

Klausur Datenbanken
Wintersemester 2015/2016
Prof. Dr. Wolfgang May
2. März 2016, 14-16 Uhr
Bearbeitungszeit: 90 Minuten

Vorname:

Nachname:

Matrikelnummer:

Bei der Klausur sind **keine Hilfsmittel** (Skripten, Taschenrechner, etc.) erlaubt. Handies müssen ausgeschaltet sein. Papier wird gestellt. Benutzen Sie nur 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.

- meine Note soll mit Matrikelnummer so bald wie möglich auf der Vorlesungs-Webseite veröffentlicht werden.
- meine Note soll nicht veröffentlicht werden; ich erfahre sie dann aus FlexNever oder beim zuständigen Prüfungsamt.

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (ER-Modell)	18	
Aufgabe 2 (Transformation in das Relationale Modell)	18	
Aufgabe 3 (SQL und Relationale Algebra)	40	
Aufgabe 4 (Verschiedenes)	14	
Summe	90	

Note:

Themenstellung: Gemälde in Museen

Alle Klausuraufgaben basieren auf einem gemeinsamen "Auftrag": In der Klausur soll eine Datenbank über Maler und Gemälde entworfen werden.

1. Maler haben –wie alle Personen– einen Namen (meistens Vorname+Nachname; es gibt keine zwei Personen mit demselben Namen). Ausserdem ist ihr Geburtsdatum, ihr Todesdatum (falls Sie nicht mehr leben), und ihr Geburtsort gespeichert.

Über jeden Maler ist ausserdem gespeichert, wo er in welchem Zeitraum gelebt hat (wobei nur die Jahreszahlen berücksichtigt werden).

2. Zu jedem Ort/Stadt, der in der Datenbank vorkommt, muss das Land angegeben werden, da es mehrere gleichnamige Städte in verschiedenen Ländern geben kann. Innerhalb eines Landes ist der Stadtname eindeutig (ggf. durch Ergänzung wie "Frankfurt (Oder)").

Paul Cézanne wurde am 19. Januar 1839 in *Aix-en-Provence* in *Frankreich* geboren, wo er auch bis zu seinem Tod am 22. Oktober 1906 lebte.

Vincent van Gogh wurde am 30. März 1853 in *Groot-Zundert* in den *Niederlanden* geboren und ist am 29. Juli 1890 in *Paris (Frankreich)* gestorben. Unter anderem hat er von 1885-1887 in *Paris* und 1888 bis 1889 in *Arles (Frankreich)* gelebt.

Neo Rauch wurde am 18. April 1960 in *Leipzig (Deutschland)* geboren, wo er auch zur Zeit noch lebt.

3. Zu jeder Stadt ist gespeichert, in welcher (geografischen) Landschaft sie liegt. Es gibt keine zwei Landschaften mit demselben Namen. Jede Landschaft liegt in genau einem Land.

Aix-en-Provence und *Arles* liegen in der *Provence*, die komplett in Frankreich liegt.

4. Kunstmuseen haben einen (eindeutigen) Namen und befinden sich in einer Stadt.

Das *Musée d'Orsay* befindet sich in *Paris*.

5. Zu jedem Gemälde ist der Titel gespeichert, und von welchem Maler es gemalt wurde (mehrere Maler können Gemälde mit demselben Titel gemalt haben). Für jedes Bild ist ausserdem gespeichert, wann es gemalt wurde (nur das Jahr), und ggf. in welchem Museum es ausgestellt ist (manche Bilder sind nicht in einem Museum, sondern in Privatbesitz oder verschollen).

Vincent van Gogh malte 1885 das Bild "*Die Kartoffeleesser*", das sich im *Van Gogh Museum* in *Amsterdam (Niederlande)* befindet.

1888 malte er "*Der Maler auf dem Weg nach Tarascon*", das verschwunden ist (vermutlich im 2. Weltkrieg vernichtet).

6. Auf einigen Bildern sind eine oder mehrere (bekannte) Personen, z.B. Maler (ggf. der Maler selber), aber auch andere bekannte Personen dargestellt. Dieses soll auch – sinnvoll abfragbar, siehe Aufgabe 2 – gespeichert werden:

1870 hat *Paul Cezanne* ein Bild "*Paul Alexis liest Emile Zola vor*" gemalt, auf dem die beiden Schriftsteller *Paul Alexis* und *Emile Zola* dargestellt sind. Das Bild befindet sich im *São Paulo Museum of Art* (in *São Paulo, Brasilien*).

Henri de Toulouse-Lautrec hat 1887 ein Bild mit dem Titel "*Vincent van Gogh*" gemalt, auf dem ebendieser dargestellt ist. Dieses Bild befindet sich im *Van Gogh Museum* in *Amsterdam*.

Vincent van Gogh malte 1887 das Bild "*Selbstportrait mit Strohhut*", das in der *National Gallery of Art (NGA)* in *Washington (USA)* zu sehen ist.

7. Einige Bilder stellen eine bestimmte Landschaft dar; auch dies soll in der Datenbank gespeichert werden:

Vincent van Gogh malte 1888 das Bild "*Sternennacht über der Rhone*", das sich im *Musée d'Orsay* in *Paris* befindet. Auf dem Bild ist die Landschaft in der *Provence* zu sehen.

Paul Cezanne malte 1890 das Bild "*Montagne Sainte-Victoire von Gardanne aus gesehen*", das ebenfalls die Landschaft in der *Provence* darstellt, und in der *NGA* in *Washington* hängt.

Aufgabe 1 (ER-Modell [18 Punkte])

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

Name:

MatNr.:

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 weiteren erfundenen) an.

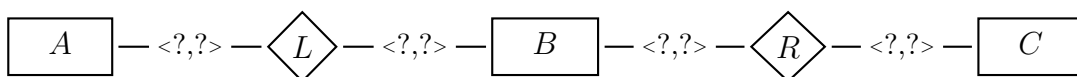
- b) Geben Sie die CREATE TABLE-Statements für diejenige Tabelle, in der gespeichert ist, in welchen Museen sich die Bilder befinden, so vollständig wie möglich an (6 P).

Aufgabe 3 (SQL und Relationale Algebra [40 Punkte])

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

- Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, der die Titel und Maler aller Gemälde ausgibt, die sich in einem Museum in derjenigen Landschaft befinden, die auf dem Gemälde dargestellt ist. (2+2 P)
- Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, der die Namen aller Maler ausgibt, von denen kein Selbstbildnis in der Datenbank gespeichert ist. (2+2 P)
- Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die alle Tripel (Malername, Geburtsdatum, Landschaft) angibt, so dass der Maler in (einer oder mehreren Städten in) dieser Landschaft einige Zeit gelebt hat, aber nie ein Bild gemalt hat, das diese Landschaft darstellt. (3+4 P)
- Geben Sie eine SQL-Anfrage an, die für jede Stadt ausgibt, wieviele Bilder, die zwischen 1870 und 1949 gemalt wurden, sich in den Museen dieser Stadt befinden. Es sollen nur Städte ausgegeben werden, in denen sich mindestens 10 solche Bilder befinden. (3 P)
- Wäre Aufgabenteil (d) einfacher oder schwieriger, wenn die Bedingung "Es sollen nur Städte ausgegeben werden, in denen sich mindestens 10 solche Bilder befinden" fehlen würde? Begründen Sie Ihre Antwort. (3 P)
- Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die/der die Namen aller Museen ausgibt, die von jedem Maler, der ein Bild, das die Landschaft der *Provence* darstellt, gemalt hat, ein solches Bild (d.h., eines, auf dem auch die *Provence* abgebildet ist) besitzen. (4+4 P)
- Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die die Titel aller Bilder und den Namen des jeweiligen Malers ausgeben, auf denen eine Landschaft **und** eine oder mehrere bekannte Personen dargestellt ist. (2+2 P)
- Ein bisschen Theorie (kann auf Basis des vorherigen Aufgabenteils (g) auch real veranschaulicht werden):

Seien $L(A, B)$ und $R(B, C)$ zwei Tabellen, die aus zwei Beziehungen entstanden sind, d.h.



(Die Aufgabe geht rechts oben weiter ...)

Betrachten Sie den folgenden Ausdruck der relationalen Algebra:

$$(L \bowtie R) \setminus (L \ltimes R)$$

- Geben Sie das Format dieses Ausdrucks an (d.h., welche Spalten hat er (1 P))
- Charakterisieren Sie die Menge der Ergebnistupel (textuell oder formal (3 P)).
- Geben Sie eine (sinnvolle) Bedingung an ein oder mehrere der Fragezeichen in dem obigen kleinen ER-Diagramm an, unter der das Ergebnis garantiert in allen erlaubten Datenbankzuständen leer sein muss (mit kurzer Begründung) ? (3 P)

Name:

MatNr.:

Aufgabe 4 (Verschiedenes [14 Punkte])

- a) In welchen Spalten treten in Ihrer Datenbank Nullwerte auf? Welche (wahrscheinlich unterschiedliche) Bedeutung repräsentieren sie in der Realität? (4 P.)
- b) In der *National Gallery of Art (NGA)* wird zunehmend der Platz knapp. Deswegen wird umstrukturiert:
- Eine *National Gallery of Modern Painting (NGMP)* wird gegründet, die sich ebenfalls in Washington befindet.
 - Alle Gemälde aus der *National Gallery of Art*, die zwischen 1870 und 1949 entstanden sind, werden dorthin abgegeben.

Geben Sie die entsprechenden SQL-Befehle an (3 P).

- c) Die Transaktion in Aufgabenteil (b) betrifft einen “größeren” Datenbestand (die Datenbank enthält ein paar hunderttausend Einträge, und es sind evtl. mehrere hundert Bilder davon betroffen). Da lohnt sich schon ein Blick auf die Auswertung und Effizienz. (4 P)
- Skizzieren Sie, wie das Statement algorithmisch auf der in Aufgabe 2 erstellten Datenbank abläuft;
 - Welche internen, datenstrukturbasierten Hilfsmittel der Datenbank können dabei die Effizienz erhöhen (Hinweis: eines ist sicher da, weitere kann der Admin angelegt/konfiguriert haben)? Beschreiben Sie kurz deren Nutzung.
- d) Während der Transaktion in Aufgabenteil (b) stürzt die DB softwaremäßig ab. Einige Bilder “sind” schon in die NGMP umgetragen worden, andere sind noch in der NGA.

Skizzieren Sie, wie die Datenbank nach einem Neustart in einen zulässigen Zustand zurückfindet. (3 P)