



Übung zur Vorlesung „Informationssysteme: Verteilte und Web-Datenbanken“

Richard Kuntschke (richard.kuntschke@in.tum.de)

Blatt 6

Abgabe: Dienstag, 30. November 2004 bis 12.00 Uhr im Raum MI 02.11.018 im entsprechend beschrifteten Fach.

Aufgabe 1 - Integration heterogener Datenquellen (10 Punkte)

Bei der Integration von heterogenen, aber ähnlichen Datenquellen ergibt sich das Problem, dass diese sich in ihren Fähigkeiten (z. B. Umfang der gelieferten Daten, Repräsentation der Daten) unterscheiden. Im Zuge der Integration werden daher für externe Datenquellen Views auf einem globalen (integrierten) Schema definiert, um dadurch deren Fähigkeiten bezüglich des globalen Schemas zu beschreiben. Diese Art der Integration kann auf verschiedene Arten durchgeführt werden.

- In der Arbeit „Information Integration Using Logical Views“ von Jeffrey D. Ullman werden zwei verschiedene Ansätze erläutert (Ein Link auf diese Arbeit ist auf der WWW-Seite zu diesen Übungen vorhanden, siehe http://www-db.in.tum.de/teaching/WS_2004-2005/VWDB_WS0405/exercises/). Integrieren Sie mit beiden Verfahren die Online-Gebrauchtwagenmärkte www.autoscout24.de und www.autoinfo.de in ein integriertes Schema.
- Da in der Arbeit von Ullman häufig auf die Logikprogrammiersprache Datalog zurückgegriffen wird, sollten Sie sich zuvor das Kapitel über deduktive Datenbanken in dem Buch „Datenbanksysteme, Eine Einführung“ von A. Kemper und A. Eickler durchlesen.

Aufgabe 2 - Bestimmung horizontaler Partitionierung (10 Punkte)

Ein Unternehmen habe die Bereiche „Spiele“ (S), „Werkzeuge“ (W) und „Verleih“ (V). Die Abteilungen seien jeweils einem Bereich primär zugeordnet: Die Abteilungen 100–250 dem Bereich S, die Abteilungen 251–400 dem Bereich W und die Abteilungen 401–499 dem Bereich V.

Die Analyse der Anwendungen ergebe folgende Zugriffsbereiche bzgl. der Relation ABT (siehe Übungsblatt 4, Aufgabe 2):

- A_1 : Zugriff auf alle Tupel mit Bereich = S
- A_2 : Zugriff auf alle Tupel mit Bereich = W
- A_3 : Zugriff auf alle Tupel mit Bereich = V
- A_4 : Zugriff auf Tupel mit AbtNr $\in [100..150]$
- A_5 : Zugriff auf Tupel mit AbtNr $\in [151..299]$
- A_6 : Zugriff auf Tupel mit AbtNr $\in [300..499]$

Die ABT-Relation sei geeignet (horizontal) zu partitionieren. Ermitteln Sie die Relevanz der Prädikate bzw. der möglichen Partitionierungen mit Hilfe des Algorithmus HORIZ_PART und ggf. durch logische Schlußfolgerung (durch scharfes Hinsehen).

Allgemeine Hinweise

- Die Aufgaben der Übungsblätter sollen gemeinsam von den Mitgliedern eines Teams bearbeitet werden. Die Teameinteilung (je 3 oder 4 Teilnehmer) ist während des Semesters fest!
- Schreiben Sie bitte auf jedes Blatt die Namen aller Team-Mitglieder und die Team-Nummer. Jedes Team-Mitglied sollte die bearbeiteten Aufgaben des Teams „vorrechnen“ können (d. h. der Lösungsansatz des Teams und evtl. dabei auftretende Probleme sollten bekannt und verstanden sein).
- Die bearbeiteten Aufgaben können handschriftlich (bitte leserlich schreiben!) oder in getippter Form abgegeben werden. Bitte heften Sie alle Blätter einer Abgabe zusammen.
- Bitte beachten Sie, dass Abgaben nur dann korrigiert werden können, wenn Sie rechtzeitig zum Abgabetermin abgegeben wurden, auf der Abgabe die Namen aller Team-Mitglieder und die Team-Nummer angegeben sind und die Team-Einteilung eingehalten wurde.