

## Modul 100: Daten analysieren und strukturieren

### Datentypen

Formal wird in der Informatik ein **Datentyp** als bestimmtes **Datenobjekt mit dazugehörigen Operationen** verstanden.

Im Beispiel aus AB100-03 haben wir gesehen, dass der Primärschlüssel vom Datentyp `INTEGER` (natürliche Zahlen ohne 0, da Schlüsselattribut) ist.

Die Grundoperationen die auf einen `INTEGER` angewandt werden, sind Addition, Subtraktion, Multiplikation und Division. Es gibt auch Aggregatsfunktionen wie `Min`, `Max`, `Avg`, usw.

Beispiele:

<code>Min (5, 2, 8, 9, 3, 7) = 2</code>	<code>Min</code> ermittelt den kleinsten Wert einer Zahlenreihe.
<code>Max (5, 2, 8, 9, 3, 7) = 9</code>	<code>Max</code> ermittelt den grössten Wert einer Zahlenreihe.
<code>Avg (5, 2, 8, 9, 3, 7) = 5.666</code>	<code>Avg</code> ermittelt den Durchschnitt einer Zahlenreihe.
<code>Sum (5, 2, 8, 9, 3, 7) = 34</code>	<code>Sum</code> ermittelt die Summe einer Zahlenreihe.
<code>Count (5, 2, 8, 9, 3, 7) = 6</code>	<code>Count</code> ermittelt die Anzahl Elemente in einer Zahlenreihe.

Ein Datentyp muss einem definierten Wertebereich (Domäne) entsprechen, damit die zulässigen Operationen definiert werden können.

#### Beobachtung I (Attribut)

In einer Tabelle muss jedes Attribut einen eindeutigen Datentypen haben oder jede Datenzeile eines Attributs hat einen Wert aus einer einzigen Domäne.

#### Beobachtung II (Primärschlüssel)

Die Primärschlüsseleigenschaft definiert, dass die Datenwerte des Primärschlüssels vom gleichen Datentypen sind (Beobachtung I) und nur einmal im Primärschlüsselattribut vorkommen dürfen.

Beispiel:

```
#ID PRIMARY KEY = {1, 2, iii, 4, fünf, vi, 4, iii; 0110,...}
```

→ Verletzung Beobachtung I.

```
{0,1,2,3} := INTEGER;
```

```
{iii, fünf, vi, 0110,...} := TEXT;
```

→ Verletzung Beobachtung II.

{4,4}: Datenwert 4 aus der Domäne `INTEGER` ist redundant.

{iii, iii}: Datenwert iii aus der Domäne `TEXT` ist redundant.

#### Definition der Datentypen (Domänen) aus SQLite:

Domäne	Beschreibung
<code>NULL</code>	Der Wert von <code>NULL</code> ist leer. Leer heisst also nicht „“ sondern <code>NULL</code>
<code>INTEGER</code> (unsigned)	0,1,2,3,..., 18'446'744'073'709'551'615 = 2 <sup>64</sup>
<code>REAL</code>	8 Byte grosse Gleitkommazahl (Floating point)
<code>TEXT</code>	Ein Textwert kann alle Zeichen (Zahlen, Buchstaben und Sonderzeichen) beinhalten. Die Kodierung ist UTF-8, UTF-16BE oder UTF-16LE.
<code>BLOB</code>	Binary Large Object. Der Datentyp ist genau, was eingegeben wird.

Denken wir zurück an die Definition aus Arbeitsblatt M100-02:

### Definition Tabelle

1. Eine Entitätsmenge (Tabelle) besteht aus einem eindeutigen Tabellennamen.
2. Innerhalb einer Tabelle ist jeder Attributname eindeutig und bezeichnet eine bestimmte Spalte.
3. Ein Attribut identifiziert eindeutig die Tupel in einer Tabelle und wird als Primärschlüssel bezeichnet.

*Die Bedingung 3 der Tabellendefinition entspricht der Beobachtung II:  
Jeder Datenwert eines Primärschlüssels ist aus der gleichen Domäne und identifiziert eindeutig eine Zeile der Tabelle.*

### NULL Datentyp

Der NULL Datentyp wird dann angewandt, wenn eine Datenzelle keinen Datenwert beinhaltet. Es gibt Prädikatsoperationen, die es erlauben zu fragen, ob eine Datenzelle NOT NULL ist. Das Prädikat returniert dann 0 oder 1.

NULL := {null}

### INTEGER Datentyp

Sind die ganzen Zahlen:

INTEGER (unsigned) := {0 | 1 | 2 | 3 | , . . . , n | n+1}

| steht für logisches ODER

n ist ein Element der natürlichen Zahlen inkl. 0.

### ACHTUNG:

In der offiziellen Dokumentation von SQLite Version 3 ist der Datentyp INTEGER wie folgt definiert:

*„INTEGER: The value is a signed integer, stored in 1, 2, 3, 4, 6 or 8 bytes depending on the magnitude of the value.“*

Die Definition schliesst die Zahl „0“ aus. Ein Test auf der SQLite Version 3 unter Linux Ubuntu 12.1 hat ergeben, dass der Wert „0“ mit einem SQL Insert Statement in eine INTEGER formatierte Datenzelle eingeführt werden kann!

Stellt man beim Erzeugen einer Tabelle die Primärschlüsseleigenschaft INTEGER PK AUTOINCREMENT (automatisches, fortlaufendes Aufzählen des Primärschlüssels PK = 1, 2, 3, 4 . . .) ein, beginnt die Aufzählung des Primärschlüsselwertes ohne 0, sie beginnt mit der Zahl 1.

### REAL Datentyp

Sind reelle Zahlen oder Gleitkommazahlen. Man unterscheidet Gleitkommazahlen:

- mit Vorzeichen (signed) - damit sind negative Zahlen möglich
- ohne Vorzeichen (unsigned) - rein positive Zahlen.

Domäne:

<b>signed</b>	$\approx -1,70141 \cdot 10^{38}$	$\approx 1,70141 \cdot 10^{38}$
<b>unsigned</b>	0	$\approx 3,40282 \cdot 10^{38}$

**TEXT Datentyp**

Sind Reihen von Unicode-Zeichen aus der UTF-8 Kodierung o.a. Bei der UTF-8-Kodierung wird jedem Unicode-Zeichen eine speziell kodierte Zeichenkette variabler Länge zugeordnet. Dabei unterstützt UTF-8 Zeichenketten bis zu einer Länge von vier Byte, auf die sich – wie bei allen UTF-Formaten – alle Unicode-Zeichen abbilden lassen.

**BLOB Datentyp**

Der BLOB Datentyp ist genial. Er ist alles was man eingibt und umschliesst alle anderen Datentypen. Er kann auch aus Video, Bild oder aus Tondateien bestehen.

BLOB steht für Binary Large Object.

Binärer Datentyp mit einer Grösse von bis zu 2GiByte.

### Weitere Datentypen aus handelsüblichen Datenbanksystemen, die in Modul 100 vereinfacht und abstrahiert sind.

Datentypen / Domäne	Vereinfachung in SQLITE
INT INTEGER TINYINT SMALLINT MEDIUMINT BIGINT UNSIGNED BIG INT INT2 INT8	<b>INTEGER</b>
CHARACTER(20) VARCHAR(255) VARYING CHARACTER(255) NCHAR(55) NATIVE CHARACTER(70) NVARCHAR(100) TEXT CLOB	<b>TEXT</b>
BLOB <i>no datatype specified</i>	<b>BLOB</b>
REAL DOUBLE DOUBLE PRECISION FLOAT	<b>REAL</b>
NUMERIC DECIMAL(10,5) BOOLEAN DATE DATETIME	<b>NUMERIC</b>

Jede Spalte einer Tabelle hat eine eindeutige Domäne zugewiesen und kann nur Werte aus dieser Domäne annehmen.

## Beziehungen

Wie wir in dem Beispiel in Aufgabenblatt 3 gesehen haben, stehen Tabellen untereinander in Beziehung.

Im Beispiel hat jeder Freund aus der Tabelle MYFRIENDS einen Ort aus der Tabelle ORT zugewiesen.

Dies geschieht über die Schlüsselbeziehung (#ORTSID; #FKORTSID).

### Definition Beziehung (Wiederholung)

Eine Beziehung zwischen zwei Tabellen ist durch eine Schlüsselbeziehung zweier Attribute definiert. Es gibt eine Primärtabelle mit einem Primärschlüssel und eine Fremdtabelle mit einem Fremdschlüssel, wobei Primär- und Fremdschlüssel vom gleichen Datentypen sind und jeder Fremdschlüssel als Primärschlüssel in der Primärtabelle existieren muss.

#### Fremdschlüsseltabelle:

MYFRIENDS(#ID PRIMARY KEY, Name, Vorname, Strasse, email, #FKORTSID FOREIGN KEY)

#ID	Name	Vorname	Strasse	email	#FKORTSID
1	Turner	Tina	Goldstreet	<a href="mailto:ttt@pop.com">ttt@pop.com</a>	1
2	Frei	Beat	Bluestreet	<a href="mailto:beat.frei@gibb.ch">beat.frei@gibb.ch</a>	2
3	Maurer	Ralph	Greenroad	<a href="mailto:ralph.maurer@gibb.ch">ralph.maurer@gibb.ch</a>	3
4	Bernanke	Ben	Blackstreet	<a href="mailto:bb@fed.us.gov">bb@fed.us.gov</a>	4
5	Meier	Claudia	Brownhouse	<a href="mailto:mc@meier.ch">mc@meier.ch</a>	3

#### Primärschlüsseltabelle:

ORT(#ORTSID PRIMARY KEY, PLZ, Ort)

#ORTSID	PLZ	Ort
1	8001	Zürich
2	8002	Zürich
3	3005	Bern
4		New York
5	3001	Bern

## Beziehungstypen

Wir unterscheiden folgende Beziehungstypen:

### Einfache Beziehung

1: Jede Entität (Zeile) der ersten Entitätsmenge (Tabelle) steht mit **genau einer** Entität (genau 1) der zweiten Entitätsmenge in Beziehung, und umgekehrt.

### Konditionelle Beziehung

c: Jede Entität der ersten Entitätsmenge kann mit **höchstens einer** Entität (0 oder 1) der zweiten Entitätsmenge in Beziehung stehen.

### Mehrfach Beziehung

m: Jede Entität der ersten Entitätsmenge steht mit **mindestens einer** Entität (1 oder mehrere) der zweiten Entitätsmenge in Beziehung.

### Mehrfachkonditionelle Beziehung

mc: Jede Entität der ersten Entitätsmenge kann mit **beliebig vielen** Entitäten (0, 1 oder mehrere) der zweiten Entitätsmenge in Beziehung stehen.

## Beziehungstypen werden pro Tabelle immer paarweise angewandt:

### 1. MYFRIENDS → ORT

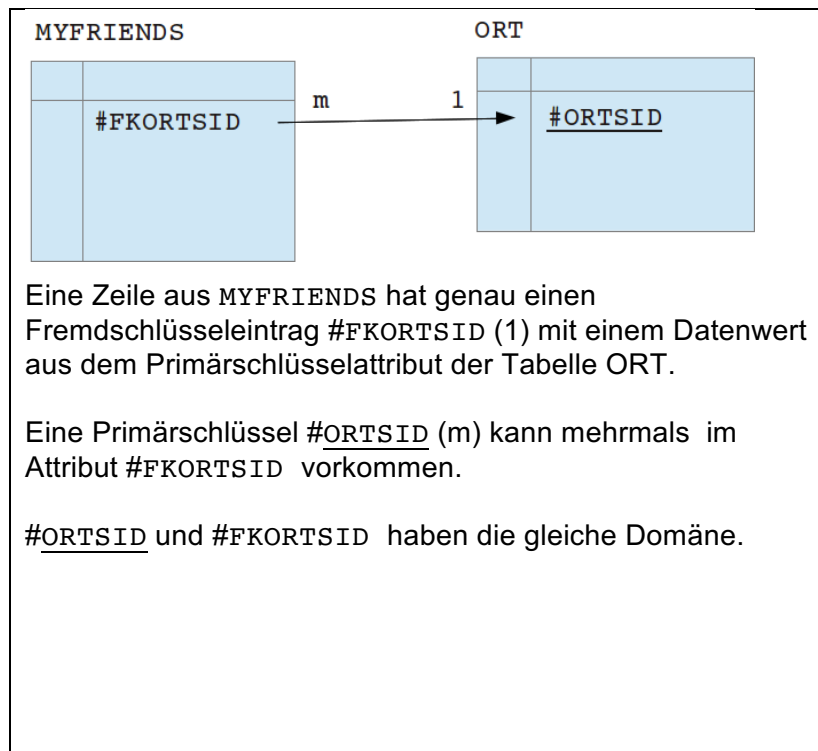
Die Entitätsmenge MYFRIENDS steht in einer bestimmten Beziehung zur Entitätsmenge ORT.

### 2. ORT → MYFRIENDS

Umgekehrt steht die Entitätsmenge ORT in einer bestimmten Beziehung zur Entitätsmenge MYFRIENDS.

#### Beispiel:

- Jeder Freund wohnt an genau einem Ort.
- Es gibt Orte, wo mehrere Freunde wohnhaft sind.



→ Das Beziehungstypenpaar der Entitätsmenge MYFRIENDS zu ORT heisst m:1 ausgesprochen „m zu 1“.

**Repetitorium AB100-04:****Aufgabe 1**

Datentypen, Domänen und Beziehungen

Begriffe:

<b>A</b>	Primärschlüssel	<b>B</b>	Fremdschlüssel
<b>C</b>	Primärtabelle	<b>D</b>	Fremdtabelle
<b>E</b>	INTEGER	<b>F</b>	REAL
<b>G</b>	UTF-8	<b>H</b>	TEXT
<b>I</b>	Wertebereich	<b>J</b>	Attribut

Welcher der oben aufgeführten Begriffe passt jeweils am besten zu den unten aufgeführten Beschreibungen? Mehrfachnennungen sind möglich (ein Begriff passt zu mehreren Beschreibungen).

Beschreibung:

Element einer Tabelle, dem genau eine Domäne zugeordnet ist.

A  B  C  D  E  F  G  H  I  J 

Hat genau einen Wert aus einer Domäne.

A  B  C  D  E  F  G  H  I  J 

Hat genau die Werte des Primärschlüssels.

A  B  C  D  E  F  G  H  I  J 

Negative und positive Gleitkommazahlen.

A  B  C  D  E  F  G  H  I  J 

Ein Fremdschlüssel einer Tabelle muss in der .... existieren.

A  B  C  D  E  F  G  H  I  J 

INTEGER hat die ganzen Zahlen als ...

A  B  C  D  E  F  G  H  I  J 

Text ist in Westeuropa meistens als .... kodiert.

A  B  C  D  E  F  G  H  I  J 

Eine Beziehung besteht aus einer Primärtabelle und ...

A  B  C  D  E  F  G  H  I  J 

UTF-8 Kodierungen sind im Datentypen ... abgebildet.

A  B  C  D  E  F  G  H  I  J 

Primärschlüssel und Fremdschlüssel sind ... einer Entitätsmenge.

A  B  C  D  E  F  G  H  I  J

**Aufgabe 2**

Bestimmen Sie das Beziehungstypenpaar:

*Beispiel:*

<i>Mensch</i>	<b>C:C</b>	<i>Lesebrille</i>	<i>tragen</i>
---------------	------------	-------------------	---------------

- Jeder Mensch trägt 0 oder 1 Lesebrille
- Jede Lesebrille wird von 0 oder 1 Mensch getragen

<b>Entitätsmenge 1</b>	<b>Beziehungstypenpaar</b>	<b>Entitätsmenge 2</b>	<b>Beziehung</b>
Buch		Fachgebiet	beschreibt
Abteilung		Angestellte	Abteilungsleiter
Abteilung		Angestellte	unterstellt
Kinder		Eltern	Familienzugehörigkeit
Standorte		Standorte	Distanz
Projekt		Projekt	Ist Unterprojekt
Frauen		Männer	Ehe
Schüler		Kurs	anmelden
Personen		Personen	Freundschaften
Website		Webserver	Hosting

**Aufgabe 3**

- a) Wählen Sie für jedes Attribut eine geeignete Domäne.
- b) Zeigen Sie die Fehler hinsichtlich der vorgestellten Definitionen (Tabellen, Datentypen, Beziehungen) auf.

#ID	Name	Vorname	Strasse	email	#FKORTSID
1	Turner	Tina	Goldstreet	ttt@pop.com	0
3,2	Frei	Beat	Bluestreet	beat.frei@gibb.ch	2
4	Maurer	Ralph	Greenroad	ralph.maurer@gibb.ch	7
6	Bernanke	Ben	Blackstreet	bb@fed.us.gov	4
6	Meier	Claudia	Brownhouse	mc@meier.ch	5

#FKORTSID	PLZ	Ort
1	8001	Zürich
2	8002	Zürich
3	3005	Bern
		New York
9	3001	Bern