UNIVERSITÄT LEIPZIG

Lese- und Übungsbuch Datenbanken:

Die Relationenalgebra

Dieter Sosna
Institut für Informatik
Abt. Datenbanken

pre-Version 0.9 vom 23. Dezember 2008

Vorwort [1]

Für das Lesen des Buches werden Grundkenntnisse des Relationenmodells erwartet, sie werden im Kapitel 1 kurz zusammengestellt. Eine ausführliche Diskussion wird in dem Band „E/R- und Relationenmodell“ dieser Reihe gegeben, vgl. [7]. In den folgenden Kapiteln 2 und 3 werden die Elemente der Relationenalgebra vorgestellt, in Kapitel 4 Grenzen und Beschränkungen angegeben und Erweiterungen vorgeschlagen. Das Kapitel 5 enthält Beispiele. Den Abschluß bildet Kapitel 6 mit Aufgaben aus Klausuren, deren Lösungen im Anhang angegeben werden. Auch für diesen Band des Lese- und Übungsbuches gilt: er ist als Ergänzung zur Vorlesung und zum Übungsbetrieb gedacht. Er soll anregen, über einige Fragen weiter nachzudenken, die dort nur angerissen werden. Keinesfalls soll er den Besuch der Lehrveranstaltungen und das Studium der Literatur ersetzen.

Dieter Sosna

Leipzig, im Dezember 2008.

---

[1] Entwurf 18 vom 23. Dezember 2008

Inhaltsverzeichnis

1 Grundbegriffe des relationalen Modells

Eine Datenbank nach dem relationalen Modell besteht aus mehreren Relationen, die die in der Datenbank enthaltenen Sachdaten darstellen. Die Art und die Zahl dieser Relationen hängt von dem in der Datenbank modellierten Sachverhalt ab. Folglich kommen zu den Sachdaten notwendigerweise Informationen, die die Struktur der Relationen beschreiben - die sogenannten Metadaten. Die Kenntnis der Metadaten ist notwendig, um die Daten korrekt und sinnvoll zu bearbeiten.
Das Ziel dieses Kapitels ist es, die benötigte Begriffswelt bereitzustellen sowie Metainformationen und ihre Beschreibung zu erläutern ([1]).

1.1 Wertebereiche, Attribute

In einer Datenbank werden Beschreibungen von Objekten der Realität oder von Beziehungen zwischen diesen Objekten abgelegt. Die Beschreibung erfolgt stets durch die Angabe von charakteristischen Eigenschaften dieser Objekte und Beziehungen. Dies geschieht so, dass im Modellierungsprozess für jedes Objekt festgestellt wird, welche Eigenschaften es charakterisieren, welche Werte der charakterisierenden Eigenschaften für dieses Objekt zutreffen und welche Werte generell möglich sind. So könnte ein Auto zum Beispiel durch die Eigenschaften „Hersteller“, „Typ“, „Fahrgestellnummer“ und „Farbe“ beschrieben werden, die Farbe hat beispielsweise bei einem konkreten Objekt den Wert „rot“, wobei insgesamt die Farben „rot“, „schwarz“ und „silber“ möglich sind.
Die Eigenschaften werden in Form von sogenannten Attributen modelliert:

Zur **Beschreibung eines Attributs** gehören die Festlegung seines Namens A, seines Wertebereichs und seiner Semantik.

- Der *Name eines Attributs* repräsentiert das Attribut.
- Die *Semantik eines Attributs* erklärt die Bedeutung des Attributs im Kontext der in der Datenbank zu modellierenden Realität. Sie wird mit Mitteln außerhalb der Relationenalgebra, meist als Text, beschrieben und bildet die Grundlage jeder sinnvollen Verwendung eines Modells.
- Der *Wertebereich* ist die Menge aller Attributwerte, die in der zu beschreibenden Miniwelt für die durch das Attribut modellierte Eigenschaft auftreten können.
  Der Wertebereich $\mathcal{D}$ eines Attributs A wird mit $dom(A)$ bezeichnet.

[1] Weitere Details in [7]

Attribute sind die die elementaren Bestandteile des Relationenmodells.

Definition 1.1: Atomarität von Attributen
Ein Attribut heißt atomar, wenn es in dem betrachteten Kontext
- weder aus kleineren Einheiten zusammengesetzt
- noch mehrwertig (mengenwertig) ist.

Festlegung:
Im Relationenmodell sind nur atomare Attribute erlaubt ([2]).

Atomarität der Attribute ist grundlegend für die gesamte theoretische Untermauerung des Relationenmodells. Mit der Atomarität erfolgen gleichzeitig semantische Festlegungen. Wird beispielsweise für Personen ein Attribut *Adresse* festgelegt, dann bedeutet dies, dass auf eine explizite Gliederung der Adresse nach Ort, Straße und Hausnummer verzichtet wird. Auch wird damit festgelegt, dass eine Person nur eine Adresse haben kann. Gibt es inhaltliche Gründe, die eine feinere Gliederung wünschenswert machen oder mehrere Adressen erfordern, muss anders modelliert werden.

Beispiel 1.2: *Das Attribut „Farbe“ hat bei einem fiktiven Autotyp X die Werte „schwarz“, „tornadorot“, „divingblue“ und „bambusgrey“.*

Definition 1.3: *Zwei Attribute heißen* **(semantisch) äquivalent***, wenn sie in Wertevorrat und Semantik übereinstimmen.*

Bei der Umsetzung eines relationalen Modells in eine Datenbank unter einem Datenbankverwaltungssystem muss eine Einbettung dieses Wertebereichs in die Datentypen des Datenbankverwaltungssystems erfolgen. Moderne Systeme erlauben dabei die Einführung neuer, nutzerdefinierter Datentypen , die sich genau der benötigten Spezifikation anpassen lassen, so dass der Datentyp eine genaue Kontrolle der Attributwerte erlaubt, und Werte, die nicht in das Modell passen, abgewiesen werden können. Im oben gebenen Beispiel 1.2: wäre das ein Aufzählungstyp mit genau den angegebenen Werten. Gibt es diese Möglichkeit nicht, muss ein den Wertevorrat umfassender vorhandener Typ genutzt werden. Im vorliegenden Beispiel könnte das Attribut Farbe als Wertebereich in der Datenbank eine Zeichenkette variabler Länge erhalten. Allerdings kann dann die Kontrolle der Korrektheit der eingegeben Werte nicht durch den benutzten Datentyp gesteuert werden. So könnte dem Attribut auch der Wert „tiefschwarz“ zugewiesen werden.

---

[2] Im NF^2-Modell als Verallgemeinerung des Relationenmodells wird diese Festlegung aufgehoben.

Unbestimmte Werte:
Häufig tritt der Fall auf, dass der Wert eines Attributs nicht bekannt ist oder evt. gar nicht existiert. Dies kann bei der Festlegung des Wertevorrats eines Attributs berücksichtigt werden, so dass er solche Werte wie „Wert nicht bekannt“ oder „Wert existiert nicht“ umfasst. Diese Werte sind im relationalen Modell durchaus legitim, verhindern aber beispielsweise bei numerischen Werten die Verwendung einer numerischer Datentypen. Zur Behebung solcher Probleme werden in relationalen Datenbankverwaltungssystem bei den eingebauten Typen die natürliche Wertebereiche um den Wert „NULL“ ergänzt, der genau die beschriebene Bedeutung hat. Man sagt dann, dass für dieses Attribut NULL-Werte erlaubt werden. Es muss betont werden, dass dieser spezielle Attributwert, der so nur in Datenbanken auftritt, semantisch nichts mit der Zahl „Null“ zu tun hat. Auch unterliegt er Einschränkungen z.B. hinsichtlich der Vergleichbarkeit. Es ist nicht möglich zu sagen, dass zwei Werte, die unbekannt sind, einander gleich sind, folglich ist NULL nicht gleich NULL. Genaugenommen sind NULL-Werte ein Bestandteil der technischen Realisierung von Datenbanken nach dem Relationenmodell. Im Prozess der Modellierung sollte die Einführung von NULL-Werten so spät wie möglich erfolgen, da sie einen Informationsverlust bedeuten: Die Attributwerte „Wert nicht bekannt“ oder „Wert existiert nicht“ eines Attributs sind eben eine präzisere Beschreibung eines Zustandes.

1.2 Relationsschemata, Relationen

Anschaulich kann eine Relation als eine Tabelle interpretiert werden. Sie besteht aus einem Tabellenkopf, der seinerseits den Tabellennamen und die Spalteneinteilung mit den Spaltennamen enthält, und dem Tabellenkörper, in dem jede Zeile einen konkreten Sachverhalt beschreibt, indem dort die Werte in den Spalten eingetragen wurden.
Diese anschauliche Deutung vor Augen kann der Begriff „Relation“ abstrakt definiert werden. Dies beginnt mit dem Relationsschema, das dem Tabellenkopf entspricht.

Definition 1.4: Relationssschema
Die Beschreibung eines Relationssschema $\mathbf{R}(A_1, A_2, A_3, ..., A_n)$ *besteht aus*
- einem Relationsnamen $\mathbf{R}$
- einer Liste von n paarweise verschiedenen Namen von Attributen
$(A_1, A_2, A_3, ..., A_n)$, $n > 0$ *und*
- der Erklärung der Semantik.

Die Anzahl n der Attributenamen heißt **Grad** *des Relationschemas.*

Das Relationschema beinhaltet alle strukturellen Metainformationen einer Relation.

Die Elemente der Liste $(A_1, A_2, A_3, ..., A_n)$ heißen *Attribute des Relationsschemas* **R**, sie stellen Attribute im Sinne des Attributbegriffs auf Seite 7 dar und müssen vor ihrer Benutzung erklärt werden.

Beispiel 1.5: *Die Professoren einer Universität sollen erfaßt werden und zwar die Namen, die Personal-Nummer, das Arbeitsgebiet und die Fakulät, an der ein Professor tätig ist.*

Dies wird durch das folgende Relationsschema beschrieben:
Professoren$(Personal\text{-}Nr, Name, Arbeitsgebiet, Fakultät)$ ■

In dem Beispiel werden die Attribute nur durch ihren Namen spezifiziert, die Namen beschreiben hier die Semantik grob, der Wertevorrat wurde nicht angegeben. Ein solches vereinfachtes Vorgehen kann für einen groben Entwurf akzeptiert werden, bei einer Implementierung oder bei der Modellierung komplexer Sachverhalte sind an dieser Stelle Konkretisierungen nötig, die zum Beispiel die verwendeten Datentypen betreffen. Um mit der Relationenalgebra zu arbeiten, sind solche vereinfachten Darstellungen häufig durchaus akzeptabel und ausreichend.

Wenn ein Relationsschema wie in Definition 1.4: durch die Liste seiner Attribute definiert ist, so beinhaltet der Begriff „Liste“ eigentlich, dass die Reihenfolge der Attribute wesentlich ist. Im Relationenmodell trägt diese Reihenfolge jedoch keine Information, deshalb sind auch alternative, mengenorientierte Definitionen anstelle von Definition 1.4: möglich, in denen von einer *Menge von Attributen* gesprochen wird. Die hier gegebene Definition ist praktisch leichter handhabbar, sie widerspiegelt die Tatsache, dass auch die Spalten einer Tabelle in einer Reihenfolge angegeben werden (müssen). Um die Unabhängigkeit von der Reihenfolge der Attribute zu modellieren, wird der Begriff „äquivalente Relationsschemata“ definiert. Mit diesem Begriff soll eine Klasseneinteilung aller Relationsschemata so bewirkt werden, dass in einer Äquivalenzklasse genau die Schemata liegen, die sich nur durch eine unterschiedliche Reihenfolge der Attribute unterscheiden.

Definition 1.6: Äquivalenz von Relationsschemata
Zwei Relationsschemata **R** *und* **S** *heißen* äquivalent, *wenn sie in den Attributen, als Menge von Attributen betrachtet, übereinstimmen und die Anordnung der Attribute in* **S** *durch eine Permutation der Reihenfolge der Attribute in* **R** *entsteht.*

Offensichtlich ist die Reihenfolge der Attribute in **S** genau dann eine Permutation der Anordnung in **R**, wenn umgekehrt die Reihenfolge in **R** eine Permutation der Reihenfolge in **S** ist (Symmetrie). Weiter ist jede Relation zu sich selbst äquivalent (Reflexivität). Und wenn **R** äquivalent **S** und **S** äquivalent zu einer weiteren Relation **T** ist, ist auch **R** äquivalent **T** (Transitivität), da dies für Permutationen allgemein gilt. Wegen dieser Eigenschaften wird durch Definition 1.6: tatsächlich eine Äquivalenzrelation in der Menge der Relationsschemata definiert.

Definition 1.7: Relation
Eine Relation $\mathbf{r}(\mathbf{R})$, *abkürzend* $\mathbf{r}$, *zu einem Relationsschema* $\mathbf{R}(A_1, A_2, ..., A_n)$ *ist eine Menge von Tupeln* $t_i, i = 1, ..., k$:

$$\mathbf{r} = \{t_1, t_2, t_3, ..., t_k\}$$

mit

$$t_i = (w_{i,1}, w_{i,2}, w_{i,3}, ..., w_{i,n}) \in dom(A_1) \times dom(A_2) \times dom(A_3) \times ... \times dom(A_n).$$ ([3])

Der **Grad der Relation** *ist gleich dem Grad des zugehörigen Relationsschemas (also n). Die Anzahl k der Tupel heißt* **Kardinalität der Relation**.

Es ist üblich, dass **R** auch als *Name* einer zum Relationsschema $\mathbf{R}(...)$ entstandenen Relation dient. Für den Begriff „Relation“ sind als alternative Bezeichnungen *Relationsinstanz* oder *Relationszustand* gebräuchlich. Jedes Tupel einer Relation beschreibt eine konkrete Entität oder eine konkrete Beziehung, die in der modellierten Miniwelt existiert.
Im folgenden wird eine Instanz der Relation **Professoren** zu dem gleichnamigen Schema aus Beispiel 1.5: gegeben:

Professoren

PNR	Name	Arbeitsgebiet	Fakultät
2364	Rahm	Datenbanken	Math.u.Informatik
0342	Beyer	Analysis	Math.u.Informatik
9645	Wartenberg	Kirchengeschichte	Theologie
4714	Cain	Klass. Archäologie	Geschichte

Tabelle 1.1: Beispieltabelle „Professoren“

Jede Spalte der Tabelle besitzt einen Eintrag im Tabellenkopf, den Spaltennamen, dieser korrespondiert zu dem entsprechenden Attributnamen des Relationsschemas. Vielfach werden die Namen so gewählt, dass aus dem Namen die Semantik erschlossen werden kann, aber dies ist nicht zwingend vorgeschrieben. Im Beispiel hat das Attribut mit dem Namen *PNR* die Bedeutung „Personalnummer“. Es erlaubt, jeden Professor eindeutig zu identifizieren. Da die Attribute als atomar vorausgesetzt wurden, werden dem entsprechend nur Tabellen betrachtet, deren Spalten nicht aus Subspalten zusammengesetzt sind.
Jedes Tupel einer Relation entspricht genau einer Zeile in der Tabelle. Eine Relation **r** wurde als Menge von Tupeln definiert, folglich: die **Reihenfolge der Tupel trägt im Relationenmodell keine Information**, m.a.W. die Reihenfolge der Tupel einer Relation ist ohne Bedeutung. In der anschaulichen Interpretation als Tabelle ist also die Reihenfolge der Zeilen als zufällig anzusehen.

Um die Komponenten eines Tupels t anzusprechen, wird im Rahmen des Relationenmodells der Name des entsprechenden Attributs in folgender Form verwendet:

[3] $\mathcal{X} \times \mathcal{Y}$ bezeichnet das Kreuzprodukt der Mengen $\mathcal{X}$ und $\mathcal{Y}$ (siehe Definition 3.9:, Seite 28

$t\,[A_i]$ und $t.A_i$ beschreiben gleichermaßen den Wert des Attributs A_i im Tupel t ([4]). Diese Notation ist vergleichbar mit der *call-by-name*-Parameterübergabe bei Programmiersprachen und begründet die oben gestellte Forderung, dass die Namen der Attribute in einem Schema alle verschieden sein müssen. Gleichzeitig kann zugelassen werden, dass ein Attributname in unterschiedlichen Relationsschemata auftreten kann.
Durch die Angabe des *qualifizierten Attributnamens* in der Form **Relationenname**.*Attributname* ($\mathbf{R}.A$) wird die eindeutige Zuordnung gesichert, wenn ein Attributname in zwei Relationsschemata auftritt.

1.3 Schlüsselkandidaten, Primärschlüssel

Entsprechend der Mengendefinition in der Mathematik sind die Elemente einer Menge unterscheidbar. Relationen wurden als Mengen von Tupeln definiert. Die Unterscheidung der Tupel kann nur auf Grund von Attributwerte erfolgen, da die Tupel keine anderen Bestandteile haben. Folglich kann es in einer Relation keine zwei gleichen Tupel geben. Zwei Tupel einer Relation müssen sich in mindestens einem Attributwert unterscheiden.
Es ist sinnvoll, solche Tupel als Elemente einer Relation auszuschließen, bei denen alle Attribute unbekannte oder unbestimmte Werte haben, denn für sie kann die Gültigkeit dieser Forderung nicht sicher verifiziert werden.
Es stellt sich die Frage, ob alle Attribute eines Relationsschema zur Unterscheidung der Tupel herangezogen werden müssen oder ob schon eine Teilmenge der Attribute dazu ausreicht. Diese Frage wird in den meisten Fällen mit ja zu beantworten sein. Attributmengen mit dieser Eigenschaft heißen Schlüsselkandidaten. Bei ihrer Festlegung zeigt sich, dass im wesentlichen die Semantik der Attribute darüber entscheidet, welche Teilmengen der Menge aller Attribute eines Relationsschemas einen Schlüsselkandidaten bilden.

Definition 1.8: Schlüsselkandidat
Sei $\mathbf{R}(\mathcal{A})$ *ein Relationsschema,* $\mathcal{A}$ *sei die Menge seiner Attribute.* $\mathcal{S}$ *sei eine Teilmenge von* $\mathcal{A}$ $(\mathcal{S} \subseteq \mathcal{A})$.
$\mathcal{S}$ *heißt* Schlüsselkandidat *von* $\mathbf{R}$*, wenn*
(1) für zwei beliebige unterschiedliche Tupel t_1*,* t_2 *aus einer beliebigen Relation (Belegung)* $\mathbf{r}(\mathbf{R})$ *stets* $t_1\,[\mathcal{S}] \neq t_2\,[\mathcal{S}]$ *gilt und*
(2) es keine echte Teilmenge $\mathcal{S}_1$ *von* $\mathcal{S}$ *gibt, so dass (1) auch für* $\mathcal{S}_1$ *zutrifft*

Die zweite in der Definition genannte Eigenschaft des Schlüsselkandidaten wird **Minimalität eines Schlüsselkandidaten** genannt.

[4]Wenn $\mathcal{B} = (A_{l_1}, A_{l_2}, ..., A_{l_j})$ eine Teilmenge aller Attribute eines Relationsschema ist, dann sei $t\,[\mathcal{B}] = (t.A_{l_1}, t.A_{l_2}, ..., t.A_{l_j})$.

Aus der Menge aller Schlüsselkandidaten wird für jedes Relationsschema einer durch Auswahl zum **Primärschlüssel** erklärt. Theoretisch ist jeder Schlüsselkandidat gleich gut geeignet, egal aus wievielen Attributen er zusammengesetzt ist. Aus praktischer Sicht wird ein Primärschlüssel mit wenigen Attributen bevorzugt, da auf die Primärschlüssel häufig bezug genommen wird (Fremdschlüssel).

Es soll ausdrücklich festgehalten werden, dass
zur Beschreibung eines Relationsschema die Festlegung des Primärschlüssels gehört.

Es ist üblich, dass in der Notation von Relationsschemata Primärschlüsselattribute unterstrichen werden. ([5])
Primärschlüssel und Werte, die eine Unbestimmtheit ausdrücken.: Hat ein Attributwert eine Semantik wie „Wert unbekannt", ist u.U. die Bedingung (1) Definition 1.8: nicht sicher verifizierbar. Deshalb werden **Attribute, in deren Wertebereich Werte auftreten, die eine Unbestimmtheit ausdrücken, von der Bildung von Schlüsselkandidaten ausgeschlossen.** Attributwerte mit der Bedeutung einer Unbestimmtheit, ganz gleich, ob sie NULL-Werte oder anders genannt werden, erfordern auch an anderer Stelle eine besondere Behandlung, wie am Beispiel der Aggregatfunktionen (Seite 43) deutlich wird.

1.4 Der Begriff „Relationenalgebra"

Ausgehend von der Veranschaulichung der Relationen als Tabellen kann die Frage beantwortet werden, wie die Ergebnisse von Anfragen an eine Datenbank veranschaulicht werden können. Das Ergebniss einer Datenbankanfrage ist wieder eine Relation. Anschaulich wird bei einer Anfrage eine neue Tabelle aus den vorhandenen Tabellen konstruiert. Abstrakter kann formuliert werden: Aus einer Tabelle oder aus mehreren Tabellen entstehen durch Abbildungen neue Tabellen.
Diese Abbildungen bilden den Gegenstand des vorliegenden Buches. Grundlegend für den Abbildungsprozess und Namensgeber für das Buch ist die folgende Definition.

Definition 1.9: Algebra
Eine (nichtleere) Menge $\mathcal{M}$ mit einen (endlichen oder unendlichen) System $\mathcal{V}$ von partiellen algebraischen Verknüpfungen in ihr heißt Algebra *$\mathcal{A}$. Sind für $\mathcal{A}$ noch Operatorenbereiche gegeben, spricht man von einer* Operatoralgebra.

Die Verknüpfungen aus $\mathcal{V}$ werden auch als **mehrstellige Operatoren** bezeichnet. (s. [5], Stichworte: Algebraische Struktur bzw. Operatorenbereiche). In dem Bereich der Datenbanktheorie ist diese Bezeichnung die übliche, sie wird auch in dem vorliegenden Lesebuch übernommen.

[5] Auf diese Auszeichnung wird bei der Arbeit mit der relationalen Algebra gelegentlich verzichtet.

Die Menge $\mathcal{M}$ aus der allgemeinen Definition einer Operatoralgebra ist in der Relationenalgebra eine Menge von Relationen, d.h. eine Menge von Tupelmengen.

1.5 Symbolische Bezeichnungen

Für die weitere Arbeit in diesem Lesebuch werden die schon in diesem Kapitel genutzten Darstellungsformen für bestimmte mathematische Objekte durchgängig weiterverwendet. Die folgende Übersicht fasst diese Darstellung zusammen.

Relationsschemata werden mit fetten Großbuchstaben oder Namen $\mathbf{A}, \mathbf{B}, \mathbf{C}, ..., \mathbf{R}, ...,$ **Professoren** bezeichnet,
Attribute mit Großbuchstaben oder Namen in kursiver Normalschrift wie im Formalsatz $A, B, C, ..., Farbe$,
Relationen mit fetten Kleinbuchstaben $\mathbf{r}, \mathbf{s}, \mathbf{t}, ...$, ggf. unter Nennung des Relationsschemas $\mathbf{r}(\mathbf{R})$.
Die **Tupel einer Relation** werden mit Kleinbuchstaben im mathematischen Formelsatz $s, t, u, ...$ bezeichnet.

Wenn das Relationsschema $\mathbf{A}$ das Attribut D hat, dann kann das Attribut mit dem qualifizierten Namen $\mathbf{A}.D$ angesprochen werden, auf die Nennung des Relationsnamens $\mathbf{A}$ kann verzichtet werden, wenn durch den Attributnamen die Relation eindeutig bestimmt ist.

Zur Vereinfachung wird vereinbart:
Der Name $\mathbf{R}$ des Relationsschemas $\mathbf{R}(A_1, ..., A_n)$ allein wird auch als Bezeichnung der Relation $\mathbf{r}(\mathbf{R})$ verwendet , während $\mathbf{R}(A_1, ..., A_n)$ sich stets auf das Schema bezieht.

Zur Bezeichnung von Mengen, die keine Relationen sind, werden kaligraphische Großbuchstaben $\mathcal{A}, \mathcal{B}, \mathcal{C}, ...$ verwendet.

2 Einstellige Operatoren

Der Inhalt dieses und des folgenden Kapitels ist es, die Verknüpfungen von Relationen und die Operatoren der Algebra zu beschreiben. Der Definition 1.9: folgend, müßten zunächst die Verknüpfungen oder mehrstelligen Operatoren definiert werden. Für das Verständnis ist es jedoch angeraten, zunächst die einstelligen einzuführen.
Diese bilden eine Relation auf eine Relation ab. In der Relationenalgebra sind zwei wesentliche einstellige Operatoren definiert: Selektion und Projektion. Die weiteren Operatoren Zuweisungsoperator, Einbettungsoperatoren und sein Inverses sind technischer Natur.

2.1 Der Selektionsoperator

Die *Selektion* wird benutzt, um aus den Tupeln einer Relation die Tupel auszuwählen, bei denen bestimmte Attribute eine Auswahlbedingung erfüllen. Dies wird allgemein mit dem folgenden Symbol beschrieben

$$\sigma_{\langle Bedingung \rangle}(\mathbf{R}),$$

dabei ist $\langle Bedingung \rangle$ ein boolscher Ausdruck, der auf den Attributen von $\mathbf{R}$ definiert ist. In die Ergebnismenge werden alle Tupel von $\mathbf{r}(\mathbf{R})$ übernommen, für die dieser Ausdruck zu wahr evaluiert wird.

Definition 2.1: Selektionsoperator
Sei $\mathbf{R}$ *eine Relation,* B *ein boolscher Ausdruck, der auf den Attributen von* $\mathbf{R}$ *definiert ist und dessen Grundbausteine Vergleiche der Attributwerte sind.*
Die Selektion *von Tupeln aus* $\mathbf{R}$ *hinsichtlich der Bedingung* B *ergibt sich zu*

$$\sigma_{\langle Bedingung \rangle}(\mathbf{R}) := \{t | t \in \mathbf{R}(\mathbf{r}) \wedge (B(t) = wahr)\}, \tag{2.1}$$

dabei bedeute $B(t)$ *den Wahrheitswert von* B *bei der Wertebelegung von* t*.*

Beispiel 2.2: *Gesucht werden alle Daten der Professoren in der Relation „Professoren“ (Tabelle 1.1, S. 11), die in der Fakultät für Mathematik und Informatik tätig sind.*

Lösung:
Die Auswahlbedingung lautet $Fakult\ddot{a}t =$ „Math.u.Informatik “.
Damit wird die Anfrage mit Hilfe der Selektion ausgedrückt:

$$\sigma_{Fakult\ddot{a}t=\text{„Math.u.Informatik“}}(\textbf{Professoren})$$

Als Ergebnis ergibt sich die Relation:

PNR	*Name*	*Arbeitsgebiet*	*Fakultät*
2364	Rahm	Datenbanken	Math.u.Informatik
0342	Beyer	Analysis	Math.u.Informatik

■

Natürlich kann das Ergebnis einer Selektion nur aus einem Tupel bestehen oder es kann der Fall eintreten, dass kein Tupel der Relation **R** das Auswahlkriterium erfüllt. Die Ergebnisrelation ist dann leer.

In dem Beispiel 2.2: ist die Bedingung von der Form „$Attribut = \text{Wert}$“, hier wird getestet, ob ein Attribut einen vorgegebenen Wert hat. Eine andere übliche Form ist „$Attribut_1 = Attribut_2$“. So führt die Frage „Welche Professoren heißen so wie ihr Arbeitsgebiet?“ auf die Bedingung „$Name = Arbeitsgebiet$“, in der die Werte zweier Attribute verglichen werden. Der boolsche Ausdruck im Selektionsoperator besteht aus einer Verknüpfung von mehreren Bedingungen der vorgestellten Formen durch die logischen Operatoren bzw. Verknüpfungen *Negation, Konjunktion und Disjunktion*:
Welche Professoren heißen nicht wie ihr Arbeitsgebiet und sind an der theologischen Fakultät?
$B = \neg(Name = Arbeitsgebiet) \wedge (Fakult\ddot{a}t = \text{„Theologie“})$ ([1]).

2.2 Der Projektionsoperator

Mit Hilfe der Projektion ist, bildlich gesprochen, es möglich, aus einer Tabelle einzelne Spalten auszuwählen und in die Ergebnis- oder in eine Zwischentabelle zu übernehmen. Die zu übernehmenden Spalten werden als Parameter des Projektionsoperators aufgeführt.

Definition 2.3: Projektionsoperator
Sei $\mathbf{r}$ *eine Relation zum Relationsschema* $\mathbf{R}(A_1, A_2, ..., A_k)$, *sei weiter* $\{B_1, B_2, ..., B_l\} \subseteq \{A_1, A_2, ..., A_k\}, l > 0$. *Die Projektion von* $\mathbf{R}$ *auf die Attribute* $\{B_1, B_2, ..., B_l\}$ *ist die Relation*

$$\pi_{B_1, B_2, ..., B_l}(\mathbf{R}) := \{(t.B_1, t.B_2, ..., t.B_l) | t \in \mathbf{R}\} \tag{2.2}$$

Die Projektion ist *nicht definiert*, wenn die in der Definition angegebene Voraussetzung $\{B_1, B_2, ..., B_l\} \subseteq \{A_1, A_2, ..., A_k\}$ nicht zutrifft oder $l = 0$ gilt.

Die Ergebnisrelation einer Projektion hat genau die Attribute $\{B_1, B_2, ..., B_l\}$, die im Projektionsoperator als Parameter angegeben wurden, damit bestimmt diese Para-

[1] Für Negation, Konjunktion und Disjunktion werden in Formeln die Symbole $\neg, \wedge$ und $\vee$ benutzt.

metermenge das Schema der Ergebnisrelation, ihr Grad ist folglich l. Die Bedingung $l > 0$ bedeutet, dass die Ergebnisrelation nach Anwendung eines Projektionsoperators mindestens ein Attribut hat.

Beispiel 2.4: *Wie lauten die Namen der Professoren in Tabelle 1.1?*

Lösung: $\pi_{Name}(\mathbf{Professoren})$

Ergebnis:

Name
Rahm
Beyer
Wartenberg
Cain

Das Ergebnis der Anwendung des Projektionsoperators auf eine Relation ist per definitionem wieder eine Relation, darf als keine Duplikate enthalten. Damit das Ergebnis der Anwendung des Projektionsoperators diese Eigenschaft hat, darf jedes Ergebnistupel nur einmal in die Ergebnisrelation aufgenommen werden (*Duplikatelimination*).

Beispiel 2.5: *Aus welchen Fakultäten kommen die Professoren der Tabelle 1.1 ?*

Lösung: $\pi_{Fakultät}(\mathbf{Professoren})$
Die Fakultät „Math.u.Informatik" erscheint nur einmal im Ergebnis:

Fakultät
Math.u.Informatik
Theologie
Geschichte

Der folgende Satz zeigt, dass der Name „Projektion" entsprechend mathematischen Begriffsbildung gewählt wurde.

Satz 2.6: *Sei* $\mathbf{R}$ *eine Relation mit der Menge* $\mathcal{A}$ *seiner Attribute und sei* $\mathcal{B} \subseteq \mathcal{A}$*. Der Projektionsoperator ist idempotent.*

$$\pi_{\mathcal{B}}(\pi_{\mathcal{B}}(\mathbf{R})) = \pi_{\mathcal{B}}(\mathbf{R}). \tag{2.3}$$

Beweis: Dem Beweis liegt die folgende Idee zu Grunde ([2]): Aus der Menge $\mathcal{A}$ aller Attribute von $\mathbf{R}$ werden die von $\mathcal{B}$ ausgewählt, anschließend aus dem Ergebnis der

[2] Entgegen der traditionellen Arbeitsweise der Mathematik werden Beweise nur soweit angegeben, wie die Details für das weitere Verständnis unumgänglich sind.

Auswahl nochmal die von $\mathcal{B}$, das ändert die Auswahl aber nicht.
Die Durchführung des Beweises folgt dem Schema, nach dem die meisten Beweise von Mengengleichheit verlaufen. Es sind die beiden Inklusionen
(a) $\pi_{\mathcal{B}}(\pi_{\mathcal{B}}(\mathbf{R})) \supseteq \pi_{\mathcal{B}}(\mathbf{R})$ und
(b) $\pi_{\mathcal{B}}(\pi_{\mathcal{B}}(\mathbf{R})) \subseteq \pi_{\mathcal{B}}(\mathbf{R})$
zu beweisen. Wir beschränken uns auf (a), um daran das typische Vorgehen zu zeigen:
Sei $x \in \pi_{\mathcal{B}}(\mathbf{R})$, dann existiert ein Tupel $q \in \mathbf{R}$ mit $x = \pi_{\mathcal{B}}(\mathbf{R})$. Für dieses q wird der Ausdruck der Ausdruck der linken Seite berechnet: $x = \pi_{\mathcal{B}}(\mathbf{q})$ und $x = \pi_{\mathcal{B}}(\mathbf{x})$, d.h. $x \in \pi_{\mathcal{B}}(\pi_{\mathcal{B}}(\mathbf{R}))$, was (a) beweist.
Der Beweis von (b) bringt keine grundsätzlich neuen Erkenntnisse. ■

2.3 Hintereinanderausführung von Operatoren

Entsprechend der Eigenschaft aller Operationen der Algebra, dass sie Abbildungen von $\mathcal{M}$ in $\mathcal{M}$ darstellen, sind sie unter Beachtung ihrer Definitionsbereiche miteinander kombinierbar. Die Kombination zweier Operatoren ist deren *Hintereinanderausführung*. So bedeutet (wenn $\mathbf{R}(A,B,C)$ eine Relation mit den Attributen A,B,C ist) $\pi_A(\pi_{A,B}(\mathbf{R}))$, dass zuerst eine Projektion auf die Attribute A und B erfolgt, anschließend wird auf das Zwischenresultat eine Projektion auf das Attribut A durchgeführt.
Diese Kombinationsmöglichkeiten verschiedener Operatoren legt die Frage nach Rechenregeln nahe, d.h. formalen Umformungsregeln für Kombinationen von Operatoren. Eine solche Regel wurde bei der Projektion schon angegeben (Formel in Satz 2.6:), weitere enthalten die folgenden Sätze.

Satz 2.7: Hintereinanderausführung von Projektionen
Sei $\mathbf{R}$ *eine Relation mit der Menge* $\mathcal{A}$ *ihrer Attribute,* $\mathcal{B} \subseteq \mathcal{A}$. *Dann gilt*

$$\pi_{\mathcal{B}}(\pi_{\mathcal{A}}(\mathbf{R})) = \pi_{\mathcal{B}}(\mathbf{R}) \tag{2.4}$$

Beweis: Der Beweis folgt unmittelbar aus der Definition. ■

Falls $\mathcal{B} \subseteq \mathcal{A}$ nicht gilt, enthält $\mathcal{B}$ ein Attribut, welches nicht in $\mathcal{A}$ ist. Dann ist $\pi_{\mathcal{A}}(\mathbf{R})$ nicht im Definitionsbereich von $\pi_{\mathcal{B}}$ (vgl. Seite 16).

Satz 2.8: Hintereinanderausführung von Selektionen
Sei $\mathbf{R}$ *eine Relation mit der Menge* $\mathcal{A}$ *ihrer Attribute,* $B_i(\mathcal{A}), i = 1,2$ *zwei Bedingungen auf den Attributen von* $\mathbf{R}$. *Dann gilt*
- *die Äquivalenz zur konjunktiven Verknüpfung der Bedingungen*

$$\sigma_{B_2}(\sigma_{B_1}(\mathbf{R}) = \sigma_{B_1 \wedge B_2}(\mathbf{R}) \tag{2.5}$$

- *die Projektionseigenschaft (Idempotenz):*

$$\sigma_{B_1}(\sigma_{B_1}(\mathbf{R}) = \sigma_{B_1}(\mathbf{R}) \qquad (2.6)$$

- *die Kommutativität*

$$\sigma_{B_1}(\sigma_{B_2}(\mathbf{R}) = \sigma_{B_2}(\sigma_{B_1}(\mathbf{R}) \qquad (2.7)$$

Beweis:
Beweisidee zu Gleichung 2.5 : Werden aus $\mathbf{r}(\mathbf{R})$ Tupel ausgewählt, die die Bedingung B_1 erfüllt, und dann aus diesen noch die, die auch B_2 erfüllen, so erüllen die Tupel dieser Ergebnismenge B_1 und B_2, also $B_1 \wedge B_2$. Mit dieser Erkenntnis folgt nun der Beweis unmittelbar aus der Definition.
Zu Gleichung 2.6: Werden aus einer Menge von Tupeln Tupel nach einem gewissen Kriterium ausgewählt und aus dem Ergebnis nochmal nach dem gleichen Kriterium ausgewählt, ändert sich an der ersten Auswahl nichts.
Ein formaler Beweis ergibt sich sofort aus Gleichung 2.5, indem dort $B_1 = B_2 = B$ gesetzt wird: $\sigma_B(\sigma_B(\mathbf{R}) = \sigma_{B\wedge B}(\mathbf{R}) = \sigma_B(\mathbf{R})$.
Zu Gleichung 2.7: Diese Gleichung ist eine unmittelbare Folge von 2.5 und der Kommutativität der Konjunktion. ■

Nach Gleichung 2.5 kann jede mehrfache Selektion in eine einzelne Selektion umgeformt werden. Dies ist bei Implementierungen vorteilhaft, da es das mehrfache Lesen einer Relation vermeidet.

Die Projektionen und Selektionen können im Sinne der Hintereinanderausführung wie alle Operatoren unter Beachtung der Definitionsgebiete verschachtelt genutzt werden.

Beispiel 2.9: *Welche Namen haben die Professoren der Theologie in Tabelle 1.1 (S. 11)?*

Lösung: $\pi_{Name}(\sigma_{Fakultät=\text{„Theologie“}}(\mathbf{Professoren}))$
Das Ergebnis ist eine einspaltige Relation mit dem Attribut „Name“ und dem Inhalt „Wartenberg“ . ■

Wird versucht, auf der Relation **Professoren** den Ausdruck
$\sigma_{Fakultät=\text{„Theologie“}}(\pi_{Name}(\mathbf{Professoren}))$
auszuwerten, ist die Selektion nicht definiert, da es in dem Zwischenresultat nach Ausführung der Projektion kein Attribut „Fakultät“ gibt.
Mit anderen Worten:

Satz 2.10: *Sei* **R** *eine Relation, $\mathcal{A}$ eine Teilmenge der Attribute des zu* **R** *gehörigen Relationsschemas, B eine Bedingung auf den Attributen von* **R**. *Die Operationen Selektion und Projektion sind nicht vertauschbar:* $\pi_{\mathcal{A}}(\sigma_B(\mathbf{R})) \neq \sigma_B(\pi_{\mathcal{A}}(\mathbf{R}))$.

Sollen diese Operationen vertauscht werden, ist die Korrektheit der Vertauschung im Einzelfall nachzuweisen, wenn z.B. eine Projektion auf eine Teilmenge $\mathcal{B}$ aller Attribute erfolgt und die Bedingung auch nur auf $\mathcal{B}$ Bezug nimmt.

2.4 Der Zuweisungsoperator

Der Zuweisungsoperator soll dazu dienen, Relationen oder Attribute umzubenennen. Er wird für eine mathematische Beschreibung der Algebra nicht zwingend benötigt, ist eher als technisches Hilfsmittel einzuordnen. Mit seiner Hilfe können Zwischenergebnisse aus Ausdrücken wiederverwendet werden. Mit

$$\mathbf{Q} \Leftarrow \sigma_{Personal\text{-}Nr=4714}(\mathbf{Professoren})$$

wird das Ergebnis der Selektion in die Relation **Q** übernommen. Das zugehörige Relationsschema wird dabei aus dem Ausdruck implizit bestimmt. Auf **Q** kann nun bezug genommen werden, z.B. mit: $\pi_{Name}(\mathbf{Q})$.
Nach diesem einführenden Beispiel zur Anwendung die Definition. Diese bestimmt zwei Verwendungsmuster des Zuweisungsoperators
(a) zur Umbenennung der Relation und der Attribute
(b) zur Fixierung von Zwischenergebnissen.

Definition 2.11: *Seien* $\mathbf{A}(A_1, A_2, ..., A_l)$ *ein Relationsschema,* **A** *die entsprechende Relation vom Grade l.*
(a) Dann definiert:

$$\mathbf{Q} \Leftarrow \mathbf{A} \tag{2.8}$$

eine Relation **Q** *mit dem Relationsschema* $\mathbf{Q}(A_1, A_2, ..., A_3)$. **Q** *enthält genau die Tupel, die* **A** *enthält.*
(b) Sei $(B_1, B_2, ..., B_l)$ *eine Liste von l Attributnamen. Dann definiert*

$$\mathbf{Q}_{(B_1,B_2,...,B_l)} \Leftarrow \mathbf{A} \tag{2.9}$$

eine Relation **Q** *mit dem Relationsschema* $\mathbf{Q}(B_1, B_2, ..., B_l)$. **Q** *enthält genau die Tupel, die* **A** *enthält.*
In jedem Fall bleibt die ursprüngliche Relation erhalten.

Bemerkung zur Definition: Mit dem Zuweisungsoperator kann keine Projektion vorgenommen werden, d.h. die Liste $(B_1, B_2, ..., B_l)$ muss genau so viele Elemente enthalten wie der Grad von **A** angibt. Andernfalls ist dieser Operator nicht definiert. Durch die Anwendung des Zuweisungsoperator wie in Gleichung 2.8 werden alle Attributnamen geändert $B_j \leftarrow A_j$, $j \in [1, l]$.

2.5 Der Einbettungsoperator

Der Einbettungoperator ordnet einem Wert x aus dem Wertevorrat eines Attributs A eine Relation vom Grade 1 zu, die genau ein Tupel hat, bei dem der Wert des Attributs gerade die Zahl x ist. Er ist kein Operator der Algebra, da sein Definitionsgebiet nicht in der Menge der Relationen liegt. Er wird u.a. bei der Definition der äußeren Verbünde (siehe Abschnitt 3.2.4, Seite 35) benötigt.

Definition 2.12: Einbettungsoperator
Sei x ein Wert aus dem Wertevorrat dom(A) eines Attributs A. Dann bezeichnet

$$\mathbf{R} \Leftarrow \chi_A(x) \tag{2.10}$$

eine Relation mit dem Relationsschema $\mathbf{R}(A)$ *und genau einem Tupel* t*, welches die Eigenschaft* $t.A = x$ *hat.*
Der **zu** χ_A **inverse Operator***, der der einstelligen Relation* $\mathbf{r}$ *vom Grade eins mit dem Relationsschema* $\mathbf{R}(A)$ *den Wert* $t.A, t \in dom(A)$ *für* $t \in \mathbf{r}$ *zuordnet, wird mit* ω *bezeichnet.*

Auch ω ist kein Element der Algebra, da hier der Wertebereich der Abbildung aus der Algebra herausführt.

In den Abschnitten 3.1.3 (S. 28) und 4.2.2 (S. 43) wird die Verwendung der Operatoren χ und ω vorgestellt.

Der Einbettungsoperator erscheint exotisch, jedoch hat er - fast unbemerkt - implizit Einzug in die Datenbankabfragesprache SQL gehalten, nämlich in Verbindung mit den Aggregatfunktionen avg, max, min , die in eine SELECT-Anweisung gekapselt werden und so ggf. den Zahlenwert der mathematischen Berechnung in einer Relation darstellen.

Beispiel 2.13: *Gegeben sei eine Relation* **Pers***(...,Lebensalter,....), in der u.a. für Personen das Lebensalter enthalten ist. Die folgende SQL-Anfrage berechnet das Durchschnittsalter:*

SELECT avg(*Lebensalter*) AS Durchschnittsalter FROM **Pers**

und erzeugt aus dem Ergebniswert eine Relation:

Durchschnittsalter
<*Wert*>

Aber auch der Operator ω *lässt sich nachweisen, indem diese SQL-Anfrage als Subquery benutzt wird:*

```
SELECT *
FROM Pers
WHERE Lebensalter > (SELECT avg(Lebensalter) AS Durchschnittsalter
                     FROM Pers
                     )
```

In der WHERE-*KLausel wird das Ergebnis des inneren Statements, da es nur aus einer Zeile mit einem numerischer Wert besteht, implizit in die dem Wert entsprechende Zahl konvertiert und steht so für den Vergleich zur Verfügung.*

■

3 Zweistellige Operatoren

3.1 Verknüpfungen aus der Mengenlehre

Die Menge $\mathcal{M}$ aus Definition 1.9: ist eine Menge von Relationen, d.h. eine Menge von Tupelmengen. Damit ist es logisch, Verknüpfungen in der Relationenalgebra zu definieren, die auf den Verknüpfungen von Tupelmengen beruhen, also u.a. auf deren Vereinigung, Durchschnitt und Differenzbildung. Weitere Verküpfungen basieren auf dem Mengenprodukt zweier Relationen, sie konstruieren aus diesem neue Relationen. Wie schon bei den einstelligen Operatoren ist auch das Ergebnis einer Verknüfung zweier Relationen wieder eine Relation, damit können sie, ggf. unter Beachtung ihrer Definitionsbereiche oder anderer in ihrer Definition gegebener Bedingungen, miteinander oder mit einstelligen Operatoren geschachtelt genutzt werden (Hintereinanderausführung).

3.1.1 Verknüpfungen der Mengentheorie: Durchschnitt, Vereinigung, Differenz

In der mathematischen Mengenlehre ist der Durchschnitt zweier Mengen $\mathcal{A}$,$\mathcal{B}$ definiert als die Menge aller Elemente, die sowohl in $\mathcal{A}$ als auch in $\mathcal{B}$ sind. Im konkreten Fall sind diese Mengen jetzt Relationen zu jeweils einem Relationsschema, die Elemente Tupel. Was heißt also, dass ein Tupel t die Eigenschaften $t \in \mathbf{r}$ und $t \in \mathbf{s}$ hat? Die Veranschaulichung der Relation als Tabelle zeigt die Lösung auf. Was soll der Durchschnitt zweier Tabellen sein? Eine sinnvolle Antwort auf diese Frage lautet: Es soll der Durchschnitt der Tabelleninhalte konstruiert werden, genauer der Einheiten, die dort dargestellt werden. Die Werte in einer Zeile entsprechen gerade diesen Einheiten. Um zu entscheiden, ob eine bestimmte Zeile aus einer Tabelle auch in einer anderen vorkommt, ist eine sinnvolle Voraussetzung zu fordern, dass die Tabellen dieselbe Struktur haben, dass sie in der Anzahl und der Art ihrer Spalten übereinstimmen. Diese Struktur wird gerade durch die Relationsschemata beschrieben.

Definition 3.1: Vereinigungskompatibilität
Zwei Relationen $\mathbf{r}(\mathbf{R})$ *und* $\mathbf{s}(\mathbf{S})$ *heißen* vereinigungskompatibel, *wenn ihre Relationsschemata* $\mathbf{R}(...)$ *und* $\mathbf{S}(...)$ *äquivalent* ([1]) *sind.*
Symbolisch wird dies durch

$$\mathbf{r}(\mathbf{R}) \cong \mathbf{s}(\mathbf{S})$$

ausgedrückt.

[1] siehe Abschnitt 1.6:, Seite 10

Vereinigungskompatible Relationsschemata stimmen in der Zahl und der Art der Attribute überein, sie unterscheiden sich höchstens wegen einer Permutation der Anordnung ihrer Attribute. Die Art der Attribute wird durch ihren Wertebereich und ihre Bedeutung bestimmt. Damit bedarf auch die Bestimmung der Vereinigungskompatibilität dieser semantischen Metainformationen. Namensunterschiede sind wegen der semantischen Äquivalenz der Attribute nicht wirklich relevant.

Satz 3.2: *Die Vereinigungskompatibilität hat die drei Eigenschaften:*
(1) $\mathbf{r}(\mathbf{R}) \cong \mathbf{r}(\mathbf{R})$
(2) $\mathbf{r}(\mathbf{R}) \cong \mathbf{s}(\mathbf{S}) \Leftrightarrow \mathbf{s}(\mathbf{S}) \cong \mathbf{r}(\mathbf{R})$
(3) *Aus* $\mathbf{r}(\mathbf{R}) \cong \mathbf{s}(\mathbf{S})$ *und* $\mathbf{s}(\mathbf{S}) \cong \mathbf{u}(\mathbf{U})$ *folgt* $\mathbf{r}(\mathbf{R}) \cong \mathbf{u}(\mathbf{U})$.

Der Beweis folgt unmittelbar aus der Definition des Äquivalenz für Relationsschemata. ■

Die Verknüpfungen *Vereinigung, Durchschnitt, Differenz* zweier Relationen werden nur unter der Voraussetzung der Vereinigungskompatibilität der Relationen definiert.

Definition 3.3: *Seien* **R** *und* **S** *vereinigungskompatible Relationen.*
Vereinigung:

$$\mathbf{R} \cup \mathbf{S} := \{t | t \in \mathbf{R} \vee t \in \mathbf{S}\} \tag{3.1}$$

Durchschnitt:

$$\mathbf{R} \cap \mathbf{S} := \{t | t \in \mathbf{R} \wedge t \in \mathbf{S}\} \tag{3.2}$$

Differenz:

$$\mathbf{R} \setminus \mathbf{S} := \{t | t \in \mathbf{R} \vee t \notin \mathbf{S}\} \tag{3.3}$$

Zur Illustration wird die Relation **Professoren** aus Beispiel 1.5: betrachtet

Professoren$(Personal\text{-}Nr, Name, Arbeitsgebiet, Fakultät)$

mit der Belegung:

Professoren

PNR	Name	Arbeitsgebiet	Fakultät
2364	Rahm	Datenbanken	Math.u.Informatik
0342	Beyer	Analysis	Math.u.Informatik
9645	Wartenberg	Kirchengeschichte	Theologie
4714	Cain	Klass. Archäologie	Geschichte

Tabelle „Professoren“nach Beipiel 1.1

und nach Beipiel 2.2: wird **Q** gebildet:

$\mathbf{Q} \Leftarrow \sigma_{Fakultät=\text{„Math.u.Informatik“}}(\mathbf{Professoren})$
mit der Belegung:

Q

PNR	Name	Arbeitsgebiet	Fakultät
2364	Rahm	Datenbanken	Math.u.Informatik
0342	Beyer	Analysis	Math.u.Informatik

Tabelle „Professoren der Fakultät für Mathematik und Informatik“

Die Veinigungskompatibilität ist offensichtlich, folglich kann die Mengendifferenz $\mathbf{PQ} \Leftarrow \mathbf{Professoren} \setminus \mathbf{Q}$ gebildet werden:

PQ

PNR	Name	Arbeitsgebiet	Fakultät
9645	Wartenberg	Kirchengeschichte	Theologie
4714	Cain	Klass. Archäologie	Geschichte

Als weiteres Beispiel dienen die folgenden Relationsschemata ([2]);

Biertrinken:
Gast$(Person, Kneipe)$ ([3])
Vorlieben$(Person, Biersorte)$
Angebot$(Kneipe, Biersorte)$

mit der Semantik ([4]):
Gast: Welche Person besucht welche Gaststätte (Kneipe);
Vorlieben: Welche Person trinkt gern welche Biersorten;
Angebot: Welche Gaststätte hat welche Biersorten im Angebot.

Beispiel 3.4: *Mit Hilfe des Bierschema sollen folgende Fragen beantwortet werden:*
(1) Welche Biersorten spielen in der Miniwelt des Biertrinkens eine Rolle?
(2) Welche angebotenen Biersorten gehören zu den Vorlieben der Personen?
(3) Welche der angebotenen Biersorten gehört nicht zur den Vorlieben der Personen?

[2] Das Beispiel wurde den Autoren von Studenten ohne Quellenangabe genannt. Falls ein Leser die Quelle kennt, wird eine Nachricht erbeten, damit sie korrekt zitiert werden kann.

[3] Das Attribut *Kneipe* bezeichne im Folgenden völlig wertfrei eine Gaststätte, in der Bier angeboten wird.

[4] Das Modell erscheint merkwürdig, der Grund wird klar, wenn versucht wird, es als E/R-Entwurf zu notieren. Entitätsmengen für „Personen“, „Biere“ und „Kneipen“ fehlen.

Lösung: Offensichtlich müssen die Biersorten aus **Vorlieben** und **Angebot** zusammengefasst werden. Dabei ist zu beachten, dass die beiden Relationen a priori nicht vereinigungskompatibel sind. Durch eine Projektion beider auf das Attribut *Biersorte* wird die Zusammenfassung durch eine Vereinigungs, Durchschnitts- bzw. Differenzbildung ermöglicht.

(1) $\pi_{Biersorte}(\mathbf{Angebot}) \cup \pi_{Biersorte}(\mathbf{Vorlieben})$
(2) $\pi_{Biersorte}(\mathbf{Angebot}) \cap \pi_{Biersorte}(\mathbf{Vorlieben})$
(3) $\pi_{Biersorte}(\mathbf{Angebot}) \setminus \pi_{Biersorte}(\mathbf{Vorlieben})$ ■

Die im Beispiel genutzte Idee, nicht vereinigungsverträgliche Mengen durch Projektionen kompatibel zu machen, kann häufig genutzt werden.

3.1.2 Sätze über die Grundverknüpfungen, die aus der Mengenlehre abgeleitet wurden.

An Hand der Definitionen 3.3:...3.3: ergibt sich, dass die Ergebnisse jeweils vereinigungskompatibel zu den Ausgangsrelationen sind, dies ergibt die Möglichkeit wiederholter Anwendung dieser Operatoren.

Für die Verknüpfungen Vereinigung, Durchschnitt und Relationendifferenz gelten im wesentlichen dieselben Aussagen wie für die gleichnamigen Operationen der Mengenlehre. Die zugehörigen Beweise können nahezu wörtlich übernommen werden. Das leuchtet sofort ein, sind doch die definierenden Gleichungen 3.1 bis 3.3 dieselben wie in der Mengenlehre, es ist ist zusätzlich nur die einschränkende Voraussetzung der Vereinigungskompatibilität hinzugekommen. Satz 3.2: sichert, dass diese Bedingung keine Probleme bereitet.

Satz 3.5: *Seien* **R**, **S** *und* **T** *vereinigungskompatible Relationen. Dann gelten:*
(a) die Kommutativität des Vereinigungsoperators:

$$\mathbf{R} \cup \mathbf{S} = \mathbf{S} \cup \mathbf{R} \tag{3.4}$$

(b) die Kommutativität des Durchschnittsoperators:

$$\mathbf{R} \cap \mathbf{S} = \mathbf{S} \cap \mathbf{R} \tag{3.5}$$

(c) die Assoziativgesetze für Vereinigung und Durchschnitt:

$$(\mathbf{R} \cup \mathbf{S}) \cup \mathbf{T} = \mathbf{S} \cup (\mathbf{R} \cup \mathbf{T}) \tag{3.6}$$

$$(\mathbf{R} \cap \mathbf{S}) \cap \mathbf{T} = \mathbf{S} \cap (\mathbf{R} \cap \mathbf{T}) \tag{3.7}$$

(d) die beiden Distributivgesetze:

$$(\mathbf{R} \cup (\mathbf{S} \cap \mathbf{T}) = (\mathbf{R} \cup \mathbf{S}) \cap (\mathbf{R} \cup \mathbf{T}) \tag{3.8}$$

$$(\mathbf{R} \cap (\mathbf{S} \cup \mathbf{T}) = (\mathbf{R} \cap \mathbf{S}) \cup (\mathbf{R} \cap \mathbf{T}) \tag{3.9}$$

(e) für die Kombination von Differenz mit Vereinigung

$$\mathbf{R} \setminus (\mathbf{S} \cup \mathbf{T}) = (\mathbf{R} \setminus \mathbf{S}) \cap (\mathbf{R} \setminus \mathbf{T}) = (\mathbf{R} \setminus \mathbf{S}) \setminus \mathbf{T} \tag{3.10}$$

(f) für die Kombination von Differenz mit Durchschnitt

$$\mathbf{R} \setminus (\mathbf{S} \cap \mathbf{T}) = (\mathbf{R} \setminus \mathbf{S}) \cup (\mathbf{R} \setminus \mathbf{T}) \tag{3.11}$$

Der Beweis folgt unmittelbar aus den Definitionen. ■

Die Liste der Beziehungen zwischen Vereinigung, Durchschnitt und Differenz ist mit den Beispielen keines ausgeschöpft, siehe [3] bzw. andere Lehrbücher der Mengenlehre.

Die Assoziativgesetze ermöglichen die Anwendung der vereinfachten Schreibweise $\mathbf{R} \cup \mathbf{S} \cup \mathbf{T}$ bzw. $\mathbf{R} \cap \mathbf{S} \cap \mathbf{T}$, die ohne Klammern auskommt. Wie auch in der Mengenlehre, ist in der Relationenalgebra die Differenzbildung nicht kommutativ, d.h. allgemein gilt

$$\mathbf{R} \setminus \mathbf{S} \neq \mathbf{S} \setminus \mathbf{R}.$$

Satz 3.6: *Die Verknüpfungen Vereinigung, Durchschnitt und Differenz vereinigungskompatibler Relationen sind nicht unabhängig voneinander, es gilt die Beziehung*

$$\mathbf{R} \cap \mathbf{S} = (\mathbf{R} \cup \mathbf{S}) \setminus ((\mathbf{R} \setminus \mathbf{S}) \cup (\mathbf{S} \setminus \mathbf{R})) \tag{3.12}$$

m.a.W.: der Durchschnitt zweier vereinigungskompatibler Relationen kann durch Vereinigung und Differenz der Relationen dargestellt werden.

Für die Hintereinanderausführung von Projektion / Selektion und den Verknüpfungen Vereinigung, Durchschnitt und Differenz können weitere Rechenregeln angegeben werden.

Satz 3.7: *Seien* $\mathbf{R}$ *und* $\mathbf{S}$ *vereinigungskompatible Relationen.*
$\mathcal{X}$ *sei eine Teilmenge der Menge* $\mathcal{A}$ *aller Attribute.*

$$\pi_{\mathcal{X}}(\mathbf{R} \cap \mathbf{S}) \subseteq \pi_{\mathcal{X}}(\mathbf{R}) \cap \pi_{\mathcal{X}}(\mathbf{S}) \tag{3.13}$$

$$\pi_{\mathcal{X}}(\mathbf{R} \cup \mathbf{S}) = \pi_{\mathcal{X}}(\mathbf{R}) \cup \pi_{\mathcal{X}}(\mathbf{S}) \tag{3.14}$$

$$\pi_{\mathcal{X}}(\mathbf{R} \setminus \mathbf{S}) \supseteq \pi_{\mathcal{X}}(\mathbf{R}) \setminus \pi_{\mathcal{X}}(\mathbf{S}) \tag{3.15}$$

Sei $\mathbf{B}$ *eine Bedingung auf* $\mathcal{A}$.

$$\sigma_{\mathbf{B}}(\mathbf{R} \cap \mathbf{S}) = \sigma_{\mathbf{B}}(\mathbf{R}) \cap \sigma_{\mathbf{B}}(\mathbf{S}) \tag{3.16}$$

$$\sigma_{\mathbf{B}}(\mathbf{R} \cup \mathbf{S}) = \sigma_{\mathbf{B}}(\mathbf{R}) \cup \sigma_{\mathbf{B}}(\mathbf{S}) \tag{3.17}$$

$$\sigma_{\mathbf{B}}(\mathbf{R} \setminus \mathbf{S}) = (\sigma_{\mathbf{B}}\mathbf{R}) \setminus \mathbf{S} \tag{3.18}$$

Der Beweis erfolgt durch Anwendung der Definitionen.
Dabei ergibt sich, dass in den Inklusionen 3.13 und 3.15 die Gleichheit im allgemeinen nicht zutrifft. Dies wird auch durch ein Beispiel deutlich.

Beispiel 3.8: *In den Formeln 3.13 und 3.15 tritt eine echte Inklusion auf, wenn die beiden Relationen*

$$\mathbf{R} = \{(2,3)\} \textit{ und } \quad \mathbf{S} = \{(1,3)\}$$

vorliegen.
Sei $\mathcal{X}$ das zweite Attribut, so ist die linke Seite in 3.13 leer, während die rechte Seite ein Tupel, bestehend aus dem Attributwert 3, enthält:

$$\pi_{\mathcal{X}}(\mathbf{R} \cap \mathbf{S}) = \emptyset \subset \{(3)\} = \pi_{\mathcal{X}}(\mathbf{R}) \cap \pi_{\mathcal{X}}(\mathbf{S}).$$

In 3.15 ist $\mathbf{R} \setminus \mathbf{S} = \mathbf{R}$, also

$$\pi_{\mathcal{X}}(\mathbf{R} \setminus \mathbf{S}) = \{(3)\} \supset \emptyset = \{(3)\} \setminus \{(3)\} = \pi_{\mathcal{X}}(\mathbf{R}) \setminus \pi_{\mathcal{X}}(\mathbf{S})$$ ■

Es existiert eine Vielzahl weiterer Regeln, die Teilmengenbeziehungen nachweisen wie schon 3.13 und 3.15. Sie können bei der Anfrageoptimierung in verteilten Datenbanken eine Rolle spielen. Im allgemeinen werden jedoch Anfrageumformungen betrachtet, bei denen die Ergebnismenge nicht verändert wird.

3.1.3 Modifiziertes Kreuzprodukt

In der Mathematik wird die folgende Definition des Kreuzprodukts, auch äußeres Produkt genannt, gegeben:

Definition 3.9: *Seien $\mathcal{A}, \mathcal{B}$ zwei Mengen. Die Menge*

$$\mathcal{A} \times \mathcal{B} := \{x | x = (s,t), \quad s \in \mathcal{A}, t \in \mathcal{B}\}. \tag{3.19}$$

heißt **Kreuzprodukt** *oder* **äußeres Produkt** *von $\mathcal{A}$ und $\mathcal{B}$.*

Wenn diese Definition unmodifiziert auf Relationen angewendet wird, ist das Ergebnis einer Kreuzproduktbildung keine Relation. Die Elemente der Ergebnismenge sind Paare von Tupeln, die ihrerseits zusammengesetzt sind. Wegen der geforderten Atomarität der Attribute (vgl. S. 8) sind sie also keine Elemente einer Relation, jede Komponente eines Paare ist in der Sprache der Relation ein nichtatomares, zusammengesetztes Attribut: Seien $\mathbf{A}(A_1, A_2, ..., A_k)$ und $\mathbf{B}(B_1, B_2, ..., B_l)$ zwei Relationsschemata. Dann hat das Kreuzprodukt der zugehörigen Relationen $\mathbf{A}, \mathbf{B}$ die formale Struktur

$$\mathbf{A} \times \mathbf{B} = \{\tilde{x} | \tilde{x} = ((s.A_1, s.A_2, ..., s.A_k), (t.B_1, t.B_2, ..., t.B_l)), \quad s \in \mathbf{A}, t \in \mathbf{B}\}.$$

Die Verknüpfung „Kreuzproduktbildung“ führt also aus der Relationenalgebra heraus. Andererseits lässt sich dem Term $((s.A_1, s.A_2, ..., s.A_k), (t.B_1, t.B_2, ..., t.B_l))$ eindeutig das Tupel $(s.A_1, s.A_2, ..., s.A_k, t.B_1, t.B_2, ..., t.B_l)$ zuordnen, d.h. jedem Element des Kreuzprodukts entspricht eindeutig ein Tupel. Die Abbildung ist sogar umkehrbar, wenn die Attribute von $\mathbf{A}$ und $\mathbf{B}$ bekannt sind. Damit kann für Relationen ein *modifiziertes Kreuzprodukt* definiert werden, welches als Ergebnis gerade die zugeordneten Tupel verwendet, also eine Relation bildet ([5]). Da es im Allgemeinen nicht mit dem Kreuzprodukt für Mengen verwechselt werden kann, wird die gleiche Bezeichnung „Kreuzprodukt“ und das gleiche Symbol $\times$ verwendet.

Definition 3.10: *Seien* **R** *und* **S** *zwei Relationen mit den Relationsschemata* $\mathbf{R}(A_1, A_2, ..., A_k)$ *und* $\mathbf{S}(B_1, B_2, ..., B_l)$.
Die Namen der Attribute von **R** *seien verschieden von den Namen von* **S**: $A_i \neq B_j, i \in [1,k], j \in [1,l]$
Dann heißt

$$\begin{aligned}\mathbf{A} \times \mathbf{B} := \{x | x = (s.A_1, s.A_2, ..., s.A_k, t.B_1, t.B_2, ..., t.B_l),\\ (s.A_1, s.A_2, ..., s.A_k) \in \mathbf{A}, (t.B_1, t.B_2, ..., t.B_l) \in \mathbf{B}\}\end{aligned} \tag{3.20}$$

das **Kreuzprodukt** *der Relationen* **A** *und* **B**.

Zum Kreuzprodukt $(\mathbf{A} \times \mathbf{B})$ gehört das Relationsschema
$(\mathbf{A} \times \mathbf{B})(A_1, A_2, ..., A_k, B_1, B_2, ..., B_l)$.

Die Kardinalität von $(\mathbf{A} \times \mathbf{B})$ ist $m \times n$, wenn m die Kardinalität von **A** und n die von **B** sind.

Leere Mengen und das Kreuzprodukt
Falls eine in das Kreuzprodukt eingehenden Relationen (z.B. **B**) leer also ist (also $n = 0$), können **keine Tupel nach Gl. 3.20 gebildet** werden, da es keine Komponenten $(B_1, B_2, ..., B_l)$ gibt. Somit gilt

$$\mathbf{A} \times \mathbf{B} = \emptyset \text{ (leere Menge)} \tag{3.21}$$

und die Formel für die Kardinalität bleibt auch in diesem Sonderfall gültig.

Ausführbarkeit der Kreuzproduktbildung:
Die in der Definition gestellte Bedingung an die Attributnamen ist insofern technischer Art, als sie jederzeit durch Umbenennen der Attribute (vgl. Abschnitt 2.8 , Seite 20 - Zuweisungsoperator) erfüllt werden kann. Damit ist die Kreuzproduktbildung **uneingeschränkt ausführbar**.

Beispiel 3.11: *Gegeben seien die beiden Relationen* **Auto** *und* **Farbe**

Auto

Typ
Touran
Sharan
Golf
Käfer

Farbe

Farbton
rot
blau
silber

[5] Es sei ausdrücklich darauf verwiesen, dass dieses Vorgehen bei der Definition des Kreuzprodukts nicht mit der Transformation zusammengesetzter Attribute des E/R-Modells verwechselt werden darf ([7], Regel 5). Dort wird stets aus einem zusammengesetzten Attribut eine eigene Relation!

mit den aufgeführten Fahrzeugtypen und den möglichen Fahrzeugfarben. Welche Kombinationen von Typ und Farbton sind möglich ?

Lösung: Das Kreuzprodukt $\mathbf{Q} \Leftarrow \mathbf{Auto} \times \mathbf{Farbe}$ hat das Relationsschema $\mathbf{Q}(Typ, Farbton)$. Die zugehörige Relation enthält die 12 möglichen Kombinationen der Tupel der Ausgangsrelationen. ■

Auch für das Kreuzprodukt bzw für Formeln, die Kreuzprodukt und Vereinigung (Durchschnitt, Differenz) enthalten, lassen sich eine Reihe von Rechenregeln angeben.

Satz 3.12: Kreuzprodukt und andere Verknüpfungen
Seien **R**, **S**, **T** *und* **U** *Relationen,* **R** *mit* **S** *bzw.* **T** *mit* **U** *vereinigungskompatibel.*
Dann gilt:

$$\mathbf{S} \times \mathbf{T} \cong \mathbf{T} \times \mathbf{S} \tag{3.22}$$

$$(\mathbf{R} \cup \mathbf{S}) \times \mathbf{T} = (\mathbf{R} \times \mathbf{T}) \cup (\mathbf{S} \times \mathbf{T}) \tag{3.23}$$

$$(\mathbf{R} \cap \mathbf{S}) \times \mathbf{T} = (\mathbf{R} \times \mathbf{T}) \cap (\mathbf{S} \times \mathbf{T}) \tag{3.24}$$

$$(\mathbf{R} \setminus \mathbf{S}) \times \mathbf{T} = (\mathbf{R} \times \mathbf{T}) \setminus (\mathbf{S} \times \mathbf{T}) \tag{3.25}$$

$$(\mathbf{R} \times \mathbf{T}) \cap (\mathbf{S} \times \mathbf{U}) = (\mathbf{R} \cap \mathbf{S}) \times (\mathbf{T} \cap \mathbf{U}) \tag{3.26}$$

Achtung, die folgende Beziehung liefert nur eine Inklusion:

$$(\mathbf{R} \times \mathbf{T}) \cup (\mathbf{S} \times \mathbf{U}) \subseteq (\mathbf{R} \cup \mathbf{S}) \times (\mathbf{T} \cup \mathbf{U}) \tag{3.27}$$

Auch hier folgen die Beweise unmittelbar aus der Definition.
Weitere Beziehungen finden sich in der schon zitierten Quelle [3].

Beispiel 3.13: *Gegeben seien die beiden Relationen* **Auto** *und* **Farbe***, jetzt haben beide Relationen ein weiteres Attribut Kennzeichen, abgekürzt als K mit der üblichen Bedeutung:*
$\mathbf{Auto}(K, Typ)$ *bzw.* $\mathbf{Farbe}(K, Farbton)$ *mit den Relationen:*

Auto

K	*Typ*
L-XX 0002	*Touran*
L-YZ 2005	*Sharan*
L-FJ 0815	*Golf*
C-AA 4711	*Käfer*

Farbe

K	*Farbton*
L-YZ 2005	*rot*
L-FJ 0815	*blau*
C-AA 4711	*rot*

Welche Farbe hat der Golf ?

Lösung: Zunächst wird eine Zwischenrelation erzeugt, die mit Hilfe des Kennzeichen als gemeinsamen Attribut beider Relationen die Informationen zusammenfügt, die zu einem Fahrzeug gehören:
$\mathbf{Q} \Leftarrow \pi_{K,Typ,Farbton}(\sigma_{\mathbf{Auto}.K=\mathbf{Farbe}.K}(\mathbf{Auto} \times \mathbf{Farbe}))$
Auf dieser Zwischenrelation erfolgt jetzt die entgültige Beantwortung:
$\pi_{Farbton}(\sigma_{Typ=\text{„Golf"}}(\mathbf{Q}))$. ■

Obwohl in der Relationenalgebra keine Relationen ohne Attribute existieren, können derartige Bildungen als Zwischenresultate entstehen. Deshalb ist die Definition des Kreuzprodukts auf diesen Fall sinnvoll zu erweitern.

Definition 3.14: Kreuzprodukt - Ergänzung. *Sei* **N** *ein (als Zwischenergebnis einer Umformung entstandenes)* **Relationsschema ohne Attribute***. Dann gilt für jede Relation* **A***:*

$$\mathbf{A} \times \mathbf{N} = \mathbf{N} \times \mathbf{A} = \mathbf{A}. \tag{3.28}$$

Von den Rechenregeln für das Kreuzprodukt in Verbindung mit Projektions- und Selektionsoperator gibt der folgende Satz einen Eindruck.

Satz 3.15: Kreuzprodukt und Selektion bzw. Projektion
(a) *Seien* **R** *und* **S** *zwei Relationen mit den Attributmengen* $\mathcal{R}$ *bzw.* $\mathcal{S}$*.* $\mathcal{X}$ *sei eine Teilmenge der Attribute von* $\mathbf{R} \times \mathbf{S}$*. Dann gilt:*

$$\pi_{\mathcal{X}}(\mathbf{R} \times \mathbf{S}) = \pi_{\mathcal{X} \cap \mathcal{R}}(\mathbf{R}) \times \pi_{\mathcal{X} \cap \mathcal{S}}(\mathbf{S}) \tag{3.29}$$

(b) *Sei* **B** *eine Bedingung über den Attributen von* $\mathbf{R} \times \mathbf{S}$*, welche sich darstellen läßt als konjunktive Verknüpfung zweier Bedingungen* $\mathbf{B_R}$ $\mathbf{B_S}$*, wobei* $\mathbf{B_R}$ *nur über den Attributen von* **R** *und* $\mathbf{B_S}$ *nur über denen von* **S** *definiert ist* $\mathbf{B} = \mathbf{B_R} \wedge \mathbf{B_S}$*. Dann gilt:*

$$\sigma_{\mathbf{B}}(\mathbf{R} \times \mathbf{S}) = \sigma_{\mathbf{B}_R}(\mathbf{R}) \times \sigma_{\mathbf{B}_S}(\mathbf{S}) \tag{3.30}$$

Beweis: ist eine freiwillige Übungsaufgabe! ■

Anmerkung zu (a): Die Ergänzung nach Gl. 3.28 zur Definition des Kreuzprodukts wird ggf. im eben gegebenen Satz 3.15: wirksam, wenn beispielsweise $\mathcal{X}$ keine Attribute von $\mathcal{S}$, d.h. nur Attribute aus $\mathcal{R}$, enthält, $\mathcal{X} \cap \mathcal{S}$ also leer ist. Formal liefert $\pi_{\mathcal{X} \cap \mathcal{S}}(\mathbf{S})$ dann ein Relationsschema ohne Attribute. Die Ergänzung sorgt nun dafür, dass Gleichung 3.29 auch für diesen Fall gilt, wie durch Ausrechnen von linker und rechter Seite bestätigt wird.

Das Kreuzprodukt wird selten allein verwendet, sondern meist in Kombination mit Selektion und Projektion wie in Beispiel 3.13:. Im folgenden Kapitel wird dies diskutiert.

3.2 Verbund-Verknüpfungen

Typisch für die Anwendung des Kreuzprodukts ist Hintereinanderausführung von Kreuzproduktbildung, Selektion und Projektion wie bei der Konstruktion der Zwischenrelation **Q** in Beispiel 3.13: (Seite 30). Da diese Kombination sehr häufig genutzt wird, wird sie als eigenständige Verknüpfung - die Verbund- oder Join-Operation - in die Algebra aufgenommen.

Die Join-Operatoren stellen das wichtigste Werkzeug bereit, um die Inhalte mehrerer Relationen zu verknüpfen. Sie gestatten, wieder bildlich und auf einen Sperialfall vereinfachend gesprochen, in einer Tabelle ein Attribut (Join-Attribut) und einen dort vorhandenen Attributwert zu nehmen, nach diesem Wert in einer anderen Tabelle in einer festgelegten Spalte zu suchen und die Zeilen, die zu dem Wert in beiden Tabellen gehören, in einer Ergebnistabelle als neue Zeile abzulegen. Dies wurde bei der Bildung von **Q** in Beispiel 3.13: durchgeführt. Dieses Vorgehen kann in verschiedener Hinsicht verallgemeinert werden, insbesondere bleibt der Vergleich nicht auf die Gleichheit beschränkt.

3.2.1 Der Gleichverbund (equi-join)

Definition 3.16: *Seien* $\mathbf{R}$ *und* $\mathbf{S}$ *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_k)$ *und* $\mathbf{S}(B_1, B_2, ..., B_l)$*. O.B.d.A seien die Attribute* $\mathbf{R}.A_k$ *und* $\mathbf{S}.B_1$ *vergleichbar, d.h. es kann entschieden werden, ob* $\mathbf{R}.A_k = \mathbf{S}.B_1$ *zutrifft. Dann definiert*

$$\mathbf{R} \bowtie_{\mathbf{R}.A_k=\mathbf{S}.B_1} \mathbf{S} := \pi_{A_1,A_2,...,A_k,B_2,...,B_l}(\sigma_{\mathbf{R}.A_k=\mathbf{S}.B_1}(\mathbf{R} \times \mathbf{S})) \tag{3.31}$$

den Gleichverbund der Relationen $\mathbf{R}$ *und* $\mathbf{S}$*.*

Diese erste Definition ist vorläufig und dient zum Verstehen.
Beim Gleichverbund wird definitionsgemäß immer die Gleichheit der Verbundattribute bewertet. Wenn keine Verwechslungen möglich sind, müssen nicht die qualifizierten Attributnamen verwendet werden. Die Definition könnte also auch
$\mathbf{R} \bowtie_{A_k=B_1} \mathbf{S} := \pi_{A_1,A_2,...,A_k,B_2,...,B_l}(\sigma_{A_k=B_1}(\mathbf{R} \times \mathbf{S}))$
lauten.

Beispiel 3.17: *Mit der Notation aus Formel 3.31 kann die Lösung des Beispiel 3.13: kurz und prägnant geschrieben werden:*
$\pi_{Farbton}(\sigma_{Typ=Golf}(\mathbf{Auto} \bowtie_{\mathbf{Auto}.K=\mathbf{Farbe}.K} \mathbf{Farbe})$ ■

Der Gleichverbund kann leicht auf mehrere Attribute ausgedehnt werden, es tritt dann in der Selektionsoperation eine konjunktive Verknüpfung der Bedingungen auf. Im Ergebnis werden alle doppelt auftretende Spalten entfernt. Dies führt zu der entgültigen Definition.

Definition 3.18: *Seien* $\mathbf{R}$ *und* $\mathbf{S}$ *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_h, A_{h+1}, ..., A_k)$ *und* $\mathbf{S}(B_1, B_2, ..., B_{k-h}, B_{k-h+1}, ..., B_l)$. *Die Attribute seien o.B.d.A. so umgeordnet, dass* A_{h+1} *und* B_1, A_{h+2} *und* B_2, ... , A_k *und* B_{k-h} *vergleichbar sind. Mit der Abkürzung*
$\mathbf{B} = (A_{h+1} = B_1) \wedge (A_{h+2} = B_2) \wedge ... \wedge (A_k = B_{k-h})$
ist der **Gleichverbund** *auf der Bedingung* $\mathbf{B}$

$$\mathbf{R} \bowtie_{\mathbf{B}} \mathbf{S} := \pi_{A_1, A_2, ..., A_k, B_{k-h+1}, ..., B_l}(\sigma_{\mathbf{B}}(\mathbf{R} \times \mathbf{S})) \tag{3.32}$$

Das Relationsschema des Ergebnisses ergibt sich implizit aus den Attributen des Projektionsoperators. Es enthält alle Attribute der Verbundrelationen, die in beiden auftretenden jedoch nur einmal. Die Kardinalität der Ergebnisrelation wird sowohl von dem Selektionsoperator als auch der Duplikatelimination bei der Projektion beeinflusst, so dass keine übersichtliche, allgemein gültige Formel angegeben werden kann.

Beispiel 3.19: *Gleichverbund über mehrere Attribute.*
In einer Bibliotheksdatenbank gibt es die Relationsschemata für

Ausleihe $\mathbf{A}(LeserNr, ISBN, ExemplarNr)$
Standort $\mathbf{S}(ISBN, ExemplarNr, RegalNr)$.

Der Leser mit der Lesernummer 0815 hatte nur ein Buch ausgeliehen, das er zurückgibt. Wo (RegalNr) wird dieses Buch im Regal eingeordnet?

Lösung: $\pi_{RegalNr}(\sigma_{LeserNr=\text{„}0815\text{“}}(\mathbf{A}) \bowtie_{(A.ISBN=S.ISBN)\wedge(A.ExemlarNr=S.ExemplarNr)} \mathbf{S})$ ■

Satz 3.20: *Der Gleichverbund ist bis auf Äquivalenz kommutativ, d.h.*

$$\mathbf{R} \bowtie_{\mathbf{B}} \mathbf{S} \cong \mathbf{S} \bowtie_{\mathbf{B}} \mathbf{R} \tag{3.33}$$

Beweis: Die Behauptung ist eine unmittelbare Folge der Tatsache, dass die Kreuzproduktbildung für Relationen bis auf Äquivalenz kommutativ ist. ■

3.2.2 Der natürliche Gleichverbund

Im Beispiel 3.13: fällt auf, dass die Attribute, die in die Auswahlbedingung des Gleichverbunds eingehen, in beiden Relationen die gleichen sind, also den gleichen Namen, den gleichen Wertevorrat und die gleiche Semantik haben. Diese Situation ist in der Praxis sehr oft gegeben, meist dann, wenn eine Fremdschlüsselbeziehung besteht. Für diesen Fall wird mit dem natürlichen Gleichverbund (*engl:* natural join) eine weitere Vereinfachung der Syntax vorgenommen. Ohne Angabe der Verbundattribute wird unterstellt, dass die gleichnamigen Attribute in den beiden Verbundrelationen auch gleiche Attribute sind und der Verbund als Gleichverbund über genau diesen Attributen ausgeführt wird.

Definition 3.21: *Seien* **R** *und* **S** *zwei Relationen mit den den Attributmengen* $\mathcal{R}$ *und* $\mathcal{S}$. *Ferner sei* $\mathcal{X} = \mathcal{R} \cap \mathcal{S}$, $\mathcal{X}$ *habe die Elemente* X_j, $j = 1,2,...l$.
Mit den Elementen von $\mathcal{X}$ *wird der boolsche Ausdruck*

$$\mathbf{B} = \bigwedge_{j=1}^{l}(\mathbf{R}.X_j = \mathbf{S}.X_j)$$

gebildet.
Dann ergibt sich als **natürlicher Gleichverbund** *von* **R** *und* **S**

$$\mathbf{R} \bowtie \mathbf{S} := \mathbf{R} \bowtie_{\mathbf{B}} \mathbf{S} \tag{3.34}$$

Beispiel 3.22: *Mit der Notation des natürlichen Gleichverbunds wird die Lösung des Beispiel 3.13:*

$$\pi_{Farbton}(\sigma_{Typ=Golf}(\mathbf{Auto} \bowtie \mathbf{Farbe})$$

Für Beispiel 3.19: ergibt sich als Lösung mit dem natürlichen Gleichverbund:

$$\pi_{RegalNr}(\sigma_{LeserNr=„0815“}(\mathbf{A}) \bowtie \mathbf{S})$$

3.2.3 Der θ-Verbund

Bei allen bisher vorgestellten Verbundverknüpfungen wurde die Gleichheit von Attributwerten als Auswahlkriterium benutzt. Dies ist sicher das in der Praxis am meisten auftretende Kriterium. Es ist jedoch möglich, die Gleichheit von Attributwerten durch eine andere Beziehung zwischen den Attributwerten zu ersetzen und die Gültigkeit dieser Beziehung zum Auswahlkriterium zu machen. Insbesondere sind die Vergleichsoperatoren $=, <, \leq, >, \geq, \neq$ von Interesse. Diese werden unter der Bezeichnung θ zuammengefasst. Man spricht von einem **θ-Verbund** (θ-Join), wenn in der Vergleichsbedingung einer der genannten Vergleichsoperatoren vorkommt. Wird eine andere Vergleichoperation als die Gleichheit gewählt, muß der Verbund neu definiert werden. Es darf dann keine Projektion zum Weglassen von Spalten erfolgen, da es i.a. keine Spalten gibt, die identisch sind.

Definition 3.23: *Seien* **R** *und* **S** *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_k)$ *und* $\mathbf{S}(B_1, B_2, ..., B_l)$. *Alle Attributnamen seinen paarweise verschieden. Die Attribute* $\mathbf{R}.A_k$ *und* $\mathbf{S}.B_1$ *seien vergleichbar, d.h. es kann entschieden werden, ob für* $\theta \in \Theta = \{=, <, >, \neq, \leq, \geq\}$ *die Bedingung* $\mathbf{R}.A_k\theta\mathbf{S}.B_1$ *zutrifft. Dann heißt*

$$\mathbf{R} \bowtie_{A_k\theta B_1} \mathbf{S} := \sigma_{A_k\theta B_1}(\mathbf{R} \times \mathbf{S}) \tag{3.35}$$

θ-Verbund *von* **R** *und* **S**.

Der θ-Verbund kann als der allgemeine Verbundoperator angesehen werden, aus dem sich Gleichverbund und natürlicher Gleichverbund als Spezialfälle ableiten lassen. Auch hier können vereinfachte Darstellungen wie beim Gleichverbund vereinbart werden: Wird nur ein Symbol θ aus der Menge Θ angegeben, so soll das bedeuten, dass ein natürlicher Verbund zu berechnen ist, bei dem anstelle des Tests auf Gleichheit jeweils der Vergleich mit dem Vergleichsoperator θ auszuführen ist.

Beispiel 3.24: *In einer Bibliothek soll festgestellt werden, welche Leser mehr als ein Buch ausgeliehen haben (dabei wird unterstellt, daß kein Leser zwei Exemplare eines Buches ausgeliehen hat).*
Für die Ausleihe existere Wieder eine Relation mit dem Schema:
Ausleihe $\mathbf{A}(LeserNr, ISBN, ExemplarNr)$

Lösung: Es wird eine Kopie **B** von **A** erstellt und dann durch eine Join-Operation geprüft, ob es zu einer Lesernummer *LeserNr* zwei verschiedene *ISBN* gibt:

$$\mathbf{B} \Leftarrow \mathbf{A}$$
$$\pi_{\mathbf{A}.LeserNr}(\mathbf{A} \bowtie_{(A.LeserNr=B.LeserNr)\wedge(\mathbf{A}.ISBN\neq\mathbf{B}.ISBN)} \mathbf{B}).$$ ■

3.2.4 Der äußere Verbund

Die bisher vorgestellten Verbundverknüpfungen sind nicht verlustfrei, d.h. ein Tupel einer Relation, das in der anderen Relation keinen „Partner" findet, erscheint nicht im Ergebnis, wie ein Beispiel zeigt:

Beispiel 3.25: *Gegeben seien die beiden Relationen* **Auto** *und* **Farbe**, *jetzt haben beide Relationen ein weiteres Attribut Kennzeichen, abgekürzt als K mit der üblichen Bedeutung:*
Auto(K, Typ) *bzw.* **Farbe**$(K, Farbton)$ *mit den Relationen:*

Auto

K	*Typ*
L-XX 0002	*Touran*
L-YZ 2005	*Sharan*
L-FJ 0815	*Golf*
C-AA 4711	*Käfer*

Farbe

K	*Farbton*
L-YZ 2005	*rot*
L-FJ 0815	*blau*
C-AA 4711	*rot*

Gesucht wird eine Tabelle der Kennzeichen, Typen und Farbtöne; wenn der Farbton nicht bekannt ist, sollen wenigstens die anderen Merkmale ausgeben werden.

Der Ausdruck **Auto** $\bowtie$ **Farbe** löst die Aufgabe nicht:

Auto $\bowtie$ **Farbe**

K	Typ	Farbton
L-YZ 2005	Sharan	rot
L-FJ 0815	Golf	blau
C-AA 4711	Käfer	rot

Das Fahrzeug L-XX 0002, dessen Farbe nicht bekannt ist, fehlt im Ergebnis, da zu dem Eintrag mit $K =$ L-XX 0002 der Relation **Auto** kein Eintrag mit diesem K in der Relation **Farbe** existiert. Es wird eine Verknüpfung gesucht, die den Verbund ausführt und anschließend das Ergebnis um alle Tupel der ersten (linken) Relation ergänzt, die keinen Verbundpartner gefunden haben. Dabei müssen evtuell fehlende Attribute der Ergebnistabelle sinnvoll ergänzt werden:

Auto ? **Farbe**

K	Typ	Farbton
L-YZ 2005	Sharan	rot
L-FJ 0815	Golf	blau
C-AA 4711	Käfer	rot
L-XX 0002	Touran	??

Ein solches Ergebnis liefert ein äußerer Verbund (*engl.* outer join). Bei der Konstruktion des äußeren Verbundes wird die Relation, in der keine Verbundpartner gefunden werden, wird um Einträge ergänzt, die genau die fehlenden Partner liefern. Auf dieser ergänzten Relation wird dann eine „normale" Verbundoperation ausgeführt. Wegen dieses Vorgehens ist zwischen linken, rechten und beidseitigem äußeren Gleichverbund zu unterscheiden. Beim linken wird die linke Relation verlustfrei in das Ergebnis übernommen, beim rechten die rechte, im dritten Fall beide. Wie bei den Verbundoperatoren sind verschiedene Vergleichsoperatoren möglich, hier wird nur eine Definition für den äußeren Gleichverbund vorgestellt.

Definition 3.26: *Seien* $\mathbf{r}(\mathbf{R})$ *und* $\mathbf{s}(\mathbf{S})$ *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_m)$ *und* $\mathbf{S}(C_1, C_2, ..., C_l)$. *Sei* $\mathcal{X}$ *die Menge der Attribute von* **R** *bzw.* **S**, *die in die Verbundbedingung* **B** *eingehen. O.B.d.A. sei* $\mathcal{X} = \{C_1, C_2, ..., C_k\}, k \leq l, k \leq j$. *Für die Attribute* $(C_{k+1}, C_{k+2}, ...C_l)$ *sei ein Wert mit der Semantik „Wert unbekannt", der hier mit ?? bezeichnet wird, definiert.*
$\mathbf{s}^* = \mathbf{s} \cup ((\pi_{\mathcal{X}}(\mathbf{r}) \setminus \pi_{\mathcal{X}}(\mathbf{s})) \times_{j=k+1}^{l} \chi_{C_j}(??))$.
Der **linke äußere Gleichverbund** *ergibt sich zu*

$$\mathbf{R} \sqsupset\!\!\bowtie_{\mathbf{B}} \mathbf{S} := \mathbf{R} \bowtie_{\mathbf{B}} \mathbf{S}^* \tag{3.36}$$

Erläuterung der Definition 3.26: Es werden die Attribute betrachtet, die in die Verbundbedingung eingehen. Alle Werte dieser Attribute, die in **R**, aber nicht in **S** vorkommen, werden durch Auffüllen mit unbestimmten Werten zu Tupeln „gemacht", die zu **S** vereinigungskompatibel sind, und zu **s** hinzugefügt. Mit dieser ergänzten Relation $\mathbf{s}^*$ finden nun alle Tupel aus **R** mindestens einen passenden Join-Partner.
Um die rechte Relation verlustfrei bei einer Join-Operation zu haben, wird der rechte äußere Verbundoperator genutzt:

Definition 3.27: *Seien* $\mathbf{r}(\mathbf{R})$ *und* $\mathbf{s}(\mathbf{S})$ *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_m)$ *und* $\mathbf{S}(C_1, C_2, ..., C_l)$. *Sei* $\mathcal{X}$ *die Menge der Attribute von* **R** *bzw.* **S**, *die in die Verbundbedingung* **B** *eingehen. O.B.d.A. sei* $\mathcal{X} = \{A_1, A_2, ..., A_k\}, k \leq l, k \leq m$. *Für die Attribute* $(A_{k+1}, A_{k+2}, ...A_m)$ *sei ein Wert mit der Semantik „Wert unbekannt",*

der hier mit ?? bezeichnet wird, definiert.
$\mathbf{r}^* = \mathbf{r} \cup ((\pi_{\mathcal{X}}(\mathbf{s}) \setminus \pi_{\mathcal{X}}(\mathbf{r})) \times_{j=k+1}^{m} \chi_{A_j}(??))$.
Der **rechte äußere Gleichverbund** *ergibt sich zu*

$$\mathbf{R} \bowtie\!\sqsubset_{\mathbf{B}} \mathbf{S} := \mathbf{R}^* \bowtie_{\mathbf{B}} \mathbf{S} \tag{3.37}$$

Durch Kombination beider Definitionen 3.26: und 3.27: kann der beidseitig verlustfreie äußere Verbundoperator erklärt werden.

Definition 3.28: *Seien* $\mathbf{r}(\mathbf{R})$ *und* $\mathbf{s}(\mathbf{S})$ *zwei Relationen mit den Relationschemata* $\mathbf{R}(A_1, A_2, ..., A_m)$ *und* $\mathbf{S}(C_1, C_2, ..., C_l)$*. Sei* $\mathcal{X}$ *die Menge der Attribute von* **R** *bzw.* **S***, die in die Verbundbedingung* **B** *eingehen. O.B.d.A. sei* $\mathcal{X} = \{A_1, A_2, ..., A_k\} = \{C_1, C_2, ..., C_k\}$*. Die erweiterten Relationen* $\mathbf{S}^*$ *und* $\mathbf{R}^*$ *werden wie in den Definitionen 3.26: und 3.27: gebildet. Dann ergibt sich der* **äußere Verbundoperator** *zu*

$$\mathbf{R} \sqsupset\!\bowtie\!\sqsubset \mathbf{S}_{\mathbf{B}}\mathbf{S} := \mathbf{R}^* \bowtie_{\mathbf{B}} \mathbf{S}^* \tag{3.38}$$

Ebenso kann ein äußerer natürlicher Gleichverbund definiert werden, indem die Verbundbedingung in der für den natürlichen Verbund typischen Weise gewählt wird, dass konjunktiv verknüpfte Gleichheitsbedingungen auf allen in beiden Relationen auftretenden Attributen gestellt werden (vgl Definition 3.21:).

Lösung zu Beispiel 3.25: (Seite 35) Der natürliche linke äußere Gleichverbund **Auto** $\sqsupset\!\bowtie$ **Farbe** löst diese Aufgabe.

3.3 Die Relationendivision

Beispiel 3.29: *In einem Unternehmen wurde ein Rechnernetz mit mehreren Servern aufgebaut. Jeder Mitarbeiter hat an seinem Arbeitsplatz einen Rechner, der als Client dient. Aus Datenschutzgründen wird in einer Zugriffstabelle festgelegt, welcher Client sich mit welchem Server verbinden darf. Das besondere Interesse von Hackern gilt natürlich den Rechnern, von denen aus alle Server erreichbar sind.*
Welche Clientcomputer dürfen sich mit allen Servern verbinden und sind deswegen als besonders sensibel einzustufen?

Diese Frage kann mit den bisher besprochenen Operatoren und Verknüpfungen der Algebra gelöst werden, aber die Lösung ist nicht offensichtlich. Deshalb wird die Verknüpfung „Relationendivision“ definiert, die genau auf diese Fragen ausgerichtet ist.

Definition 3.30: *Seien* **R** *eine Relation vom Grad* r *und* **S** *vom Grade* s *mit* $r > s$ *und* $s > 0$*. Für die entsprechenden Attributmengen* $\mathcal{S}$ *von* **S** *bzw.* $\mathcal{R}$ *von* **R** *gelte:* $\mathcal{S} \subset \mathcal{R}$*.*
Ferner bezeichne $\mathcal{Y} = \mathcal{R} \setminus \mathcal{S}$ *und* t *ein Tupel vom Grade* $(r - s)$ *mit Werten in der Produktmenge* $\times_{Y \in \mathcal{Y}} dom(Y)$*.*

$$\mathbf{R} \div \mathbf{S} := \{t | \forall u \in \mathbf{S} : (t, s) \in \mathbf{R}\}. \tag{3.39}$$

heißt **Division** *von* **R** *und* **S**.
Zur Ergebnisrelation der Division gehören genau die Attribute aus $\mathcal{Y}$

In der Definition ist wie beim modifizierten Kreuzprodukt (t,s) als Tupel zu interpretieren (vgl. Seite 28).

Lösung zu Beispiel 3.29: Mit der Definition 3.30: läßt sich das Beispiel ohne Schwierigkeiten lösen. Folgende, stark vereinfachte Relationsschemata modellieren den Sachverhalt:

$\mathbf{C}(Clientname)$	% Schema der Liste aller Clients,
$\mathbf{S}(Servername)$	% Schema der Server,
$\mathbf{R}(Clientname, Servername)$	% Schema für Zugriffserlaubnis

In **R** wird das Paar $(Clientname, Servername)$ eingetragen genau dann, wenn der Client auf diesen Server zugreifen darf.
Offensichtlich erfüllen **R** und **S** die Voraussetzungen aus Definition 3.30:. $\mathbf{R} \div \mathbf{S}$ beschreibt die gesuchte Rechnermenge und enthält genau die Clientnamen, die mit allen Servernamen aus **S** in Verbindung stehen.

Rechenregeln für den Divisionsoperator

Satz 3.31: Division und Kreuzprodukt
Wenn **R** *und* **S** *die Voraussetzungen der Definition 3.30: erfüllen, dann gilt:*

$$(\mathbf{R} \div \mathbf{S}) \times \mathbf{S} \subseteq \mathbf{R} \tag{3.40}$$

Ohne Vorbedingungen gilt stets:

$$(\mathbf{R} \times \mathbf{S}) \div \mathbf{S} = \mathbf{R} \tag{3.41}$$

Beweis:
Die Voraussetzungen aus Definition 3.30: sichern, dass in 3.40 $(\mathbf{R} \div \mathbf{S}) \times \mathbf{S}$ und **R** äquivalente Schemata besitzen. Nach Definition der Relationendivision $(\mathbf{R} \div \mathbf{S})$ sind in deren Ergebnis nur Teile t von Tupeln, für die gerade die modifizierte Paarbildung (aus dem modifizierten Kreuzprodukt) mit jedem Tupel aus **S** in **R** liegt. Dies beweist die Inklusion 3.40. Die Aussage lässt sich nicht verschärfen, da sowohl Beispiele für die strenge Inklusion als auch für die Gleichheit gibt.
Die Beziehung 3.41 ist eine unmittelbare Anwendung der Definitionen des modifizierten Kreuzprodukts und der Relationendivision. ■

Beispiel 3.32: *Sei* $\mathbf{R} = \mathbf{T} \times \mathbf{S}$, *dann gilt* $(\mathbf{R} \div \mathbf{S}) \times \mathbf{S} = \mathbf{R}$.
Beweis:

$$(\mathbf{R} \div \mathbf{S}) \times \mathbf{S} = ((\mathbf{T} \times \mathbf{S}) \div \mathbf{S}) \times \mathbf{S} =^{(nach\ 3.41)} \mathbf{T} \times \mathbf{S} = \mathbf{R}.$$ ■

Beispiel 3.33: *Aufgrund der definierenden Gleichung 3.39 gilt unmittelbar*

$$(\mathbf{R} \div \mathbf{S}) \subseteq \pi_{\mathcal{Y}}(\mathbf{R}), \tag{3.42}$$

wenn $\mathcal{Y}$ wie oben die Menge der Attribute von **R** *bezeichnet, die nicht auch Attribute von* **S** *sind.*

Satz 3.34: Darstellung der Division
R *und* **S** *erfüllen die Voraussetzungen der Definition 3.30:. Seien $\mathcal{R}$ und $\mathcal{S}$ die Mengen der Attribute von* **R** *bzw.* **S**, *$\mathcal{Y} = \mathcal{R} \setminus \mathcal{S}$ ist die Menge der Attribute von* **R**, *die nicht auch Attribute von* **S** *sind.*

$\mathbf{T}_1 \Leftarrow \pi_{\mathcal{Y}}(\mathbf{R})$
$\mathbf{T}_2 \Leftarrow \pi_{\mathcal{Y}}((\mathbf{T}_1 \times \mathbf{S}) \setminus \mathbf{R})$
$\mathbf{T} \Leftarrow \mathbf{T}_1 \setminus \mathbf{T}_2$

Es gilt: $\mathbf{T} = \mathbf{R} \div \mathbf{S}$

Beweis: Die Konstruktion der Relationen sichert jeweils die benötigten Vereinigungskompatibilitäten. $(\mathbf{T}_1 \times \mathbf{S})$ besteht aus den $\mathcal{Y}$-Spalten aller Tupel aus **R**, die jeweils mit jedem Tupel aus **S** im Sinne des modifizierten Kreuzprodukts verbunden sind. Wird davon im Sinne der Mengendifferenz **R** abgezogen, verbleiben gerade die Tupel aus dieser Verbindung, die nicht in **R** waren, $\mathbf{T}_2$ enthält also die dazugehörigen $\mathcal{Y}$-Komponenten, sie waren in **R** also nicht mit allen Tupeln aus **S** verbunden. Wegen $\mathbf{T} \Leftarrow \mathbf{T}_1 \setminus \mathbf{T}_2$ folgt somit $\mathbf{T} \subseteq \mathbf{R} \div \mathbf{S}$.
Dass umgekehrt $\mathbf{T} \supseteq \mathbf{R} \div \mathbf{S}$ zutrifft, folgt aus der Definition des Divisionsoperators. Für ein Tupel t aus $\mathbf{R} \div \mathbf{S}$ wird die Definition angewendet, für die nach der Definition dazu existierenden Tupel $(t,s) \in \mathbf{R}$ werden die Konstruktionen des Satzes nachgerechnet, es ergibt sich, dass $t \in \mathbf{T}$ ist. ∎

4 Weiter über Operatoren

4.1 Basisoperationen und abgeleitete Operationen

Werden die angegebenen Operationen und Verknüpfungen zusammenfassend betrachtet, kann festgestellt werden:

Die bisher beschriebenen Operationen und Verknüpfungen sind nicht alle unabhängig voneinander. Insbesondere wurden einige Verknüpfungen eingeführt, um häufig benötigte Sachverhalte kurz und übersichtlich aufschreiben zu können, die JOIN-Verknüpfungen beispielsweise. Bei weiteren Verknüpfungen wurde ihre Darstellung durch andere Operationen und Verknüpfungen angegeben. Unter Beachtung der bekannten Ergebnisse der Mengenalgebra über die Beziehungen zwischen Mengendurchschnitt, Mengenvereinigung und Mengendifferenz kann folgende Minimalitätseigenschaft formuliert werden:

Satz 4.1: *Die Grundoperationen der Relationenalgebra können auf Selektion, Projektion, Relationenvereinigung, Relationendifferenz und modifiziertes Kreuzprodukt beschränkt werden.*

Beweis: Wie in der Mengenalgebra gilt auch unter der Bedingung, dass die Relationen vereinigungskompatibel sind:

$$\mathbf{R} \cap \mathbf{S} = \Big(\mathbf{R} \cup \mathbf{S}\Big) \Big\backslash \Big((\mathbf{R} \setminus \mathbf{S}) \cup (\mathbf{S} \setminus \mathbf{R})\Big) \tag{4.1}$$

Die Bildung der Durchschnittsmenge zweier Relationen kann also durch Vereinigungs- und Differenzbildungen ausgerückt werden. Die anderen in Satz behaupteten Abhängigkeiten ergeben sich direkt aus den Definitionen der mehrstelligen Operatoren.
Der Einbettungsoperator χ gehört nicht zur Algebra, da sein Definitionsgebiet nicht in der Relationenmenge liegt. ■

Die im Satz genannte Menge von Operationen und Verknüpfungen ist minimal in dem Sinne, dass beim Weglassen weiterer Bestandteile die Funktionalität eingeschränkt wird. Für Vereinigung und Differenz folgt dies wieder aus entsprechenen Sätzen der Mengenlehre. Die Selektion trifft als einzige Operation eine Auswahl nach Attributwerten, nur die Projektion verringert den Grad einer Relation, während nur das Kreuzprodukt den Grad vergrößert. Beim Weglassen eines dieser Operatoren wären also Tabellenumformungen mit der entsprechenden Eigenschaft nicht mehr möglich.

4.2 Erweiterungsmöglichkeiten

Die relationale Algebra dient nur dazu, die Informationsgewinnung aus den Relationen mit ihren Relationsschemata zu beschreiben. Sie dient nicht dazu zu modellieren, wie die Relationsschemata entstehen oder wie die Relationen entstehen. Es wird nur beschrieben, die sich die Relationsschemata der Ergebnisse der Operationen und Verknüpfungen aus den Schemata der Argumente ableiten.
Dies beachtend, bringt ein Vergleich mit den Möglichkeiten der Datenbankanfragesprache SQL die Erkenntnis, dass dennoch Defizite in der Ausdrucksmächtigkeit bestehen. Die sind
- allgemeinere Auswahlbedingungen bei Selektionen und Verbund
- das Fehlen der Aggregatfunktionen
- das Fehlen der Konstruktion der rekursiven Hülle

Der ersten beiden Kritikpunkte können behoben werden, eine Idee zum zweiten Punkt wurde in [2] gegeben.

4.2.1 Erweiterte Auswahlbedingungen

Der Selektionsoperator σ kann verallgemeinert werden, indem eine allgemeine boolsche Funktion auf den Attributen zur Auswahl zugelassen wird, die dem Wertetupel der Attribute einen Wahrheitswert zuordnet. Am Beispiel eines verallgemeinerten Verbundoperators von **R** und **S** soll dies demonstriert werden. Hier dient dann die allgemeine boolsche Funktion, die nicht nur konjunktive, disjunktive und negierende Verknüpfungen von Vergleichsoperationen auf den Attributen enthält, als Kriterium für die Selektion.

Beispiel 4.2: *Sei*

$$\mathbf{V}(\textit{vek-id}, \textit{vonx}, \textit{vony}, \textit{vonz}, \textit{nachx}, \textit{nachy}, \textit{nachz})$$

ein Relationsschema, die zugehörige Relation **V** *beinhalte Vektoren im Raum, von denen jeweils eine Zweipunktedarstellung in kartesischen Koordinaten sowie eine Identifikationsnummer vek-id gegeben sind. Alle Vektoren mögen eine Länge größer Null haben.*
Sei **U** *eine umbenannte Kopie von* **V***:*

$$\mathbf{U}(\textit{Uid}, \textit{Uvonx}, \textit{Uvony}, \textit{Uvonz}, \textit{Unachx}, \textit{Unachy}, \textit{Unachz}) \Leftarrow \mathbf{V}$$

Auf $\mathbf{V} \times \mathbf{U}$ *wird die Bedingung* **B** *definiert:*

$$\begin{aligned} \mathbf{B} = (\quad & (\textit{nachx} - \textit{vonx})(\textit{Unachx} - \textit{Uvonx}) \\ + \quad & (\textit{nachy} - \textit{vony})(\textit{Unachy} - \textit{Uvony}) \\ + \quad & (\textit{nachz} - \textit{vonz})(\textit{Unachz} - \textit{Uvonz}) \\ = \quad & 0 \\ ) \quad & \end{aligned}$$

Der Ausdruck $\pi_{\textit{vek-id},\textit{Uid}}(\sigma_{\mathbf{B} \wedge (\textit{Uid} > \textit{vek-id})}(\mathbf{V} \times \mathbf{U}))$ *liefert alle Vektorpaare, die* **B** *erfüllen (anschaulich: deren Vektoren aufeinander senkrecht stehen). Die Komponente* $(\textit{Uid} >$

vek-id) der Auswahlbedingung ist technisch, sie erreicht, daß zu einem Vektorpaar mit den Indizes (i,j) nicht noch das Paar (j,i) im Ergebnis auftritt.

4.2.2 Aggregatfunktionen

Die Aggregatfunktionen des SQL-Anfragesprache zählen die Anzahl von Tupeln oder berechnen Maximum, Minimum und Durchschnitt der in einer Relation vorkommenden Werte eines Attributs. Sie ordnen einer Relation (oder einem Teil einer Relation) eine reelle Zahl zu ([1]). Folglich sind die Aggregatfunktionen keine Operatoren in der Relationenalgebra, sondern Funktionale ([2]) auf den Relationen. Ihr Wertebereich liegt im Gegensatz zu den bisher besprochenen Operationen oder Verknüpfungen nicht in der Relationenmenge. Diese Eigenschaft läßt sich durch Kombination mit dem Einbettungsoperator (Definition 2.12:, Seite 21) verändern. Um konzeptionel klar zwischen einer einstelligen Relation der Kardinalität eins und dem einzigen darin eingetragenen Wert zu interscheiden, wird neben dem Einbettungsoperator χ noch sein Inverses ω eingeführt (vgl. Seite 21).

In Anlehnung an die SQL-Aggregatfunktionen werden folgende Funktionale auf den Relationen der Relationenalgebra definiert:

- Das **Kardinalitätsfunktional : $\sharp(\mathbf{R})$**
  Es ist definiert auf einer beliebigen Relation $\mathbf{R}$. Sein Wert ist die Kardinalität dieser Tupelmenge, d.h. die Anzahl der Tupel.
- Das **Zählfunktional für das Attribut A: $\sharp_A(\mathbf{R})$:**
  Es zählt, wie oft das Attribut A der Relation $\mathbf{R}$ einen Wert hat, der nicht die Semantik „Wert unbekannt“ oder eine ähnliche Semantik trägt. Es erfolgt keine Eliminierung von Wertduplikaten.
- Das **Minimumfunktional für das Attribut A: $MIN_A(\mathbf{R})$**
  Es ist für e in Attribut A einer beliebigen Relation $\mathbf{r}(\mathbf{R})$ mit einem Wertebereich in den reellen Zahlen definiert und bestimmt das Minimum aller Werte von A, die in der Relation auftreten:
  $MIN_A(\mathbf{R}) := min_{t\in\mathbf{r}}(t.A)$.
  Es versteht sich, dass Werte mit der Semantik „Wert unbekannt“ keinen Sinn für die Minimumbildung machen und folglich nicht beachtet werden.
- Das **Maximumfunktional für das Attribut A: $MAX_A(\mathbf{R})$**
  Analog zum Minimumbildung, jedoch mit
  $MAX_A(\mathbf{R}) := max_{t\in\mathbf{r}}(t.A)$.
- Das **Summenfunktional für das Attribut A: $\Sigma_A(\mathbf{R})$**
  Es ist ist für ein Attribut A einer beliebigen Relation $\mathbf{R}$ mit einem Wertebereich

[1] Die Bedeutung der Tatsache, dass in einem SQL-Statement die Aggregatfunktionen in einer SELECT-Klausel gekapselt sind, wird noch diskutiert (Seite 44)

[2] Ein Funktional auf einer Menge von Objekten ist eine Abbildung, die jedem Objekt eine reelle (oder komplexe) Zahl zuordnet.

in den reellen Zahlen definiert und bestimmt die Summe aller Werte von A, die in den Tupeln der Relation auftreten:
$\Sigma_A(\mathbf{R}) := \Sigma_{t \in \mathbf{r}}(t.A)$.
Es versteht sich, dass Werte mit der Semantik „Wert unbekannt" keinen Sinn bei der Summenbildung machen, also nicht beachtet werden, Duplikate gehen in die Rechnung ein.

- Das **Durchschnittsfunktional für das Attribut** A**:** $\Delta_A(\mathbf{R})$
  Es ist ist für ein Attribut A einer beliebigen Relation **R** mit einem Wertebereich in den reellen Zahlen definiert und bestimmt den Durchschnitt aller Werte von A, die in der Relation auftreten. Bezüglich unbestimmter Werte und Duplikate verhält es sich wie das Summenfunktional. Es gilt:
  $\Delta_A(\mathbf{R}) = \Sigma_A(\mathbf{R}) / \sharp_A(\mathbf{R})$

Mit diesen Funktionalen wurden die Aggregatfunktionen aus der SQL-Anfragesprache beschrieben, selbstverständlich sind weitere Definitionen möglich, z.B. die Berechnung der Varianz und anderer Statistikfunktionen. Da diese in den Datenbankverwaltungssystemen (noch) nicht realisiert sind, wird dieser Weg hier nicht weiter verfolgt.

Beispiel 4.3: *Vorgelegt sei nochmals die Professorentabelle von Seite 11:*

Professoren

PNR	*Name*	*Arbeitsgebiet*	*Fakultät*
2364	*Rahm*	*Datenbanken*	*Math.u.Informatik*
0342	*Beyer*	*Analysis*	*Math.u.Informatik*
9645	*Wartenberg*	*Kirchengeschichte*	*Theologie*
4714	*Cain*	*Klass. Archäologie*	*Geschichte*

Beispieltabelle „Professoren"

Wieviele Professoren sind in der Tabelle erfasst?

Lösung: $\chi_{Anzahl}(\sharp(\mathbf{Professoren}))$

mit der Ergebnisrelation

Anzahl
4

■

An der Lösung ist zu sehen, wie aus dem Wert der Kardinalität mittels $\chi_{Anzahl}(.)$ eine Relation mit dem Attribut *Anzahl* erzeugt wird. Diese Vorgehen mag umständlich erscheinen, es entspricht jedoch der Anfragekonstruktion , wie sie auch im SQL-Standard vorgesehen ist, auch dort kann auf die Aggregatfunktionen nur im Rahmen einer SELECT-Klausel, also der Konstruktionsanweisung für eine Relation, zugegriffen werden.

Beispiel 4.4: *Aus wievielen Fakultäten kommen die Professoren in der Relation des vorigen Beispiels?*

Lösung: Der Ansatz $\sharp_{Fakultät}(\textbf{Professoren})$ wäre falsch, da Duplikate bei den Fakultätsnamen mitgezählt würden. Korrekt ist

$$\chi_{FakZahl}(\sharp(\pi_{Fakultät}(\textbf{Professoren}))).$$

Duplikate in den Fakultätsnamen werden durch Projektion auf das Attribut *Fakultät* eliminiert, somit ist die Kardinalität des Zwischenresultats $\pi_{Fakultät}(\textbf{Professoren})$ die Lösung. ■

In den Beispielen bewirkt die Anwendung des Einbettungsoperators χ, dass das Ergebnis eine Relation ist. Soll das Ergebnis jedoch numerisch weiterverwendet werden, wird gerade der Zahlenwert benötigt.

Beispiel 4.5: *In einer Datenbank sei die folgende Relation* **Mitarbeiter**. *Das Attribut Gehalt steht für das Jahresgehalt in Euro.*
Wie groß ist das Durchschnittsgehalt der Mitarbeiter?

Mitarbeiter

PNR	Name	Abteilung	Gehalt
1000	Star	A1	55.000
1001	Spatz	A1	23.000
1002	Meise	B	34.000
1003	Vogel	C	36.000
1004	Amsel	C	unbekannt

Tabelle 4.1: Beispieltabelle „Mitarbeiter“

Lösung: $\chi_{DGehalt}(\Delta_{Gehalt}(\textbf{Mitarbeiter}))$ ergibt eine Relation mit dem Attribut *DGehalt*. Es wird dabei der Mitarbeiter Amsel korrekterweise nicht in die Berechnung einbezogen, da sein Gehalt nicht bekannt ist. Der Wert des Attributs *DGehalt* ist 148.000/4 = 37.000 (Euro).
Ohne Benutzung des Einbettungsoperators liefert $\Delta_{Gehalt}(\textbf{Mitarbeiter})$ die Zahl 37.000. Das numerisch gleiche Ergebnis ergeben die Berechnungen

$$\frac{\Sigma_{Gehalt}(\textbf{Mitarbeiter})}{\sharp_{Gehalt}(\textbf{Mitarbeiter})}$$

oder

$$\frac{\omega(\chi_{Summe}(\Sigma_{Gehalt}(\textbf{Mitarbeiter})))}{\omega(\chi_{Anzahl}(\sharp_{Gehalt}(\textbf{Mitarbeiter})))},$$

da $\sharp_{Gehalt}(\textbf{Mitarbeiter})$ nur die Anzahl der Einträge in der Spalte *Gehalt* zählt, die einen Wert repräsentieren. Der Ausdruck

$$\frac{\Sigma_{Gehalt}(\textbf{Mitarbeiter})}{\sharp(\textbf{Mitarbeiter})}$$

ergibt wegen der Zählung aller Mitarbeiter den falschen Wert 29.000 (Euro). ■

Die Aggregatfunktionen in ihrer SQL-Definition haben noch eine andere, bisher unerwähnte Eigenschaft. Durch Angabe einer `GROUP-BY`-Klausel, in der ein oder oder mehrere Attribute genannt werden, wird das Ergebnis einer `SELECT`-Anweisung in Gruppen angeordnet, wobei die Aggregatfunktionen für jede Gruppe neu berechnet werden. Auch um diese Eigenschaft kann die Relationenalgebra erweitert werden, indem eine Erweiterung des Einbettungsoperators definiert wird. Dabei wird zusätzlich eine Liste der Gruppierungsattribute als Parameter an den Einbettungsoperator übergeben.

Definition 4.6: Einbettungsoperator mit Gruppierung
Sei $\mathbf{r}(\mathbf{R})$ *eine Relation zum Relationsschema* $\mathbf{R}(A_1, A_2, ..., A_n)$. *Sei* $B_1, B_2, ..., B_l$ *eine Teilmenge der Attribute von* $\mathbf{R}$, A *ein Attribut mit* $A \in (A_1, A_2, ..., A_n)$, *aber* $A \notin (B_1, B_2, ..., B_l)$.
ϕ *sei eine der oben definierten Aggregatfunktionen.*
X *sei ein Attributname.*
Dann ist

$$\chi_{B_1,B_2,...,B_l;X}(\phi_A(\mathbf{R})) := \{t | t = (t.B_1, ..., t.B_l, X)\} \tag{4.2}$$

mit

$(t.B_1, ..., t.B_l) \in \pi_{B_1,...,B_l}(\mathbf{R})$ *und*
$t.X = \phi_A(\sigma_{B_1=t.B_1 \wedge ... \wedge B_l=t.B_l}(\mathbf{R}))$,

wobei ϕ *für jede Wertekombination* $(t.B_1, ..., t.B_l)$ *neu berechnet wird und zwar über genau den Tupeln, die bei* $\pi_{B_1,...,B_l}(\mathbf{R})$ *auf die Wertekombination* $(t.B_1, ..., t.B_l)$ *abgebildet werden.*

Beispiel 4.7: *Gegeben sei eine Relation mit dem Relationsschema*
$\mathbf{Buecher}(ISBN, Autor, Titel, Verlag, Preis)$
dabei sei ISBN die eindeutige ISBN-Nummer jedes Buches, Preis der Preis in Euro. Das Attribut Verlag bezeichne in eindeutiger Weise den Verlag.
Gesucht wird eine Relation $\mathbf{Q}$*, die die Verlage und die Durchschnittspreise der bei ihnen erschienen Bücher enthält.*

Lösung: Die gesuchte Relation wird aus **Buecher** mit Hilfe der Aggregatfunktion Durchschnitt und des erweiterten Einbettungsoperators konstruiert:
$\mathbf{Q} \Leftarrow \chi_{Verlag;Durchschnittspreis}(\Delta_{Preis}(\mathbf{Buecher}))$.
Entsprechend der allgemeinen Beschreibung, wird für jeder Wert des Attributs *Verlag* der Durchschnitt der Preise der Bücher ermittelt, die diesen Wert des Verlagsattributs haben und zusammen mit dem Verlag als Tupel im der Ergebnisrelation abgelegt. ■

Die folgende Aussage zeigt, dass sich hinter der Definition eine prozedurale Schleifenkonstruktion verbirgt.

Satz 4.8: Einbettungsoperator mit Gruppierung
Sei $\mathbf{r}(\mathbf{R})$ *eine Relation zum Relationsschema* $\mathbf{R}(A_1, A_2, ..., A_n)$. *Sei* $B_1, B_2, ..., B_l$ *eine Teilmenge der Attribute von* $\mathbf{R}$, A *ein Attribut mit* $A \in (A_1, A_2, ..., A_n)$, *aber* $A \notin (B_1, B_2, ..., B_l)$.
ϕ *sei eine der oben definierten Aggregatfunktionen,*
$\mathbf{s}$ *eine Relation , X sei ein Attributname.*
Bilde:

$\mathbf{s} := \emptyset;$
$\mathbf{FOR}\ \mathbf{i} \in \pi_{B_1, B_2, ..., B_l}(\mathbf{r})\ \{$
$w = \Phi_A(\sigma_{B_1 = \mathbf{i}.B_1, ..., B_l = \mathbf{i}.B_l}(\mathbf{r}));$
$\mathbf{s} \Leftarrow \mathbf{s} \cup (\mathbf{i}.B_1, \mathbf{i}.B_2, ..., \mathbf{i}.B_l, w)\ \}$.

Dann gilt $\mathbf{s} = \chi_{B_1, B_2, ..., B_l; X}(\phi_A(\mathbf{R}))$.

Auf den Beweis wird hier verzichtet.

4.2.3 Die rekursive Hülle

Zur Einführung in das Problem wird ein Relation mit dem Relationsschema

$$\mathbf{Nachfahren}(Name, Kind)$$

betrachtet. Dabei wird vereinfachend unterstellt, dass der Name eine Person eindeutig beschreibt. Das Attribut *Kind* hat als Wert den Namen einer anderen Person, dabei soll zwischen der durch *Name* bezeichneten und der durch *Kind* bezeichneten Person das Eltern-Kind-Verhältnis bestehen. Wie können mit der Relationenalgebra solche Anfragen wie „Welche Nachfahren von J.W. von Goethe sind bekannt?“, „Welche Vorfahren hat Dieter Sosna?“ bearbeitet werden. Würde die Anfrage auf eine bestimmte Generationenzahl beschränkt werden, z.B. auf Kinder und die Enkel von Goethe, so könnte die Antwort mit Hilfe einer geeigneten festen Anzahl von JOIN-Ausdrücken und deren Vereinigung gegeben werden:

$$\mathbf{N} \Leftarrow \mathbf{Nachfahren}; \qquad \mathbf{N1} \Leftarrow \mathbf{Nachfahren};$$

$$\pi_{\mathbf{N1}.Kind}((\sigma_{Name=\text{‘Goethe’}}(\mathbf{N})) \bowtie_{\mathbf{N}.Kind=\mathbf{N1}.Name} \mathbf{N1})$$
$$\cup\ \pi_{Kind}(\sigma_{Name=\text{‘Goethe’}}(\mathbf{N}))$$

Indem für jede Generation eine weitere Verbundoperation und eine Vereinigung ausgeführt wird, kann jede vorgegebene Rekursionstiefe erreicht werden. Das beantwortet aber nicht die Frage, wie *alle* Nachfahren einer Person ermittelt werden können. In der Anfragesprache SQL besteht die Lösung aus wiederholte Vereinigungen von JOIN-Ergebnissen, wobei ein JOIN-Partner gerade das Resultat der letzten Vereinigungsoperation ist. Der Zyklus bricht ab, wenn sich die Ergebnismenge nicht mehr verändert. Für das Beispiel mit den Nachfahren ergibt sich:

$\mathbf{N} \Leftarrow \mathbf{Nachfahren}$;
h sei eine Relation mit dem Relationsschema von **Nachfahren**;

Konstruktion der Hülle:

$\mathbf{h} = \emptyset$; $\quad \mathbf{S} \Leftarrow \mathbf{N}$;
WHILE ($\mathbf{S} \neq \mathbf{h}$){
$\mathbf{h} \Leftarrow \mathbf{S}$;
$\mathbf{S} \Leftarrow (\mathbf{S} \cup (\pi_{S.Name,N.Kind}(\mathbf{S} \ltimes_{S.Kind=N.Name} \mathbf{N})))$
}.

Der Ausdruck $\pi_{Kind}(\sigma_{Name=Goethe}(\mathbf{S}))$ liefert als Ergebnismenge die Namen aller Nachfahren von Goethe (immer unter der Annahme, dass der Name die Person eindeutig beschreibt). Das Beispiel kann zu einer allgemeinen Bildungsvorschrift für die rekursive Hülle ausgebaut werden. Für diese ist charakteristisch, dass sie eine Wiederholung (Schleifenkonstruktion) und damit einen prozeduralen Aspekt in die Definition einbringt.

5 Anwendungsbeispiele

In Vorlesungen zur Datenbanktheorie werden eine Reihe von Beispielen gegeben, wie auf einem Datenbankschema verbale Anfragen mit Hilfe der Relationenalgebra semiformal beschrieben werden.
Auch die Umformung in der umgekehrten Richtung spielt eine Rolle: Zu semiformalen Anfragen wird eine umgangssprachliche, aber exakte Beschreibung gesucht.
Diese Fragenstellungen sind meist leicht zu bearbeiten. Fehler entstehen durch Nichtbeachten der Tatsache, dass die Resultatmenge im Allgemeinen mehrere Datensatze umfassen kann, also den Plural in der verbalen Formulierung erfordert.
Weitere Aufgaben beinhalten die Umformung von Anfragen mit dem Ziel der Optimierung, z.B. mit dem Ziel, eine möglichst geringe Netzlast zu erzeugen, wenn zum Beispiel die einzelnen Relationen auf verschiedenen Rechner verteilt sind.
In diesem Abschnitt des Buches soll durch Beispiele aus diesen drei Bereichen das Zusammenwirken der Operationen der Algebra bei Anfragen vorgestellt werden.

5.1 Anfragen auf Relationsschemata

Aufgabenstellung:
Bei den Aufgaben dieses Typs sind mehrere Relationsschemata gegeben und dazu die verbale Beschreibung der Eigenschaften von zu findenden Tupeln. Diese Eigenschaften werden durch die Vorgabe von Werten von Attributen festgelegt. Als

Lösung
einer solchen Aufgabe soll die Angabe einer Vorschrift verstanden werden, die beschreibt, wie mit Hilfe von Operationen der Algebra aus den Relationen zu den gegebenen Schemata eine neue Relation konstruiert wird, die genau die Tupel mit den in der Aufgabestellung geforderten Attributen und den vorgeschriebenen Eigenschaften (Attributwerten) enthält.

Im Allgemeinen erfordert die Bearbeitung solcher Aufgaben die Kenntnis der Semantik der Relationen und der Attribute.

Für die Beispiele dieses Abschnitts wird ein Datenbankschema ([1]) unterstellt, welches die Dramen der deutschen Klassik zum Inhalt hat und aus folgenden Relationen besteht.

[1]Quelle: [6]

Drama$(Autor, \underline{Titel}, U\text{-}Ort, U\text{-}Jahr)$
Rolle$(\underline{Figur}, Titel, Typ)$
Schauspieler$(\underline{PNR}, Name, W\text{-}Ort)$
Darsteller$(\underline{PNR, Figur}, Theater, A\text{-}Ort, A\text{-}Jahr)$

Zur Abkürzung werden die Relationen **Darsteller**, **Schauspieler** und **Rolle** mit **D**, **S** bzw. **R** abgekürzt. Die Namen sind zum Teil selbsterklärend, folgende Erläuterungen beschreiben die Semantik ergänzend:

- Relation **Drama**:
  U-Ort,*U-Jahr* bezeichnen den Uraufführungsort und das Uraufführungsjahr des Dramas.
- Relation **Rolle**:
  Figur ist eine Figur aus einem Drama,
  Titel ein Fremdschlüssel auf die Relation **Drama**,
  Typ beschreibt den Typ der Figur, z.B. „jugendlicher Held".
  Er wird unterstellt, dass es in jedem Drama mindestens eine Figur gibt.
- Relation **Schauspieler**:
  PNR ist eine fiktive eindeutige Personalnummer für jeden Schauspieler.
  W-Ort bezeichne den Wohnort eines Schauspielers.
- Zur Relation **Darsteller**:
  PNR ist Fremdschlüssel auf **Schauspieler**, *Figur* auf **Rolle**,
  Theater bezeichnet das Theater, an dem der Schauspieler mit der Personalnummer *PNR* die Rolle *Figur* gespielt hat,
  A-Ort und *A-Jahr* bezeichnet Ort und Jahr dieser Aufführung.

Zur Vereinfachung wird festgelegt, dass Jahreszahlen wie ganze Zahlen behandelt werden.

Mit Hilfe der Schemadefinitionen und der zusätzlich gegebenen Metainformationen können die Aufgaben zur Umformung von umgangsprachlichen Beschreibungen in semiformale Beschreibungen bearbeitet werden.

Beispiel 5.1: *Formulieren Sie mit Hilfe der Operatoren der Relationenalgebra: Welche Darsteller (PNR) haben im Schauspielhaus Leipzig gespielt?*

Lösung: Die für die Anfrage relevanten Informationen finden sich in der Relation **Darsteller**, die Auswahlbedingung lautet $Theater =$ „Schauspielhaus Leipzig", als Ergebnis wird nur die Auflistung der Personalnummern *PNR* erwartet.
$\pi_{PNR}(\sigma_{Theater=\text{„Schauspielhaus Leipzig"}}(\mathbf{D}))$ ■

Beispiel 5.2: *Zusätzlich zu den Anforderungen aus Beispiel 5.1: soll der Name des Schauspielers mit ausgegeben werden.*

Lösung: Diese zusätzliche Information *Name* befindet sich in der Relation **Schauspieler**. Die Verbindung zur Relation **Darsteller** wird über die in beiden Relationen vorhandenen Attribute *PNR* hergestellt.

$\pi_{PNR,Name}(\mathbf{S} \bowtie \sigma_{Theater=\text{„Schauspielhaus Leipzig“}}(\mathbf{D}))$ ■

Die Namensgebung für die Attribute ermöglicht in dem Beispiel den Einsatz des Natürlichen-Gleichverbund-Operators, alternativ können qualifizierte Attributnamen eingesetzt werden. Dies führ zu einer Lösung mit Hilfe des Gleichverbund-Operators:

$\pi_{PNR,Name}(\mathbf{S} \bowtie_{\mathbf{S}.PNR=\mathbf{D}.PNR} \sigma_{Theater=\text{„Schauspielhaus Leipzig“}}(\mathbf{D}))$

Beide bisher vorgestellten Lösungen dieser Aufgabe führen die Selektion ‚$\sigma_{Theater=\text{„Schauspielhaus Leipzig“}}$‘ unmittelbar auf der Grundrelation **D** aus und nehmen ihr Ergebnis als einen JOIN-Partner. Diese Relation enthält nur noch die für die Anfrage relevanten Datensätze. Das JOIN-Ergebnis hat damit im allgemeinen eine wesentlich kleinere Kardinalität als die Relation **D**.
Auch die folgende Lösung der Aufgabe ist korrekt:

$\pi_{PNR,Name}(\sigma_{Theater=\text{„Schauspielhaus Leipzig“}}(\mathbf{S} \bowtie \mathbf{D}))$

Sie unterscheidet sich von den anderen Lösungen durch die Reihenfolge der Operatoren. Der JOIN-Operator bezieht alle Tupel der Relation **D**, seine Ergebnisrelation hat die gleiche Kardinalität wie die Relation **D**. Dies verlangt in einer realen Datenbank mehr Ressourcen. Deshalb zeugt es von gutem Stil, wenn die Umsetzung von Anfragen so erfolgt, dass Projektionen und Selektionen möglichst auf den Basisrelationen ausgeführt werden und nur die für das Weitere wesentlichen Attribute enthält. In diesem Sinn kann auch die Lösung aus Beispiel 5.2: noch verbessert werden. Der Wohnort eines Schauspielers wird nicht benötigt und kann entfernt werden:

$\pi_{PNR,Name}(\pi_{PNR,Name}(\mathbf{S}) \bowtie \sigma_{Theater=\text{„Schauspielhaus Leipzig“}}(\mathbf{D}))$

Beispiel 5.3: *Finde alle Schauspieler (Name, Wohnort), die (mindestens) einmal im „Faust“ mitgespielt haben.*

Lösung: „Faust“ist der Titel eines Dramas. Das Attribut *Titel* gehört zur Relation **Drama**, tritt aber in der Relation **Rolle** (**R**) wieder auf. Insbesondere enthält **R** alle Titel von Dramen, da jedes Drama mindestens eine Figur hat. Die Verbindung zwischen **S** mit den Attributen *Name* und *W-Ort* und **R** erfordert die Einbeziehung der Relation **D**. Gleichverbünde über die jeweils gemeinsamen Attribute schaffen den Zusammenhang:

$\pi_{Name,W\text{-}Ort}(\mathbf{S} \bowtie \mathbf{D} \bowtie \sigma_{Titel=\text{„Faust“}}(\mathbf{R}))$ ■

Auch in dieser Lösung können noch nicht benötigte Attribute vor der Joinausführung durch Projektionen entfernt werden, im Sinne des besseren Übersicht wurde darauf verzichtet.

In den bisher vorgestellten Beispielen wurden die Join-Operatoren auf Relationen ausgeführt, in denen in natürlicher Weise die Join-Attribute die gleiche Bedeutung haben. Die Semantik von Attributen ist jedoch für den Gleichverbund nicht wesentlich. Es ist ausreichend, wenn eine Vergleichoperation für die entsprechenden Attribute existiert.

Beispiel 5.4: *Finde alle Schauspieler (Name, Wohnort), die bei in Weimar uraufgeführten Dramen an ihrem Wohnort als „Held“ mitgespielt haben.*

Lösung: Die in der Aufgabe gestellten Bedingungen an die Attribute lassen sich in zwei Kategorien einteilen. Einmal sind dies Bedingungen, in denen für Attribute Werte festgelegt werden: $U\text{-}Ort =$ „Weimar“ usw. , diese Bedingungen werden in Selektionen realisiert. Neu ist eine zweite Kategorie, die Gleichheit von Werten verlangt, die zu verschiedenen Attributen verschiedener Relationen gehören. Auf Grund der Semantik kann erwartet werden, dass die Attribute $W\text{-}Ort$ und $A\text{-}Ort$ als Unterkategorien einer allgemeinen Kategorie „Ort “ vergleichbar sind. Diese Annahme führt zu der Join-Bedingung: $\mathbf{S}.W\text{-}Ort = \mathbf{D}.A\text{-}Ort$. Die Definition 3.18: der Joinoperatoren zeigt, dass auch diese Bedingung sich in einer Selektion niederschlägt, die beim Ausführen der Verbundoperatoren realisiert werden.

$$\pi_{Name,W\text{-}Ort}(\mathbf{S} \bowtie_{\mathbf{S}.W\text{-}Ort=\mathbf{D}.A\text{-}Ort} (\mathbf{D} \bowtie \sigma_{R\text{-}Typ=\text{„Held“}}(\mathbf{R}) \bowtie \sigma_{U\text{-}Ort=\text{„Weimar“}}(\mathbf{Drama}))) \qquad \blacksquare$$

Beispiel 5.5: *Finde alle Schauspieler (Name,PNR), die nie eine Rolle in einem klassischen Drama gespielt haben.*

Lösung: Alle Schauspieler sind in der Relation **S** erfasst. Für jeden Schauspieler, der jemals in einem der Dramen mitgewirkt hat, gibt es für jede seiner Rollen einen Eintrag in **D**. Für die die Antwort qualifizieren sich die Schauspieler aus **S**, die nicht in **D** vorkommen. Die Differenzbildung von Relationen liefert ein Ergebnis mit dieser Semantik. Zu ihrer Ausführung ist Vereinigungskompatibilität der beteiligten Mengen notwendige Voraussetzung. Diese Kompatibilität wird erreicht, indem störende Attribute durch Projektionen entfernt werden: $\pi_{PNR}(\mathbf{S}) \setminus \pi_{PNR}(\mathbf{D})$. Leider musste durch die Projektion auch das für das Ergebnis benötigte Attribut $\mathbf{S}.Name$ mit entfernt werden. Eine Joinoperation mit **S** auf dem einzigen Attribut PNR des Zwischenergebnisses sorgt dafür, dass der Name für die abschließende Projektion wieder verfügbar ist:

$$\pi Name, PNR(\mathbf{S} \bowtie_{\mathbf{S}.PNR=PNR} (\pi_{PNR}(\mathbf{S}) \setminus \pi_{PNR}(\mathbf{D}))) \qquad \blacksquare$$

Beispiel 5.6: *Finde alle Schauspieler (Name), die alle Rollen gespielt haben ?*

Lösung: Auf Grund der Semantik der Begriffe ist in der Praxis klar, dass es keinen Schauspieler gibt, der diese Bedingung erfüllt, ein Mephistodarsteller hat sicher nie das Gretchen gespielt. Dennoch kann diese Frage formal korrekt gestellt werden.
Die Frage erfordert festzustellen, welche Personalnummer aus **D** in dieser Relation mit allen Werten des Attributs *Figur* aus **R** verbunden ist. Diese Frage wird durch die Operation der Relationendivision beantwortet. Durch Projektionen wird wieder dafür gesorgt, dass die beiden in die Division eingehenden Relationen die an ihre Relationsschemata zu stellenden Bedingungen erfüllen (vgl. Definition des Divisionsoperators, Seite 37): $\mathbf{D} \div \pi_{Figur}(\mathbf{R})$
oder besser $\pi_{PNR,Figur}(\mathbf{D}) \div \pi_{Figur}(\mathbf{R})$.
Diese Zwischenergebnismenge enthält das Attribut *Name* nicht, deshalb ist eine weitere Joinoperation mit **S** notwendig:

$$\pi_{Name}(\mathbf{S} \bowtie (\pi_{PNR,Figur}(\mathbf{D}) \div \pi_{Figur}(\mathbf{R})))$$ ■

Beispiel 5.7: *Welche Schauspieler (Name, Personalnummer) haben in einer Aufführung eines Dramas eine Doppelrolle gespielt?*

Lösung Der Begriff „Doppelrolle“ soll bedeuten, dass der Schauspieler in einer Aufführung, d.h. im gleichen Jahr am gleichen Theater am gleichen Ort mindestens zwei (verschiedene) Rollen gespielt hat.
Die hier vorgestellte Lösung verzichtet auf die im Abschnitt 4.2.2 vorgestellten Zählfunktionen. Der Kern der Anfrage besteht darin festzustellen, ob es zu einer Personalnummer in der Relation **Darsteller** Paare von Einträgen gibt, die sich nur in dem Attribut *Figur* unterscheiden. Technisch wird dies durch eine Joinoperation auf zwei Kopien von **Darsteller** realisiert. $\mathbf{X} \Leftarrow \mathbf{D}$ erzeugt die Kopie von **D**. Die Bedingung

$$\begin{aligned} B = ((&\mathbf{D}.PNR = \mathbf{X}.PNR) \wedge (\mathbf{D}.A\text{-}Jahr = \mathbf{X}.A\text{-}Jahr) \wedge \\ &(\mathbf{D}.A\text{-}Ort = \mathbf{X}.A\text{-}Ort) \wedge \\ &(\mathbf{D}.Theater = \mathbf{X}.Theater) \wedge \\ &(\mathbf{D}.Figur \neq \mathbf{X}.Figur)) \end{aligned}$$

wird für den Verbund der beiden Kopien von Darsteller benutzt, die zum Ergebnis fehlenden Attribute werden durch eine Join-Operation mit **S** bereitgestellt.

$$\pi_{Name,PNR}(\mathbf{S} \bowtie \pi_{PNR}(\mathbf{D} \bowtie_B \mathbf{X}))$$ ■

Das folgende Beispiel erscheint wesentlich schwieriger. Die Ursache liegt darin, dass die hier angegebenen Relationen nur (binäre) Beziehungen zwischen den Kategorien **Person**, **Biersorte** und **Kneipe** beschreiben, die Kategorien selbst im Datenbankschema nicht auf jedoch nicht auftreten.

Beispiel 5.8: Biertrinken *(*[2]*)*
$\mathbf{Gast}(Person, Kneipe)$
$\mathbf{Vorlieben}(Person, Biersorte)$
$\mathbf{Angebot}(Kneipe, Biersorte)$
Für die Relationen benutzen wir die Abkürzungen $\mathbf{G}, \mathbf{V}, \mathbf{A}$.
Welche Personen besuchen besuchen mindestens eine Kneipe, die mindestens eines ihrer Lieblingsbiere hat?

Lösung: Die Lösung wird schrittweise mit Hilfe des Zuweisungsoperators aufgebaut:

$$\mathbf{Q}_{QPerson,QKneipe,QBier} \Leftarrow \pi_{Person,Kneipe,Bier}(\sigma_{\mathbf{G}.Kneipe=\mathbf{A}.Kneipe}(\mathbf{G} \times \mathbf{A}))$$

Q enthält zu einer Person, welche Biere die von ihr besuchten Kneipen führen. Mit der Einführung von **Q** wurden die Attribute umbenannt. Diese Umbenennung bereitet folgende Produktbildung vor: $\mathbf{P} \Leftarrow \mathbf{V} \times \mathbf{Q}$
P hat das Relationsschema

$\mathbf{P}(Person, Bier, QPerson, QKneipe, QBier)$.

Die folgende Relation **R** enthält die gesuchten Personen:

$$\mathbf{R} = \pi_{Person}(\sigma_{Person=QPerson \wedge Bier=QBier}(\mathbf{P}))$$ ■

5.2 Anfragen unter Nutzung der Algebraerweiterungen

Das vorliegende Kapitel soll demonstrieren, wie insbesondere die Zählfunktionen zur Beantwortung von Anfragen nach quantitativen Eigenschaften genutzt werden.

Beispiel 5.9: *Wieviele Dramen gibt es?*

Lösung: Es ist die Kardinalität der Relation **Drama** zu ermitteln. Entsprechend der auf Seite 49 getroffenen Vereinbarung, was unter einer Lösung einer Aufgabe zu verstehen ist, wird mit Hilfe des Einbettungsoperators das numerische Ergebnis der Anwendung des Zählfunktionals in eine Relation transformiert.

$$\chi_{Anzahl}(\sharp(\mathbf{Drama}))$$ ■

Beispiel 5.10: *Welches ist das älteste Drama von Schiller?*

Lösung: Da die Daten nur die Ermittlung des Jahres der Uraufführung gestatten, kann das Ergebnis nur eine Liste der Titel der Dramen Schillers sein, die alle in dem Jahr ihre Uraufführung hatten, in dem auch das älteste Drama dieses Autors erstmals gespielt wurde.

---

[2] Zur Semantik vgl. Seite 25.

Die Anfrage wird in mehreren Schritten beantwortet. Zuerst wird in einer Zwischenrelation abgelegt, wann welcher Autor sein ältestes Werk geschrieben hat:
$\mathbf{Q} \Leftarrow \chi_{Autor;M\text{-}Jahr}(MIN_{A\text{-}Jahr}(\mathbf{Drama}))$
Diese Relation wird mit der Relation **Drama** verbunden, dabei wird die Bedingung $\mathbf{Drama}.U\text{-}Jahr = \mathbf{Q}.M\text{-}Jahr$ gestellt. Durch Selektionen wird auf $Autor =$ „Schiller“ eingeschränkt:

$$\pi_{Titel}(\sigma_{Autor=\text{„Schiller“}}(\mathbf{Drama}) \\ \bowtie_{\mathbf{Drama}.U\text{-}Jahr=\mathbf{Q}.M\text{-}jahr} \sigma_{Autor=\text{„Schiller“}}(\mathbf{Q}))$$

Eine alternative Lösung ergibt sich mit der Join-Bedingung
$B = (\mathbf{Drama}.Autor = \mathbf{Q}.Autor) \wedge$
$(\mathbf{Drama}.U\text{-}Jahr = \mathbf{Q}.M\text{-}Jahr)$ und dem Ausdruck

$$\pi_{Titel}(\mathbf{Drama} \bowtie_B \sigma_{Autor=\text{„Schiller“}}(\mathbf{Q}))$$

In dieser Lösung kommt der Name des Autors nur einmal explizit vor, die Umformulierung für andere Autoren ist einfacher.

■

Beispiel 5.11: *Welcher Schauspieler (Name) spielte häufiger den „Faust“ als den “Ferdinand“, wobei er aber beide Rollen gespielt haben soll?*

Lösung: Es wird unterstellt, dass in jeder Spielzeit (A-Jahr) ein Eintrag erfolgt, wenn ein Schauspieler eine Rolle spielt. Damit können die Anzahlen durch Zählen auf der Relation **D** ermittelt werden. Da der Selektionsoperator nur Bedingungen in Form boolscher Ausdrücke auf den Attributen einer Relation erlaubt, müssen die Anzahlen als Attribute einer Relation realisiert werden:
$\mathbf{Q} \Leftarrow \chi_{PNR,Figur;Anzahl}(\sharp(\mathbf{D}))$
Diese Relation enthält für jede Person und jede Rolle (*Figur*) die Anzahl, wie oft der Schauspieler diese Rolle in verschiedenen Jahren, Orten und Theatern gespielt hat. Damit läßt sich die Aufgabe lösen: (Abkürzungen wie oben)

$$\mathbf{P} \Leftarrow \mathbf{Q}$$
$$\mathbf{R}(PNR, F_1, A_1, F_2, A_2) \Leftarrow$$
$$\pi_{\mathbf{Q}.PNR,\mathbf{Q}.Figur,\mathbf{Q}.Anzahl,\mathbf{P}.Figur,\mathbf{P}.Anzahl}(\mathbf{Q} \bowtie_{\mathbf{Q}.PNR=\mathbf{P}.PNR} \mathbf{P})$$

$$\pi_{Name}(\mathbf{S} \bowtie \\ \pi_{PNR}(\\ \sigma_{(F_1=\text{„Faust“})\wedge(F_2=\text{„Ferdinand“})\wedge(A_1>A_2)\wedge(A_2>0)}(\mathbf{R}) \\ ) \\ )$$

■

Beispiel 5.12: *Wer (Autor) schrieb das älteste Drama?*

Lösung: Auf Grund der Semantik der Daten muss die Frage präzisiert werden. Es liegen nur Informationen über das Jahr der Uraufführung in der Datenbank vor.

Mit diesen kann die folgende Frage beantwortet werden:
Von welchen Autoren gibt es mindestens ein Drama, das in dem Jahr uraufgeführt wurde, in dem erstmalig eine Uraufführung stattfand?

$\mathbf{Q} \Leftarrow \chi_{Erste}(Min_{U\text{-}Jahr}(\mathbf{Drama})$

$\mathbf{P} \Leftarrow \mathbf{Drama} \times \mathbf{Q}$

Da **Q** in jedem Falle nur die Ordnung 1 und die Kardinalität 1 hat, hat **P** im Vergleich zu **Drama** ein zusätzliches Attribut *Erste*, welches in allen Tupeln den Wert der Jahreszahl der ältesten Uraufführung hat.
Damit ergibt sich als Lösung:

$\pi_{Autor}(\sigma_{U\text{-}Jahr=Erste}(\mathbf{P}))$ ■

5.3 Interpretation von Algebraausdrücken

Jetzt werden (mehrere) Relationsschemata und ein auf diesen definierter Ausdruck mit Operatoren der Relationenalgebra.

Als **Lösung** im Sinne dieses Abschnitts gilt eine umgangssprachliche Beschreibung, welche **genau die Tupelmenge charakterisiert**, die sich als Ergebnisrelation bei Anwendung des Ausdrucks ergäbe.

Typische Stolpersteine liegen dabei in der korrekten Benutzung von Singular und Plural. Als Antwort wird meist eine Menge unbestimmter Kardinalität erwartet, also ist dann der Plural zu nutzen: „Welche Schauspieler haben ...?“ statt „Welcher Schauspieler hat ...?“. Ein weiterer Punkt ist die häufig unpräzise Semantik der Umgangssprache. So bedeutet „Es gibt einen Schauspieler mit dem Namen ‚Schmidt‘“ , dass es mindestens einen Schauspieler mit diesem Namen gibt. Diese Probleme sind in der Mathematik bekannt und haben zur Einführung solcher Sprachausdrücke wie „Es gibt *genau einen* Schauspieler ...“ oder „Es gibt *einen und nur einen* Schauspieler ...“. Die exakte Formulierung in der Umgangssprache erfordert häufig die Benutzung dieses mathematischen Schrachgebrauchs.

Für die folgenden Beispiele wird wieder die Datenbank mit den klassischen Dramen unterstellt (siehe Seite 49). Auch werden die Abkürzungen **Schauspieler S**, **Darsteller D** und **Rolle R** wieder verwendet.

Beispiel 5.13:

$\pi_{Name}(\mathbf{S} \bowtie \sigma_{A\text{-}Jahr=2004 \wedge A\text{-}Ort=„Leipzig“}(\mathbf{D}))$

Lösung: Welche Schauspieler (Name) haben im Jahre 2004 in Leipzig (mindestens einmal) gespielt? ■

Beispiel 5.14:

$$\pi_{Wohnort}(\mathbf{S} \bowtie \pi_{PNR,Figur}(\mathbf{D} \bowtie \pi_{Figur}(\sigma_{R\text{-}Typ=„Liebhaber“}(\mathbf{R})))$$

Lösung: Welches sind die Wohnorte von Schauspielern, die mindestens einmal eine Rolle als „Liebhaber“ gespielt haben ? ■

Beispiel 5.15:

$$\pi_{Name,PNR}(\mathbf{S} \bowtie (\pi_{PNR}(\sigma_{A\text{-}Ort=„Weimar“}(\mathbf{D})) \setminus \pi_{PNR}(\sigma_{W\text{-}Ort=„Leipzig“}(\mathbf{S}))))$$

Lösung: Welche Schauspieler (Name, Personalnummer), die mindestens einmal in Weimar gespielt haben, wohnen nicht in Leipzig? Die Negation „wohnen nicht in Leipzig“ ergibt sich als Folge der Mengendifferenzbildung. ■

5.4 Operatorbäume und Anfrageoptimierung

Abschnitt in Vorbereitung

6 Testaufgaben

In diesem Kapitel werden einige Aufgaben aus Klausuren zur Bearbeitung angeboten. Die Lösungen werden im Anhang vorgestellt. Für diese Aufgaben wird ein neues Datenbankschema eingeführt, dessen Semantik zunächst beschrieben wird.

6.1 Zum Bibliotheksschema

Die Bibliotheksdatenbank wird durch folgendes Schema beschrieben.

Verlage$(\underline{Name}, Ort)$
Autoren$(\underline{ANR}, Name, Vorname)$
Buecher$(\underline{ISBN}, Titel, VName, Jahr, Preis)$
Rolle$(\underline{ANR, ISBN}, Rolle, Rang)$
Exemplare$(\underline{ISBN, ExNR}, Standort)$
Leser$(\underline{LNR}, Name, Ort, Anschrift)$
Ausleihe$(LNR, \underline{ISBN, ExNR}, Datum)$

Diese Relationen sollen bei Bedarf mit **V**, **A**, **B**, **R**, **E** , **L** und **AL** abgekürzt werden. Ihre Bedeutung wird im einzelnen noch durch einige Angaben erläutert.

Verlage: Es wird unterstellt, daß der Name (Attribut *Name*) eines Verlags diesen eindeutig identifiziert. Das Attribut *Ort* kennzeichne den Ort, an dem der Verlag seine Erzeugnisse herausgibt. Es wird vereinfachend angenommen, dass es zu einem Verlag nur einen Ort gibt.

Autoren: Da die Attribute *Name* und *Vorname* eine Person nicht eindeutig kennzeichnen, wird eine eindeutige Autorennummer (*ANR*) vergeben. In der Relation **Autoren** werden neben Autoren auch Herausgeber erfaßt. Die konkrete Tätigkeit einer Person an einem bestimmten Buch wird in dem Attribut *Rolle* in der Relation **Rolle** beschrieben.

Buecher: Ein Tupel in dieser Relation beschreibt ein Verlagsprodukt, in der Fachsprache der Bibliothekare Buchtitel genannt, aber nicht das konkrete Buchexemplar, welches im Regal steht. das Attribut ISB-Nummer (*ISBN*) ist ein im Verlagswesen gebräuchlicher Code, er identifiziert ein Buch eindeutig. *VName* gibt den Namen des Verlages an, in dem das Buch erschienen ist (Fremdschlüssel). Es wird hier unterstellt, daß jedes Buch in genau einem Verlag verlegt wurde. *Jahr* ist das Erscheinungsjahr, *Preis* sei ein Preis in Euro.

Rolle: An einem Buch können mehrere Autoren oder Herausgeber beteiligt sein. Das Attribut *rang* gibt in Form einer natürlichen Zahl an, an welcher Position eine Person in der Liste der Autoren bzw. Herausgeber steht.
Das Attribut *Rolle* habe die Werte „a“ für einen Autor und „h“ für einen Herausgeber.

Exemplare: In der Bibliothek befinden sich Exemplare eines Buchtitels. Diese werden mit dem Attribut *ExNr* durchnummeriert und zwar so, daß die Numerierung für jeden Buchtitel von Eins beginnt. Somit kann ein konkretes Buch durch die gemeinsame Angabe von *ISBN* und *ExNr* bestimmt werden. Der Standort beschreibt, wo das Buch seinen Platz im Magazin der Bibliothek hat.

Leser: Wie oben schon bemerkt, identifiziert ein Name eine Person nicht eindeutig. Um diesem Umstand Rechnung zu tragen, wird für jeder Leser eine eindeutige Lesernummer *LNR* vergeben. *Ort* sei der Wohnort des Lesers.

Ausleihe: Hier wird erfaßt, welcher Leser welches Buch(exemplar) wann ausgeliehen hat. Bei der Rückgabe wird der entsprechende Eintrag gelöscht, so daß stets nur die im Augenblick ausgeliehenen Bücher erfaßt sind. Das Attribut *Datum* gibt das Datum an, an dem die Ausleihe erfolgt. Für dieses Attribut wird unterstellt, daß Datumsangaben wie Zahlen verglichen werden können, ein Datum D_1, welches vor einem Datum D_2 liegt, erfüllt „$D_1 < D_2$“. Der Ausdruck „23.8.04 < 3.9.04“ ist also „wahr“ ([1]).

Unter Benutzung des Relationenschema „Bibliothek“sind für die folgenden Fragen Ausdrücke in der Relationenalgebra anzugeben.

Aufgabe 6.1: *An welchem Ort verlegt der Verlag „Edition am Gutenbergplatz“?*

Aufgabe 6.2: *Welche Anschriften haben die Leipziger Leser?*

Aufgabe 6.3: *Welche Bücher (ISBN), die in der Bibliothek vorhanden sind, wurden im Verlag „Edition am Gutenbergplatz“ verlegt?*

Aufgabe 6.4: *Geben Sie eine Liste der Bücher (Titel, ISBN, Erscheinungsort) aus. Dabei soll zu jedem Buchtitel die Angabe des Verlages vorhanden sein.*

---

[1] In Datenbankverwaltungssystemen wird ein spezieller Datentyp mit sehr detailierten Eigenschaften für das Datum bereitgestellt. Aus Gründen einer einfachen Darstellung wird er hier nicht vorgestellt. Um eine Vorstellung von der Vergleichbarkeit mit dem „<“-Zeichen zu erhalten, kann angenommen werden, daß ein Datum intern in Form des Julianischen Datums der Astronomie als reelle Zahl dargestellt und verglichen wird.

Aufgabe 6.5: *Geben Sie eine Liste der Bücher (Titel, ISBN, Erscheinungsort) aus. Es gibt Buchtitel, zu denen kein Verlagsname angegeben ist. Diese sollen mit ausgegeben werden.*

Aufgabe 6.6: *Wie Aufgabe 6.5:, jedoch soll der OUTER-JOIN-Operator nicht verwendet werden und in der Ausgabe soll bei den Buchtiteln mit fehlendem Verlag der Hinweis „Verlag unbekannt" erscheinen.*

Aufgabe 6.7: *Welche Leser (Name) haben ein Buch mit dem Titel „Relationenalgebra" ausgeliehen?*

Aufgabe 6.8: *Welche Autoren (Name, Vorname) waren nur als Herausgeber tätig, jedoch nie als Autor im engeren Sinn des Wortes?*

Aufgabe 6.9: *Welche Leser wohnen an dem Ort, an dem der B.G.-Teubner-Verlag seine Bücher veröffentlicht?*

Aufgabe 6.10: *Welche Leser (Name, Lesernummer) haben derzeit kein Buch ausgeliehen?*

Aufgabe 6.11: *Welche Leser (Lesernummer) haben alle Bücher ausgeliehen, an denen Erhard Rahm als Herausgeber beteiligt ist?*

Aufgabe 6.12: *Aufgabenstellung wie eben, jedoch sollen die Namen der Leser ausgeben werden.*

Aufgabe 6.13: *Welche Leser (Name, Lesernummer) haben das Buch „Relationenalgebra" (mindestens) zweimal aufgeliehen? Es wird dabei unterstellt, daß es nur ein Buch mit diesem Titel, aber in mehreren Exemplaren gibt. (Hinweis: Diese Frage kann und soll hier ohne Benutzung der Algebraerweiterung durch Zählfunktionen gelöst werden.)*

6.2 Aufgaben auf dem Bierschema

Für die folgenden Aufgaben wird das Bierschema vorausgesetzt **Bierschema:**
$\mathbf{Gast}(Person, Kneipe)$
$\mathbf{Vorlieben}(Person, Biersorte)$
$\mathbf{Angebot}(Kneipe, Biersorte)$
Für die Relationen benutzen wir die Abkürzungen $\mathbf{G}, \mathbf{V}, \mathbf{A}$.
Erläuterungen:
Gast: Die Tupel dieser Relation drücken aus, welche Person welche Gaststätte besucht.
Vorlieben: Welche Person bevorzugt welche Biersorte.
Angebot: Welche Gaststätte hat welche Biersorte im Angebot.

Aufgabe 6.14: *Welche Biersorten sind in der Datenbank erfaßt?*

Aufgabe 6.15: *Welche Biersorten im Angebot sind nicht Lieblingsbier irgendeiner Person?*

Aufgabe 6.16: *Welche Gäste finden keines ihrer bevorzugten Biere im Angebot?*

Aufgabe 6.17: *Welche Gäste sind in allen Kneipen gewesen?*

Aufgabe 6.18: *Welche Personen sind sind in einer Kneipe gewesen, die eines der Lieblingsbiere der Person führt?*

Aufgabe 6.19: *Welche Personen sind nur in Kneipen gewesen, die eines der Lieblingsbiere der Person führt?*

Aufgabe 6.20: *Welche Personen sind in einer Kneipe gewesen, die nur die Lieblingsbiere der Person haben?*

Aufgabe 6.21: *Welche Personen sind nur in Kneipen gewesen, die nur die Lieblingsbiere der Person haben?*

6.3 Aufgaben zu den Aggregatfunktionen

Für die folgenden Aufgaben wird das Bibliotheksschema (Seite 59) vorausgesetzt.

Aufgabe 6.22: *Auf dem Bibliotheksschema ist die Gesamtzahl der Buchexemplare zu ermitteln.*

Aufgabe 6.23: *Wieviele Buchtitel sind ausgeliehen? (Ein Buchtitel hat seine eigene ISBN.)*

Aufgabe 6.24: *Welchen Durchschnittspreis hat ein Buchtitel?*

Aufgabe 6.25: *Welchen Wert stellen die Bücher dar?*

Aufgabe 6.26: *Welche Bücher (ISBN) sind am ältesten?*

Die letzten Aufgaben greifen nochmals auf das Bierschema (Seite 61) zurück.

Aufgabe 6.27: *In wievielen Kneipen war jede Person, die in mindestens einer Kneipe war?*

Aufgabe 6.28: *In wievielen Kneipen war jede im Schema erfasste Person?*

Literaturverzeichnis

[1] Codd, E.: *A Relational Model for Large Shared Data Banks*. CACM, 13:6, Juni 1970.

[2] Elmasri, E. und Navathe, R.E.: *Grundlagen von Datenbanksystemen*. 3.Auflage,dt.,Pearson Studium, München, 2002.
ISBN 3-8273-7021-3

[3] Klaua, D.: *Allgemeine Mengenlehre*. Akademie-Verlag, Berlin, 1964.

[4] Lausen, G. und Vossen, G.: *Objektorientierte Datenbanken: Modelle und Sprachen*. R. Oldenburg Verlag München, 1996.

[5] Naas, J. und Schmid, H.L. (Herausgeber): *Mathematisches Wörterbuch, Bd. 1 und 2*. Akademieverlag Berlin und B.G. Teubnerverlagsgesellschaft Leipzig, 1961.

[6] Rahm, E.: *Datenbanksysteme 1*.
Vorlesung an der Universität Leipzig.
URL: http://dbs.uni-leipzig.de/en/lehre/db-lernmaterial-vorl.html

[7] Sosna, D.: *Lese- und Übungsbuch Datenbanken:*
E/R- und Relationenmodell
URL:

7 Anhang: Lösung der Testaufgaben

Viele Aufgaben im Bereich der Anwendung der Relationenalgebra können auf unterschiedlichen Wegen gelöst werden, so dass die in diesem Kapitel vorgestellten Lösungen nur als Beispiel anzusehen sind.

7.1 Zum Bibliotheksschema

Für die ersten Aufgaben wurde folgendes Schema verwendet (vgl. Seite 59):

$\mathbf{Verlage}(\underline{Name}, Ort)$
$\mathbf{Autoren}(\underline{ANR}, Name, Vorname)$
$\mathbf{Buecher}(\underline{ISBN}, Titel, VName, Jahr, Preis)$
$\mathbf{Rolle}(\underline{ANR, ISBN}, Rolle, Rang)$
$\mathbf{Exemplare}(\underline{ISBN, ExNR}, Standort)$
$\mathbf{Leser}(\underline{LNR}, Name, Ort, Anschrift)$
$\mathbf{Ausleihe}(LNR, \underline{ISBN, ExNR}, Datum)$

Diese Relationen sollen bei Bedarf mit **V**, **A**, **B**, **R**, **E** , **L** bzw. **AL** abgekürzt werden.

Aufgabe 7.1: *An welchem Ort verlegt der Verlag „Edition am Gutenbergplatz“?*

Lösung: $\pi_{Ort}(\sigma_{Name=\text{'Edition am Gutenbergplatz'}}(\mathbf{V}))$ ■

Aufgabe 7.2: *Welche Anschriften haben die Leipziger Leser?*

Lösung: Die Frage wird als Frage nach Name und Anschrift der Leser, die in Leipzig wohnen, interpretiert, denn es macht wenig Sinn, nur eine Liste ohne Namen zu erhalten.

$\pi_{Name,Anschrift}(\sigma_{Ort=\text{'Leipzig'}}(\mathbf{L}))$

Um beispielsweise die Ausgabe zum Adressdruck zu nutzen, wäre die vollständige Anschrift mit Nennung des Ortsnamens besser geeignet, auch wenn dieser durch die Aufgabenstellung festgelegt ist:

$\pi_{Name,Ort,Anschrift}(\sigma_{Ort=\text{'Leipzig'}}(\mathbf{L}))$ ■

Aufgabe 7.3: *Welche Bücher (ISBN), die in der Bibliothek vorhanden sind, wurden im Verlag „Edition am Gutenbergplatz“ verlegt?*

Lösung: Entgegen dem ersten Anschein muss zur Beantwortung der Frage nur eine Relation genutzt werden. Der Name der Verlage kommt als Fremdschlüssel in der Relation **Buecher** vor.

$\pi_{ISBN}(\sigma_{VName='\text{Edition am Gutenbergplatz}'}(\mathbf{B}))$ ■

Aufgabe 7.4: *Geben Sie eine Liste der Bücher (Titel, ISBN, Erscheinungsort) aus. Dabei soll zu jedem Buchtitel die Angabe des Verlages vorhanden sein.*

Lösung: Die benötigten Attribute sind in den Relationen **B** und **V**, Ihre Verbindung erfolgt über den Fremdschlüssel $VName$ in **B** auf **V**:

$\pi_{Titel,ISBN,Ort}(\mathbf{B} \bowtie_{\mathbf{B}.VName=\mathbf{V}.Name} \mathbf{V})$ ■

Aufgabe 7.5: *Geben Sie eine Liste der Bücher (Titel, ISBN, Erscheinungsort) aus. Es gibt Buchtitel, zu denen kein Verlagsname angegeben ist. Diese sollen mit ausgegeben werden.*

Lösung: Gegenüber der vorangehenden Lösung 7.4: ergibt sich nur eine Modifikation, der JOIN-Operator ist durch den linken äußeren Gleichverbund (LEFT-OUTER-JOIN-Operator) zu ersetzen:

$\pi_{Titel,ISBN,Ort}(\mathbf{B} ⟕_{\mathbf{B}.VName=\mathbf{V}.Name} \mathbf{V})$ ■

Aufgabe 7.6: *Wie Aufgabe 6.5:, jedoch soll der OUTER-JOIN-Operator nicht verwendet werden und in der Ausgabe soll bei den Buchtiteln mit fehlendem Verlag der Hinweis „Verlag unbekannt“ erscheinen.*

Lösung: Der Lösungsweg wird schrittweise aufgebaut:

$\mathbf{T_1} \Leftarrow \pi_{Titel,ISBN,Ort}(\mathbf{B} \bowtie_{\mathbf{B}.VName=\mathbf{V}.Name} \mathbf{V})$

enthält die gewünschten Angaben für alle Bücher, die einen Verlag haben (vgl. Lösung 7.4:.

$\mathbf{H} \Leftarrow \pi_{Titel,ISBN}(\mathbf{B}) \setminus \pi_{Titel,ISBN}(\mathbf{T_1})$

enthält die *Titel* und *ISBN* der Bücher, die keinen Verlag haben. Mit Hilfe von Einbettungsoperator und Kreuzprodukt werden hier die fehlende Information ergänzt:

$\mathbf{T_2} \Leftarrow \mathbf{H} \times \chi_{Ort}(\text{„Verlag unbekannt“})$

Durch Vereinigung von $\mathbf{T_1}$ und $\mathbf{T_2}$ ergibt sich die Gesamtanfrage:

$\mathbf{T_1} \cup \mathbf{T_2}$. ■

Aufgabe 7.7: *Welche Leser (Name) haben ein Buch mit dem Titel „Relationenalgebra“ ausgeliehen?*

Lösung: Zunächst wird eine Lösung angegeben, die das Ergebnis schrittweise aufbaut:

$\mathbf{Q}_1 \Leftarrow \pi_{ISBN}(\sigma_{Titel='\text{Relationenalgebra}'}(\mathbf{B}))$

$\mathbf{Q}_2 \Leftarrow \pi_{LNR}(\mathbf{A} \bowtie \mathbf{Q}_1)$

$\mathbf{Q}_3 \Leftarrow \pi_{Name}(\mathbf{L} \bowtie \mathbf{Q}_3)$

$\mathbf{Q}_3$ beinhaltet das Ergebnis, die Projektionen bewirken, dass nur die für die weitere Konstruktion der Antwort benötigten Attribute im jeweils folgenden Schritt auftauchen.

Selbstverständlich läßt sich alles ohne Zuweisungsoperator ausdrücken, wobei auch auf die Projektion zum Entfernen nicht weiter benötigter Spalten verzichtet wird:

$\pi_{Name}(\mathbf{L} \bowtie (\mathbf{A} \bowtie (\sigma_{Titel='\text{Relationenalgebra}'}(\mathbf{B}))))$

Wegen der Assoziativität der JOIN-Operation kann auf die Klammersetzung bei der JOIN-Benutzung verzichtet werden:

$\pi_{Name}(\mathbf{L} \bowtie \mathbf{A} \bowtie \sigma_{Titel='\text{Relationenalgebra}'}(\mathbf{B}))$ ■

Aufgabe 7.8: *Welche Autoren (Name, Vorname) waren nur als Herausgeber tätig, jedoch nie als Autor im engeren Sinn des Wortes?*

Lösung: Aus der Relation **Rolle** werden die Herausgeber extrahiert und vom Ergebnis die Autoren durch Differenbildung entfernt, die mindestens einmal mit den Attribut $Rolle = '\text{a}'$ eingetragen waren. Die Differenzbildung erfordert Vereinigungskompatibilität, die durch Projektion erzwungen wird. Zum Schluss sorgt ein JOIN mit **Autoren** für das Bereitstellen der Attribute *Name*, *Vorname*:

$\pi_{Name,Vorname}(\mathbf{A} \bowtie_{ANR} (\pi_{ANR}(\sigma_{Rolle='\text{h}'}(\mathbf{R})) \setminus \pi_{ANR}(\sigma_{Rolle='\text{a}'}(\mathbf{R}))))$ ■

Aufgabe 7.9: *Welche Leser wohnen an dem Ort, an dem der B.G.-Teubner-Verlag seine Bücher veröffentlicht?*

Lösung: In der ersten Teilanfrage werden die Orte ermittelt, an denen der Verlag mit dem Namen „B.G.-Teubner“ publiziert:

$\mathbf{Q}_1 \Leftarrow \pi_{Ort}(\sigma_{Name='\text{B.G.-Teubner}'}(\mathbf{V}))$

Dann erfolgt der Vergleich dieser Orte mit den Wohnorten der Leser, da keine Einschränkungen vorgeschrieben wurden, werden alle Attribute in das Ergebnis aufgenommen. Wegen der Eigenschaften des Gleichverbundoperators erscheint Attribut *Ort* nur einmal im Relationsschema des Ergebnisses $\mathbf{Q}_2$.

$\mathbf{Q}_2 \Leftarrow (\mathbf{L} \bowtie \mathbf{Q}_1)$

Offensichtlich $\mathbf{Q}_2$ hat dasselbe Relationsschema wie **Leser**, da das Join-Attribut *Ort* nur einmal im Ergebnis erscheint. $\mathbf{Q}_2$ enthält genau die gesuchten Leser. Nachdem die Lösung schrittweise aufgebaut wurde, kann sie auchin einem Ausdruck angegeben werden:

$\mathbf{L} \bowtie \pi_{Ort}(\sigma_{Name='\text{B.G.-Teubner}'}(\mathbf{V}))$ ■

Aufgabe 7.10: *Welche Leser (Name, Lesernummer) haben derzeit kein Buch ausgeliehen?*

Lösung:Zuerst wird die Anfrage logisch äquivalent umgeformt:*Welche Leser stehen mit ihrer Lesernummer nicht in der Ausleihe.* Solche Umformulierungen bieten häufig den Ansatz zur Bearbeitung, wenn sich die neue Formulierung durch Operatoren der Algebra ausdrücken lässt:

$\pi_{Name,LNR}(\mathbf{L} \bowtie (\pi_{LNR}(\mathbf{L}) \setminus \pi_{LNR}(\mathbf{A})))$

Um die Differenz zu bilden, ist zunächst die Vereinigungsverträglichkeit durch Projektionen zu erzwingen. Die weiter benötigten Attribute der Relation **Leser**, die durch die Projektion zunächst verloren waren, werden danach durch einen natürlichen Gleichverbund auf dem Attribut *LNR* wieder hinzugefügt. ■

Aufgabe 7.11: *Welche Leser (Lesernummer) haben alle Bücher ausgeliehen, an denen* Erhard Rahm *als Herausgeber beteiligt ist?*

Lösung: Das Wort *alle* in der Aufgabenstellung ist ein sicheres Indiz für Benutzung des Divisionsoperators. Die Ausführung der Relationendivision ist an Bedingungen gebunden (vgl. Seite 37), deren Einhaltung durch die Konstruktion geeigneter Zwischenergebnisse vorbereitet wird:

$\mathbf{Q}_1 \Leftarrow \pi_{ISBN}(\sigma_{Rolle='\text{h}'}(\mathbf{R}) \bowtie_{ANR} \sigma_{Name='\text{Rahm}' \wedge Vorname='\text{Erhard}'}(\mathbf{A}))$

$\mathbf{Q}_1$ enthält die ISBN aller Bücher, bei denen Erhard Rahm Herausgeber ist. Damit ergibt sich das Ergebnis $\mathbf{Q}_2$ zu:

$\mathbf{Q}_2 \Leftarrow \pi_{LNR}(\mathbf{A} \div \mathbf{Q}_1)$ ■

Aufgabe 7.12: *Aufgabenstellung wie eben, jedoch sollen die Namen der Leser ausgeben werden.*

Lösung: $\mathbf{Q}_1$ und $\mathbf{Q}_2$ werden wie eben konstruiert und mit der Relation **Leser** verbunden, so dass die Lesernummer *LNR* für die Projektion bereitsteht:

$\pi_{Name}(\mathbf{L} \bowtie_{LNR} \mathbf{Q}_2)$ ■

Aufgabe 7.13: *Welche Leser (Name, Lesernummer) haben das Buch „Relationenalgebra" (mindestens) zweimal aufgeliehen? Es wird dabei unterstellt, dass es nur ein Buch mit diesem Titel, aber in mehreren Exemplaren gibt. (Hinweis: Diese Frage kann und soll hier ohne Benutzung der Algebraerweiterung mit ihren Zählfunktionen gelöst werden.)*

Lösung: Es werden zwei Instanzen der Relation **Ausleihe** benutzt und mit einem Verbundoperator dort Einträge mit gleichen Lesernummern und ISBN, aber verschiedenen Exemplarnummern gesucht:

$\mathbf{R} \Leftarrow \mathbf{A}$
$\mathbf{P} \Leftarrow \pi_{LNR}(\mathbf{R}$
$\bowtie_{\mathbf{R}.LNR=\mathbf{A}.LNR \wedge \mathbf{R}.ISBN=\mathbf{A}.ISBN \wedge \mathbf{R}.ExNR<\mathbf{A}.ExNR}$
$\mathbf{A})$

Die JOIN-Bedingung, insbesondere $\mathbf{R}.ExNR < \mathbf{A}.ExNR$, sorgt für die richtige Auswahl. Die Angaben zum Leser ergeben sich nach einer weiteren Verbundoperation:

$\pi_{Name,LNR}(\mathbf{L} \bowtie \mathbf{P})$ ■

7.2 Zum Bierschema

Bierschema:(1)

$\mathbf{Gast}(Person, Kneipe)$
$\mathbf{Vorlieben}(Person, Biersorte)$
$\mathbf{Angebot}(Kneipe, Biersorte)$

Für die Relationen benutzen wir die Abkürzungen $\mathbf{G}, \mathbf{V}, \mathbf{A}$.

Aufgabe 7.14: *Welche Biersorten sind in der Datenbank erfaßt?*

Lösung:Biersorten können sowohl in der Relation **Angebot** als auch in **Vorliebe** auftreten, beide Relationen können Einträge enthalten, die in der jeweils anderen nicht vorkommen, folglich ergeben sich alle Biersorten aus einer Vereinigung, für deren Ausführbarkeit die Vereinigungskompatibilität durch Projektion erreicht wird:

$\pi_{Biersorte}(\mathbf{Angebot}) \cup \pi_{Biersorte}(\mathbf{Vorlieben})$

■

Aufgabe 7.15: *Welche Biersorten im Angebot sind nicht Lieblingsbier irgendeiner Person?*

Lösung:

$\pi_{Biersorte}(\mathbf{Angebot}) \setminus \pi_{Biersorte}(\mathbf{Vorlieben})$ ■

[1] Semantische Erläuterungen auf Seite 61

Aufgabe 7.16: *Welche Gäste finden keines ihrer bevorzugten Biere im Angebot?*

Lösung: Die Lösungsmenge bilden alle Gäste, die ein Lieblingsbier haben, mit Ausnahme derer, die ihr Lieblingsbier in keinem Angebot finden. $\mathbf{Q} \Leftarrow \mathbf{V} \times \mathbf{A}$ liefert alle Kombinationen der Paare $(Person, Biersorte)$ mit allen Paaren $(Kneipe, Biersorte)$. Werden aus diesen Quadrupeln diejenigen selektiert, bei denen die Werte in den beiden Biersortenattributen übereinstimmt, enthält das Ergebnis die Menge von Personen, die in mindestens einer Kneipe eine ihrer Lieblingssorten finden können:

$\mathbf{P} \Leftarrow \sigma_{\mathbf{V}.Biersorte=\mathbf{A}.Biersorte}(\mathbf{V} \times \mathbf{A})$

Als Endergebnis ergibt sich:

$\pi_{Person}(\mathbf{V}) \setminus \pi_{Person}(\mathbf{P})$ ■

Aufgabe 7.17: *Welche Gäste sind in allen Kneipen gewesen?*

Lösung: Auch die Menge aller Kneipen muss in diesem Schema erst durch eine Vereinigung ermittelt werden, dann folgt die Lösung nach einer Relationendivision:

$\mathbf{G} \div (\pi_{Kneipe}(\mathbf{A}) \cup \pi_{Kneipe}(\mathbf{G}))$ ■

Aufgabe 7.18: *Welche Personen sind sind in einer Kneipe gewesen, die eines der Lieblingsbiere der Person führt?*

Lösung:Das Wort *ein* wird, wie in der Mathematik üblich, im Sinne von *mindestens ein* gebraucht. $\mathbf{Q} \Leftarrow \mathbf{V} \bowtie_{Biersorte} \mathbf{A}$ beschreibt, in welchen Kneipen die Personen ihre Lieblingssorten finden, diese Information ist mit der Relation **Gast** abzugleichen:

$\pi_{Person}(\mathbf{G} \cap \pi_{Person,Kneipe}(\mathbf{Q}))$ ■

Aufgabe 7.19: *Welche Personen sind nur in Kneipen gewesen, die eines der Lieblingsbiere der Person führt?*

Lösung: Die Lösung wird von allen Personen gebildet mit Ausnahme derer, für die es eine besuchte Kneipe gibt, die keines der Lieblingsbiere der Person führt. Wie in der vorigen Aufgabe wird **Q** gebildet. $\pi_{Person,Kneipe}(\mathbf{Q})$ enthält alle Kombinationen $(Person, Kneipe)$, wo in der Kneipe die Person mindestens eines ihrer Lieblingsbiere vorfindet. Somit wird die Aufgabe durch

$\pi_{Person}(\mathbf{G} \setminus \pi_{Person,Kneipe}(\mathbf{Q}))$

gelöst. ■

Aufgabe 7.20: *Welche Personen sind in einer Kneipe gewesen, die nur die Lieblingsbiere der Person haben?*

Lösung: Auch bei dieser Aufgabe wird zunächst eine logisch äquivalente Umformung der Aufgabenstellung gegeben: Die Lösung wird von allen Personen gebildet mit Ausnahme derer, wo in den besuchten Kneipen ein Bier im Angebot ist, welches nicht zu den Lieblingsbieren der Person zählt. Durch $\mathbf{Q} \Leftarrow \mathbf{G} \bowtie_{Kneipe} \mathbf{A}$ werden die Angebote der Kneipen ermittelt, die die Person besucht hat. $\mathbf{P} \Leftarrow (\pi_{Person,Biersorte}\mathbf{Q}) \setminus \mathbf{V}$ beschreibt die Tupel, für die giltdassß die von der Person besuchten Kneipen mehr Biersorten im Angebot haben als die Lieblingssorten der Person ausmachen. Folglich ist die Lösung:

$\pi_{Person}(\mathbf{G}) \setminus \pi_{Person}(\mathbf{P})$ ■

Aufgabe 7.21: *Welche Personen sind nur in Kneipen gewesen, die nur die Lieblingsbiere der Person haben?*

Lösung: Die Umformung der Aufgabenstellung führt auf die Formulierung:
Gesucht werden alle Personen, die in mindestens einer Kneipe waren, mit Ausnahme derer, wo es in den besuchten Kneipen mindestens eine gibt, die ein Bier im Angebot hat, was nicht in der Menge der Lieblingsbiere dieser Person ist:

$\pi_{Person}(\mathbf{G}) \setminus \pi_{Person}(\pi_{Person,Biersorte}(\mathbf{G} \bowtie_{Kneipe} \mathbf{A}) \setminus \mathbf{V})$ ■

7.3 Zu Aggregatfunktionen

Für die folgenden Aufgaben wird das Bibliotheksschema (Seite 59) vorausgesetzt.

Aufgabe 7.22: *Auf dem Bibliotheksschema ist die Anzahl der Buchexemplare zu ermitteln.*

Lösung: Der Zugang zum Ergebnis erfolgt durch Zählung der Einträge des Attributs *ISBN* oder, da *ISBN* Primärschlüssel ist, entspricht dies der Ermittlung der Kardinalität der Relation **Buecher**:

$\chi_{Zahl}(\sharp(\mathbf{Buecher}))$ oder $\chi_{Zahl}(\sharp_{ISBN}(\mathbf{Buecher}))$

Das Ergebnis ist bei diesen Lösungen eine Relation mit dem Attribut *Zahl*, der Ausdruck $\omega(\chi_{Zahl}(\sharp(\mathbf{Buecher})))$ wandelt dieses Ergebnis in eine Zahl um. ■

Aufgabe 7.23: *Wieviele Buchtitel sind ausgeliehen? (Zur Begriffsbildung: Ein Buchtitel hat seine eigene ISBN.)*

Lösung: Es sind die verschiedenen Werte des Attributs *ISBN* in **Ausleihe** zu zählen:

$\chi_{Zahl}(\sharp(\pi_{ISBN}(\mathbf{Ausleihe})))$

Die Projektion auf *ISBN* bewirkt, da das Ergebnis eine Menge ist, die Elimination von Duplikaten, wenn von einem Buchtitel mehrere Exemplare ausgeliehen sind.
■

Aufgabe 7.24: *Welchen Durchschnittspreis hat ein Buchtitel?*

Lösung: $\chi_{D\text{-}Preis}(\Delta_{Preis}(\mathbf{Buecher}))$ ■

Aufgabe 7.25: *Welchen Wert stellen die vorhandenen Bücher dar?*

Lösung: Hier muss der Preis jedes Buchtitels so oft summiert werden, wie Exemplare vorhanden sind. $\mathbf{B} \bowtie_{ISBN} \mathbf{E}$ hat ein Tupel für jedes Exemplar.
$\chi_{Ges\text{-}Preis}(\Sigma_{Preis}(\mathbf{B} \bowtie_{ISBN} \mathbf{E}))$ ■

Aufgabe 7.26: *Welche Bücher (ISBN) sind am ältesten?*

Lösung: Am ältesten sind die Bücher mit dem kleinsten Erscheinungsjahr. Diese Zahl wird mit dem Einbettungsoperator χ in eine Relation überführt und mit Hilfe eines Gleichverbundes mit jedem Tupel aus **Buecher** verglichen:
$\pi_{ISBN}(\mathbf{B} \bowtie_{\mathbf{B}.Jahr=MinJahr} (\chi_{MinJahr}(MIN_{Jahr}(\mathbf{B}))))$ ■

Aufgabe 7.27: *In wievielen Kneipen war jede Person, die in mindestens einer Kneipe war?*

Lösung: Hier muss offensichtlich ein Zähler für jede Person neu gestartet werden, dies bewirkt gerade der Einbettungsoperator mit Gruppierung in Analogie zur group by- Klausel aus SQL:
$\chi_{Person;Anzahl}(\sharp_{Kneipe}(\mathbf{Gast}))$ ■

Aufgabe 7.28: *In wievielen Kneipen war jede im Schema erfasste Person?*

Lösung: Zur Ergebnismenge der vorigen Aufgabe kommen jetzt noch die Personen hinzu, deren Vorlieben zwar erfaßt wurden, die aber keine Kneipe besucht haben. Für diese muss die Anzahl gleich Null (0) gesetzt werden.
$\mathbf{Q} \Leftarrow \pi_{Person,Anzahl}(\chi_{Person;Anzahl}(\sharp_{Kneipe}(\mathbf{Gast})))$
$\mathbf{P} \Leftarrow (\pi_{Person}(\mathbf{V}) \setminus \pi_{Person}(\mathbf{G})) \times \chi_{Anzahl}(0)$
Die Lösung lautet dann: $\mathbf{Q} \cup \mathbf{P}$ ■

Index