

Informationssysteme

SS 2002



UNIVERSITÄT
DES
SAARLANDES

Prof. Dr.-Ing. G. Weikum
Dipl.-Inform J. Graupmann



Zweite Teilklausur

Name, Vorname	Übungsgruppe x
Matr.-Nr.	

Aufgabe	1	2	3	4	Gesamt
Punkte					

Dienstag, 09.07.2002

Dauer: 90 Minuten

Generelle Hinweise:

Hilfsmittel: alle, außer Rechner, Handies, PDAs, etc.

- Schreiben Sie bitte auf jedes Blatt, das Sie abgeben, deutlich Ihren Namen, Ihre Matrikelnummer und Ihre Übungsgruppe.
- Sie können zur Bearbeitung der Klausur beliebige Unterlagen außer elektronischen Geräten benutzen.
- Die Klausur besteht aus fünf Aufgaben. Sie können insgesamt 12 Punkte erreichen.
- Sie haben zur Bearbeitung aller Aufgaben insgesamt 90 Minuten Zeit.
- Die Lösungen sind im – vollständig ausgefüllten – Deckblatt abzugeben.



Aufgabe 1: Relationale Entwurfstheorie (4 = 1+3 Punkte)

Sie sollen eine Datenbank für einen Internet-Buchhändler entwerfen. Diese soll die folgenden Attribute enthalten:

Titel	Titel eines Buchs
Autor	Autor eines Buchs
Verlag	Verlag, in dem ein Buch erschienen ist
Ort	Ort des Verlags
Jahr	Jahr, in dem ein Buch erschienen ist
Preis	Preis eines Buchs
Kunde	Kunden-Nr. einer Person, die mindestens ein Buch gekauft hat
Datum	Datum, an dem ein Kunde ein Buch gekauft hat
Leser	E-Mail Adresse einer Person, die ein Buch gelesen und dazu eine Bewertung abgegeben hat (diese Person muss das Buch aber nicht notwendigerweise bei dem Internet-Buchhändler gekauft haben)
Bewertung	Bewertung eines Buchs (von 0 bis 5 Sternen) durch einen Leser
Name	Name einer Person

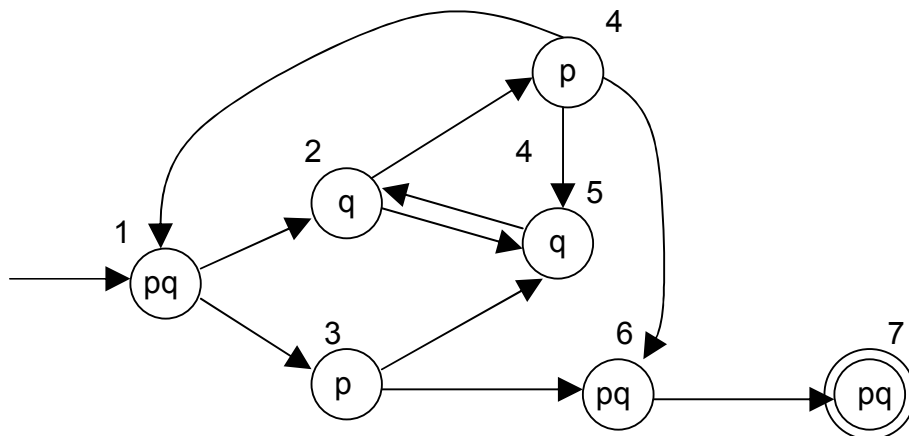
Im Gespräch mit dem Betreiber der Anwendung wurden außerdem die folgenden Funktionalabhängigkeiten erhoben:

- F1 $T, V \rightarrow P, A$
- F2 $T, V, K \rightarrow D, P$
- F3 $V \rightarrow O$
- F4 $T, A \rightarrow O, J$
- F5 $K, N \rightarrow L$
- F6 $L \rightarrow N, K$
- F7 $T, V, L \rightarrow B$
- F8 $K \rightarrow N$

- a) In welcher Normalform - 3NF, BCNF oder nur 1NF - ist die Relation, die alle Attribute enthält? Begründen Sie Ihre Antwort!
- b) Führen Sie eine Relationendekomposition durch, bei der alle Relationen des resultierenden Schemas in BCNF sind. Ist Ihre Zerlegung verlustfrei und abhängigkeitsbewahrend? Begründen Sie Ihre Antwort!

Aufgabe 2: CTL und Modellprüfen (2 Punkte, je 0.5 pro Teilaufgabe)

Gegeben ist das folgenden Transitionssystem. Die Zustände, in denen die elementare Aussage p bzw. q gilt, sind mit p bzw. q markiert.





Ermitteln Sie mittels Modellprüfer alle Zustände, in denen die folgenden Eigenschaften gelten. Benutzen Sie hierfür den Markierungs-Algorithmus aus dem Vorlesungsskript und geben die markierten Zustände für jeden Schritt an.

- a) $AG (p \vee q)$
- b) $EF (p \wedge q)$
- c) $EG (p \wedge q)$
- d) $EF (AG (p \wedge q))$

Aufgabe 3: Serialisierbarkeit (2 = 1+1 Punkte)

Betrachten Sie folgenden Schedule der 5 Transaktionen t_1, \dots, t_5 :

$r_1(e) \ r_2(a) \ r_3(b) \ r_2(e) \ r_3(a) \ r_3(c) \ w_3(c) \ r_1(b) \ w_2(e) \ r_4(a) \ r_5(e) \ r_4(b) \ r_1(d) \ w_4(b) \ r_5(b) \ w_1(d)$

- a) Zeigen Sie, dass der Schedule konflikt-serialisierbar ist.
- b) Bestimmen Sie alle möglichen seriellen Ausführungen, die zu dem Schedule äquivalent sind.

Aufgabe 4: ERM (2 Punkte)

Modellieren Sie die folgenden Informationszusammenhänge als ERM-Diagramm.

Die Universität besteht aus 8 Fakultäten. Jede Fakultät hat einen Dekan und einen Studiendekan, die beide Professoren sein müssen, sowie einen Fakultätsrat, der sich aus 6 Professoren, 2 Vertretern des wissenschaftlichen Personals, 2 Vertretern der Studierenden und 1 Vertreter des Verwaltungspersonals (z.B. Sekretariats- oder Bibliothekspersonal) zusammensetzt.

Jede Fakultät bietet einen oder mehrere Studiengänge an; bei jedem Studiengang hat genau eine Fakultät die Federführung (organisiert z.B. das zuständige Prüfungssekretariat). Studierende sind in mindestens einem Studiengang eingeschrieben. Jeder Student hat eine universitätsweit eindeutige Matrikelnummer. Ein Student ist –falls er in mehreren Studiengängen verschiedener Fakultäten eingeschrieben ist, in nur einer Fakultät wahlberechtigt.

Jeder Mitarbeiter der Universität kann über Namen und Telefonnummer im Personalverzeichnis gefunden werden; bei wissenschaftlichem Personal und Professoren ist dort auch der Titel aufgeführt sowie bei Professoren zusätzlich die Bezeichnung des jeweiligen Lehrstuhls.

Aufgabe 5: Information Retrieval (2 Punkte)

Gegeben sei ein Korpus mit den folgenden 6 kurzen Dokumenten:

- d1:** *Ein guter Schüler ist kein Tor und wird zu einem guten Studenten.*
- d2:** *An der Universität sind nur Studenten.*
- d3:** *Hier sind nur Schüler.*
- d4:** *Durch dieses Tor sind wir Weltmeister geworden.*
- d5:** *Das war nie ein Tor.*
- d6:** *Wir sind Weltmeister.*



Im Korpus kommen also nach Stopwortelimination und Stammformreduktion die folgenden Terme vor

t1: *Schüler*

t2: *Universität*

t3: *Student*

t4: *Tor*

t5: *Weltmeister*

Die $m \times n$ - Term-Dokument-Matrix A ($m=5$, $n=6$) mit den absoluten Termhäufigkeiten pro Dokument sieht wie folgt aus:

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Die SVD zerlegt $A = U \times \Delta \times V^T$ folgendermaßen in

$$U = \begin{pmatrix} -0.44 & -0.30 & 0.57 & 0.58 & 0.25 \\ -0.13 & -0.33 & -0.59 & 0.00 & 0.73 \\ -0.48 & -0.51 & -0.37 & 0.00 & -0.61 \\ -0.70 & 0.35 & 0.15 & -0.58 & 0.16 \\ -0.26 & 0.65 & -0.41 & 0.58 & -0.09 \end{pmatrix} \quad \Delta = \begin{pmatrix} 2.16 & 0 & 0 & 0 & 0 \\ 0 & 1.59 & 0 & 0 & 0 \\ 0 & 0 & 1.28 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 0 & 0.391 \end{pmatrix}$$

$$V^T = \begin{pmatrix} -0.75 & -0.28 & -0.20 & -0.45 & -0.33 & -0.12 \\ -0.29 & -0.53 & -0.19 & 0.63 & 0.22 & 0.41 \\ 0.28 & -0.75 & 0.45 & -0.20 & 0.12 & -0.33 \\ 0.00 & 0.00 & 0.58 & 0.00 & -0.58 & 0.58 \\ -0.53 & 0.29 & 0.63 & 0.19 & 0.41 & -0.22 \end{pmatrix}$$

Betrachten Sie im folgenden nur die zwei größten Singulärwerte, also die approximative SVD mit $k=2$!

Bestimmen Sie die beiden relevantesten Dokumente (also die Top 2 des Resultats-Rankings) für die Anfrage "*Schüler Universität*" mit den Termen t_1 , t_2 , wobei Sie als Ähnlichkeitsmaß das Skalarprodukt verwenden sollen.