

Datenbanksysteme

herausgegeben von der Fachschaft Mathematik/Informatik, umbrochen von Sabine

Aufgabe 1 – ER-Modellierung (14 Punkte)

- a) Vervollständigen Sie auf dem Lösungsblatt das ER-Modell aus Abbildung 1 für ein Fahrzeugszenario, indem Sie die korrekten Typkonstruktoren des EER-Modells einzeichnen und die Attribute unterstreichen, die Sie mindestens als Schlüssel benötigen.

Das Szenario ist durch die folgenden Eigenschaften gekennzeichnet: Fahrzeuge lassen sich in Land- und Wasserfahrzeuge untergliedern. Fahrzeuge dürfen dabei auch gleichzeitig Land- und Wasserfahrzeuge sein. Es darf auch Fahrzeuge geben, die weder Land- noch Wasserfahrzeuge sind.

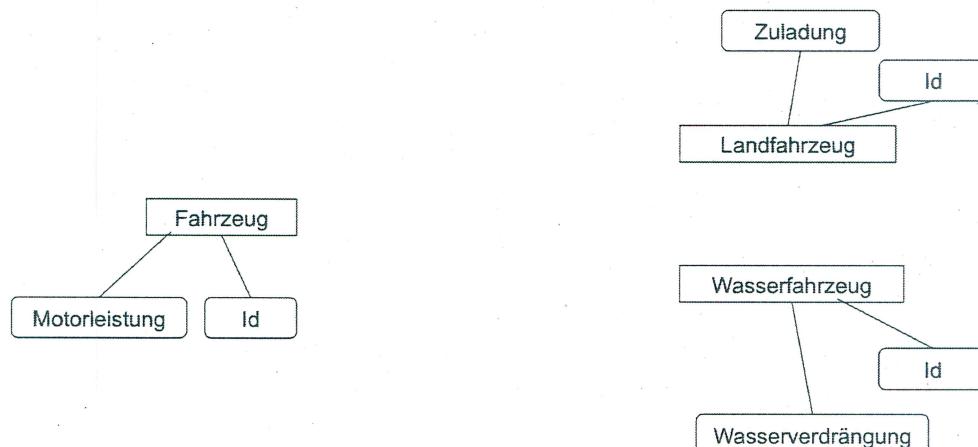


Abbildung 1: EER-Diagramm Fahrzeuge

(3 Punkte)

Gehen Sie für die folgenden Teilaufgaben von dem ER-Modell in Abbildung 2 aus. Die Schlüsselattribute sind durch Unterstreichen gekennzeichnet, die Kardinalitätsangaben sind Teilnehmerkardinalitäten. Das Modell stellt einen Ausschnitt eines Universitätsszenarios dar.

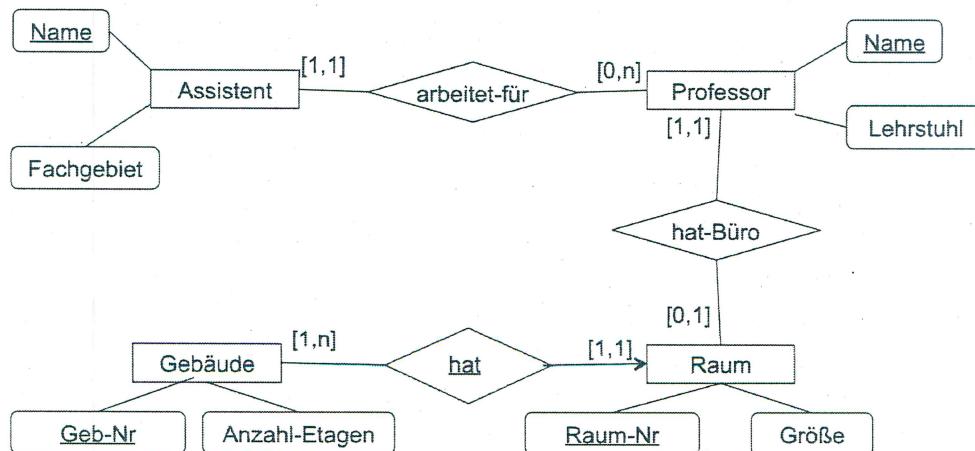


Abbildung 2: ER-Diagramm Universität

- b)** Geben Sie bei den folgenden Aussagen mit „wahr“ oder „falsch“ an, ob sie sich genau so im ER-Modell mit den Teilnehmerkardinalitäten aus Abbildung 2 widerspiegeln. Für jede richtige Antwort gibt es 1/2 Punkt für jede falsche wird 1/2 Punkt abgezogen. Diese Teilaufgabe kann nicht mit weniger als 0 Punkten bewertet werden.

	RICHTIG	FALSCH
Für einen Professor muss immer genau ein Assistent arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>
Ein Gebäude hat mindestens einen, kann aber unendlich viele Räume besitzen.	<input type="checkbox"/>	<input type="checkbox"/>
Ein Professor kann nur dann ein Büro in einem Raum haben, wenn dieser Raum auch einem Gebäude zugeordnet ist.	<input type="checkbox"/>	<input type="checkbox"/>
Ein Professor kann ein Büro haben, muss aber nicht.	<input type="checkbox"/>	<input type="checkbox"/>
Das Modell erlaubt eine eindeutige Zuordnung zwischen jedem Assistenten in der Datenbank und dem Büro seines Professors.	<input type="checkbox"/>	<input type="checkbox"/>
Es ist möglich herauszufinden, in welchem Gebäude ein Professor sein Büro hat.	<input type="checkbox"/>	<input type="checkbox"/>

(3 Punkte)

- c)** Überführen Sie das ER-Modell aus Abbildung 2 kapazitätserhaltend in ein relationales Modell. Optimieren Sie die Relationen dabei soweit wie möglich wie in der Vorlesung besprochen. Verwenden Sie SQL-DDL-Anweisungen. Bitte geben Sie nur die Attribute an, nicht aber deren Datentypen. Geben Sie die Schlüssel und Fremdschlüssele an.

(6 Punkte)

- d)** Modellieren Sie die Entitäten Gebäude und Raum und ihre Beziehung aus Abbildung 2 durch ein EER-Modell mit den erweiterten Möglichkeiten zur Modellierung von Schlüsseln.

(2 Punkte)

Aufgabe 2 – Outlier und Clustering (16 Punkte)

- a)** Was ist ein Outlier? Geben Sie die distanzbasierte Definition aus der Vorlesung an und verwenden Sie genau die gleichen Abkürzungen wie in der Vorlesung, nämlich O für das Outlier-Objekt, T für den Datenbestand sowie p und D exakt im in der Vorlesung vorgestellten Sinn.

(2 Punkte)

- b)** In der Vorlesung wurde ein Algorithmus für die Suche nach dem nächsten Nachbarn vorgestellt, der eine räumliche Indexstruktur nutzt. In diesem Aufgabenteil geht es darum, nicht einen, sondern die k nächsten Nachbarn auszugeben. Schreiben Sie den Algorithmus aus der Vorlesung zusammen mit den dafür erforderlichen Erweiterungen in Pseudocode auf. (Andere Algorithmen, auch wenn sie das richtige Ergebnis generieren, führen zu Punktabzug.)

Hilfestellung: Der Algorithmus aus der Vorlesung hat den Namen und die Parameter nächsterNachbar(Baum, Anfrage), und er ruft eine Funktion Traversierung(Element, Queue, Anfrage) auf. Ihr Algorithmus soll den Namen und die Parameter k-nächsterNachbar(Baum, Anfrage, k) haben und auch die o. g. Funktion Traversierung mit der o. g. Signatur einbinden (die Sie als Teil Ihrer Lösung ebenfalls hinschreiben sollen). Verwenden Sie zum Einfügen in eine Priority Queue PQ bitte die Funktion Einfügen(PQ, Priorität, einzufügendesObjekt). Um sich das erste Element der Queue ausgeben zu lassen und dieses gleichzeitig zu löschen, verwenden Sie bitte die Notation e ← Head von PQ.

(5 Punkte)

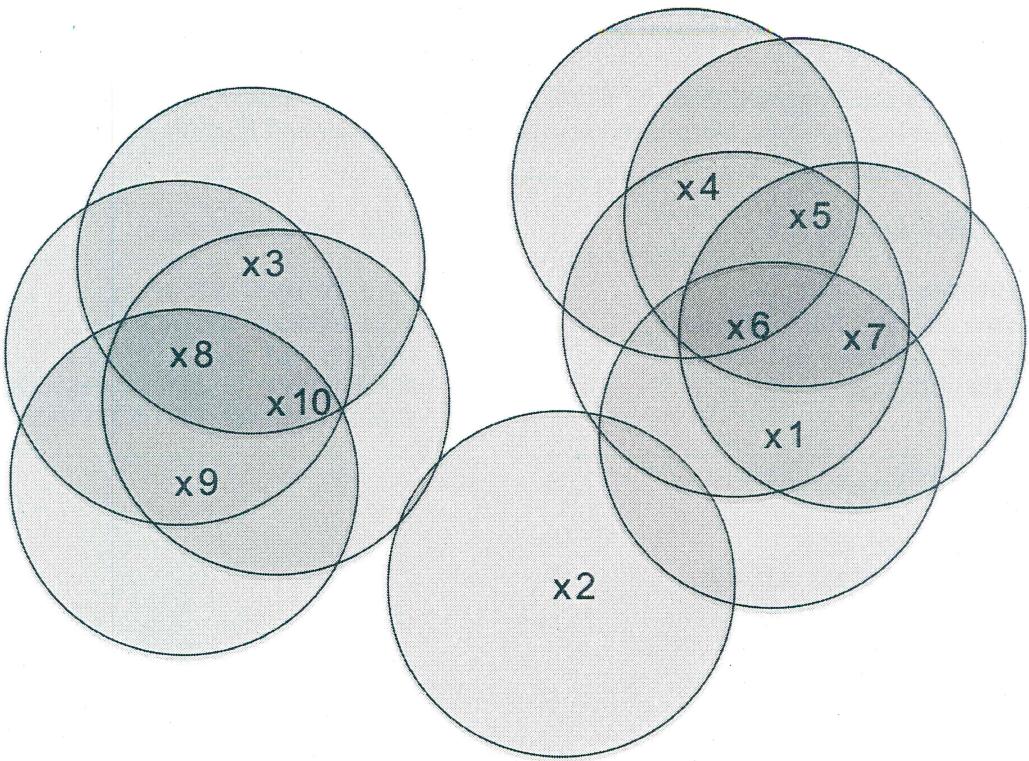


Abbildung 3: Datenbestand mit Objekten 1, ..., 10.

- c) In diesem Aufgabenteil geht es um den Algorithmus DBSCAN; verwenden Sie genau die Variante aus der Vorlesung und wenden Sie ihn auf den kleinen Datenbestand gemäß Abbildung 3 an. Um jedes Datenobjekt, gekennzeichnet durch "x" und benannt mit einer Nummer, ist ein Kreis mit Radius ϵ eingezeichnet. Der Algorithmus betrachtet das Objekt mit der kleinsten Nummer, das noch nicht angefasst wurde, als nächstes.
- c1) Sei $MinPts = 4$. Welche Objekte sind dicht? Welche Objekte sind Dichte-erreichbar? Wie genau sieht das Clustering-Resultat mit DBSCAN aus? Schreiben Sie das Resultat wie folgt hin: $C_1 = \{U, V, \dots\}, C_2 = \{X, Y, \dots\}, \dots$ Dabei sind U, V, X, Y usw. Nummern von Objekten. Die Reihenfolge der Cluster muss die sein, in der DBSCAN sie generiert. Bitte lassen Sie Noise weg, falls DBSCAN eine derartige Ausgabe generieren würde.
 (3 Punkte)
- c2) Sei $MinPts = 2$. Welche Objekte sind dicht? Welche Objekte sind Dichte-erreichbar? Wie genau sieht das Clustering-Resultat mit DBSCAN aus? Schreiben Sie das Resultat wie folgt hin: $C_1 = \{U, V, \dots\}, C_2 = \{X, Y, \dots\}, \dots$ Dabei sind U, V, X, Y usw. Nummern von Objekten. Die Reihenfolge der Cluster muss die sein, in der DBSCAN sie generiert. Bitte lassen Sie Noise weg, falls DBSCAN eine derartige Ausgabe generieren würde.
 (3 Punkte)
- c3) Für den Fall, dass $MinPts = 1$, lässt sich stets eine Formel angeben für die Anzahl der Dichte-erreichbaren Objekte. Schreiben Sie diese Formel hin. Erklären Sie ggf. neue Parameter. Erklären Sie kurz in natürlicher Sprache, warum Ihre Formel richtig ist.
 (1.5 Punkte)
- d) Angenommen, ein Datenbestand bestehend aus N Datenobjekten ist im Einheitsraum Ω der Dimensionalität d gleichverteilt. Wie groß ist die Wahrscheinlichkeit, dass die NN-Distanz einer Position Q in diesem Raum kleiner ist als ein gegebenes r ? Verwenden Sie die folgende Notation (andernfalls Punkt abzugrenzen):

- $sphere^d(center, rad)$ ist die Kugel um die Position $center$ mit Radius rad im d -dimensionalen Raum.
- $Volume(Obj)$ ist das Volumen des Objekts Obj .

Hinweis: Bitte geben Sie nur die Formel an – versuchen Sie nicht, die Formel auszurechnen.

(1.5 Punkte)

Aufgabe 3 – SQL (15 Punkte)

Ein Konsortium von Fluggesellschaften "FlyAirKIT" gestaltet seine Datenhaltung mithilfe eines SQL-basierten Datenbanksystems. Dort seien Passagiere, Flugstrecken und Buchungen durch folgende Relationen modelliert:

```

CREATE TABLE Passagiere
(
    id INTEGER NOT NULL,
    vorname VARCHAR2(20) NOT NULL,
    name VARCHAR2(20) NOT NULL,
    PRIMARY KEY (id)
);

CREATE TABLE Flugstrecken
(
    id INTEGER NOT NULL,
    firma VARCHAR2(20) NOT NULL,
    abflug VARCHAR2(20) NOT NULL,
    flugziel VARCHAR2(20) NOT NULL,
    preis INTEGER NOT NULL,
    dauer INTEGER NOT NULL,
    PRIMARY KEY (id)
);

```

```

CREATE TABLE Buchungen
(
    p_id INTEGER NOT NULL,
    f_id INTEGER NOT NULL,
    PRIMARY KEY (p_id, f_id),
    FOREIGN KEY (p_id) REFERENCES Passagiere(id),
    FOREIGN KEY (f_id) REFERENCES Flugstrecken(id)
);

```

Die Inhalte der Relationen sind wie folgt:

Passagiere		
id	vorname	name
1	'Mike'	'Salzberg'
2	'Melissa'	'Mueller'
3	'Stephan'	'Berufsky'
4	'Hellen'	'Likelace'

Flugstrecken					
id	firma	abflug	flugziel	preis	dauer
1	'AeroInfo'	'Karlsruhe'	'Palo Alto'	500	11
2	'LuftBW'	'Freiburg'	'San Diego'	650	12
3	'AirDurlach'	'Palo Alto'	'Berlin'	550	10
4	'SQLJet'	'San Diego'	'Freiburg'	750	12

Buchungen	
p_id	f_id
1	1
1	4
2	2
2	3
2	4
3	3
4	1
4	4

Erläuterung:

- preis ist der Preis eines Fluges, den eine Fluggesellschaft anbietet, in Euro (ganzzahlig).
- dauer ist die gesamte Dauer eines Fluges in Stunden (ganzzahlig).

Alle Aufgaben lassen sich mit SQL-Konstrukten lösen, die in der Vorlesung vorgestellt wurden. Die Lösungen müssen dem SQL-Standard folgen und unabhängig vom Datenbankinhalt sein.

Hinweis: Diese Teilaufgaben sind voneinander unabhängig.

- a) Der Passagier Hans Schnee führt mehrere berufliche Flüge durch, die er von seinem Arbeitgeber bezahlt haben möchte. Dafür hat er um eine Rechnung für alle seine Buchungen gebeten. Formulieren Sie eine SQL-Anfrage, die den Namen, den Vornamen, die Anzahl der Buchungen dieses Passagiers und die Gesamtkosten seiner Buchungen ausgibt. Vergeben Sie aussagekräftige Attributnamen.

Hinweis: Sie können annehmen, dass die Tabelle 'Passagiere' nur einen Kunden mit dem Vornamen 'Hans' und dem Namen 'Schnee' enthält.

(2,5 Punkte)

- b)** Der Geschäftsführer des Konsortiums möchte eine neue Werbung erstellen, in der die billigsten Flüge angezeigt werden. Erstellen Sie eine Sicht namens 'Lowcost', die alle Flugstrecken (Abflug, Flugziel und Preis) enthalten, die weniger als 50 Euro kosten und länger als 2 Stunden dauern.
(2 Punkte)
- c)** Das Luftfahrtbundesamt hat der Firma LuftBW die Lizenz entzogen. Das Konsortium hat den Passagieren mit Buchungen dieser Firma die Kosten erstattet. Jetzt möchte das Konsortium alle Flugstrecken dieser Firma und die entsprechenden Buchungen aus der Datenbank löschen. Formulieren Sie eine Folge von SQL-Anweisungen, die die Datenbank entsprechend bereinigt.
(2,5 Punkte)
- d)** Der Geschäftsführer des Konsortiums möchte sparen. Er hat deswegen entschieden die Städte mit den wenigsten Kunden zu identifizieren, um Flugstrecken abzulösen, die zu diesen hin oder von diesen weg führen. Formulieren Sie eine SQL-Anfrage, die die Städte und die Gesamtanzahl an Passagieren berechnet, die zu den Städten hin oder von den Städten wegreisen. Die Städte sollen nach der Anzahl an Passagieren aufsteigend sortiert werden.
(4 Punkte)
- e)** Das Konsortium möchte potentiellen Passagieren sogenannte Multihop-Flüge anbieten. Solche Flüge bestehen aus zwei oder drei Flügen, wobei das Ziel eines Fluges dem Abflug des nächsten Fluges entspricht. Hier ist zu beachten, dass diese Flüge keine Rundfahrten enthalten (das heißt, jeder Abflug-Flughafen darf zu keinem späteren Zeitpunkt ein Flugziel im Multihop-Flug sein). Formulieren Sie eine SQL-Anfrage, die alle Multihop-Flüge ausgibt. Das Ergebnis soll die folgenden Informationen enthalten: Abflug und Flugziel des Multihop-Fluges, Anzahl an Zwischenhalten und Gesamtkosten des Multihop-Fluges. Vergeben Sie aussagekräftige Attributnamen.
(4 Punkte)

Aufgabe 4 – Funktionale Abhängigkeiten & Histories (15 Punkte)

Gegeben sei eine Relation $R(A, B, C, D, E, F, G)$ mit der folgenden Menge an funktionalen Abhängigkeiten:

$$F = \{ \begin{array}{l} A \rightarrow B, \\ B C \rightarrow D, \\ D E \rightarrow A, \\ F \rightarrow C E, \\ G \rightarrow F, \\ A C E \rightarrow G \end{array} \}$$

Alle Attribute der Relation sind atomar.

- a)** Geben Sie für folgende funktionalen Abhängigkeiten f an, ob diese aus F ableitbar sind, d. h. ob $F \models f$ gilt. Für jede richtige Antwort gibt es 1/2 Punkt für jede falsche wird 1/2 Punkt abgezogen. Diese Teilaufgabe kann nicht mit weniger als 0 Punkten bewertet werden.
(2,5 Punkte)

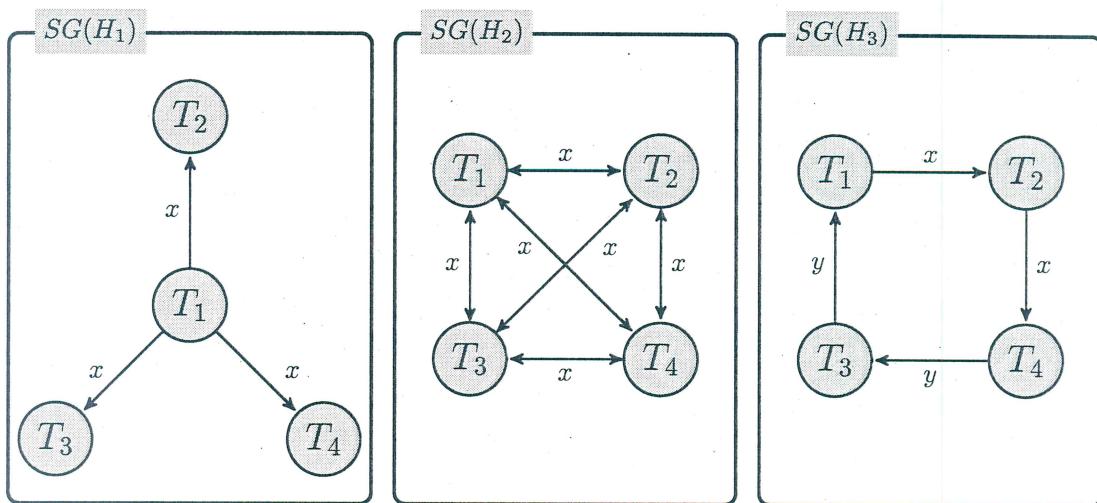
FUNKTIONALE ABHÄNGIGKEIT f	ABLEITBAR	NICHT ABLEITBAR
$A B D E \rightarrow C$	<input type="checkbox"/>	<input type="checkbox"/>
$B G \rightarrow A E$	<input type="checkbox"/>	<input type="checkbox"/>
$C D E \rightarrow A G E$	<input type="checkbox"/>	<input type="checkbox"/>
$F \rightarrow G$	<input type="checkbox"/>	<input type="checkbox"/>
$F G \rightarrow A$	<input type="checkbox"/>	<input type="checkbox"/>

- b)** Bestimmen Sie die Menge aller Schlüssel der Relation R . Für jeden richtigen Schlüssel gibt es 1/2 Punkt für jeden falschen wird 1/2 Punkt abgezogen. Diese Teilaufgabe kann nicht mit weniger als 0 Punkten bewertet werden.

In welcher höchsten Normalform (1NF, 2NF, 3NF, BCNF) befindet sich R ? Begründen Sie kurz Ihre Antwort.

(6,5 Punkte)

c) Gegeben seien die folgenden Serialisierbarkeitsgraphen $SG(H_1)$, $SG(H_2)$ und $SG(H_3)$:



Hinweis: Eine Kante, die von Transaktion T_i zu Transaktion T_j führt und mit x beschriftet ist, besagt, dass es zwei konfigurierende Operationen o_i und o_j gibt, die beide auf x zugreifen und o_i in der History vor o_j steht.

Ermitteln Sie für jeden Serialisierbarkeitsgraphen jeweils eine History, aus der der entsprechende Serialisierbarkeitsgraph erzeugt werden kann. Hierbei soll

- H_1 aus exakt 4 Lese-/Schreiboperationen,
 - H_2 aus exakt 7 Lese-/Schreiboperationen und
 - H_3 aus exakt 6 Lese-/Schreiboperationen

bestehen.

(6 Punkte)

$$H_1 = \begin{array}{|c|c|c|c|} \hline & & & \\ \hline \end{array} \quad c_1 \quad c_2 \quad c_3 \quad c_4$$

$$H_2 = \begin{bmatrix} & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \end{bmatrix} \quad c_1 \quad c_2 \quad c_3 \quad c_4$$

$$H_3 = \begin{array}{c|c|c|c|c|c} & & & & & \\ \hline & & & & & \\ \hline \end{array} \quad c_1 \quad c_2 \quad c_3 \quad c_4$$