

Einführung in Datenbanken

— Übungsblatt 1 (Web-Quellen, Tabellen, CSV) —

Ihre Lösungen laden Sie bitte in die Übungsplattform in StudIP hoch ([StudIP-Eintrag der Vorlesung], Reiter „Übungsplattform“, dann auf „Weiter zu ... Übungsplattform“).

Einsendeschluss ist Montag, der 25.10.2021, 18⁰⁰.

Hausaufgaben müssen einzeln bearbeitet werden. „Zu ähnliche Lösungen“ führen automatisch zu 0 Punkten für alle Beteiligten. Das gilt auch dann, wenn Sie nicht direkt abgeschrieben haben, sondern nur zufällig die gleiche Quelle benutzt haben.

Hausaufgabe 1 (10 Punkte)

Es soll gemeinsam eine Liste nützlicher Webseiten zum Thema der Vorlesung („Datenbanken“) erstellt werden. Die Informationen sollen in zwei Tabellen festgehalten werden.

a) (5 Punkte) Die erste Tabelle “WEBSEITEN“ soll folgende Struktur haben:

WEBSEITEN			
STUD	NR	URI	TITEL
abcde	1	http://www.informatik.uni-halle.de/~brass/db21/	Vorlesung
abcde	2	https://www.postgresql.org/	PostgreSQL

In die erste Spalte fügen Sie Ihr 5-stelliges Login ein. Es sollen später die von allen Studenten gesammelten Webseiten zusammen in eine Tabelle eingefügt werden. Damit die Einträge eindeutig referenziert werden können, sollen die Spalten **STUD** und **NR** zusammen dafür genutzt werden. Die Nummer wählen Sie fortlaufend, von 1 beginnend.

Sie müssen eine Liste von fünf Webseiten abgeben. Die beiden Webseiten im Beispiel oben sind dabei ausgeschlossen. Wir werden kontrollieren (natürlich mit SQL), dass keine zwei Studenten genau die gleichen fünf Webseiten abgegeben haben (möglicherweise prüfen wir auch eine Überlappung von drei oder vier Webseiten). Bei zu starker Überlappung gibt es weniger Punkte (möglicherweise auch 0). Das gilt auch, wenn alle Ihre Webseiten schon von der Vorlesungsseite und ihren Unterseiten aus verlinkt sind (eine gewisse Überlappung ist natürlich möglich).

Wir möchten auch, dass die von Ihnen abgelieferten Webseiten wirklich nützliches Material zum Erlernen von Datenbanken enthalten. Verstehen Sie die Eintragung von Webseiten in die Datenbank als eine Art Empfehlung der Webseite. Wenn Sie irgendwelche ganz unpassenden Webseiten in Ihren Daten haben, bekommen Sie 0 Punkte.

Es ist möglich, dass wir am Ende die komplette Tabelle mit allen Abgaben über Adminer für alle Kursteilnehmer zur Verfügung stellen. Schauen Sie sich die Seiten also ein wenig an, bevor Sie sie in Ihre Liste aufnehmen. (Es ist natürlich auch ein Ziel dieser Hausaufgabe, dass Sie für sich nützliche Webseiten entdecken.)

Sie müssen die Daten im CSV-Format („comma-separated values“) abgeben. Dieses Format wird häufig benutzt, um Tabellen auszutauschen. Es ist auch ein Export-Format von Tabellen-Kalkulationen (wie Excel). Sie sollten die Datei aber besser mit einem normalen Texteditor erstellen, den Sie auch zur Programmierung verwenden würden (z.B. Notepad++ [<https://notepad-plus-plus.org/>]). Mindestens sollten Sie die erzeugte Datei in einem Editor anschauen, und sich nicht darauf verlassen, dass Excel sie schon korrekt erzeugt. Es gibt bei CSV recht viele Abweichungen vom offiziellen Format nach [RFC 4180]. Z.B. werden teils auch andere Trennzeichen als Kommata für die Spalten verwendet, wir akzeptieren nur Kommata. Zu der Hausaufgabe gehört auch, dass Sie sich in dieses Format einarbeiten. In einfachsten Fall gibt es eine Zeile in der Datei pro Tabellenzeile (Datensatz), und die Werte der Spalten sind durch Kommata getrennt. Man könnte die obige Tabelle also so aufschreiben:

```
abcde,1,http://www.informatik.uni-halle.de/~brass/db21/,Vorlesung
abcde,2,https://www.postgresql.org/,PostgreSQL
```

Wenn ein Spalteneintrag ein Komma enthält (oder einen Zeilenumbruch) funktioniert das so natürlich nicht, dann muss man den Datenwert in `"..."` einschließen. Man darf die Datenwerte immer so einschließen, auch wenn sie nicht Kommata oder Zeilenumbrüche enthalten. Hier ist das exemplarisch für die dritte und vierte Spalte des ersten Datensatzes gemacht, beim zweiten nur für die dritte:

```
abcde,1,"http://www.informatik.uni-halle.de/~brass/db21/","Vorlesung"
abcde,2,"https://www.postgresql.org/","PostgreSQL"
```

Wenn ein Datenwert ein Anführungszeichen `"` enthält, muss es verdoppelt werden. Es gibt noch eine optionale Kopfzeile mit den Spaltennamen, die verwenden Sie bitte nicht (wir wollen alle CSV-Dateien zusammenfügen).

- Der Internet-Standard „RFC 4180“ für den MIME-Typ `text/csv`:

[<https://www.rfc-editor.org/info/rfc4180>]

- Englischsprachige Wikipedia:

[https://en.wikipedia.org/wiki/Comma-separated_values]

- Deutsche Wikipedia:

[[https://de.wikipedia.org/wiki/CSV_\(Dateiformat\)](https://de.wikipedia.org/wiki/CSV_(Dateiformat))]

Schließen Sie bitte auch die letzte Zeile mit einem Zeilenumbruch ab, so dass wir alle abgegebenen Dateien leicht zusammenfügen können. Verwenden Sie die UTF-8 Codierung für Zeichen (das ist nur relevant, wenn Ihr Text Umlaute enthält). Schreiben Sie die Tabellendaten in eine Datei „webseiten.csv“, und laden Sie diese in die Übungsplattform hoch.

- b) **(5 Punkte)** In der zweiten Tabelle sollen Schlagworte/Suchbegriffe („Tags“) zu den Webseiten erfasst werden.

SUCHBEGRIFFE		
STUD	NR	BEGRIFF
abcde	1	Vorlesung
abcde	1	Uni Halle
abcde	1	Logik
abcde	1	SQL
abcde	2	PostgreSQL
abcde	2	DBMS
abcde	2	Open Source

Insgesamt soll Ihre Tabelle mindestens 10 und maximal 30 Einträge (Zeilen) haben. Natürlich sollen die Werte `STUD` und `NR` eine Webseite aus der Tabelle `WEBSEITEN` identifizieren, d.h. die Kombination dieser Werte muss auch dort auftreten. Diese Daten schreiben Sie bitte in eine zweite Datei „suchbegriffe.csv“, die Sie auch in die Übungsplattform hochladen.

Bonusaufgabe (5 Punkte)

- c) Entwickeln Sie das Programm von Folie 2-38 (in Java). Dieses Programm soll Daten der folgenden Form aus einer Datei `abgaben.csv` einlesen:

Vorname,Nachname,Aufgabennummer,Punkte

und dann die Gesamtpunktzahl für jeden Studierenden ausgeben (alphabetisch geordnet). Sie können voraussetzen, dass keine „...“ in der CSV-Datei verwendet werden. Schreiben Sie ganz oben in einen Kommentar, wieviel Zeit Sie dafür gebraucht haben. Bitte vergessen Sie das nicht: Ziel dieser Aufgabe ist ja auch, die Zeitersparnis durch die Verwendung von SQL beziffern zu können.