

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	Schätzung für "4"
Aufgabe 1 (ER-Modell)	15	12
Aufgabe 2 (Transformation in das Relationale Modell)	18	14
Aufgabe 3 (SQL und Relationale Algebra)	46	24
Aufgabe 4 (Verschiedenes)	11	6
Summe	90	56

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 2018.¹ 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
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 Wähler³ hat zwei Stimmen, eine *Erststimme* und eine *Zweitstimme*. Mit der *Erststimme* kann man eine Stimme für einen *Wahlkreiskandidaten* einer Partei abgeben. 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.

¹Einige Namen und konkrete Werte wurden zum einfacheren Verständnis verändert.

²Im Text dieser Aufgabe werden ausschliesslich Parteien verwendet, die eine aus drei Großbuchstaben bestehende Abkürzung haben, um umständliche Formulierungen wie “die Partei “die Grünen”” zu vermeiden.

³In diesem Text wird zur besseren Verständlichkeit die kurze, traditionelle Sprachform verwendet. Es sind alle Wähler*innen, Kandidat*innen gemeint.

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).

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 darf angenommen werden, dass Gemeinden komplett in einem einzigen *Wahlkreis* liegen (in Wirklichkeit können große Gemeinden (wie Frankfurt) auch mehrere Wahlkreise umfassen⁴).

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 z.B. 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*).⁵

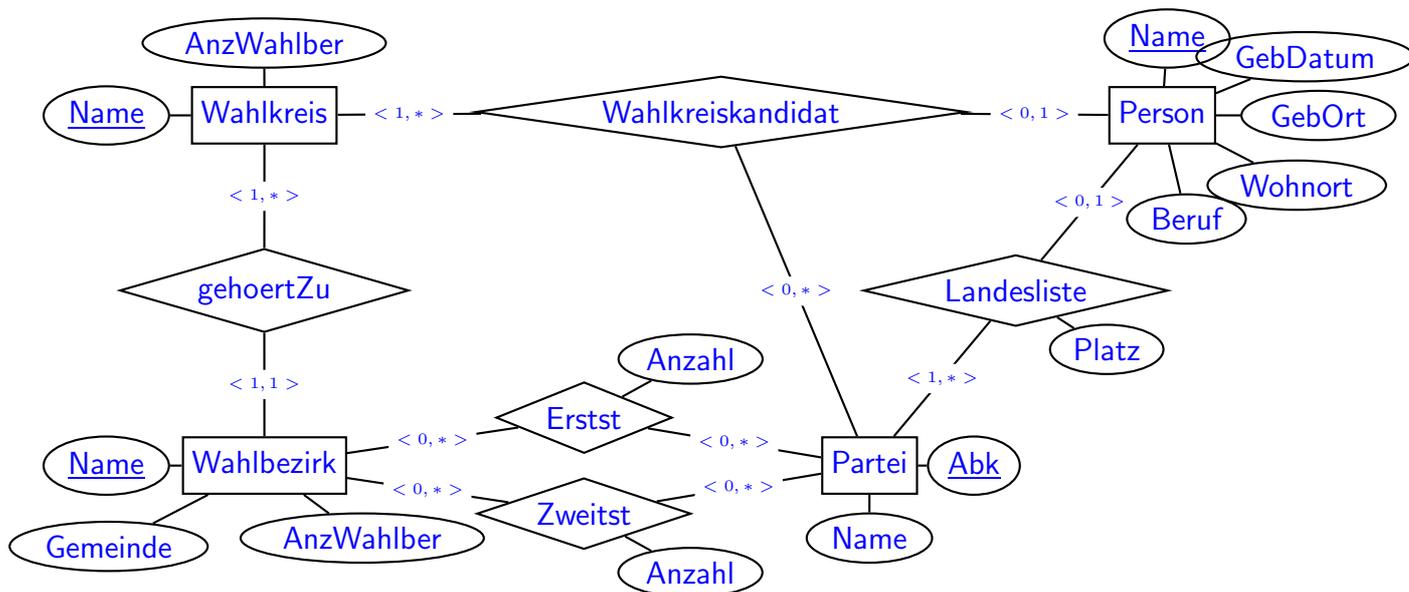
⁴Die Musterlösung ist für den allgemeinen Fall geeignet.

⁵In Wirklichkeit gibt es noch Überhangmandate und Ausgleichsmandate, die in dieser Aufgabenstel-

Aufgabe 1 (ER-Modell [15 Punkte])

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

Lösung



Wichtig:

- $\langle -, * \rangle$ bei Partei \rightarrow Wahlkreiskandidat und bei Wahlkreis \rightarrow Wahlkreiskandidat. Mit einer solchen dreistelligen Beziehung kann man im ER-Modell nicht ausdrücken, dass es höchstens einen Kandidaten pro Paar (Wahlkreis, Partei) gibt. Dies geht erst bei der Erstellung des relationalen Schemas (siehe Aufgabe 2b).

Alternativen und Kommentare:

- Nur eine Beziehung *Ergebnis*(Wahlbez, Partei; Erst, Zweit) Dies bringt aber zwei Dinge zusammen, die zwar einen semantischen Zusammenhang haben, aber besser in zwei "atomare" Einheiten aufgespalten werden. Bei der Auszählung sind ja meistens nicht beide Werte gleichzeitig fertig (vgl. Aufgaben 4b und 4c).
- *Ergebnis* (Wahlbezirk, Partei; erststimmen, zweitstimmen) kann auch als schwacher Entitätstyp modelliert werden.
- *Landeslistenplatz*(Partei, Person; platz) als schwacher Entitätstyp; dann ist entweder Partei identifizierender Entitätstyp und platz lokaler Schlüssel, oder nur Person der identifizierende Entitätstyp (und es gibt keinen lokalen Schlüsselanteil).
- *Wahlkreiskandidatur*(Partei, Person, Wahlkreis) als schwacher Entitätstyp. Dann kann man entweder Person oder (Wahlkreis, Partei) als identifizierende Entitätstypen nehmen (vgl. auch Aufgabe 2b).

lung ignoriert werden. Die CDU hat 40 Direktmandate gewonnen, aber nur 27% der Zweitstimmen, womit ihr nur 30 Mandate zustehen würden. Dies führte dazu, dass der Landtag mit "Ausgleichmandaten" auf 137 Sitze vergrößert wurde (und damit die CDU davon genau ihre 40 Sitze und damit 27% hat), und die SPD damit 29 Sitze bekam, also noch 7 weitere Plätze von der Landesliste einen Sitz bekamen.

- Man kann *Gemeinde* auch als Entitätstyp modellieren, der dann nur das Attribut *Name* hat. Da man (als Geburtsorte) dann eigentlich auch nicht-hessische Orte und sogar ggf. nicht-deutsche Orte aufnehmen müsste, ist das aber nicht sehr sinnvoll. Dann hat man eine $\langle 0, * \rangle$ -Kardinalität zwischen Gemeinde und Wahlkreis/-Bezirk, da solche Orte keinem Wahlkreis/-bezirk angehören.
- Die Kardinalitäten $\langle 0, * \rangle$ zwischen Partei/Wahlbezirk und Erst-/Zweitstimme repräsentieren die "leere" Datenbank vor der Stimmenauszählung.
- Man kann "Listenkandidat" und "Wahlkreiskandidat" als (nicht-disjunkte) Spezialisierungen von Person modellieren. Das bringt aber nicht viel, weil diese keine eigenen Attribute haben.

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, Schlüssel 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.

Lösung

Wahlkreis	
<u>Name</u>	AnzWahlber
Gießen-Stadt	98075
Gießen-Land	93639
:	:

Wahlbezirk			
<u>Name</u>	Gemeinde	<u>Wahlkreis</u>	AnzWahlber
Lich-Kirche	Lich	Gießen-Land	828
Lich-Rathaus	Lich	Gießen-Land	963
:	:	:	:

Person				
<u>Name</u>	GebDat	GebOrt	Wohnort	Beruf
V.Bouffier	18.12.1951	Gießen	Gießen	Rechtsanwalt
T.S.-G.	1.10.1969	Oberstdorf	Lich	PolitikWiss.
E.K.-H.
:	:	:	:	:

Partei	
<u>Abk</u>	Name
CDU	Christl-Demokr. U.
SPD	Sozialdem. P. Dtsch.
:	:

Wahlkreiskandidat		
<u>Name</u>	<u>Partei</u>	<u>Wahlkreis</u>
V.Bouffier	CDU	Gießen-Land
T.S.-G.	SPD	Gießen-Land
:	:	:

Landesliste		
<u>Name</u>	<u>Partei</u>	<u>Platz</u>
V.Bouffier	CDU	1
E.K.-H.	SPD	2
T.S.-G.	SPD	1
:	:	:

Erst		
<u>Wahlbez</u>	<u>Partei</u>	Anzahl
Lich-Kirche	CDU	200
Lich-Kirche	SPD	187
Lich-Kirche	FDP	16
:	:	:

Zweit		
<u>Wahlbez</u>	<u>Partei</u>	Anzahl
Lich-Kirche	CDU	179
Lich-Kirche	SPD	148
Lich-Kirche	FDP	32
:	:	:

- Bei der Tabelle *Wahlkreiskandidat* kann auch (*Person*) als Primary Key verwendet werden; bei der Tabelle *Landesliste* kann auch (*Person*) als Primary Key verwendet werden. Die oben dargestellte Variante bildet die Anwendungssemantik besser ab. Man hat man hier also jeweils zwei *Candidate Keys* zur Auswahl.
- Man kann auch die Beziehungen *Wahlkreiskandidat* und *Landesliste* mit in die Tabelle für *Person* aufnehmen. Jede Person kann ja maximal in einem Wahlkreis Wahlkreiskandidat

sein und höchstens auf einem Landeslistenplatz stehen (ggf auch beides). Man hat dann ggf. Nullwerte (und sollte auch einige UNIQUE-Constraints):

Person							
Name	GebDat	GebOrt	Wohnort	Beruf	Partei	WahlkrKandFuer	Listenpl.
V.Bouffier	18.12.1951	Gießen	Gießen	Rechtsanwalt	CDU	Gießen-Land	1
E.K.-H.	CDU	Kassel-Stadt	2
:	:	:	:	:	:	:	:
Karl Napf	CDU	<i>null</i>	123
T.S.-G.	1.10.1969	Oberstdorf	Lich	PolitikWiss.	SPD	Gießen-Land	1
:	:	:	:	:	:	:	:

Weniger vorteilhaft (vgl. auch Kommentar zu Aufgabe 1, und Aufgaben 4b und 4c) ist es, die Erst- und Zweitstimmenergebnisse in einer gemeinsamen Tabelle zu speichern:

Ergebnis			
Wahlbez	Partei	Erst	Zweit
Lich-Kirche	CDU	200	179
Lich-Kirche	SPD	187	148
Lich-Kirche	FDP	16	32
:	:	:	:

- 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).

Lösung

```
CREATE TABLE Wahlkreiskandidat                               Basis 3.5P,
( Name          VARCHAR2(50) REFERENCES Person(Name) NOT NULL UNIQUE, 1/2+1/2 P
  Partei        VARCHAR2(40) REFERENCES Partei(Name),      1/2P
  Wahlkreis     VARCHAR2(40) REFERENCES Partei(Name),      1/2P
  PRIMARY KEY (Partei, Wahlkreis))                          1/2P
```

```
CREATE TABLE Wahlkreiskandidat                               Basis 3P,
( Name          VARCHAR2(50) REFERENCES Person(Name) PRIMARY KEY, 1/2+1/2 P
  Partei        VARCHAR2(40) REFERENCES Partei(Name) NOT NULL, 1/2+1/2P
  Wahlkreis     VARCHAR2(40) REFERENCES Partei(Name) NOT NULL, 1/2+1/2P
  UNIQUE (Partei, Wahlkreis))                                1/2P
```

Bei der ersten Variante könnte man argumentieren, dass "Person" NULL sein darf, und bedeutet, dass eine Partei in einem Wahlkreis (noch) keinen Listenkandidaten aufgestellt hat, bzw. dass evtl. am Anfang dort Einträge noch fehlen. Dann muss man allerdings in Aufgabe (3g) mit der Division aufpassen, und dort "AND name IS NOT NULL" prüfen. Normalerweise wird man Tripel in diese Tabelle immer dann eintragen, wenn ein Kandidat festgelegt wurde.

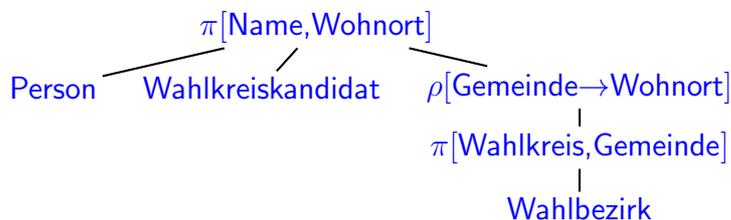
Aufgabe 3 (SQL und Relationale Algebra [46 Punkte])

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)

Lösung

```
SELECT Person.Name, Wohnort
FROM Person, Wahlkreiskandidat, Wahlbezirk
WHERE Person.Name = Wahlkreiskandidat.Name
      AND Wahlbezirk.Wahlkreis = Wahlkreiskandidat.Wahlkreis
      AND Person.Wohnort = Wahlkreis.Gemeinde
```



- 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)

Lösung

```
SELECT z1.wahlbez, z1.partei
FROM Zweit z1
WHERE Anzahl = (
    SELECT MAX(anzahl)
    FROM Zweit z2
    WHERE z2.wahlbez = z1.wahlbez)
```

```
SELECT z1.wahlbez, z1.partei
FROM Zweit z1
WHERE NOT EXISTS (SELECT *
    FROM Zweit z2
    WHERE z2.wahlbez = z1.wahlbez
    AND z2.anzahl > z1.anzahl)
```

```
SELECT z1.wahlbez, z1.partei
FROM Zweit z1
WHERE Anzahl >= ALL (
    SELECT Anzahl
    FROM Zweit z2
    WHERE z2.Wahlbez = z1.Wahlbez)
```

```
SELECT z1.wahlbez, z1.partei
FROM Zweit z1,
    (SELECT z2.wahlbez, MAX(anzahl) as max
    FROM Zweit z2
    GROUP BY z2.Wahlbez) maxe
WHERE z1.wahlbez = maxe.wahlbez
AND z1.anzahl = maxe.max
```

```
SELECT wahlbez, partei --- und so braucht man nicht mal mehr z1, z2:
FROM Zweit
WHERE (wahlbez, anzahl)
```

```

IN ( SELECT wahlbez, MAX(anzahl)
      FROM Zweit
      GROUP BY Wahlbez )

```

- 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)

Lösung

```

SELECT DISTINCT gemeinde
FROM Wahlbezirk g
WHERE NOT EXISTS
  (SELECT *
   FROM Wahlbezirk w, Erst e1, Erst e2
   WHERE g.gemeinde = w.gemeinde
        AND w.name = e1.wahlbez
        AND w.name = e2.Wahlbez
        AND e1.partei = 'CDU'
        AND e2.partei = 'SPD'
        AND e2.anzahl > e1.anzahl)

```

```

(SELECT gemeinde
 FROM Wahlbezirk g)
MINUS
(SELECT gemeinde
 FROM Wahlbezirk w
 WHERE g.gemeinde = w.gemeinde
 AND (SELECT anzahl FROM Erst e1
      WHERE w.name = e1.wahlbez
      AND e1.partei='SPD')
 > (SELECT anzahl FROM Erst e2
    WHERE w.name = e2.wahlbez
    AND e2.partei='CDU')) )

```

```

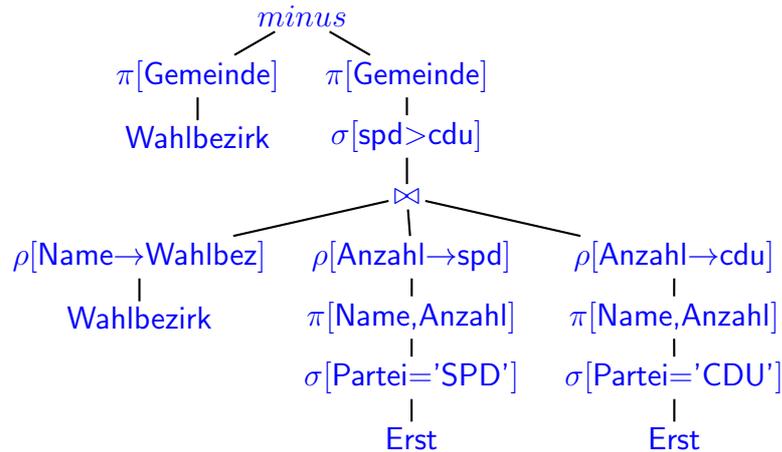
(SELECT gemeinde
 FROM Wahlbezirk g)
MINUS
(SELECT gemeinde
 FROM Wahlbezirk w,
  (SELECT * FROM Erst WHERE partei='CDU') cdu,
  (SELECT * FROM Erst WHERE partei='SPD') spd,
 WHERE w.name = cdu.wahlbez AND w.name = spd.wahlbez
 AND spd.anzahl > cdu.anzahl)

```

-- wenn man eine Variante mit "MINUS" waehlt, wird kein DISTINCT benoetigt

(Wenn man ganz spitzfindig ist, müsste man noch einbeziehen, dass eine der beiden Parteien in einem Wahlbezirk keine Erststimmen bekommen hat, und kein Tupel für sie existiert.)

Wenn man eine Tabelle *Gemeinde(Name)* hat, kann man die verwenden, muss aber ggf. auf Hessen einschränken



d) Geben Sie eine SQL-Anfrage an, die die *Erststimmen*-Wahlergebnisse für die *Wahlkreis* 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)

Lösung

```
SELECT wahlkreis, partei, SUM(anzahl) AS anzahl
FROM Wahlbezirk, Erst
WHERE Wahlbezirk.name = Erst.wahlbez
GROUP BY wahlkreis, partei
```

Das Ergebnis der Anfrage (und das View/virtuelle Tabelle in der nächsten Teilaufgabe) sieht dann so aus (V. Bouffier hat das Duell der Landesspitzenkandidaten im Wahlkreis Gießen-Land mit 34% zu 27% der abgegebenen Stimmen gewonnen):

WahlkreisErst		
Wahlkreis	Partei	Anzahl
Gießen-Stadt	CDU	...
Gießen-Stadt	SPD	...
Gießen-Stadt	FDP	...
:	:	:
Gießen-Land	CDU	21946
Gießen-Land	SPD	17245
Gießen-Land	FDP	3392
:	:	:

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)

Lösung

```
CREATE VIEW WahlkreisErst AS (
    SELECT wahlkreis, partei, SUM(anzahl) AS anzahl
    FROM Wahlbezirk, Erst
    WHERE Wahlbezirk.name = Erst.wahlbez
    GROUP BY wahlkreis, partei )
```

- 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)

Lösung Hier wird das View "WahlkreisErst" aus der vorigen Teilaufgabe verwendet:

```
SELECT d.wahlkreis, d.name, d.partei
FROM WahlkreisErst e1, Wahlkreiskandidat d
WHERE e1.wahlbez = d.wahlkreis
    AND e1.partei = d.partei
    AND e1.anzahl = (
        SELECT MAX(anzahl)
        FROM WahlkreisErst e2
        WHERE e2.wahlkreis = e1.wahlkreis)
```

```
SELECT d.wahlkreis, d.name, d.partei
FROM WahlkreisErst e1, Wahlkreiskandidat d
WHERE e1.wahlbez = d.wahlkreis
    AND e1.partei = d.partei
    AND NOT EXISTS (SELECT *
        FROM WahlkreisErst e2
        WHERE e2.wahlkreis = e1.wahlkreis
            AND e2.anzahl > e1.anzahl)
```

... oder z.B. eine der anderen Anfragestrukturen aus Aufgabe 3b.

- 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)

Lösung

```
SELECT abk
FROM Partei
WHERE NOT EXISTS
    (SELECT name
     FROM Wahlkreis w
     WHERE anzWahlb >= 100000
        AND NOT EXISTS
            (SELECT *
             FROM Wahlkreiskandidat wkk
```

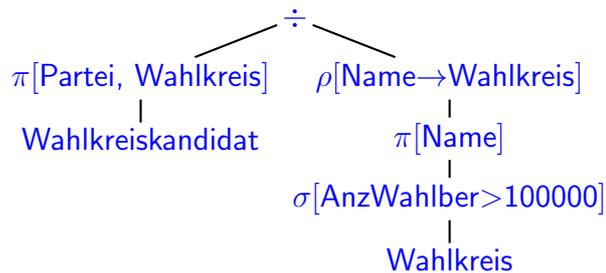
```
SELECT abk
FROM Partei
WHERE NOT EXISTS
    (SELECT name
     FROM Wahlkreis
     WHERE AnzWahlb >= 100000
        AND NOT (Partei.abk, Wahlkreis.name)
            IN (SELECT Partei, Wahlkreis
                FROM Wahlkreiskandidat))
```

```

WHERE w.partei = Partei.abk
      AND wkk.wahlkreis = w.name))

(SELECT abk
 FROM Partei)
MINUS
(SELECT abk
 FROM ((SELECT DISTINCT abk, w.name
        FROM Partei, Wahlkreis w
        WHERE AnzWahlber >= 100000 )
 MINUS
 (SELECT partei, wahlkreis
  FROM Wahlkreiskandidat)
 ))

```



- 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)

Lösung Das ist offensichtlich die Negation von Aufgabenteil (g):

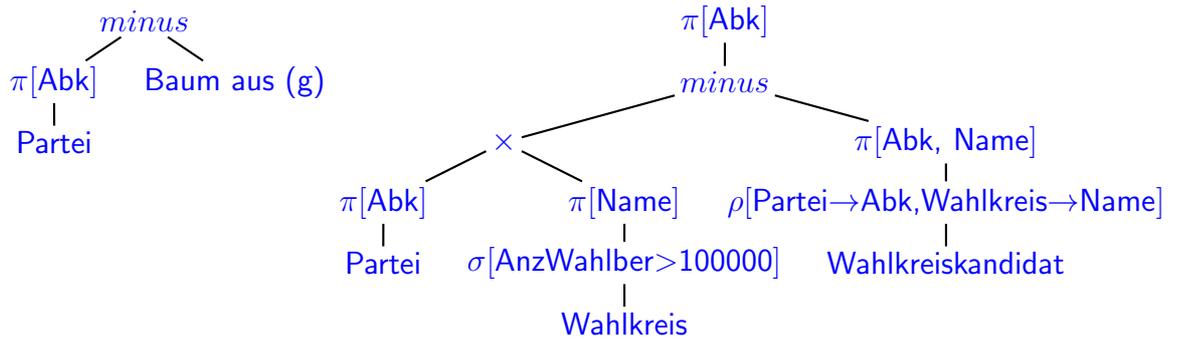
```

-- das aeussere NOT weglassen:
SELECT abk
FROM Partei
WHERE EXISTS
  (SELECT name
   FROM Wahlkreis
   WHERE AnzWahlb >= 100000
    AND NOT (Partei.abk, Wahlkreis.name)
             IN (SELECT partei, wahlkreis
                 FROM Wahlkreiskandidat))

-- einfach die Parteien nehmen,
-- die in (g) nicht herauskamen:
SELECT abk
FROM Partei
WHERE abk NOT IN (g)

-- die vorhandenen von allen (partei,wahlkreis>100000)-Paaren abziehen
SELECT partei
FROM (SELECT abk FROM partei) x,
     (SELECT name FROM Wahlkreis
      WHERE AnzWahlb >= 100000) y
WHERE NOT (x.abk, y.name) IN
  (SELECT partei, wahlkreis
   FROM Wahlkreiskandidat)

```



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

Lösung Die Äquivalenz gilt: Beide Terme beschreiben "alle $a \in S$, die nicht mit jedem $b \in T$ als (a, b) in R vorkommen".

Man kann beide Terme im obigen Szenario anhand des Beispiels aus Aufgabenteil (h) "Alle Parteien, die *nicht* in allen Wahlkreisen mit mehr als 100.000 Wahlberechtigten einen Wahlkreiskandidaten aufgestellt haben" erklären:

- i) ist genau diese Anfrage unter Verwendung der Division, der erste bei (h) angegebene Algebra-Baum;
- ii) ist genau die dritte der SQL-Formen oben (bzw. der zweite bei (h) angegebene Algebra-Baum), nämlich die Parteien der "fehlenden" Einträge (Partei, Wahlbezirk)

Die relationale Division entspricht dem Allquantor:

$$R \div T = \{a \in \pi[A](R) : \forall b \in T : (a, b) \in R\}.$$

Damit ist

$$\begin{aligned} (i) \quad S - (R \div T) &= \{a \in S : a \notin R \div T\} = \\ &= \{a \in S : a \notin \pi[A](R) \text{ oder } \exists b \in T : (a, b) \notin R\} \\ &\quad \text{wenn } a \notin \pi[A](R), \text{ da } T \neq \emptyset, \text{ gibt es ein } b \in T \text{ so dass } (a, b) \notin R \\ &= \{a \in S : \exists b \in T : (a, b) \notin R\} . \end{aligned}$$

Noch einfacher ist auch direkt

$$(ii) \quad \pi[A]((S \times T) - R) = \{a \in S : \exists b \in T : (a, b) \notin R\}$$

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)

Lösung Im ER-Diagramm kommen nur Entitätstypen, Beziehungstypen und Attribute vor. Nur diese werden bei der Umsetzung in das relationale Modell als Tabellen- und Spaltennamen verwendet. Konkrete Werte (Namen von Parteien, "CDU", "SPD") etc. eines Anwendungsfalles können daher nicht im relationale Schema auftreten. In einem guten Entwurf darf dies auch nicht der Fall sein.

- Welche Nachteile hat dieser Entwurf? (2P)

Lösung

- * Es wäre sehr umständlich, eine Anfrage zu formulieren, die die stärkste Partei feststellt.
Die Anfrage aus Aufgabe (3b) (stärkste Partei im Wahlbezirk) wäre sehr umständlich zu formulieren; ebenso wäre die Anfrage aus Aufgabe (3d) (Wahlkreissumme für alle Parteien) sehr lang (jede Partei ein einzelnes Teilstatement über die entsprechende Spalte). (Die Anfrage aus (3c) (SPD stärker als CDU) wäre jedoch sogar einfacher zu formulieren).
- * Wenn bei einer anderen Wahl andere Parteien antreten, müßte man die Spaltennamen ändern bzw. Spalten löschen oder hinzufügen, und die o.g. Anfragen (3b, 3d) ändern.

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)

Lösung

```
INSERT INTO Zweit VALUES('Frankfurt-Oberrad I', 'CDU', 10);
INSERT INTO Zweit VALUES('Frankfurt-Oberrad I', 'SPD', 28);
INSERT INTO Zweit VALUES('Frankfurt-Oberrad I', 'FDP', 24);
```

- 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)

Lösung

```
UPDATE Zweit
SET Anzahl = 200
WHERE Wahlbez = 'Frankfurt-Oberrad I' AND Partei = 'CDU';
```

```
UPDATE Zweit
SET Anzahl = 183
WHERE Wahlbez = 'Frankfurt-Oberrad I' AND Partei = 'SPD';
```

Weiteres kann man in den folgenden FAZ-Artikeln nachlesen:

<http://www.faz.net/aktuell/rhein-main/nach-hessischer-landtagswahl-wahlbezirke-muessen-neu-ausgezaehlt-werden-15871496.html>

<http://www.faz.net/aktuell/rhein-main/region-und-hessen/in-hessen-koennte-es-jetzt-doch-eine-ampelkoalition-geben-15881424.html>

<http://www.faz.net/aktuell/rhein-main/frankfurter-reaktionen-auf-wahlpanne-15882870.html>

[...] Die Wahlpanne in ihrem Stadtteil ärgert viele Oberräder maßlos. 364 Stimmen wurden im Wahlbezirk 38004 rund um die Gruneliuschule nicht gewertet, weil ein Schriftführer nur ein Teilergebnis ans Wahlamt durchgab und der Fehler wegen einer Computerpanne nicht bemerkt wurde. [...] Der Schriftführer hatte dem Wahlamt versehentlich aber nur jene Wahlkreis- und Landesstimmen durchgegeben, die auf verschiedene Parteien verteilt worden waren. [...]

<http://www.faz.net/aktuell/rhein-main/hessischer-landeswahlleiter-spricht-ueber-panne-15894731.html>

- 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)

Lösung Das klingt nicht gut. Wie im Kapitel “Transaktionen” behandelt, arbeiten die üblichen Datenbanksysteme mit Transaktionen. Alle *committeten* Transaktionen (d.h. die in ihnen durchgeführten Inserts, Updates und Deletions) sind bei geeigneter Konfiguration und Installation des Datenbanksystems automatisch sicher und ihre Persistenz ist garantiert. Im Fall eines Absturzes kann mit Hilfe des Server-Logs (und ggf. eines zusätzlichen externen Logs) der entsprechende konsistente Zustand wieder hergestellt werden.

Zusätzliche Speicherungen zur Laufzeit sind also nicht notwendig.

Entweder liegt dem "WahlWeb" gar keine Datenbank zugrunde (sondern eine hauptspeicherbasierte Implementierung in einer "normalen" Programmiersprache, bei der tatsächlich explizit ein Dump gespeichert werden muss, oder die Begründung für den Zusammenbruch des Systems ist falsch.

Weitere Gedanken dazu:

Performance: es sind 6381 Wahlbezirke (mit 4,4 Millionen Wahlberechtigten), 23 Parteien, insgesamt aber nur 413 Wahlkreiskandidaten (also 7,5 pro Wahlkreis). Damit je 6381 Einzelübermittlungen von 23-Tupeln und 7- bzw 8-Wertepaaren per Web-Service, die intern in entsprechend viele Einzeleinträge in "Erststimme" / "Zweitstimme" bzw. "Ergebnis" umgesetzt werden. Da darf die Datenbank nicht der Engpass sein – zumal es keine Schreib-Konflikte geben kann, da jeder Wahlkreis seine eigenen Tupel erzeugt.

Der Web-Service muss gerade mal soviele Verbindungen aufmachen können wie mehrere Wahlbezirke in derselben Sekunde übermitteln wollen. Wenn "alle" innerhalb zwei Stunden fertig werden, sind das 100 Verbindungen pro Minute. Naja. Selbst wenn man den HTTPS-Aufwand dazurechnet, ist das nicht viel. Als Default bei tomcat ohne weitere Zusatzpakete sind max. 200 Verbindungen gleichzeitig möglich.

Den Artikel findet man unter

<http://www.faz.net/aktuell/rhein-main/spd-hessen-koennte-nach-wahlpanne-zweitstaerkste-kraft-werden-15879431.html>.

Er ist –nicht nur aus IT-Sicht– lesenswert.

[...] Eine knappe Stunde nach Schließen der Wahllokale sei das "Wahlweb Hessen", das an diesem Abend zum ersten Mal zum Einsatz kam, so überlastet gewesen, dass sich die Mitarbeiter des Frankfurter Wahlamts nicht mehr hätten anmelden können. "System derzeit nicht erreichbar", habe auf den Monitoren gestanden.

Nach Darstellung Grochockis ist so eine Art Rückstau bis in die Wahllokale entstanden, der sich "äußerst negativ in zeitlicher und qualitativer Hinsicht" auf die Erfassung ausgewirkt habe. Die Mitarbeiter des Wahlamts seien im Lauf des Abends dazu übergegangen, die Ergebnisse aus den Wahllokalen von Hand zu notieren. Dadurch sei es zu Übertragungsfehlern gekommen, zudem habe keine automatische Plausibilitätsprüfung stattgefunden. Komplette ausgebliebene Ergebnisse seien aufgrund der Resultate in benachbarten Stimmbezirken geschätzt worden.

[...]

[...] das "Wahlweb" als ein modernes und bewährtes System zur Eingabe, Kontrolle, Korrektur und Freigabe von Wahlergebnissen. Es sei in Hessen schon seit 2002 im Einsatz, allerdings hätten die Wahlämter die Zahlen bei dieser Landtagswahl erstmals direkt eingeben müssen, anstatt ein System der Kommune vorzuschalten. Weil am Wahlabend aus Sicherheitsgründen zu häufig abgespeichert worden sei, habe das System von etwa 20 bis 21 Uhr zu langsam gearbeitet.

[...]