

Übungen zu Multimedia-Datenbanken

Aufgabenblatt 3

Prof. Dr. Steffen Staab,
Olaf Görlitz, Christoph Ringelstein
Fachbereich Informatik, Universität Koblenz–Landau

Ausgabe: 09.05.2006

Abgabe: 18.05.2006

Aufgabe 1: Datenbankmanagement mit IBM DB2 (11 Punkte)

Das im 2. Aufgabenblatt modellierte relationale Datenbankschema soll nun auch praktisch in einer Datenbank implementiert werden. Dazu steht für alle Gruppen eine IBM DB2 Instanz auf koenig.uni-koblenz.de zur Verfügung.

Die Zugangsdaten sind für jede Gruppe die selben wie für den FTP-Server allerdings mit der Einschränkung, dass der Loginname nur 8 Zeichen lang sein und nur Buchstaben enthalten darf. Deshalb muss jeder längere Gruppenname ab der 9. Stelle abgeschnitten werden. *Details in der Übung.*

1. Datenbankzugriff (5 Punkte)

Schreibt eine eigene Java-Klasse zum Ausführen von SQL-Statements auf der Datenbank und laded sie auf den FTP-Server hoch. (Der Klassen-/Dateiname sollte für sich sprechen; z.B. 'DBClient.java')

Hinweise zur Programmierung der Java-Klasse werden in der Übung gegeben und danach auch in der Newsgroup gepostet.

Tipp: Alle für Aufgabe 1.2 erforderlichen SQL-Statements können entweder fest in die Klasse einprogrammiert oder als Parameter an das Programm übergeben werden. Letzteres vermeidet ständiges Neu-kompilieren.

2. Tabellen anlegen (6 Punkte)

Legt auf Basis der im Aufgabenblatt 2, Aufgabe 2 angegebenen SQL-Statements, die dem Schema entsprechenden Tabellen an und füllt sie mit Daten. Es sollen mindestens 6 der 12 Austragungsorte der Fußball-WM und Daten von mindestens 6 Spielern (inkl. Mannschaft) eingefügt werden. Die benötigten Informationen sind selbstständig im Internet zu sammeln. Überprüft hinterher die Daten in der Datenbank.

Hinweis: Die SQL-Statements dürfen auch angepasst werden, falls inzwischen eine Verbesserungsmöglichkeit der Modellierung entdeckt wurde.

3. (Bonusaufgabe) Datenbank abfragen (3 Punkte)

Fügt zusätzlich mindestens 6 Vorrundenspiele in die Datenbank ein. Formuliert das SQL-Statement, das zu einem gegebenen Spieler die Namen der Stadien zurück gibt, in denen die Spiele stattfinden, an denen der Spieler teilnimmt. Welches Ergebnis wird von der Datenbank zurückgegeben?

Aufgabe 2: Information Retrieval (19 Punkte)

1. Indexmengen (3 Punkte)

Definiert 6 Indexmengen $d_1..d_6$, bestehend aus mehreren Keywords, die von den sechs in Eurer Datenbank enthaltenen Spielern stammen. (z.B. Diego Maradona: $d_1 = \{\text{Stürmer, Napoli, Weltmeister, UEFA Champion}\}$). Die Keywords sollen so gewählt sein, dass mindestens fünf der Keywords in wenigstens 2 verschiedenen Indexmengen auftauchen.

2. Boole'sches Modell (4 Punkte)

Gegeben sei die Anfrage $(q_1 \text{ or } q_2) \text{ and } (q_3 \text{ or } q_4) \text{ and } q_5$ (in konjunktiver Normalform) für eine Suche über die Keywords der Indexmengen. Folgendes gilt für $q_1..q_5$: $q_1, q_2 =$ ein Fußballverein (z.B. Bayern München, Real Madrid), $q_3, q_4 =$ ein Titelgewinn (z.B. Europameister, UEFA-Cup Champion) und $q_5 =$ Spielerposition (z.B. Stürmer, Torwart).

1. Wählt für $q_1..q_5$ passende Suchterme, die in den Indexmengen enthalten sind, und notiert die entsprechende Suchanfrage.
2. Wandelt die gegebene Anfrage in die Disjunktive Normalform um.
3. Warum ist die disjunktive Normalform (aus Datenbanksicht) besser als die konjunktive Normalform?
4. Wertet die Anfrage auf den Indexmengen aus und gebt das Ergebnis an.

3. Fuzzy Modell (12 Punkte)

1. Erstellt mit Hilfe des Ansatzes von Ogawa, Morita, Kobayashi eine Term-zu-Term Korrelationsmatrix für die Keywords $q_1..q_5$. (5 Punkte)
2. Errechnet für vier von sechs Indexmengen aus $d_1..d_6$ die resultierenden Relevanzwerte und stellt sie in Tabellenform dar. (5 Punkte)
3. Für diese vier Indexmengen soll erneut die Anfrage aus Aufgabe 2 unter Anwendung des Fuzzy-Modells ausgeführt werden. Wie lauten die Ergebnisse, wenn anstelle von *and* und *or* die Funktionen *min* und *max* und ein Threshold von 0,3 verwendet werden? (2 Punkte)

Aufgabe 3: FTP upload

Laded das fertig bearbeitet Übungsblatt auf den FTP-Server hoch.