

Klausur Datenbanken
Wintersemester 2018/2019
Prof. Dr. Wolfgang May
21. Februar 2019, 14-16:30 Uhr
Bearbeitungszeit: 120 Minuten

Vorname:

Nachname:

Matrikelnummer:

Bei der Klausur sind **keine Hilfsmittel** (Skripten, Taschenrechner, etc.) erlaubt. Mobiltelefone müssen ausgeschaltet sein. Papier wird gestellt. Benutzen Sie die **ausgeteilten**, zusammengehefteten **Blätter** für Ihre Antworten. Schreiben Sie mit blauem/schwarzem Kugelschreiber, Füller, etc.; Bleistift ist nicht erlaubt.

Zum **Bestehen** der Klausur sind **45** Punkte hinreichend.

(Unterschrift (DSVGO))

Meine Note soll mit meiner persönlichen Codezahl: so bald wie möglich auf der Vorlesungs-Webseite veröffentlicht werden.

Meine Note soll nicht veröffentlicht werden; ich erfahre sie dann aus FlexNow oder beim zuständigen Prüfungsamt.

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (ER-Modell)	15	
Aufgabe 2 (Transformation in das Relationale Modell)	18	
Aufgabe 3 (SQL und Relationale Algebra)	46	
Aufgabe 4 (Verschiedenes)	11	
Summe	90	

Note:

Themenstellung: Landtagswahl

Alle Klausuraufgaben basieren auf einem gemeinsamen “Auftrag”: In der Klausur soll eine Datenbank für die Erfassung der Kandidaturen bei der Landtagswahl sowie der Wahlergebnisse am Wahltag entworfen werden. Als Vorbild für die Beispiele dient die Wahl in Hessen vom 28. Oktober 2018. Das System soll bei allen zukünftigen Landtagswahlen einsetzbar sein.

Im folgenden enthalten die Items die eigentliche Aufgabenstellung, der Fließtext ist als ergänzende Erklärung gedacht.

1. Es treten mehrere *Parteien* zu einer Wahl an. Jede Partei hat einen Namen und eine Abkürzung. Z.B. “Christlich-Demokratische Union” bzw. “CDU”, oder “Sozialdemokratische Partei Deutschlands”, “SPD”.
2. Das (Bundes)land ist in *Wahlkreise* aufgeteilt (bei der Hessenwahl waren es 55). Jeder Wahlkreis hat einen Namen, außerdem ist gespeichert, wieviele Wahlberechtigte dort gemeldet sind.

Im Wahlkreis *Gießen-Stadt* gibt es 98075 Wahlberechtigte; im Wahlkreis *Gießen-Land* gibt es 93639 Wahlberechtigte.

Der übliche Wahlmodus ist grob folgender: Jeder Wähler⁵ hat zwei Stimmen, eine *Erststimme* und eine *Zweitstimme*. Mit der *Erststimme* kann man eine Stimme für einen *Wahlkreiskandidaten* einer Partei abgegeben. Der Wahlkreiskandidat mit den meisten Stimmen ist dann als Abgeordneter für den Wahlkreis in den Landtag gewählt (auf diese Weise wurden also 55 von 110 Sitzen besetzt, sogenannte “Direktmandate”). Mit der *Zweitstimme* (die die “wichtigere” Stimme ist) wählt man eine der angetretenen Parteien. Die Sitze im Landtag werden *insgesamt* (d.h., in Hessen die restlichen 55 Sitze) dann prozentual nach dem Stimmenanteil der Zweitstimmen mit Kandidaten der (geordneten) Landeslisten der einzelnen Parteien aufgefüllt.

3. Zu allen gespeicherten *Personen* (dies sind die Wahlkreiskandidaten und die Listenkandidaten) sind Name, Geburtsort, Beruf, Geburtsdatum und Wohnort (formal: Wohngemeinde) gespeichert. In dieser Aufgabe wird angenommen, dass es keine zwei namensgleichen Orte/Gemeinden, und keine zwei namensgleichen Kandidaten gibt.

Volker Bouffier wurde am 18.12.1951 in *Gießen* geboren, ist Rechtsanwalt, und wohnt in *Gießen*. *Torsten Schäfer-Gümbel* (“TSG”) wurde am 1.10.1969 in *Oberstdorf* geboren, ist Politikwissenschaftler und wohnt in *Lich*.

4. Für jeden *Wahlkreis* und jede *Partei* ist gespeichert, wer dort für diese Partei als *Wahlkreiskandidat antritt*. In jedem Wahlkreis kann jede Partei höchstens einen solchen Wahlkreiskandidaten benennen (meistens tun dies aber nur die aussichtsreichen, großen Parteien). Eine Person kann höchstens in einem Wahlkreis Wahlkreiskandidat sein.

Im Wahlkreis *Gießen-Land* sind *Volker Bouffier* für die *CDU* und *Torsten Schäfer-Gümbel* für die *SPD* als Wahlkreiskandidaten angetreten (und drei weitere Kandidaten anderer Parteien).

⁵in diesem Text wird zur besseren Verständlichkeit die kurze, traditionelle Sprachform verwendet. Es sind alle Wähler*innen gemeint.

5. Jeder Wahlkreis ist zur Organisation der Stimmabgabe in viele einzelne *Wahlbezirke* eingeteilt. Jeder Wahlbezirk hat einen eindeutigen Namen. Auch zu jedem Wahlbezirk ist gespeichert, wieviele Personen dort wahlberechtigt sind, und zu welcher *Gemeinde* er gehört.

Z.B. gibt es im Wahlkreis *Gießen-Land* die Wahlbezirke *Lich-Kirche* mit 828 Wahlberechtigten und *Lich-Rathaus* mit 963 Wahlberechtigten. Beide gehören zu der Gemeinde *Lich*.

Wie man am Beispiel schon sieht, gibt es in den meisten Gemeinden mehrere *Wahlbezirke*. In dieser Aufgabe wird angenommen, dass Gemeinden komplett in einem einzigen *Wahlkreis* liegen (in Wirklichkeit können große Gemeinden (wie Frankfurt) auch mehrere Wahlkreise umfassen).

Nach Beendigung der Wahl (18:00 Uhr) werden die Stimmen in den einzelnen Wahlbezirken gezählt und in die Datenbank eingetragen:

6. Für jeden *Wahlbezirk* wird eingetragen, wieviele Erststimmen die einzelnen Parteien bekommen haben, und wieviele Zweitstimmen die einzelnen Parteien bekommen haben.

Z.B. hat die CDU im Wahlbezirk *Lich-Kirche* 200 Erststimmen bekommen. Im selben Wahlbezirk hat die SPD 187 Erststimmen und die FDP 16 Erststimmen bekommen (und drei weitere Parteien den Rest, die anderen hatten keinen Wahlkreiskandidaten in diesem Wahlkreis aufgestellt). Im Wahlbezirk *Lich-Rathaus* hat die SPD 239 Erststimmen bekommen, und die CDU 184, usw.

Die Zweitstimmen-Ergebnisse im Wahlbezirk *Lich-Kirche* sind CDU: 179, SPD: 148, FDP: 32 und weitere ...

7. Außerdem sind für jede Partei die *Kandidaten auf den Plätzen der jeweiligen Landesliste* (mit denen der Landtag entsprechend der Zweitstimmen aufgefüllt wird) gespeichert. Auf Platz 1 der Landesliste der CDU steht *Volker Bouffier*, auf Platz 2 steht *Eva Kühne-Hörmann*. Auf Platz 1 der Landesliste der SPD steht *Torsten Schäfer-Gümbel*. In beiden Fällen umfasst die Liste mehr als 100 Plätze; die anderen Parteien haben weniger Plätze vergeben.

Am Ende ging es dann so aus:

Volker Bouffier (CDU) gewann den Wahlkreis Gießen-Land mit 21946 Erststimmen (Anteil der Erststimmen über alle Wahlbezirke dieses Wahlkreises) vor *Torsten Schäfer-Gümbel* (SPD) mit 17245 Erststimmen. *Volker Bouffier* erhielt somit das Direktmandat als Wahlkreiskandidat für die CDU.

Landesweit hat die SPD 19,8% der Zweitstimmen erhalten. Sie ist in 10 Wahlkreisen stärkste Partei bei den Erststimmen geworden, d.h. hat 10 Direktmandate gewonnen. Mit 19,8% der Zweitstimmen stehen ihr insgesamt 22 der 110 Landtagsmandate zu, also kommen die ersten 12 Kandidaten der SPD-Landesliste, die kein Direktmandat erhalten haben, in den Landtag (darunter auf Listenplatz 1 *Torsten Schäfer-Gümbel*).⁶

⁶In Wirklichkeit gibt es noch Überhangmandate und Ausgleichsmandate, die in dieser Aufgabenstellung ignoriert werden.

Name:

MatNr.:

Aufgabe 1 (ER-Modell [15 Punkte])

Entwickeln Sie ein ER-Modell für das Szenario. Geben Sie darin die Schlüsselattribute sowie die Beziehungskardinalitäten an.

Aufgabe 2 (Transformation in das Relationale Modell [18 Punkte])

- a) **Lösen Sie diesen Aufgabenteil auf dem *letzten* Blatt und trennen dieses ab (und geben es am Ende mit ab!).** Dann haben Sie dieses Blatt separat zugreifbar um später damit die Aufgaben 2b, 3 und 4 (SQL, Relationale Algebra+SQL, Diverses) zu lösen.

Geben Sie an, welche Tabellen (mit Attributen, Schlüsseln etc.) Ihre Datenbank enthält (keine SQL CREATE TABLE-Statements, sondern einfach grafisch). (12 P)

Markieren Sie dabei auch Schlüssel (durch unterstreichen) und Fremdschlüssel (durch überstreichen).

Geben Sie die Tabellen mit jeweils mindestens zwei Beispieldupeln (z.B. denen, die sich aus dem Aufgabentext ergeben, und/oder weiteren erfundenen) an.

- b) Geben Sie das CREATE TABLE-Statement für diejenige Tabelle, in der die Wahlkreis-kandidaturen gespeichert sind, so vollständig wie möglich (d.h. mit allen notwendigen Constraints) an (6 P).

Aufgabe 3 (SQL und Relationale Algebra [46 Punkte])

Verwenden Sie für diese Aufgabe die von Ihnen entworfene relationale Datenbasis. Keine der Antworten soll Duplikate enthalten.

- a) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck oder -Baum an, die die Namen und den Wohnort derjenigen *Wahlkreiskandidaten* ausgeben, die in dem Wahlkreis, zu dem ihr Wohnort gehört, kandidieren. (3+3P)
- b) Geben Sie eine SQL-Anfrage an, die für jeden *Wahlbezirk* den Namen des Wahlbezirks sowie die Partei (Abkürzung) ausgibt, die in diesem Wahlbezirk die meisten Zweitstimmen bekommen hat. (3P)
- c) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck oder -Baum an, die die Namen derjenigen *Gemeinden* ausgeben, in denen es keinen Wahlbezirk gibt, in denen die SPD mehr Erststimmen erhalten hat, als die CDU. (4+4P)
- d) Geben Sie eine SQL-Anfrage an, die die *Erststimmen*-Wahlergebnisse für die *Wahlkreise* berechnet: Für jeden *Wahlkreis* soll für jede Partei die Summe ihrer Erststimmen in den *Wahlbezirken* dieses Wahlkreises angegeben werden. (3P)
Geben Sie einen kleinen Auszug der auszugebenden Tabelle als Beispiel (wie z.B. in Aufgabe 2a) an. (1P)
- e) Definieren Sie ein SQL-View als Ergebnis der Anfrage aus Teil (d) und geben ihm einen Namen, mit dem es im weiteren als Tabelle verwendet werden kann. (1P)
- f) Geben Sie eine SQL-Anfrage an, die für jeden Wahlkreis den Namen und die Partei der gewählten Wahlkreiskandidaten ausgibt (d.h. der Kandidaten, deren Partei in dem entsprechenden Wahlkreis am meisten Erststimmen erhalten hat): (3P)
- g) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck oder -Baum an, die die *Abkürzungen* derjenigen Parteien ausgeben, die in allen Wahlkreisen, in denen mindestens 100.000 Wahlberechtigte leben, einen Wahlkreiskandidaten aufgestellt haben. (5+5P)
- h) Geben Sie eine SQL-Anfrage *und* einen Algebra-Ausdruck oder -Baum an, die die *Abkürzungen* derjenigen Parteien ausgeben, die *NICHT* in allen Wahlkreisen, in denen mindestens 100.000 Wahlberechtigte leben, einen Wahlkreiskandidaten aufgestellt haben. (3+3P)
- i) Seien $R(A, B)$, $S(A)$ und $T(B)$ Relationen, $T \neq \emptyset$.
Sind die beiden folgenden Ausdrücke äquivalent oder nicht (mit Beweis oder nachvollziehbarer Begründung)? (5P)

i) $S - (R \div T)$

ii) $\pi[A]((S \times T) - R)$

Name:

MatNr.:

Aufgabe 4 (Verschiedenes [11 Punkte])

- a) Jemand hat die Tabelle für die Wahlbezirksergebnisse (Erststimmen) folgendermaßen erstellt:

Erststimmen										
Wahlbezirk	CDU	SPD	FDP	Grüne	Linke	AfD	Piraten	PARTEI
Lich-Kirche	200	187	16
Lich-Rathaus
:	:	:	:	:	:	:	:	:	:	:

(Dies ist nebenbei auch ein Hinweis, dass die richtige Lösung zu Aufgabe 2 nicht so aussieht.)

- Warum kann dieser Entwurf nicht entstehen, wenn man von einem guten ER-Diagramm ausgeht? (2P)
- Welche Nachteile hat dieser Entwurf? (2P)

In den folgenden Aufgabenteilen wird wieder das von Ihnen in den Aufgaben 1 und 2 entwickelte Datenbankschema verwendet:

- b) Geben Sie mindestens eines der (evtl. mehreren) SQL-Statements an, mit denen das folgende Ergebnis (Zweitstimmen) des *Wahlbezirks Frankfurt-Oberrad I* in die Datenbank eingefügt wird: CDU 10, SPD 28, FDP 24, (2P)
- c) In den Tagen nach der Wahl hat sich herausgestellt, dass in manchen Wahlbezirken falsche Ergebnisse in das System eingegeben wurden.

Geben Sie mindestens eines der (evtl. mehreren) SQL-Statements an, mit denen das Ergebnis (Zweitstimmen) des *Wahlbezirks Frankfurt-Oberrad I* wie folgt *korrigiert* wird: CDU 200, SPD 183, FDP 35, (2P)

- d) Am Wahlabend, als die in den einzelnen Wahlbezirken ausgezählten Ergebnisse von den Wahlhelfern über die Anwendung “WahlWeb” in die zentrale Datenbank eingegeben wurden, fiel das System zeitweise aus. Begründung (aus einem Artikel der FAZ): “Weil am Wahlabend aus Sicherheitsgründen zu häufig abgespeichert worden sei, habe das System von etwa 20 bis 21 Uhr zu langsam gearbeitet.”

Wie ist diese Begründung mit den in der Vorlesung erworbenen Kenntnissen zu Datenbanktechnologie zu bewerten? (3P)