

Datenbanken, Hands-on 1

HENRY HAUSTEIN

Eine Vorbemerkung: Ich weiß nicht, inwieweit diese Aufgaben richtig gelöst sind, ich würde mich also freuen, wenn jemand mit einem anderen Ansatz meine Zahlen mal überprüft und es als Kommentar postet.

Aufgabe 1

Eine mögliche SQL-Abfrage ist z.B.

```
1 SELECT area, count(area) AS anzahl
2 FROM artist
3 WHERE begin_date_year < 1850
4 GROUP BY area HAVING anzahl > 10
5 ORDER BY anzahl DESC;
```

Dieser liefert eine Liste mit den Area-IDs und deren Vorkommen in der Tabelle `artist`. Zusätzlich filtern wir noch nach Geburtsdatum vor 1850 (`WHERE begin_date_year < 1850`) und es müssen mehr als 10 Vorkommen in der Tabelle sein (`HAVING anzahl > 10`). Das `HAVING`-Statement ist nötig, weil so etwas wie `WHERE anzahl > 10` einen Fehler erzeugt.

Will man noch den Namen der entsprechenden Regionen angezeigt bekommen, ist die SQL-Abfrage etwas aufwendiger:

```
1 SELECT artist.area, area.name, count(artist.area) AS anzahl
2 FROM artist, area
3 WHERE artist.begin_date_year < 1850 AND artist.area = area.id
4 GROUP BY artist.area HAVING count(artist.area) > 10
5 ORDER BY anzahl DESC;
```

Diese liefert die folgende Tabelle:

artist.area	area.name	anzahl
81	Germany	1543
73	France	1169
105	Italy	1049
221	United Kingdom	766
14	Austria	355
222	United States	344
194	Spain	270

176	Russia	191
432	England	181
56	Czech Republic	156
57	Denmark	151
202	Sweden	148
21	Belgium	130
150	Netherlands	118
214	Turkey	114
170	Poland	91
72	Finland	70
160	Norway	70
97	Hungary	64
103	Ireland	57
203	Switzerland	55
434	Scotland	39
171	Portugal	38
435	Wales	35
44	China	34
107	Japan	33
1794	Flanders	29
30	Brazil	23
67	Estonia	22
38	Canada	19
4434	Paris	19
99	India	18
117	Latvia	17
84	Greece	16
138	Mexico	15
175	Romania	12
219	Ukraine	11
1178	London	11

Mit insgesamt 38 Einträgen. Hier sieht man auch, dass in der Aufgabenstellung mit *Länder* eigentlich die Areas gemeint sind. Ich habe extra noch mal nachgefragt, *Länder* wie London sollen mitgezählt werden.

Aufgabe 2

Wir brauchen die ID der Gruppe *Arcade Fire*, das ist das passende SQL-Statement dazu:

```
1 SELECT * FROM artist WHERE name = "Arcade Fire";
```

Das liefert uns die ID 164813. Damit können wir uns anzeigen lassen, mit wem die Gruppe zusammengearbeitet hat:

```
1 SELECT * FROM artist_credit_name WHERE artist = 164813;
```

artist_credit	position	artist	name
164813	0	164813	Arcade Fire
822784	0	164813	Arcade Fire
823326	1	164813	Arcade Fire
867032	0	164813	Arcade Fire
973655	1	164813	Arcade Fire
997531	1	164813	Arcade Fire
1317313	1	164813	Arcade Fire
1230340	0	164813	The Reflektors
1540340	0	164813	Arcade Fire
1648422	1	164813	Arcade Fire
2028174	2	164813	Arcade Fire
1779951	0	164813	Arcade Fire
1779952	0	164813	Arcade Fire
1875665	0	164813	Arcade Fire
2295985	0	164813	Arcade Fire

Jetzt suchen wir uns die richtigen IDs für die in der Aufgabenstellung vorgegebenen Formate raus. Die Bedingung `name LIKE "%vinyl%"` ist dann erfüllt, wenn `name` das Wort *vinyl* an irgendeiner Stelle enthält (die % sind so eine Art Platzhalter für jede Buchstabenkombination).

```
1 SELECT *
2 FROM medium_format
3 WHERE name LIKE "%vinyl%" AND name != "VinylDisc";
```

Das liefert uns

id	name	description
29	7" Vinyl	—
30	10" Vinyl	—
31	12" Vinyl	—
7	Vinyl	—

80	VinylDisc (DVD side)	The DVD side of a vinyl + DVD VinylDisc. The vinyl side should be added as a separate medium.
81	VinylDisc (Vinyl side)	The vinyl side of a VinylDisc. The CD or DVD side should be added as a separate medium.
82	VinylDisc (CD side)	The CD side of a vinyl + CD VinylDisc. The vinyl side should be added as a separate medium.

Damit können wir alle Releases finden, die das richtige Format haben. Besonders nützlich erweist sich hier die WHERE-IN-Klausel, mit der wir überprüfen können, ob ein Attribut einen Wert aus einer Liste annimmt. Diese Liste liefern wir mit einer Subquery.

```

1 SELECT *
2 FROM medium
3 WHERE format IN
4     (SELECT id FROM medium_format
5      WHERE name LIKE "%vinyl%" AND name != "VinylDisc");

```

Das liefert 352703 Ergebnisse, unter anderem

id	release	position	format	name
384122	384122	1	31	—
369088	369088	1	31	—
948451	948451	1	31	—
600903	600903	1	31	—
47158	47158	1	31	—
588082	588082	1	31	—
26221	26221	1	31	—
709966	709966	1	31	—
709967	709967	1	31	—
593289	593289	1	7	—

Endlich können wir die finale SQL-Abfrage zusammensetzen: Dazu schauen wir in der release-Tabelle nach den Releases im richtigen Format und dem richtigen Künstler. Duplikate im Namen filtern wir mit dem GROUP-BY-Statement

```

1 SELECT *
2 FROM release
3 WHERE artist_credit IN
4     (SELECT artist_credit
5      FROM artist_credit_name
6      WHERE artist = 164813)
7 AND id IN
8     (SELECT id FROM medium WHERE format IN
9      (SELECT id

```

```

10             FROM medium_format
11             WHERE name LIKE "%vinyl%" AND name != "VinylDisc")
12         )
13     GROUP BY name;

```

Das liefert uns folgende Tabelle mit 17 Einträgen.

id	name	artist_credit	status	language	comment
2011803	Alting Nu	164813	1	120	Danish limited version
1304715	Arcade Fire	164813	1	120	remaster
2389021	Baby Mine	164813	1	120	—
57460	Cold Wind / Brazil	164813	1	120	—
2007060	Everything Now	164813	1	120	Day version
252852	Funeral	164813	1	120	—
287958	Intervention	164813	1	120	—
252655	Keep the Car Running / Broken Window	164813	1	120	—
820844	Neon Bible	164813	1	120	—
301944	No Cars Go	164813	1	120	—
312976	Poupée de cire, poupée de son / No Love Lost	997531	1	120	—
227251	Rebellion (Lies)	164813	1	120	—
1345908	Reflektor	164813	1	120	—
2330208	The Reflektor Tapes	164813	1	120	—
730810	The Suburbs	164813	1	120	—
697791	The Suburbs / Month of May	164813	1	120	—
2138539	⟨Hindi-Text⟩	164813	1	120	Hindi limited version

Aufgabe 3

Die folgende SQL-Abfrage liefert uns direkt das Ergebnis:

```

1  SELECT count(*)
2  FROM release_info
3  WHERE date_year >= 1980 AND date_year < 1990;

```

Es sind 139614 Releases.

Aufgabe 4

Zuerst brauchen wir wieder die ID von *Ariana Grande*:

```
1 SELECT * FROM artist WHERE name = "Ariana Grande";
```

Sie hat die ID 823336.

Damit können wir nach Mitproduzenten suchen und auch gleich zählen lassen:

```
1 SELECT count(*) FROM artist_credit_name WHERE artist = 823336;
```

Es sind 173.

Aufgabe 5

Wieder suchen wir uns die ID der *The Beatles*:

```
1 SELECT * FROM artist WHERE name = "The Beatles";
```

und erhalten 303.

Damit können wir die `artist-alias`-Tabelle fragen:

```
1 SELECT * FROM artist_alias WHERE artist = 303;
```

und erhalten

id	artist	name
179051	303	B
106844	303	⟨nicht darstellbar⟩
10389	303	fab four
117401	303	⟨nicht darstellbar⟩
165662	303	The Beatles
1126	303	Beetles
179055	303	Be
179056	303	Beat
130	303	Beatles
184586	303	The Savage Young Beatles
241307	303	Les Beatles
76594	303	⟨nicht darstellbar⟩
106841	303	Los Beatles
263503	303	⟨nicht darstellbar⟩
263502	303	⟨nicht darstellbar⟩
263504	303	⟨nicht darstellbar⟩
263506	303	⟨nicht darstellbar⟩

263507	303	⟨nicht darstellbar⟩
263508	303	⟨nicht darstellbar⟩
76592	303	⟨nicht darstellbar⟩
263509	303	⟨nicht darstellbar⟩
263510	303	Bítlarnir
263511	303	⟨nicht darstellbar⟩
263512	303	⟨nicht darstellbar⟩
263513	303	⟨nicht darstellbar⟩
263514	303	⟨nicht darstellbar⟩
263515	303	⟨nicht darstellbar⟩
263516	303	⟨nicht darstellbar⟩
263517	303	⟨nicht darstellbar⟩
263518	303	⟨nicht darstellbar⟩
263519	303	⟨nicht darstellbar⟩
263520	303	Die Beatles
263521	303	I Beatles
263523	303	The Silver Beatles
115264	303	⟨nicht darstellbar⟩

Da ich viele dieser Sonderzeichen nicht darstellen kann, gibt es hier ein Bild:

```

[sqlite> SELECT * FROM artist_alias WHERE artist = 303;
id|artist|name
179051|303|B
106844|303|Битлз
10389|303|fab four
117401|303|披頭四
165662|303|The Beatles
1126|303|Beatles
179055|303|Be
179056|303|Beat
130|303|Beatles
184586|303|The Savage Young Beatles
241307|303|Les Beatles
76594|303|ザ・ビートルズ
106841|303|Los Beatles
263503|303|البيتلز
263502|303|H ቢትሊስ
263504|303|ਫਾਫੈਟਲ ਸ
263506|303|بيتلز
263507|303|မ ဂီတလေ
263508|303|הביטלס
76592|303|더 비틀즈
263509|303|द बीटल्स
263510|303|Bitlarnir
263511|303|ବିଟଲ୍ସ
263512|303|ਬੀ ਟੀ ਟਲਜ਼
263513|303|Битлси
263514|303|ਬੀ ਟਲਜ਼
263515|303|ਬੀਟਲਜ਼
263516|303|ਬੀਟਲਜ਼
263517|303|เดอะบีเทิลส์
263518|303|بيتلز
263519|303|ბიტლზ
263520|303|Die Beatles
263521|303|I Beatles
263523|303|The Silver Beatles
115264|303|披头士

```

Wie das mit dem WITH-Statement funktioniert, habe ich bisher nicht rausgefunden, deswegen habe ich die einzelnen Namen händisch zusammengefügt und kam am Ende auf eine Länge von 354 Zeichen.