

# Prüfungsprotokoll Praktische Informatik

Prüfer: Prof. Seidl

Datum: WS2006/2007

Prüfungsthemen:

- Einführung in Datenbanken (nach Kemper und Eickler, 6. Auflage)
- Indexstrukturen für Datenbanken (nach Vorlesung von Prof. Seidl, WS05/06)
- Betriebssysteme (nach Tanenbaum, 2. Auflage)

Note: 1,0

Fragen und Antworten sind nach Gedächtnis wiedergegeben, die Reihenfolge stimmt wahrscheinlich nicht mit der aus der Prüfung überein (was aber auch eigentlich egal sein sollte). Es gab mit Sicherheit auch noch andere Fragen, aber die sind mir entfallen.

## 1 Datenbanken

- Was ist eine Relation? (Teilmenge des kartesischen Produkts des Wertebereichs von Attributen)
- Was braucht man alles für so ein Attribut? (Name und Wertebereich)
- Was ist eine Transaktion? (Folge von Operationen, die als atomar behandelt werden und die DB von einem konsistenten Zustand in einen konsistenten Zustand überführen)
- Was ist ACID? (Abkürzung für einige Eigenschaften, die eine Transaktion haben sollte; einzeln die Buchstaben dieser Abkürzung erklärt)
- Klappt das mit der Isolation bei ACID wirklich? (hier wußte ich nicht so genau, wie ich antworten sollte. Habe dann mit einem entschiedenen "sollte eigentlich so gehen" und einem Verweis aufs Two-Phase-Locking geantwortet (2PL kannte ich aus dem Betriebssysteme-Buch, Kapitel über Deadlocks, und dort wird es im Zusammenhang mit Transaktionen bei Datenbanken genannt). Prof. Seidl meinte dann, daß das eher Richtung Implementierung von DB ginge, und damit war die Frage überstanden ;))
- Datenmodelle für Datenbanken? (relationales, OO-Modell, objektrelationales und satzorientiertes, außerdem sind XML-Datenbanken möglich)

- Objekt-relacionales Modell, was ist da mit  $(NF)^2$  gemeint? (non-first normal form, Zulassen von mengen- und relationenwertigen Attributen)
- Die Basisoperatoren der RA? (alle 6 aufgezählt, dabei erwähnt, daß es eigentlich 7 sind (wenn man den überladenen Umbenennungsoperator als zwei Operationen zählt))
- Unterschiede zwischen relationalem und objektorientiertem DB-Modell? (Nachteile des relationalen Modells im Vergleich zum objektorientierten beschrieben: Segmentierung, keine Methoden (dazu Bsp. gegeben mit geometrischen Daten in der DB: wenn man Rotationen realisieren will, dann muß das extern im Anwendungsprog geschehen und wenn man bspw. ein Programm in Java und eines in C will, dann muß die Rotation einmal in C und einmal in Java erstellt werden statt ein einziges Mal in der DB selber))
- Nachteile, wenn man externen Programmcode benutzt? (Unentscheidbarkeit erwähnt, die man da etwa fürs Halteproblem hat. Diese Antwort hatte ich mir eigentlich für Fragen zu den Erweiterungen von SQL in Richtung Turing-Vollständigkeit überlegt, aber sie paßte hier auch ganz gut. Prof. Seidl sagte dann noch, daß dabei außerdem keine Optimierung durch den Anfrageoptimierer möglich ist. Vermutlich wäre das mit dem Anfrageoptimierer also die hier richtige Antwort gewesen)
- SQL: Anlehnung an prozedurale RA und an was noch? (relationaler Tupelkalkül)
- Allgemeine Form einer Anfrage im TK? ( $\{s | s \in S \wedge P(s)\}$ ; irgendwie hier noch auf sichere Ausdrücke zu sprechen gekommen, daß dadurch eben verhindert wird, daß unendliche Ergebnismengen entstehen)
- Obige allgemeine TK-Anfrage: Kann man da noch weitere Variablen hinzufügen? (Ja: Existenz- und Allquantoren für gebundene Variablen)
- Join von  $[A,B]$  und  $[B,C]$  im Domänenkalkül? (hingeschrieben, dabei kurz unsicher geworden, ob ich nicht den Join vergessen hätte, aber dann erwähnt, daß der Join da schon drin ist, nämlich implizit durch Verwendung der gleichen Variable in beiden Tupelkonstruktoren)
- Worauf arbeiten Tupel- und Domänenkalkül (Tupelvariablen werden an Tupel von Relationen gebunden, Domänenvariablen an den Wertebereich von Attributen)

## 2 Indexstrukturen

- Braucht man Indexstrukturen überhaupt? (theoretisch nein, sequentieller Scan wäre auch möglich. Praktisch aber will man natürlich eine möglichst gute Leistung erzielen und braucht also doch Indexstrukturen)
- Bitmap-Indexe: wie funktionieren die? (ein Bit pro Attributwert, Bitvektor ist  $n$  Bits lang für  $n$  Datensätze in der Relation)

- Bitmap-Indexe: für ein gegebenes Schema (MatrNr, Name, Semester,(?, weiß nicht mehr)): feststellen, wo Bitmaps sinnvoll wären
- Bereichskodierung bei Bitmaps? (ein Bit pro Bereich, Bsp aus der VL mit Bereichen aus Monaten eines Jahres gebracht)
- Bitmap-Indexe: Anzahl Operationen auf Index für Punktanfrage, wenn bereichskodierter Index vorliegt? (2, aber erwähnt, daß man auch ohne weiteres noch einen Index für Punktanfragen dazunehmen könnte)
- R-Baum: Eigenschaften und wie sieht so ein R-Baum aus? (einen Knoten mit Unterknoten aufgezeichnet, ein paar Datenrechtecke dazu mit MBR, dann die Existenz von sog. Index entries erwähnt, von denen jeder ein MBR und einen Zeiger auf einen Teilbaum besitzt und dann ein paar Pfeile für die Zuordnung der Datenrechtecke zum Blatt und des MBR zum IE im übergeordneten Knoten; dann noch die #IEs pro Knoten  $\frac{M}{2} \dots M$ , Balanciertheit,...)
- Kann ein R-Baum entarten? (Jain: der Baum selbst ist immer balanciert, aber bei sehr starker Überlappung der Rechtecke durchsucht man trotzdem alle Indexknoten und alle Zeiger auf Datenobjekte in den Blättern; noch erwähnt, daß das im hochdimensionalen fast immer passiert)
- Wie ist die Komplexität beim Suchen im R-Baum? (logarithmisch in der Höhe des Baumes, außer bei sehr starker Überlappung.)

Und im Worst Case? (hier habe ich erwähnt, daß ich das mal in einem der Protokolle gelesen habe und daß das angeblich in  $O(N)$  für  $N$  Datensätze liegt, was ich selber nicht glauben wollte. Daher hatte ich das tatsächlich mal nachgerechnet und kam dann über die Baumhöhe (logarithmisch in  $N$ ) und die Zahl der Knoten im Baum (#IEs pro übergeordnetem Knoten  $\Rightarrow$  #Knoten pro Ebene  $\Rightarrow$  ergibt in der Summe über alle Ebenen eine geometrische Reihe  $\Rightarrow$  per geometrischer Summenformel die Anzahl Knoten im Baum) auf irgendwas mit  $M^{\log_M(N)}$  - artiges mit konstantem Term, also gerade  $O(N)$ . Habe in der Prüfung den ungefähren Weg skizziert, das hat laut Nachbesprechung wesentlich mit zum guten Ergebnis beigetragen, weil ich erkennbar nicht nur nach Protokollen gelernt habe)

- Anzahl Knoten im Binärbaum? (war vermutlich ein Test, ob ich wirklich die Herleitung von Höhe und Knotenanzahl beim R-Baum beherrschte oder nur auswendig gelernt habe. Andererseits kommt der normale Binärbaum ja auch in Indexstrukturen vor, nämlich beim Relationalen Intervallbaum als Primärstruktur)

### 3 Betriebssysteme

- Prozesse: was sind Prozesse?
- Muß das Betriebssystem von den Prozessen wissen? (Ja, es soll sie ja scheitern und ihnen Dienste anbieten)

- Müssen die Prozesse vom Betriebssystem wissen? (Jein: wenn nur ein Prozeß läuft, könnte man auf das BS verzichten. Das geht sogar bei mehreren Prozessen, wenn das BS aus der realen Maschine mehrere virtuelle Maschinen macht und diese den Prozessen anbietet (siehe Tanenbaum). Aber das ist eher selten, normalerweise: BS bietet erweiterte Maschine und verwaltet Ressourcen. Prozesse wollen die HW benutzen und müssen daher Anfragen ans BS stellen können  $\Rightarrow$  Systemaufrufe  $\Rightarrow$  Prozesse müssen vom BS wissen)
- Multimedia-Betriebssysteme: was für besondere Eigenschaften werden hier verlangt? (Echtzeit) Was ist damit gemeint? (Bsp. mit weichen Echtzeitbedingungen für Video und Audio gegeben sowie harten Echtzeitbedingungen etwa beim Auto)
- Dateien bzw. IT-Sicherheit: Zugriffsrechte? (hier habe ich mit den rwx-Bits für Besitzer, Gruppe und Andere bei Dateien geantwortet. Aber Prof. Seidl wollte da wohl eher auf die Zugriffsrechte allgemein für Objekte hinaus)

## 4 Fazit

Die Atmosphäre in der Prüfung war recht entspannt. Prof. Seidl prüft wirklich sehr fair, stellt präzise Fragen und besitzt noch dazu Humor. Der übliche Prüfungsstress kommt daher fast gar nicht vor. Vermutlich sind das einige der Gründe, warum Prüfungen bei ihm so beliebt sind, und auch ich kann Prof. Seidl als Prüfer nur weiterempfehlen.

In dem Zusammenhang sollte ich eines aber nicht verschweigen: es wollen sich so viele bei ihm prüfen lassen, daß freie Prüfungstermine Mangelware sind. Das gilt nicht nur für die erste Terminvereinbarung, sondern auch für das Verschieben eines Termins, falls man merkt, daß es (zu) knapp wird. Um ein paar Wochen verschieben geht dann nicht, rechnet in dem Fall mal eher mit mehreren Monaten! Von daher kann ich nur davon abraten, Prüfungen bei Prof. Seidl zu unterschätzen. Und noch etwas: scheinbar lernen einige nur nach Protokollen. Aber wie aus diesem Protokoll ersichtlich sein sollte merkt Prof. Seidl, ob man selber etwas beherrscht oder nur den Fragenkatalog durchgearbeitet hat. Das läßt sich natürlich auch als Vorteil verstehen: erkennbare Eigenleistung wird belohnt.