale DB-Modell & SQL

<table>
<tr><th></th><th>Relationales Datenmodell (RDM)</th><th>Netzwerk- und Hierarchisches Datenmodell (NDM, HDM)</th><th>Objekt-orientierte Datenmodelle (OODM)</th><th>Objekt-relationale Datenmodelle</th></tr>
<tr><td>Überblick über die Konzepte</td><td rowspan="5">3.1</td><td rowspan="6">4.1
4.2</td><td rowspan="5">5.1</td><td rowspan="5">6.1</td></tr>
<tr><td>Darstellung von Assoziationen</td></tr>
<tr><td>Datendefinition</td></tr>
<tr><td>Anfragen</td></tr>
<tr><td>Aktualisierungs-operationen</td></tr>
<tr><td>Spezifika</td><td>3.2 SQL</td><td>5.2 ODMG</td><td>6.2, 6.3</td></tr>
</table>

Relationale Darstellung von Beziehungen



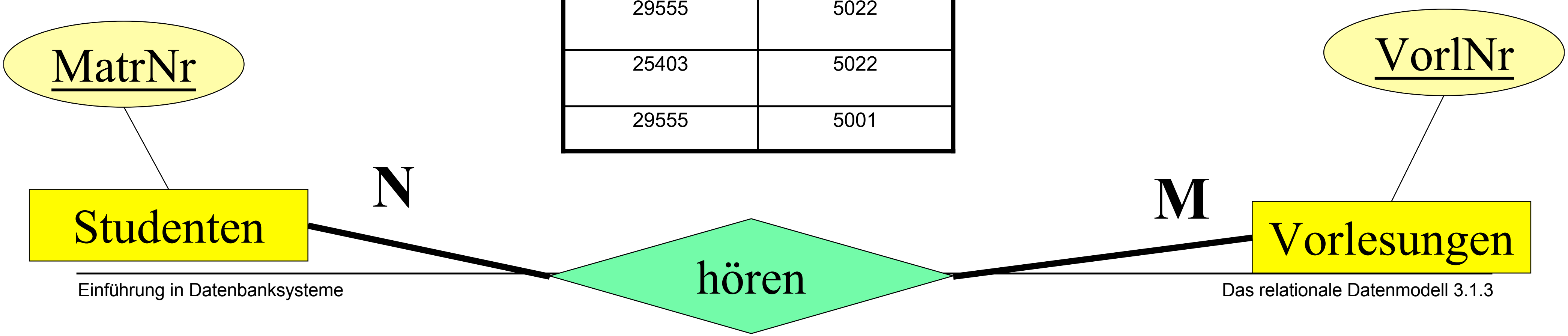
$$R: \{[\underbrace{A_{11},\ldots, A_{1k_1}}_{\text{Schlüssel von } E_1}, \underbrace{A_{21},\ldots, A_{2k_2}}_{\text{Schlüssel von } E_2}, \ldots, \underbrace{A_{n1},\ldots, A_{nk_n}}_{\text{Schlüssel von } E_n}, \underbrace{A_1^R,\ldots, A_{k_R}^R}_{\text{Attribute von } R}]\}$$

Ausprägung der Beziehung *hören*

Studenten	
MatrNr	...
26120	...
27550	...
...	...

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
29555	5001

Vorlesungen	
VorlNr	...
5001	...
4052	...
...	...



Notation für Relationenschemata

Schema: Tabellenname = {[Attr1: Typ1, Attr2: Typ2, ...]}

In eckigen Klammern [...] wird angegeben, wie die Tupel aufgebaut sind.

Die Mengenklammern sollen ausdrücken, daß es sich bei einer Relationenausprägung um eine Menge von Tupeln handelt.

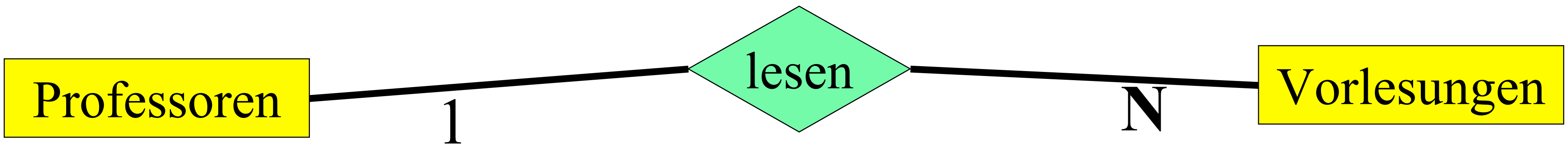
Manchmal werden die Attribute auch als Menge benötigt:

Wir schreiben für das Schema der Tabelle $\mathcal{R}$: $\mathcal{R}$ = {Attr1, Attr2, ...}.

Eine konkrete Relation R ist eine Teilmenge des Kreuzproduktes von

dom(Attr1) x dom(Attr2) x ...

Verfeinerung des relationalen Schemas



1:N-Beziehung

Initial-Entwurf

Vorlesungen : {[*VorlNr, Titel, SWS*]}

Professoren : {[*PersNr, Name, Rang, Raum*]}

lesen: {[*VorlNr, PersNr*]}

Verfeinerung des relationalen Schemas

1:N-Beziehung

Initial-Entwurf

Vorlesungen : {[*<u>VorlNr</u>, Titel, SWS*]}

Professoren : {[*<u>PersNr</u>, Name, Rang, Raum*]}

lesen: {[*<u>VorlNr</u>, PersNr*]}

Verfeinerung durch Zusammenfassung

Vorlesungen : {[*<u>VorlNr</u>, Titel, SWS,* ***gelesenVon***]}

Professoren : {[*<u>PersNr</u>, Name, Rang, Raum*]}

Regel

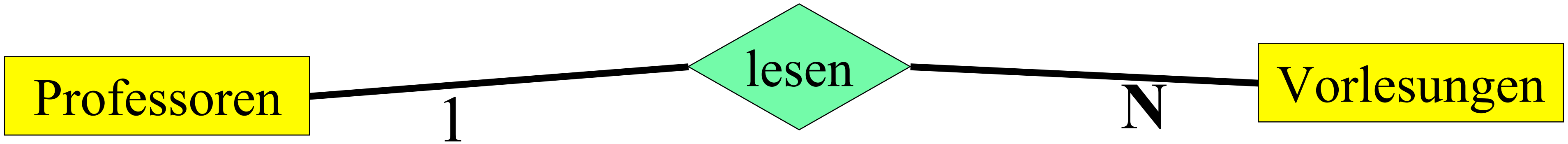
Relationen mit gleichem Schlüssel kann man zusammenfassen

aber nur diese und keine anderen!

Ausprägung von *Professoren* und *Vorlesung*

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

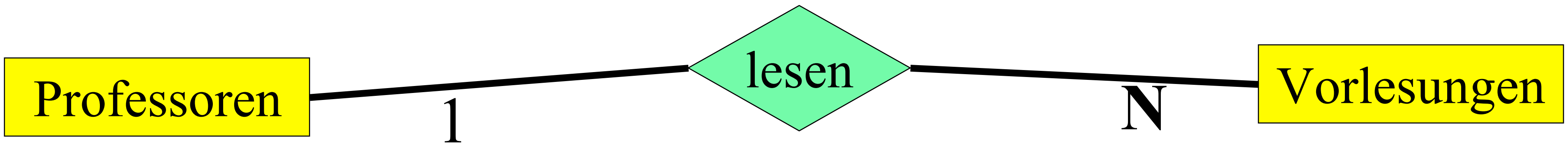
Vorlesungen			
VorlNr	Titel	SWS	Gelesen Von
5001	Grundzüge	4	2137
5041	Ethik	4	2125
5043	Erkenntnistheorie	3	2126
5049	Mäeutik	2	2125
4052	Logik	4	2125
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126
5259	Der Wiener Kreis	2	2133
5022	Glaube und Wissen	2	2134
4630	Die 3 Kritiken	4	2137



Vorsicht: So geht es NICHT

Professoren				
PersNr	Name	Rang	Raum	liest
2125	Sokrates	C4	226	5041
2125	Sokrates	C4	226	5049
2125	Sokrates	C4	226	4052
...	...	...	...	...
2134	Augustinus	C3	309	5022
2136	Curie	C4	36	??

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



Anomalien

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

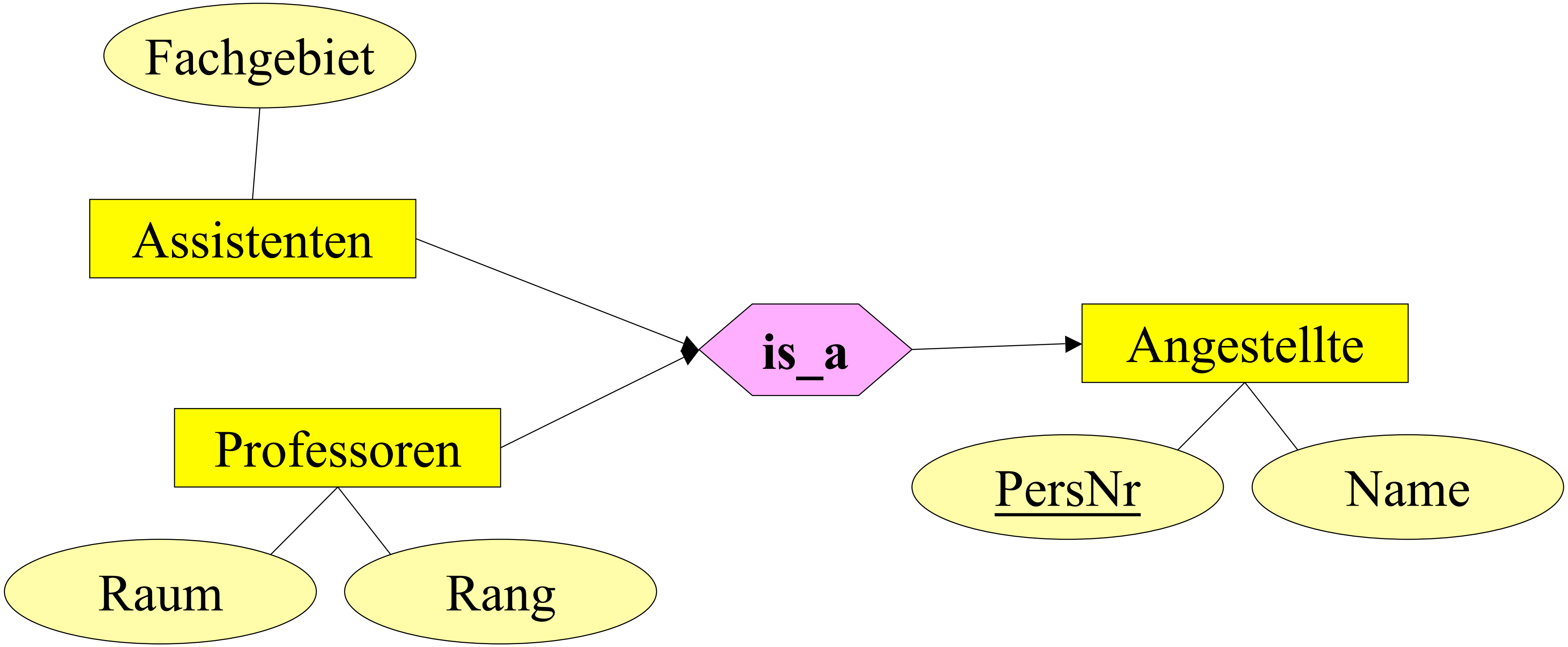
Professoren				
PersNr	Name	Rang	Raum	liest
2125	Sokrates	C4	226	5041
2125	Sokrates	C4	226	5049
2125	Sokrates	C4	226	4052
...	...	...	...	...
2134	Augustinus	C3	309	5022
2136	Curie	C4	36	??

Update-Anomalie: Was passiert wenn Sokrates umzieht

Lösch-Anomalie: Was passiert wenn „Glaube und Wissen“ wegfällt

Einfügeanomalie: Curie ist neu und liest noch keine Vorlesungen

Relationale Modellierung der Generalisierung



Angestellte: {[*<u>PersNr</u>*, *Name*]}

Professoren: {[*<u>PersNr</u>*, *Rang*, *Raum*]}

Assistenten: {[*<u>PersNr</u>*, *Fachgebiet*]}

Vereinbarung zur Notation

Sei $\mathcal{R}$ = *{A, B, C, D}* ein Relationenschema.

Seien r und t Tupel aus einer konkreten Relation R gemäß dem Schema $\mathcal{R}$.

Sei weiterhin □ □ $\mathcal{R}$.

Wir vereinbaren:

r.□ = t.□ soll heißen, daß für alle A aus □ gilt: r.A = t.A.

Funktionale Abhängigkeiten

Schema

- $\mathcal{R}$ = {*A*, *B*, *C*, *D*}

Ausprägung *R*

Seien □ □ $\mathcal{R}$, □ □ $\mathcal{R}$

□ □ □ genau dann wenn □ *r*, *s* □ *R* mit *r*.□ = *s*.□ □ *r*.□ = *s*.□

R			
A	*B*	*C*	*D*
a4	b2	c4	d3
a1	b1	c1	d1
a1	b1	c1	d2
a2	b2	c3	d2
a3	b2	c4	d3

{A} → {B}

{C, D } → {B}

Nicht: {B} → {C}

Notationskonvention:

CD → B

Beispiel

Stammbaum				
Kind	Vater	Mutter	Opa	Oma
Sofie	Alfons	Sabine	Lothar	Linde
Sofie	Alfons	Sabine	Hubert	Lisa
Niklas	Alfons	Sabine	Lothar	Linde
Niklas	Alfons	Sabine	Hubert	Lisa
...	...	...	Lothar	Martha
…	…	…	…	…

Kind → Vater,Mutter

Kind,Opa → Oma

Kind,Oma → Opa

Schlüssel

□ □ $\mathcal{R}$ ist ein Super-Schlüssel, falls folgendes gilt:

- □ □ $\mathcal{R}$

Wir nennen □ Super-Schlüssel, weil noch nichts darüber ausgesagt wird,

daß der Schlüssel □ minimal ist.

□ ist voll funktional abhängig von □ genau dann wenn gilt

- □ □ □ und
- □ kann nicht mehr verkleinert werden, d.h.
  - □*A* □ □ folgt, dass (□ □ {□}) □ □ nicht gilt, oder kürzer
  - □*A* □ □: □((□ □ {□}) □ □)

Notation für volle funktionale Abhängigkeit: □ □ $\overset{\bullet}{}$ □

□ □ $\mathcal{R}$ ist ein Kandidaten-Schlüssel, falls folgendes gilt:

- □ □ $\overset{\cdot}{}$ $\mathcal{R}$

Schlüsselbestimmung

Städte			
Name	BLand	Vorwahl	EW
Frankfurt	Hessen	069	650000
Frankfurt	Brandenburg	0335	84000
München	Bayern	089	1200000
Passau	Bayern	0851	50000
...	...	...	...

Kandidaten-schlüssel von *Städte*:

- {Name,BLand}
- {Name,Vorwahl}

Beachte, dass 2 kleinere Städte dieselbe Vorwahl haben können

Bestimmung funktionaler Abhängigkeiten

Professoren: {[PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung]}

- {PersNr} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung}
- {Ort,BLand} → {EW, Vorwahl}
- {PLZ} → {Bland, Ort, EW}
- {Bland, Ort, Straße} → {PLZ}
- {Bland} → {Landesregierung}
- {Raum} → {PersNr}

Zusätzliche Abhängigkeiten, die aus obigen abgeleitet werden können:

- {Raum} → {PersNr, Name, Rang, Raum, Ort, Straße, PLZ, Vorwahl, Bland, EW, Landesregierung}
- {PLZ} → {Landesregierung}

RDM: Relationale Algebra - Anfragen (1)

Vereinigung $R \cup S$:

- ❑ Alle Tupel zweier Relationen werden in einer Ergebnisrelation zusammengefaßt.
- ❑ Das Ergebnis enthält keine Duplikate.

$$R \cup S := \{ r \mid r \in R \lor r \in S \}$$

Relation R

ANr	*AName*	*Menge*
001	Anlasser	1.000
⋮	⋮	⋮
199	Kolben	5.000

Relation S

ANr	*AName*	*Menge*
237	Ölfilter	1.560
⋮	⋮	⋮
851	Schraube	25.000

Ergebnisrelation $R \cup S$

ANr	*AName*	*Menge*
001	Anlasser	1.000
⋮	⋮	⋮
199	Kolben	5.000
237	Ölfilter	1.560
⋮	⋮	⋮
851	Schraube	25.000

RDM: Relationale Algebra - Anfragen (2)

Differenz R \ S:

- ❑ Die Tupel zweier Relationen werden miteinander verglichen.
- ❑ Die in der ersten, nicht aber in der zweiten Relation befindlichen Tupel werden in die Ergebnisrelation aufgenommen.

$$R \setminus S := \{ r \mid r \in R \wedge r \notin S \}$$

Relation R

ANr	*AName*	*Menge*
001	Anlasser	1.000
237	Ölfilter	1.560
199	Kolben	5.000

Relation S

ANr	*AName*	*Menge*
851	Schraube	25.000
232	Gummiring	2.000
001	Anlasser	1.000

Ergebnisrelation $R \setminus S$

ANr	*AName*	*Menge*
237	Ölfilter	1.560
199	Kolben	5.000

RDM: Relationale Algebra - Anfragen (3)

Durchschnitt R $\cap$ S:

- Alle Tupel, die sowohl in der Relationen R als auch in der Relation S enthalten sind, werden in der Ergebnisrelation zusammengefaßt.

$$R \cap S := \{ r \mid r \in R \wedge r \in S \}$$

Relation R

ANr	*AName*	*Menge*
001	Anlasser	1.000
007	Zündkerze	1.380
199	Kolben	5.000

Relation S

ANr	*AName*	*Menge*
001	Anlasser	1.000
199	Kolben	5.000
237	Ölfilter	1.560

Ergebnisrelation $R \cap S$

ANr	*AName*	*Menge*
001	Anlasser	1.000
199	Kolben	5.000

RDM: Relationale Algebra - Anfragen (4)

Kartesisches Produkt R × S:

- ❑ Alle Tupel zweier Relationen R und S werden kombinatorisch miteinander verbunden. Wenn die Relation R n Spalten und die Relation S m Spalten umfaßt, dann besitzt R × S (*n+m*) Spalten.
- ❑ Wenn die Relation R k Zeilen und die Relation S l Zeilen umfaßt, dann besitzt R × S (*k*l*) Zeilen.
- ❑ Um eindeutige Attributbezeichnungen in der Ergebnisrelation zu gewährleisten, müssen Attribute, die in den Relationen R und S gleich bezeichnet sind, vor der Bildung des kartesischen Produkts umbenannt werden.

$$R \times S := \{ (r_1, ..., r_n, s_1, ..., s_m) \mid (r_1, ..., r_n) \in R, (s_1, ..., s_m) \in S \}$$

- ❑ <u>Beispiel:</u>
  - Projekte × Projektdurchführung (s. nächste Folie)

RDM: Relationale Algebra - Anfragen (5)

Beispiel: Projekte × Projektdurchführung

Projekte

Nr	*Titel*	*Budget*
100	DB Fahrpläne	300.000
200	ADAC Kundenstamm	100.000
300	Telekom Statistik	200.000

Projektdurchführung (Ausschnitt)

Nr	*Kurz*
100	MFSW
200	PERS
300	MFSW

Ergebnisrelation Projekte × Projektdurchführung

Nr	*Titel*	*Budget*	*Nr2*	*Kurz*
100	DB Fahrpläne	300.000	100	MFSW
100	DB Fahrpläne	300.000	200	PERS
100	DB Fahrpläne	300.000	300	MFSW
200	ADAC Kundenstamm	100.000	100	MFSW
200	ADAC Kundenstamm	100.000	200	PERS
200	ADAC Kundenstamm	100.000	300	MFSW
300	Telekom Statistik	200.000	100	MFSW
300	Telekom Statistik	200.000	200	PERS
300	Telekom Statistik	200.000	300	MFSW

RDM: Relationale Algebra - Anfragen (6)

Join (Verbindung) R ><$_{\theta}$ S:

- Eine Verbindung zwischen zwei Relationen wird in einer Kombination von kartesischem Produkt und nachfolgender Selektion (σ) gemäß des Prädikats θ hergestellt.
- Im allgemeinen Fall (*Theta-Join*) vergleicht ein (beliebiges) Prädikat θ mehrere Attribute aus den Relationen R und S (Spezialfall: Equi-Join).

$$R >\!<_{\theta} S := \sigma_{\theta}(R \times S)$$

- <u>Beispiele:</u>
  - *Projekte* ><$_{(\text{Nr} \neq \text{Nr})}$ *Projektdurchführung* (s. nächste Folie)
  - *Projekte* ><$_{(\text{Budget} > 150000) \wedge (\text{Nr} = \text{Nr})}$ *Projektdurchführung*
- Die Ergebnisrelation enthält die Zeilen des kartesischen Produkts der Relationen R und S, die θ erfüllen.

RDM: Relationale Algebra - Anfragen (7)

Beispiel: Projekte >< $_{(Nr \neq Nr)}$ Projektdurchführung

Projekte

Nr	*Titel*	*Budget*
100	DB Fahrpläne	300.000
200	ADAC Kundenstamm	100.000
300	Telekom Statistik	200.000

Projektdurchführung (Ausschnitt)

Nr	*Kurz*
100	MFSW
200	PERS
300	MFSW

Ergebnisrelation

Nr	*Titel*	*Budget*	*Nr2*	*Kurz*
100	DB Fahrpläne	300.000	200	PERS
100	DB Fahrpläne	300.000	300	MFSW
200	ADAC Kundenstamm	100.000	100	MFSW
200	ADAC Kundenstamm	100.000	300	MFSW
300	Telekom Statistik	200.000	100	MFSW
300	Telekom Statistik	200.000	200	PERS

RDM: Relationale Algebra - Anfragen (8)

Join (Verbindung): Fortsetzung

- ❑ Von besonderer Bedeutung im RDM ist der *Natural Join*, da er eine Verknüpfung von Tabellen über ihre Fremdschlüsselwerte erlaubt.
  - <u>Beispiel:</u>
    - *Projekte* >< *Projektdurchführung* := *Projekt* >< $_{\mathrm{Nr} = \mathrm{Nr}}$ *Projektdurchführung*
    - In diesem Fall betrachtet □ nur die Gleichheit zwischen Fremdschlüssel und Primärschlüssel, die den gleichen Attributnamen (*Nr*) besitzen.
- ❑ Weitere abgeleitete *Joinoperationen* (*Semi-Join, Outer-Join*, ...) und die Division zweier Relationen sind beschrieben in:
  - S.M. Lang, P.C. Lockemann. Datenbankeinsatz. Springer, Berlin u.a., 1995.

RDM: Relationale Algebra - Anfragen (9)

Natural Join: Projekte >< $_{(Nr = Nr)}$ Projektdurchführung

Projekte

Nr	*Titel*	*Budget*
100	DB Fahrpläne	300.000
200	ADAC Kundenstamm	100.000
300	Telekom Statistik	200.000

Projektdurchführung (Ausschnitt)

Nr	*Kurz*
100	MFSW
200	PERS
300	MFSW

Ergebnisrelation

Nr	*Titel*	*Budget*	*Nr2*	*Kurz*
100	DB Fahrpläne	300.000	100	MFSW
200	ADAC Kundenstamm	100.000	200	PERS
300	Telekom Statistik	200.000	300	MFSW

RDM: Relationale Algebra - Anfragen (10)

Projektion $\pi_{(r_{f_1}, \ldots, r_{f_n})}$ (R):

- ❑ n Spalten einer m-stelligen Relation R werden über ihren Namen ausgewählt.
- ❑ Dadurch entsteht eine n-stellige Relation *($n \leq m$).*
- ❑ Die Reihenfolge der Spalten in der Ergebnisrelation kann definiert werden.
- ❑ Duplikatelimination in der Ergebnisrelation.

$$\pi_{(r_{f_1}, \ldots, r_{f_n})}(R) := \{ (r_{f_1}, \ldots, r_{f_n}) \mid (r_1, \ldots, r_m) \in R \}$$

- ❑ <u>Beispiel:</u> $\pi_{(Nr,\ Budget)}$ (Projekte)

Projekte

Nr	*Titel*	*Budget*
100	DB Fahrpläne	300.000
200	ADAC Kundenstamm	100.000
300	Telekom Statistik	200.000

Ergebnisrelation
$\pi_{(Nr,\ Budget)}$(Projekte)

Nr	*Budget*
100	300.000
200	100.000
300	200.000

RDM: Relationale Algebra - Anfragen (11)

Selektion $\sigma_\varphi(R)$:

- Bestimmte Tupel einer Relation werden ausgewählt und in der Ergebnisrelation vereinigt.
- Zur Auswahl der zu übernehmenden Tupel dient das Prädikat φ : R ➠ { `true, false` }, in dem die Attributbezeichner als Eingabevariablen dienen.
- Anwendung dieses Prädikats auf jedes Tupel der Ausgangsrelation, indem die Werte des Tupels unter den jeweiligen Attributen für die Variablen eingesetzt werden.
- In die Ergebnisrelation werden alle Tupel übernommen, für die das Prädikat den Wahrheitswert `true` liefert.

$$\sigma_\varphi(R) := \{ r \in R \mid \varphi(r) \}$$

Relation R

ANr	*AName*
001	Anlasser
007	Zündkerze
199	Kolben

Ergebnisrelation
$\sigma_{ANr < 199}(R)$

ANr	*AName*
001	Anlasser
007	Zündkerze

RDM: Relationale Algebra: Zusammenfassung

Vorteil:

- ❑ Einfache, mathematische Behandlung, z.B. $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$
- ❑ Einfache (naive) Implementierung durch typisierte (!?) algebraische Signatur

Nachteile:

- ❑ Eingeschränkte Ausdrucksmächtigkeit auf Relationenebene (Summe, Mittelwert, Kardinalität)
- ❑ Reine Anfragesprache
- ❑ Prozedurale Formulierung des Lösungswegs (➠ Expressions) statt deklarativer Spezifikation des Ergebnisses (➠ *Kalküle*)