



## *Institut für Informatik - Lehrstuhl Datenbanken*

### **Einführung in Datenbanken**

Wintersemester 22/23 - Prof. Dr. Stefan Brass, PD Dr. Alexander Hinneburg

#### **Übung 2: Einfache SQL-Anfragen**

Abgabe bis Di. 1.11.2022, 9:00 Uhr,

Übungsplattform: Studip → Einführung in Datenbanken → Übungsplattform

Geben Sie für die folgenden Anfragen entsprechende SQL-Statements an. Für jede Anfrage soll nur ein SQL-Statement angegeben werden. Formatieren und kommentieren Sie die SQL-Statements, so dass sie gut zu lesen sind. Orientieren Sie sich an der Formatierung aus der Vorlesung. Für unkommentierte oder nicht sinnvoll formatierte SQL-Statements gibt es Punktabzug. SQL-Statements mit Syntax-Fehlern werden mit Null Punkten bewertet. Geben Sie jede Teilaufgabe als separate Text-Datei mit der Endung .sql ab (kein Word). Unter Windows können Sie Text-Dateien mit Notepad editieren. Das Programm Notepad++ (<http://notepad-plus-plus.org>) bietet mehr Funktionen.

#### **Aufgabe 2.1:**

6 Punkte

Loggen Sie sich in die PostgreSQL-Datenbank über StudIP ein und wählen Sie das Schema `president_public`. Die Tabellen haben folgenden Aufbau:

- `state(state_name, admin_entered, year_entered)`
- `president(pres_name, birth_year, years_serv, death_age, party, state_born→state)`
- `pres_hobby(pres_name→president, hobby)`
- `administration(admin_nr, pres_name→president, year_inaugurated)`
- `admin_pr_vp(admin_nr, pres_name→president, vice_pres_name)`
- `pres_marriage(pres_name→president, spouse_name, pr_age, sp_age, nr_children, mar_year)`
- `election(election_year, candidate, votes, winner_loser_indic)`

- Geben Sie alle Staaten(`state_name`) aus, deren erster Buchstabe C ist. Geben sie die Namen der Bundesstaaten aus.
- Welcher Präsident (`pres_name`) hatte ein Hobby, das irgendwo `ball` in beliebiger Gross- und Kleinschreibung in der Bezeichnung hat (d.h. `ball` oder `Ball` oder noch andere Varianten sind eine Teilzeichenkette des Hobbys)? Geben sie den Namen des Präsidenten und das Hobby aus. Nutzen Sie die Tabelle `pres_hobby`.
- Geben Sie alle Bundesstaaten aus, deren Namen aus genau vier Zeichen besteht. Nutzen Sie die Tabelle `state`. Geben sie die Namen der Bundesstaaten aus.

**Aufgabe 2.2:**

6 Punkte

In der Postgres-Datenbank `maerchen_public` sind Informationen über die Inhalte von Märchen als Tabellen verfügbar. Die Attribute, die den Primärschlüssel bilden, sind unterstrichen.

**DOCUMENT** ( DOCUMENT\_ID, TITLE )

**DOCUMENT\_TERM** ( DOCUMENT\_ID→DOCUMENT, POSITION\_OF\_TOKEN\_IN\_DOCUMENT, TERM, TOKEN, POS\_ID→POS )

**POS\_TYPE** ( POS\_ID, LOW, HIGH, DESCRIPTION, LONG\_DESCRIPTION, PARENT\_POS→POS )

Die Daten enthalten die Titel der Märchen in Tabelle `DOCUMENT`, die einzelnen Wörter der Texte in Tabelle `DOCUMENT_TERM` mit ihrer Position im Text (`POSITION_OF_TOKEN_IN_DOCUMENT`), der Wortgrundform (`TERM`), der Wortform wie sie im Text vorkommt (`TOKEN`) sowie der detailliertesten Wortart (`POS_ID`), die sich computer-linguistisch für das Wort bestimmen lässt. Die hierarchische Gliederung der Wortarten wird in Tabelle `POS_TYPE` beschrieben auf zwei verschiedene Arten beschrieben: (a) Unterwortarten verweisen auf die nächst allgemeinere Wortart in dem Attribut `PARENT_POS`. Die allgemeinsten Wortarten haben einen `NULL`-Wert in diesem Attribut. (b) Jeder Wortart ist ein numerisches Intervall [`LOW`,`HIGH`] zugeordnet in dem die Intervalle der Unterwortarten enthalten sind. Beschreibung der Abkürzungen der Wortarten: [http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/stts\\_guide.pdf](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/stts_guide.pdf).

Loggen Sie sich in die PostgreSQL-Datenbank über StudIP (Einführung in Datenbanken → Informationen → Postgres-DB) ein und wählen Sie links das Schema `maerchen_public`.

- a) Geben Sie alle Märchentitel (`title`) aus, die irgendwo den String `könig` in beliebiger Schreibweise enthalten.
- b) Geben Sie alle Märchentitel (`title`) aus, die mehr als 30 Zeichen lang sind.
- c) Geben Sie für alle Märchentitel (`title`) aus, wie viele Wörter sie enthalten. Gehen Sie davon aus, dass die Wörter durch genau ein Leerzeichen getrennt sind und keine weiteren Leerzeichen am Anfang und Ende des Titels sind. Nutzen Sie die String-Funktionen `char_length` und `replace`. Die Ausgabe soll den Märchentitel (`title`) und die Anzahl der Wörter enthalten.