

Datenbanksysteme (f. InformatikerInnen)	schriftliche Einzelprüfung	22.06.2006	1
---	----------------------------	------------	---

Aufgabe 1 [Relationale Abfragen: 25 Punkte]

Gegeben ist folgendes Relationenschema:

fussballer (name, alter, tore, teamname)

team (teamname, trainername)

begegnung (team1, team2, stadion, tore1, tore2)

fussballer.teamname \diamond *team*, *begegnung.team1* \diamond *team*, *begegnung.team2* \diamond *team*

Ein Tupel in der Relation *begegnung* der Form ('A','B','Berlin',0,3) bedeutet, dass A in Berlin gegen B gespielt hat und dabei kein Tor geschossen, aber 3 erhalten hat. Sie können davon ausgehen, dass für jede Begegnung jeweils 2 Tupel eingetragen werden. Im Falle des Beispiels existiert also auch das Tupel ('B','A','Berlin',3,0).

Formulieren Sie die folgende Abfrage in Relationenalgebra (ohne Erweiterungen für Aggregatfunktionen zu verwenden):

- (8 Punkte) Ermitteln Sie die Trainer aller Teams, die mindestens zwei Begegnungen gespielt haben.

Formulieren Sie die folgenden Abfragen in SQL99:

- (5 Punkte) Ermitteln Sie die Namen aller Fußballer, die mindestens einmal in München und in Hamburg gespielt haben, aber niemals im Berliner Stadion.
- (7 Punkte) Ermitteln Sie die Namen aller Fußballer, die gleich alt sind und die gleiche Anzahl von Toren erzielt haben, wobei die beiden Spieler niemals gegeneinander angetreten sind.
- (5 Punkte) Ermitteln Sie Name und Alter aller Fußballer, die alle Tore Ihrer Mannschaft geschossen haben.

Aufgabe 2 [Datalog: 15 Punkte]

Folgendes Relationenschema ist gegeben:

person(name, wohnort),

sendetAn(name1, name2),

sendetAn.name1 \diamond *person*,

sendetAn.name2 \diamond *person*.

Die Relation "sendetAn" beschreibt die Weiterleitung einer Nachricht. Ein Eintrag ist so zu interpretieren, dass die Person mit "name1" die Person mit "name2" benachrichtigt hat.

Finden Sie eine Datalog-Abfrage, die die Namen aller in "Wien" wohnenden Personen findet, die direkt oder indirekt eine Nachricht von "Hermine Musterfrau" erhalten haben.

Aufgabe 3 [Indexstrukturen: 15 Punkte]

Welche Arten von Indizes haben Sie kennengelernt? Wie können Indizes intern realisiert werden?

Erklären Sie insbesondere die Begriffe:

Primär- und Sekundärindex

Dichter Index

Dünner Index

Multilevel Index

Warum muss ein Sekundärindex immer ein dichter Index sein?

Aufgabe 4 [Query Optimierung: 20 Punkte]

Gegeben ist das folgende Relationenschema:

R1(A, B), R2(C, D)

Weiters ist folgende SQL Abfrage gegeben:

$\pi(A, C)(\sigma_{A>5 \wedge C<10}(R1 *_{(A=C)} R2))$

- (2 Punkte)

Stellen Sie den SQL Ausdruck grafisch dar.

- (8 Punkte)

Führen Sie eine heuristische Optimierung mit Hilfe der in der Vorlesung angegebenen Äquivalenzumformungsregeln durch, mit dem Ziel den Abarbeitungsaufwand zu minimieren.

Stellen Sie den umgeformten Ausdruck grafisch dar und geben Sie die Reihenfolge der verwendeten Äquivalenzumformungsregeln an.

Begründen Sie Ihre Vorgangsweise.

c. (10 Punkte)

Nehmen Sie an, dass die Relation R1 1000 und die Relation R2 5000 Datensätze enthält, wobei die Blockgröße für beide Relationen 10 ist.

Für den Join wird das Block-Nested-Loop Verfahren verwendet (Memorygröße 12 Blöcke) und die Selektivität der Selektionen ist jeweils 1/10 (Annahme der Unabhängigkeit).

Geben Sie die Kosten (Anzahl von Blockzugriffen) für

- den ursprünglichen SQL Ausdruck und
- den umgeformten SQL Ausdruck an.

Nehmen Sie an, dass alle Zwischenergebnisse und das Endergebnis materialisiert werden müssen.

Aufgabe 5 [XML: 15 Punkte]

Gegeben ist die folgende DTD:

```
<!DOCTYPE Sammlung [  
  <!ELEMENT Sammlung (CD*)>  
  <!ELEMENT CD (Name, Interpret, Track*)>  
  <!ELEMENT Track (Titel, Dauer, Komponist)>  
  <!ELEMENT Name( #PCDATA )>  
  <!ELEMENT Interpret( #PCDATA )>  
  <!ELEMENT Titel( #PCDATA )>  
  <!ELEMENT Dauer( #PCDATA )>  
  <!ELEMENT Komponist( #PCDATA )>  
>
```

a. (5 Punkte)

Geben Sie ein gültiges XML Dokument an, das auf dieser DTD beruht und in dem alle Elemente der DTD vorkommen.

b. (5 Punkte) Schreiben Sie eine XPath Abfrage, die die Interpret-Elemente von CD's liefert, die höchstens 10 Tracks haben.

c. (5 Punkte) Schreiben Sie eine XQuery Abfrage, die die Namen aller CD's sortiert nach Interpret liefert, deren Gesamtdauer länger als eine Stunde ist.

Aufgabe 6 [Serialisierbarkeit: 10 Punkte]

Erläutern Sie folgende Begriffe:

konflikt-serialisierbar

view-serialisierbar

recoverable schedule

cascadeless schedule