23. Datenbanken

23.0 Begründung des Themas

In den zentralen Abiturprüfungen im Fach Informatik in NRW wird zum Bereich Datenbanken jedes Jahr eine Aufgabe gestellt.

Die Gesellschaft für Informatik (GI) führt in den „Bildungsstandards Informatik“ (für die SI) u. a. aus:

Schülerinnen und Schüler aller Jahrgangsstufen

- *verstehen den Zusammenhang von Information und Daten sowie verschiedene Darstellungsformen für Daten,*
- *verstehen Operationen auf Daten und interpretieren diese in Bezug auf die dargestellte Information,*
- *führen Operationen auf Daten sachgerecht durch.*

  …
- *erstellen informatische Modelle zu gegebenen Sachverhalten,*

23.1 Einführung – Praxisbeispiele

- Telefonbuch
- Online – Bestellungen (Versandhandel)
- Google
- Handy

23.2 Einführung – Geschichte

Problem: Verwaltung großer Datenmengen

- Karteikarten (→ Entity)
- Lochkarten (Hollerith)
- COBOL
- DBase
- Access

23.3 Definitionen

(nach Horn/Kerner/Forbrig: Lehr- und Übungsbuch Informatik Bd. 1)

Eine **Datenbank** (DB) ist eine systematisch strukturierte, langfristig verfügbare Sammlung von Daten.

Beispiele für Datenbanken

- Personaldatenbanken in Firmen, Patientenkarteien in Arztpraxen, Versichertenkarteien,
- Abonnentenverzeichnisse in Verlagen, Buchhaltung der Telefon-, Stromanbieter...
- Telefonbuch, Adressverzeichnisse ...
- Speicherung wissenschaftlicher Forschungsergebnisse, (Gendatenbank)

Ein **Datenbanksystem** (DBS) setzt sich aus einer konkreten Datenbank und einem Datenbank-Managementsystem (DBMS) zusammen.

Das **DBMS** ist ein Softwaresystem zur Definition, Administration und Manipulation von Daten. Es stellt die Schnittstelle zwischen Benutzer und Datenbank dar und dient der effizienten Speicherung und Abfrage der strukturierten Daten.

Wesentliche Anforderungen an ein DBS

- keine Redundanz
- Korrektheit der Daten
- Zugriffskontrolle entsprechend der Rolle des Benutzers
- Mehrbenutzerbetrieb

Modellierung:

In einer Datenbank werden Informationen über bestimmte, aus einem Ausschnitt der realen Welt stammenden Objekte gespeichert. Ein Objekt der realen Welt, über das Informationen zu speichern oder zu verarbeiten sind, ist eine **Entität**. Dies kann z. B. eine Person, ein Subjekt, ein Ereignis, ein Prozess oder ein Begriff sein.

Eine **Entität** ist ein eindeutig identifizierbares Objekt oder ein eindeutig identifizierbarer Sachverhalt der realen Welt oder der Vorstellungswelt.

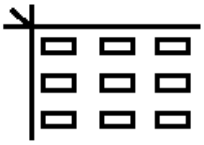
Beispiele: Heribert Schulze, ein VW, Einstellung von Heribert Schule bei VW, sein IPad

Die Entität wird durch ihre **Attribute** / Eigenschaften bestimmt. Die bei einer bestimmten Entität auftretenden Werte sind Attributwerte.

Entität: Heribert Schulz;
Attribute: Name, Vorname, Augenfarbe, Telefonnummer ...
Attributwerte: Schulz, Heribert, grau-braun, 0213 6666...

23.4 Arten von DBMS

Mögliche Modelle:

hierarchisches Modell	relationales Modell	objektorientiertes Modell
		
• Datenstruktur: Bäume • Datenobjektimplementierung: Pointer • Vorteile: einfach und effizient • Nachteile: unflexibel und teils schwierige Modellierung, keine m:n-Modellierung • Einsatz: Verzeichnis einer Festplatte (mit Unterverzeichnissen)	• siehe weiter unten	• jüngstes Modell: 1989 von C. Beeri entwickelt • Datenstruktur: Speichern der Daten in Objekten • starke Orientierung an der OOP

neu: XML - Datenbanken

Im Folgenden wird ausschließlich das 1970 von C. F. Codd entwickelte relationale Datenbankmodell genutzt, da es für die Bearbeitung der Aufgaben im Zentralabitur vorausgesetzt wird..

23.5 Merkmale des relationalen Modells

- Alle Angaben über Entitätstypen und Beziehungstypen werden in zweidimensionalen Tabellen (Relationen) festgehalten.
- Die Spalten der Tabellen entsprechen den Attributen. Die Anordnung der Spalten ist beliebig. Die Spalten sind durch ihre Namen eindeutig bestimmt.
- Die Zeilen entsprechen den Entitäten. Die Anordnung der Zeilen innerhalb der Tabelle ist beliebig.
- Jede Zeile ist über einen Schlüssel eindeutig identifizierbar.
- Jede Relation besitzt einen eindeutigen Namen.

23.6 Aufbau einer Datenbank

Jede Datenbank besteht aus einer oder (meistens) mehreren **Tabellen**.

Eine Tabelle setzt sich aus Spalten und Zeilen zusammen.

Im Gegensatz zu Excel oder Word ist jede Spalten mit einer besonderen Bedeutung verbunden:

Die **Spalten** entsprechen den **Attributen** (Eigenschaften) und werden auch **Felder** genannt; in der Tabellenüberschrift steht eine Bezeichnung für das Attribut – der Feldname, in den Zeilen darunter die Attributwerte für die einzelnen Entitäten.

Eine Tabellenzeile wird **Datensatz** genannt; sie enthält alle Attributwerte zu einer Entität. Hier: alle Eigenschaften eines Lesers.

Feld ↓

	benutzer_id	nachname	vorname	klasse	Geburtsdatum	Ort	PLZ	Strasse
	1405	Pütz	Jonathan	06D	16.08.1999	Solingen	42659	Balkhausen 74
	1406	Pütz	Nane	06D	05.04.2000	Solingen	42659	Bülowstr. 13
	1407	Richter	Erick	06D	17.10.2000	Solingen	42659	Höhrath 113a
	1408	Schüttler	Maxim	06D	22.04.2000	Solingen	42659	Ohmweg 7
	1409	Seefeld	Rebekka	06D	01.09.2000	Solingen	42659	Schlicker Weg 51
Datensatz →	1410	Semmler	Cornelius	06D	24.05.2001	Solingen	42659	Pfaffenberger Weg 113
	1411	Sliwinski	Katrin	06D	16.04.2000	Solingen	42659	Am Buschberg 27
	1412	Speitmann	Jil	06D	07.01.2000	Solingen	42659	Beckmannstr. 71
	1413	Temmel	Carolin	06D	12.09.2000	Solingen	42659	Balkhauser Weg 138
	1414	Warneke	Johanna	06D	07.04.2000	Solingen	42659	Hertzweg 1
	1415	Weber	Tristan	06D	21.04.2000	Solingen	42659	Eichenstr. 59
	1416	Wehning	Kai	06D	20.02.2000	Solingen	42659	Gabelsberger Str. 85
	1417	Wodicka	Jasmin	06D	13.10.2000	Solingen	42651	Akazienweg 13

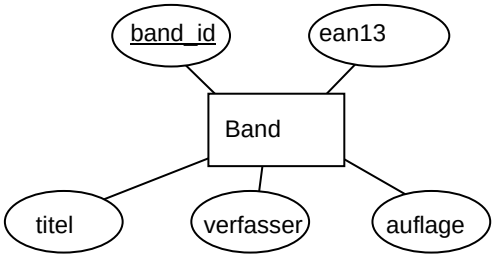
23.7 Darstellung

Am Beispiel der Büchertabelle werden die verschiedenen benutzten Darstellungsweisen benutzt und Begriffe verdeutlicht.

Entity - ein einzelnes Buch – jedes Buch → Objekt

Entitytyp Aufbau der Tabelle, Set von Eigenschaften(Attributen) → Klasse

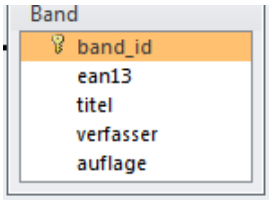
Entity – Relationship – Diagramm:



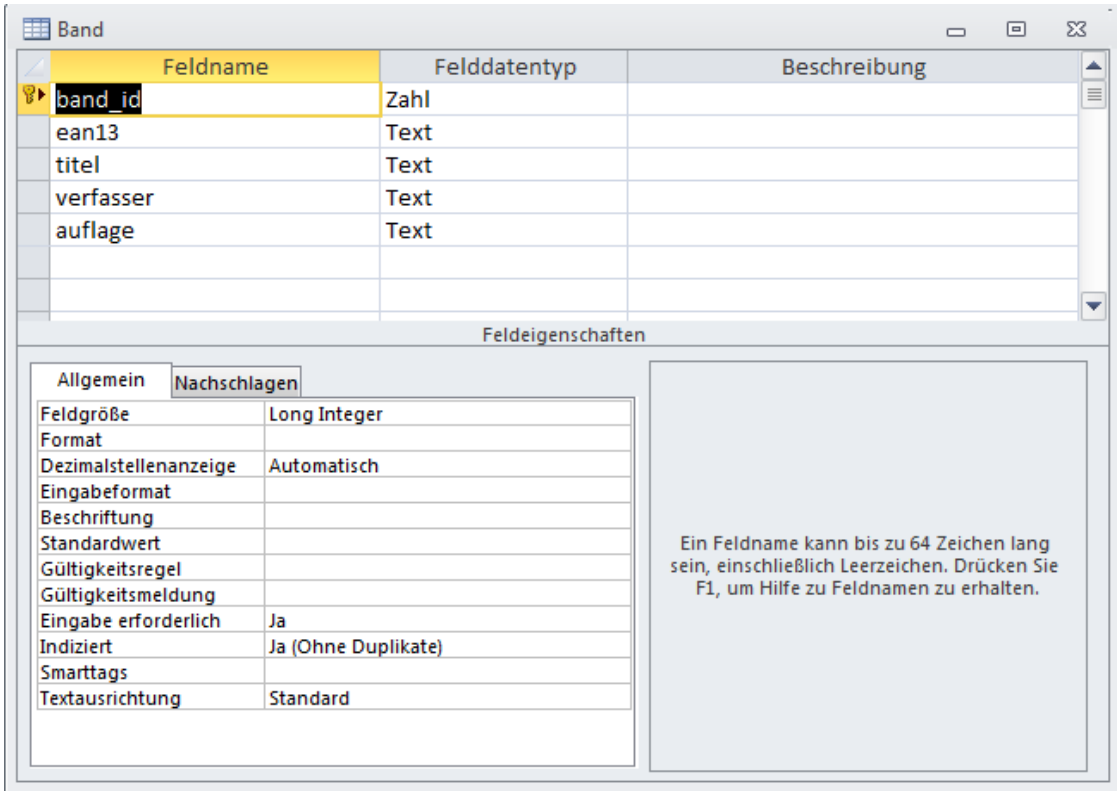
Datenbankschema

Band (band_id, ean13, titel, verfasser, auflage) **/Primärschlüssel unterstrichen**

Teil eines Beziehungsdiagramms (Open Office / Access)



Tabellenentwurf



und schließlich die Tabelle selbst

Band				
band_id	**ean13**	**titel**	**verfasser**	**auflage**
5	9783125177000	Pons Micro-Robert en poche	Robert, Paul	1. Aufl., Nachdr.1989
6	9783507107229	Physik	Dorn-Bader	2000
7	9783766157645	Der gallische Krieg	Friedrich Maier u. Heinrich Voit	1990
8	9783127322101	Lambacher-Schweizer 11	Manfred Baum	1. Aufl. 2000
9	9783791504544	Drachenreiter	Cornelia Funke	1997
10	9783491756519	Basiswissen Abitur - Philosophie	Michael Wittschier	2006
11	9783458322504	Goethes Schr. zur Naturwiss.	Goethe, Johann Wolfgang von hrsg. von Horst Günther	1. Aufl.1981
12	9783124560902	Historisch-politische Weltkunde	Asmut Brückmann	1. Aufl., 1. Dr. 1993
13	9783464410042	Texte, Themen und Strukturen		2. Aufl., 1. Dr. 2006
14	9783507107205	Physik Gymnasium Sek II 11	Ausgabe Dorn-Bader	Dr. A, 1 1998
15	9783766134110	Chemie 2000+ Bd. 1	Tausch von Wachtendonk	1. Aufl., 1. Dr. 2001

24. Abfragen

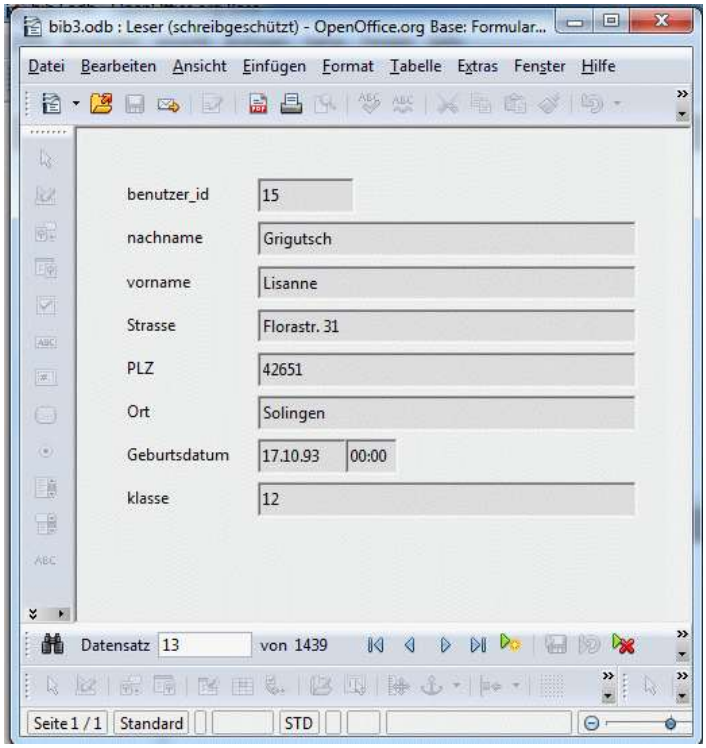
Als Fundgrube für Abfragen empfehle ich das Begleitmaterial zu „Microsoft Office – Class in a Box“ vom Cornelsen – Verlag. Eine Datenbank „Biblio“ wird dort mitgeliefert. Sie enthält über 1500 Eintragungen zu Büchern einer Schülerbücherei.

Die folgenden Angaben und Aufgaben beziehen sich auf eine Datenbank „bib“ einer Schulbuchausleihe. Sie liegt in den Formaten „Access“ und „OpenOffice Base“ vor.

24.1 Handling

Bücher, Bücher, haufenweise Bücher..

Unsere Schulbücherei enthält etwa 3.500 Exemplare von 250 Büchern. Die Angaben dazu sind in einer Tabelle erfasst Sie können die Bücher in Tabellenform auflisten oder auf einem Karteikarten-Formular der Reihe nach anzeigen lassen.

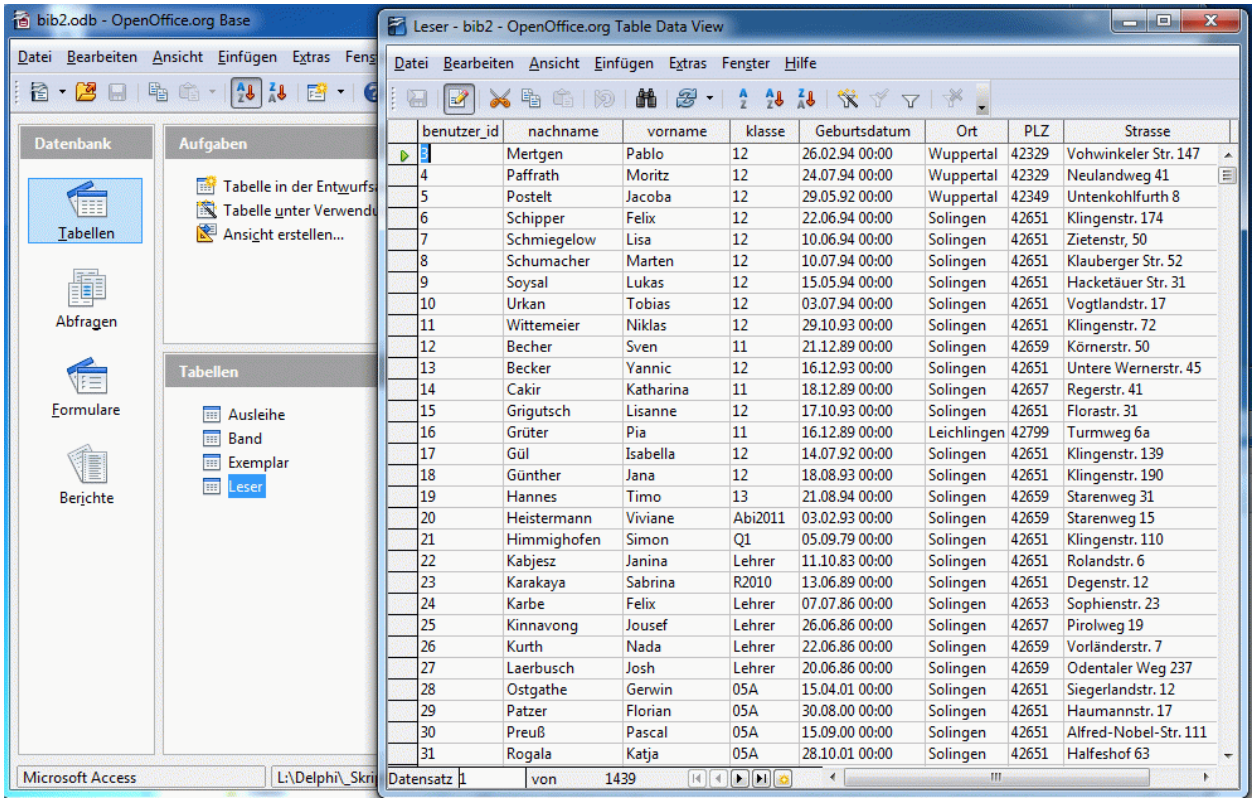


Formulardarstellung

Laden Sie die Datenbank „bib“. Das Datenbankfenster ist organisiert wie eine kleine Kartei. Durch Klick auf die Register kann man sechs Karten nach vorn holen. Die Tabelle Bücher und ein passendes Formular sind hier eingetragen.

Wählen Sie die Karte Formulare und klicken auf Öffnen.

Tabellensicht



benutzer_id	nachname	vorname	klasse	Geburtsdatum	Ort	PLZ	Strasse
3	Mertgen	Pablo	12	26.02.94 00:00	Wuppertal	42329	Vohwinkeler Str. 147
4	Paffrath	Moritz	12	24.07.94 00:00	Wuppertal	42329	Neulandweg 41
5	Postelt	Jacoba	12	29.05.92 00:00	Wuppertal	42349	Untenkohlfurth 8
6	Schipper	Felix	12	22.06.94 00:00	Solingen	42651	Klingenstr. 174
7	Schmiegelow	Lisa	12	10.06.94 00:00	Solingen	42651	Zietenstr, 50
8	Schumacher	Marten	12	10.07.94 00:00	Solingen	42651	Klauberger Str. 52
9	Soysal	Lukas	12	15.05.94 00:00	Solingen	42651	Hacketäuer Str. 31
10	Urkan	Tobias	12	03.07.94 00:00	Solingen	42651	Vogtlandstr. 17
11	Wittemeier	Niklas	12	29.10.93 00:00	Solingen	42651	Klingenstr. 72
12	Becher	Sven	11	21.12.89 00:00	Solingen	42659	Körnerstr. 50
13	Becker	Yannic	12	16.12.93 00:00	Solingen	42651	Untere Wernerstr. 45
14	Cakir	Katharina	11	18.12.89 00:00	Solingen	42657	Regerstr. 41
15	Grigutsch	Lisanne	12	17.10.93 00:00	Solingen	42651	Florastr. 31
16	Grüter	Pia	11	16.12.89 00:00	Leichlingen	42799	Turmweg 6a
17	Gül	Isabella	12	14.07.92 00:00	Solingen	42651	Klingenstr. 139
18	Günther	Jana	12	18.08.93 00:00	Solingen	42651	Klingenstr. 190
19	Hannes	Timo	13	21.08.94 00:00	Solingen	42659	Starenweg 31
20	Heistermann	Viviane	Abi2011	03.02.93 00:00	Solingen	42659	Starenweg 15
21	Himmighofen	Simon	Q1	05.09.79 00:00	Solingen	42651	Klingenstr. 110
22	Kabjesz	Janina	Lehrer	11.10.83 00:00	Solingen	42651	Rolandstr. 6
23	Karakaya	Sabrina	R2010	13.06.89 00:00	Solingen	42651	Degenstr. 12
24	Karbe	Felix	Lehrer	07.07.86 00:00	Solingen	42653	Sophienstr. 23
25	Kinnavong	Jousef	Lehrer	26.06.86 00:00	Solingen	42657	Pirolweg 19
26	Kurth	Nada	Lehrer	22.06.86 00:00	Solingen	42659	Vorländerstr. 7
27	Laerbusch	Josh	Lehrer	20.06.86 00:00	Solingen	42659	Odentaler Weg 237
28	Ostgathe	Gerwin	05A	15.04.01 00:00	Solingen	42651	Siegerlandstr. 12
29	Patzer	Florian	05A	30.08.00 00:00	Solingen	42651	Haumannstr. 17
30	Preuß	Pascal	05A	15.09.00 00:00	Solingen	42651	Alfred-Nobel-Str. 111
31	Rogala	Katja	05A	28.10.01 00:00	Solingen	42651	Halfeshof 63

24.2 Suchen in Tabellen

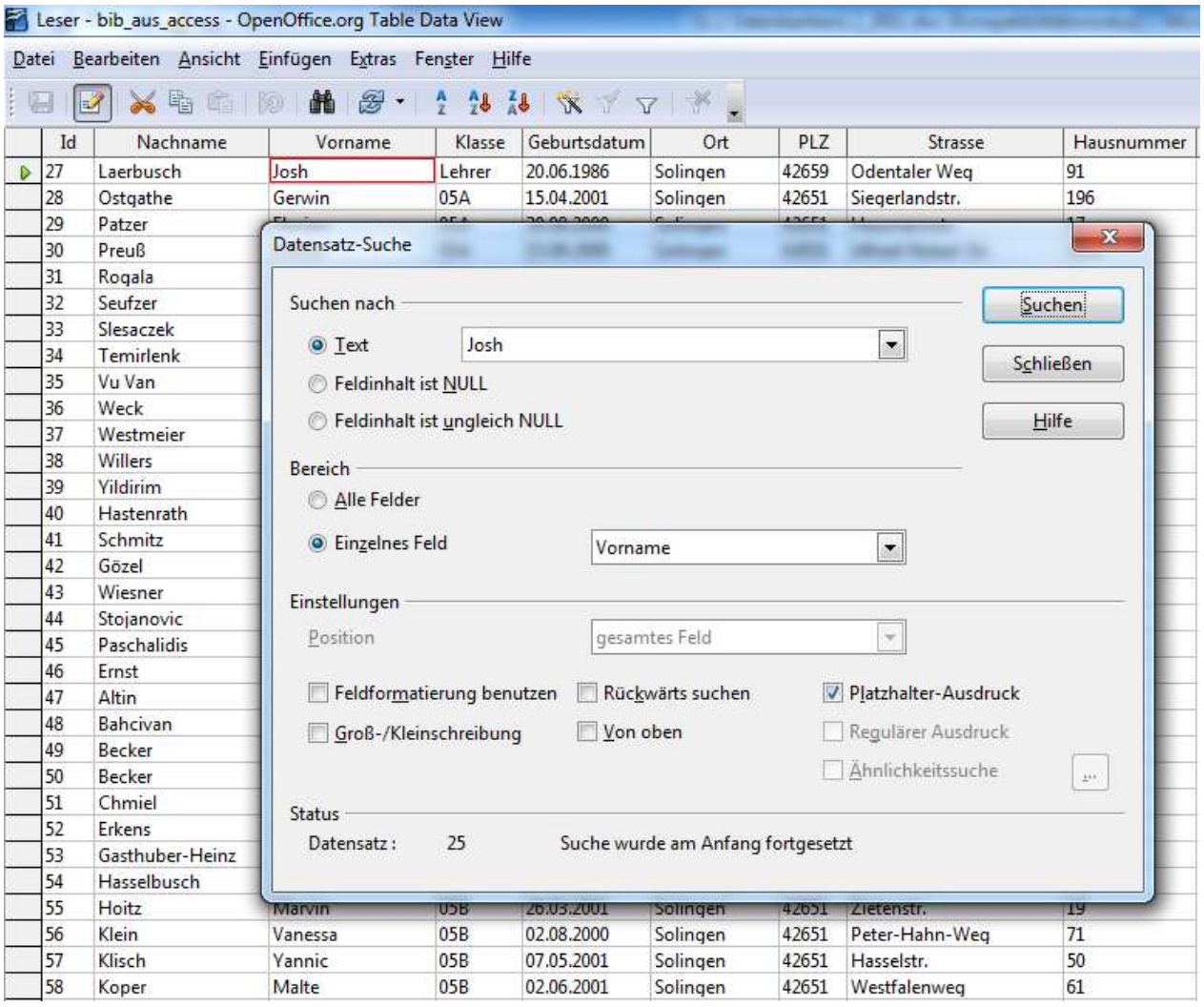
Jokerzeichen	steht für
*	beliebig viele (auch null) Zeichen
?	genau ein beliebiges Zeichen
#	genau eine Ziffer

Suchbegriff	gefundene Einträge (Tabelle Leser, Feld nachname)
Beck*	Beck, Becker
*mann	Vormann, Hoffmann
L*t	Lippert, Lucht
P*l*	Paul, Pohl, Puslat
?P*	zweiter Buchstabe= p (Spix, Opp, Apelt)
????	Namen mit vier Buchstaben (Küll)
* *	findet Namen aus zwei Wörtern (Leerzeichen im Namen bei Lo Galbo))

24.3 Suchen mit dem „Fernrohr“

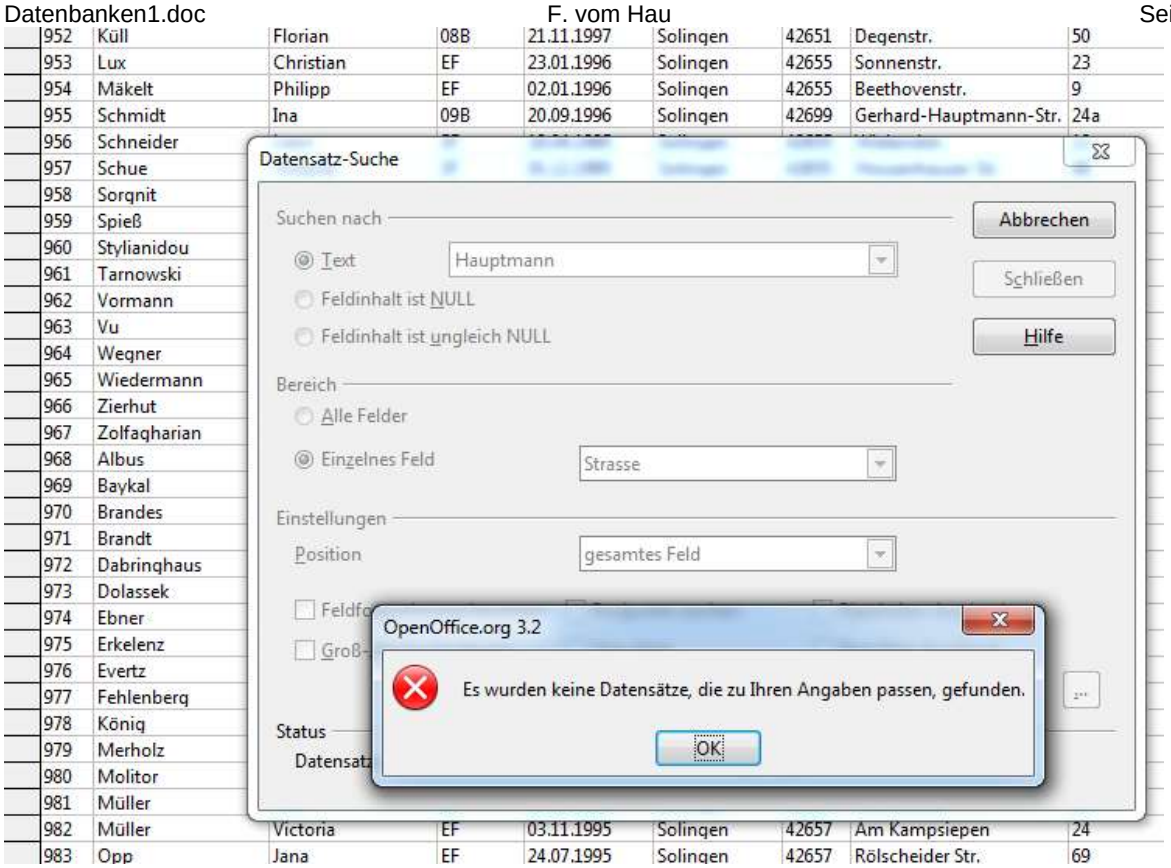
Gibt es einen Schüler mit Vornamen „Josh“

Die Tabelle „Leser“ öffnen, die Spalte „Vornamen“ anklicken und dann auf das Fernrohrsymbol klicken.

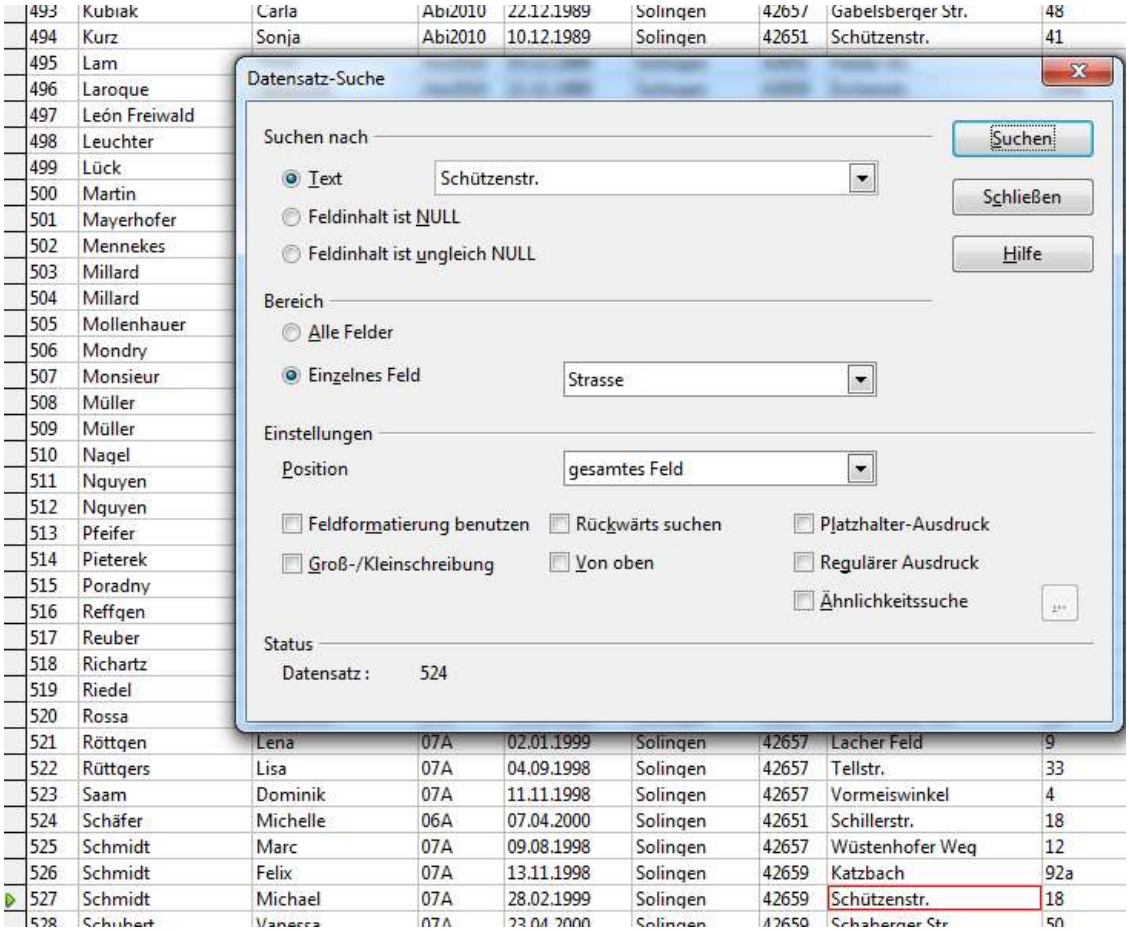


Oft erhält man eine Fehlermeldung, obwohl man den Datensatz „sieht“.

- Mögliche Gründe:
- Es ist das falsche Feld eingestellt.
- Es ist „gesamtes Feld“ eingestellt, man sucht aber oft nur nach einem Teil des Feldinhaltes.
- Schreibfehler, auch solche, die man nicht sieht: Leerzeichen



Wir suchen alle Schüler, die auf der Schützenstraße wohnen:

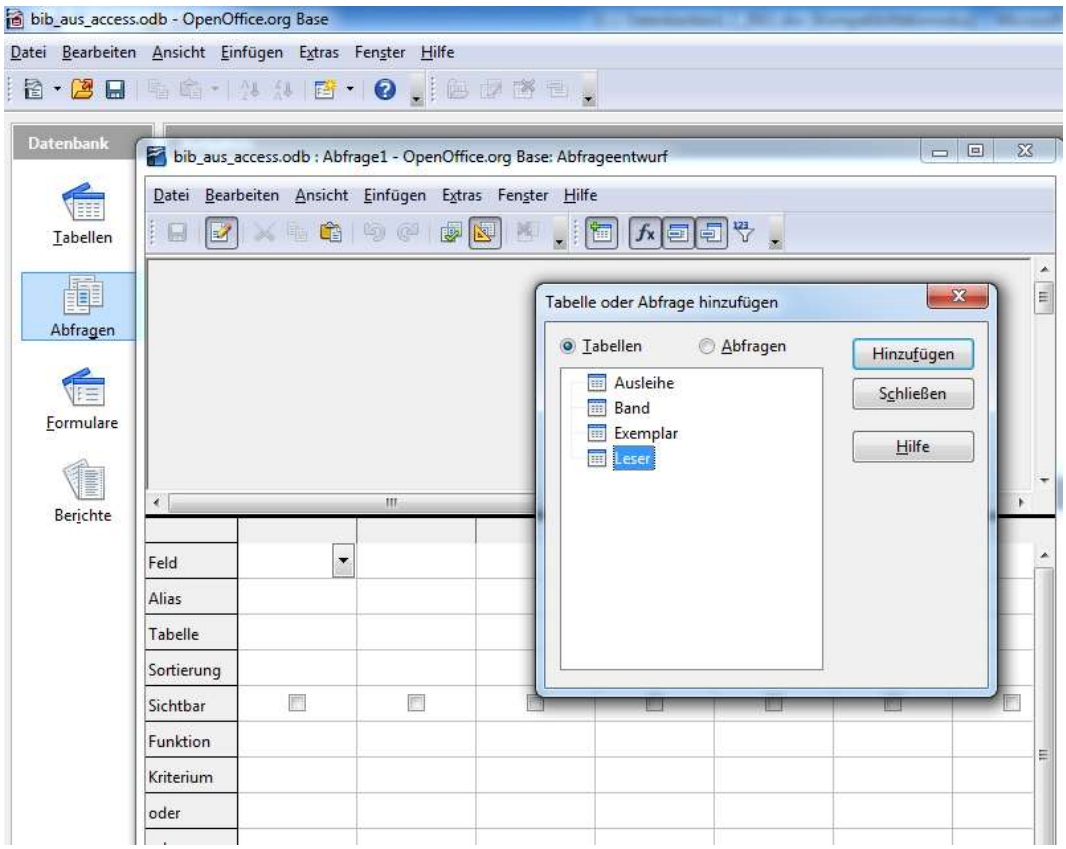


Es gibt insgesamt 16 Treffer. Die Daten abzuschreiben ist mühsam.
Es gibt aber eine Lösung für dieses Problem: Abfragen

24.4 Abfragen

Wir schließen die Tabelle „Leser“, wechseln in den Bereich „Abfragen“ und wählen „Abfrage in der Entwurfsansicht erstellen“.

Wir wählen die Tabelle „Leser“ aus und klicken auf „Hinzufügen“ und dann „Schließen“.

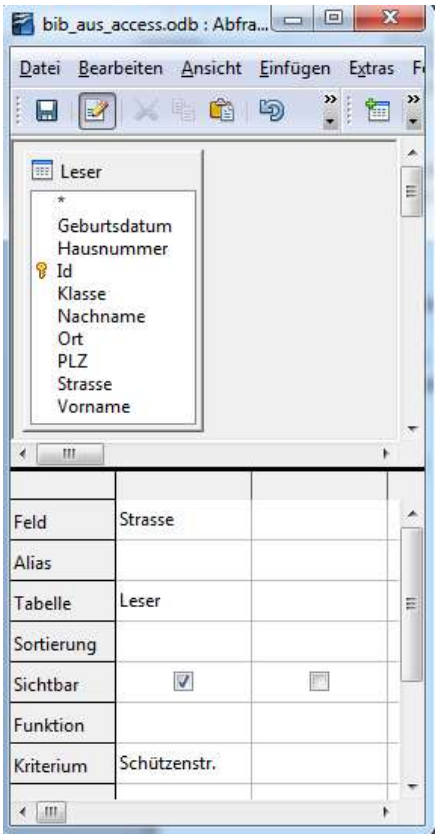


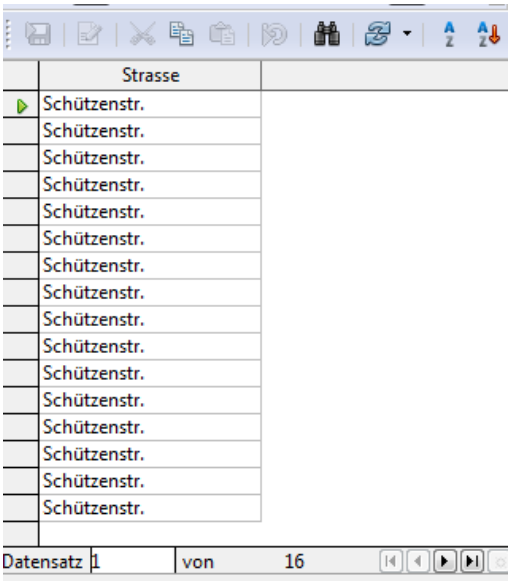
Ein Doppelklick auf „Strasse“ und dann unter „Kriterium“ Schützenstr. eingetragen.

Klick auf den Button „Abfrage ausführen“.
Das Ergebnis ist enttäuschend: 16mal Schützenstr. untereinander.

Merke: Bei einer Abfrage muss man angeben, was genau angezeigt werden soll.

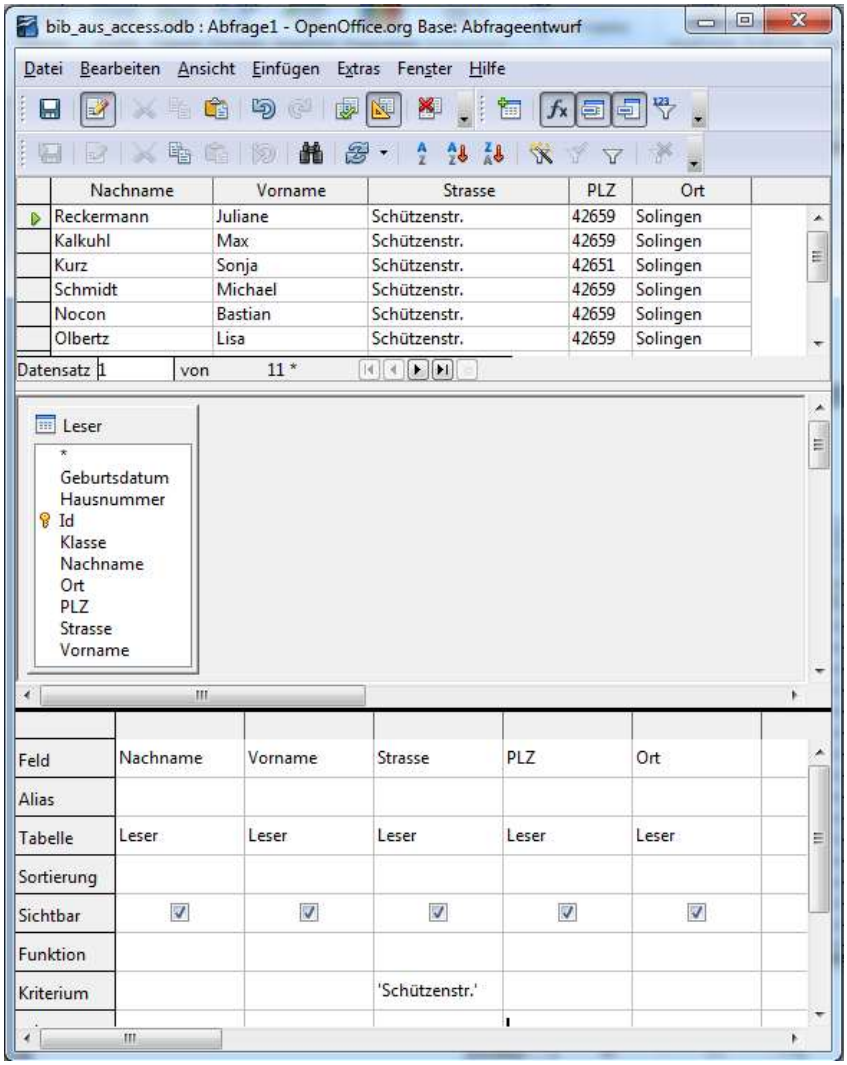
Dafür bekommt man dann aber auch eine Tabelle mit Ergebnissen, die man ausdrucken, per Email verschicken und in anderen Programmen weiter verwenden





kann.

Also etwa so:



Aufgaben:

1. Suche alle Leserinnen mit Vornamen „Lisa“. Es sind 16. Wenn man den Eintrag in „wie Lisa*“ änderst, bekommt man auch „Lisanne“ und die Doppelnamen mit Lisa am Anfang.
2. Suche alle Leserinnen und Leser, deren Nachname auf „mann“ endet. (Hoffmann, Amtmann,..)

3. Welche Bücher haben vom Titel her etwas mit Chemie zu tun?
4. Wer ist der Verfasser der „Fundgrube für den Physikunterricht“?
5. Wie viele Auflagen von „Micro-Robert“ gibt es?

24.5 Größer und Kleiner

6. Wir suchen alle Leserinnen und Leser, die höchstens 12 Jahre alt sind. (Trage bei Kriterium unter Geburtsdatum das heutige Datum – 12 Jahre ein, etwa: >#18.02.2000#)
7. Suche die ältesten Leser. (Es gibt unter den Lesern auch Abiturienten aus dem Jahr 2010)
8. Wie viele Leser haben einen Nachnamen aus der ersten Hälfte des Alphabets (A-L)?

24.6 UND- und ODER - Operator

9. Wie viele Leser haben einen Nachnamen mit „S“ am Anfang? Hier hilft ein Jokerzeichen.
10. Wie viele Leser haben einen Nachnamen mit M – R am Anfang? (>'M' und < 'S')
11. Welche Leserinnen und Leser sind in der 5. oder 6. Klasse? (wie 05* und wie 06* bei den Kriterien untereinander verknüpft die Angaben mit „oder“)

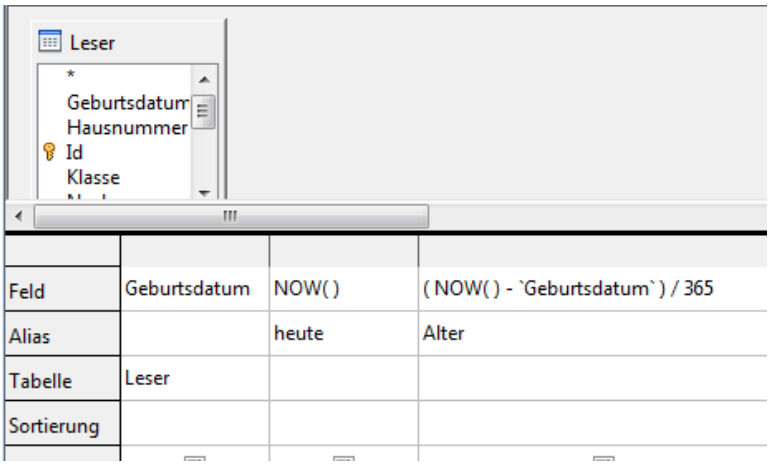
12.
13. Man weiß nur „Tim aus der 13“. Gibt es einen Leser, der zu diesen Angaben passt? (Kriterien, die in einer Zeile stehen, werden automatisch mit „und“ verknüpft.)

24.7 Berechnete Daten

Daten, die man berechnen kann, sogenannte abgeleitete Daten, gehören nicht in die Tabellen, sondern sollten immer jeweils neu berechnet werden.

Beispiel: das Alter.
Bei über 1000 Personen in unserer Datenbank hat wahrscheinlich jeden Tag jemand Geburtstag. Eine Spalte mit Altersangabe wäre somit immer irgenwo falsch.

Deshalb belässt man es mit der Spalte „Geburtsdatum“ und berechnet – wenn benötigt – in einer Abfrage das aktuelle Alter der Leser.



14. Gesucht werden alle volljährigen Leser und Leserinnen.

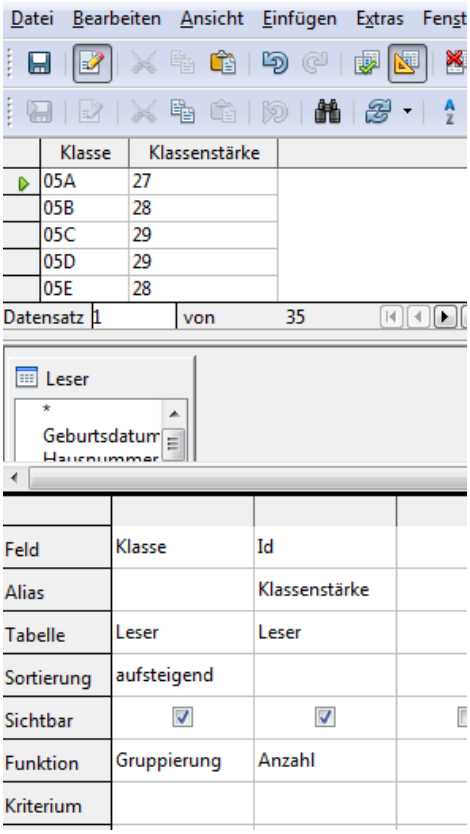
24.8 Funktionsabfragen

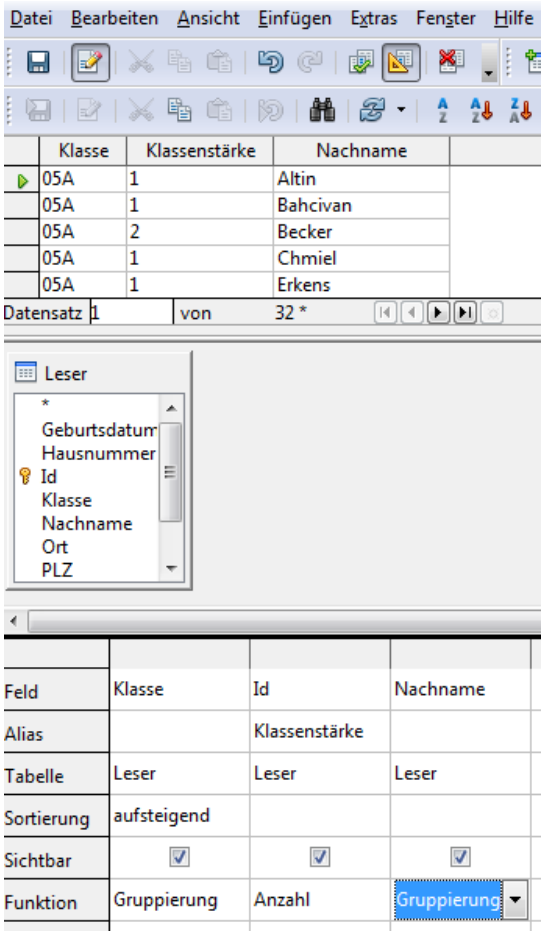
Aus den vielen Daten in der Datenbank können wir statistische Daten erheben.

Zum Beispiel die Anzahl der Schülerinnen und Schüler je Klasse, also die Klassenstärke.

Dazu müssen wir angeben, welche Datensätze zusammen gehören. In der Zeile mit der Bezeichnung „Funktion“ wählt man bei „Klasse“ den Begriff „Gruppierung“ aus, in der nächsten Spalte bei „Id“ den Begriff „Anzahl“.

Die Auswertung muss man sich jetzt so vorstellen, dass sich die Leser in „Grüppchen“ aufstellen müssen, die den jeweiligen Klassen entsprechen. „Anzahl“ gibt dann die Zahl der Gruppenmitglieder an.



In diesem Zusammenhang macht man häufiger den Fehler, zu viel anzugeben(s.nebenan). Zusätzlich die Spalte Nachname mit „Gruppierung“ führt dazu, dass innerhalb der Klassen noch einmal gruppiert wird nach den Namen. Jetzt haben wir lauter Einser-Grüppchen – bis auf die beiden Beckers in der 5a.

15. Lass das Durchschnittsalter je Klasse ausrechnen!