

Introduction to Data and Knowledge Engineering

Beispielklausur

Matrikelnummer

Nachname

Vorname

Fachbereich

Semester

Aufbaustudium

Erreichte Punkte								
1	2	3	4	5	6	7	8	Σ
								/100

Viel Erfolg!

1 Datenbanken

10 1. Erklären Sie kurz und präzise die Begriffe

1. Superschlüssel
2. Schlüsselkandidat
3. Primärschlüssel
4. Fremdschlüssel
5. partieller Schlüssel (oder Teilschlüssel).

- 12 2. Eine Stadtverwaltung möchte ihre Bürger in einer Datenbank speichern. Dazu sollen alle Personen mit Vor- und Nachnamen sowie dem Geburtsdatum erfasst werden. Falls 2 Personen miteinander verheiratet sind, soll dies inklusive dem Datum der Eheschliessung erfasst werden. Außerdem wird gespeichert, welche Personen in einer Eltern-Kind-Beziehung zueinander stehen. In der Datenbank wird jede Adresse der Stadt erfasst, auch wenn sie nicht bewohnt ist. Jede Person muss unter mindestens einer Adresse gemeldet sein und an einer Adresse können auch mehrere Personen wohnen. Ausserdem wird die Hundehaltung miterfasst. Jeder Hund hat eine Person als Halter und wird mit seinem Rufnamen und seiner Rasse gespeichert.

Identifizieren Sie die Entities und ihre Attribute und erstellen Sie einen ERM-Entwurf des Systems inklusive Komplexitäten. Ergänzen Sie, nur falls notwendig, Attribute zur eindeutigen Identifikation der Entities.

3. Gegeben seien die Relationen *Klausur*(KID, Vorlesung, Jahr, Maximalpunktzahl, Minimalpunktzahl) und *Pruefungsleistung*(PID, Student, Klausur, Punktzahl). In einer *Pruefungsleistung* wird erfasst, welche Punktzahl ein Student in einer bestimmten *Klausur* erzielt hat. Erzielt ein Student weniger als die Minimalpunktzahl, ist er durchgefallen. Das Attribut *Klausur* in der Relation *Pruefungsleistung* ist ein Fremdschlüssel auf *KID* in *Klausur*.

6

- (a) Formulieren Sie folgende Anfrage im Relationentupelkalkül, in Relationaler Algebra und in SQL: Welche Maximalpunktzahl hätte man in der Prüfungsleistung mit der PID 4711 erzielen können?

Formulieren Sie folgende Anfragen in SQL

- 3 (b) Wie viele Studenten sind im Jahr 2007 in der Klausur zur Vorlesung "DKE" durchgefallen?
- 3 (c) Erstellen Sie eine Liste mit der durchschnittlich erzielten Punktzahl aller Prüfungsleistungen eines Jahres der Vorlesung "DKE". Die Liste soll das Jahr und die Durchschnittspunktzahl dieses Jahres enthalten-
- 5 (d) Erniedrigen Sie bei allen Prüfungsleistungen zur Vorlesung "DKE" im Jahr 2007, die genau die Mindestpunktzahl erreicht haben, die Punktzahl um 1.

2 Knowledge Engineering

5. Datalog

Gegeben seien folgende Relationen:

```
zug(Zugnummer, Zugtyp)
haelt_in(Zugnummer, Ort, Bahnsteig, Ankunftszeit, Abfahrtszeit)
```

Hinweise:

- In Ihren Antworten können Sie alle Relationen durch deren Anfangsbuchstaben abkürzen.
- Die Ankunfts- und Abfahrtszeiten sind als Integer im Format *StundeStundeMinuteMinute* gespeichert, d.h. z.B. 9:07 Uhr ist repräsentiert als Zahl 907.
- Wir ignorieren das Problem, wie Anfangs- und Endbahnhöfe dargestellt werden, d.h. Sie können davon ausgehen, daß zu jedem Bahnhof eine Ankunfts- und Abfahrtszeit existiert.

3

- (a) Definieren Sie in Datalog eine Relation *direktverbindung(A, B)*, die alle Orte *A* und *B* umfaßt, zwischen denen eine direkte Verbindung (d.h. ohne Umsteigen) existiert. Ob der Zug hierbei von *A* nach *B* oder von *B* nach *A* fährt ist egal.

4

- (b) Herr Meier möchte abends nach 18:00 Uhr so schnell wie möglich raus aus Darmstadt. Schreiben Sie eine Relation *fahrplan(Abfahrtszeit, Bahnsteig)*, die einen Fahrplan aller ICE oder IC Züge angibt, die nach 18:00 Uhr von Darmstadt abfahren.

6

- (c) Frau Müller möchte gerne wissen, welche Orte sie von Darmstadt mit genau einmal Umsteigen erreichen kann. Da sie aber gebrechlich ist, interessieren sie nur solche Verbindungen, wo sie beim Umsteigen nicht auf einen anderen Bahnsteig wechseln muß. Definieren Sie in Datalog eine Relation `fahrten(Abfahrtszeit,Umsteigeort,Ankunftsort)`, die alle diese Fahrten umfaßt.

6. Fixpunktsemantik

- 5 (a) Mittels des EPP (Elementary Production Principle) wurden die folgenden Fakten erzeugt. Für jede Iteration sind jeweils nur die *neu hinzukommenden* Fakten angegeben. Nach der 3. Iteration ist der Fixpunkt erreicht.

1. Iteration: $r(a,b)$, $r(a,c)$, $r(b,c)$, $r(c,d)$, $r(d,e)$

2. Iteration: $s(a,c)$, $s(a,d)$, $s(b,d)$, $s(c,e)$

3. Iteration: $s(c,a)$, $s(d,a)$, $s(d,b)$, $s(e,c)$

Geben Sie ein Datalog-Programm an, das genau dieses Verhalten hervorruft.

- 2 (b) Nehmen Sie an, die obigen fünf Fakten für $r/2$ und acht Fakten für $s/2$ seien gegeben. Zusätzlich sei folgende Relation gegeben:

$t(A) :- s(A,B), \neg r(A,C).$

Geben Sie alle Fakten an, die mittels dieser Relation abgeleitet werden können.

Anmerkung: Wie in der Vorlesung bezeichnet \neg die Negation eines Literals.

- 4 (c) Welches grundsätzliche Problem gibt es mit Negation und der Fixpunkt-Semantik? Tritt das Problem im in Aufgabe (b) betrachteten Fall auf?

8. Semantic Web

4

(a) Erklären Sie in ein oder zwei Sätzen die unterschiedlichen Ansätze von *Web Mining* und des *Semantic Web* zur Verbesserung des Zugriffs auf die vielfältigen Informationen des Webs.

4

(b) Nennen Sie zwei Beispiele für Sachverhalte, die in OWL aber nicht in RDF Schema modelliert werden können.