

## **Gedächtnisprotokoll**

---

Prüfer: Prof. Seidl

Themen: Einführung in Datenbanken, Implementierung von Datenbanken, Data Mining Algorithmen

Dauer: ca. 60 min

Note: 1.3

---

Zum Protokoll: Ich habe versucht, so viel aufzuschreiben wie möglich, mir sind aber bestimmt ein paar Fragen entfallen. Die Reihenfolge der Fragen kommt mir mittlerweile auch etwa seltsam vor, aber das sollte nicht wirklich wichtig sein. Die einzelnen Themengebiete habe ich nicht weiter durch Überschriften abgegrenzt, da der Übergang in der Prüfung auch fließend war und Prof. Seidl auch ganz zum Schluß erst noch schnell auf Themen aus Implementierung eingegangen ist. Wir haben uns relativ lange an Data Mining Algorithmen aufgehalten, womit wir auch angefangen haben. Die Antworten sind wenn überhaupt nur angerissen um anzudeuten, wohin es gehen sollte.

Euch allen viel Glück bei der Prüfung. Prof. Seidl war super nett, und wenn ich mal nicht weiter wußte, hat er das auch einfach so stehen lassen ohne mich da weiter zu verunsichern. Sehr angenehme Prüfung, umrahmt von einem kurzen Einstieg zu Data Warehouse und einer netten Anekdote, wie weit sich Preclaiming in Datenbanksystemen durchgesetzt hat.

Andi

---

### **Was ist ein Data Warehouse?**

Von den Operationalen Datenbanken getrennt geführte Datenbank...

### **Wie kommen die Daten in ein DW?**

Normalisieren, Fehler beseitigen,....

### **Was sind DataCubes, wie sind sie aufgebaut und wie kann man damit arbeiten?**

Erklären was Data Cubes sind, wo die einzelnen Daten stehen...

### **Wie wird in einem Data Warehouse modelliert?**

Sternschema erklären, hab eins zu Verkäufer, Käufer, Produkt gemacht

### **Wie sieht denn eine Suchanfrage in einem Data Cube aus?**

SQL-Anfrage machen, erklären wo in einem Data Cube welche Daten stehen

### **Welche Themen gab es noch bei DataMining?**

Assoziationsregeln, Klassifikation, Clustering

### **Was ist das Klassifikationsproblem?**

Test&Train...

### **Was unterscheidet Klassifikation von Clustering?**

Gegebene Klassen vs Klassen müssen erst noch gefunden werden

### **Optimalen Bayes-Klassifikator erklären**

mit Formeln, Satz von Bayes, eindimensional und mehrdimensional, mit geschätztem Mittelwertvektor und Kovarianzmatrix.

### **Nächster-Nachbar-Klassifikator erklären**

mit Verbesserungen (kNN → Distanz, Gewichtung, Klassenhäufigkeit, MaxLikelihood)

### **Was ist ein Entscheidungsbaum?**

Aufbau, Splitstrategien (Gini-Index und Entropie/Wissengewinn), Pruning

### **Wie erkennt man Overfitting?**

Baum besser auf Trainingsmenge als auf Testmenge

### **Was ist der Unterschied in der Qualität von NächsterNachbar und Entscheidungsbaum.**

Entscheidungsbaum liefert explizites Wissen, warum Objekt in Klasse zugeordnet

### **Was ist die Aufgabe der Suche von Assoziationsregeln?**

Frequent-Itemsets finden, Assoziationsregeln, Cover, Support, Frequency und Konfidenz erklären

### **Wie sieht eine Assoziationsregel aus?**

Rumpf→Kopf, Support, Confidence

### **Wie werden Frequent-Itemsets gesucht?**

Apriori-Eigenschaft erklären, Algorithmus erklären, einen Schritt bei Apriori-Algorithmus machen.

### **Wie wird Support bestimmt? welche Verbesserungen gibt es ?**

Hashtabelle, Indexstruktur

### **Wie werden Assoziationsregeln aus Frequent-Itemsets gefunden?**

Teilmengenerlegung

### **Wie werden sie auf Confidence überprüft?**

Supportwerte speichern und über Formel ausrechnen

### **Wie viele ‚interessante‘ Regeln kann man aus einem Frequent-Itemset rausholen?**

IPotenzmengen-2, (leere Menge auf einer Seite ist nicht interessant)

### **Wie sieht eine SQL-Anfrage aus? (in bezug auf das eben erstellte Sternschema, wieviel hat ein Verkäufer insgesamt verkauft?)**

hingeschrieben

**Was ist bei einer GROUP BY Anweisung zu beachten?**

Alles nicht-aggregierte aus dem Select auch in die GROUP BY

**Welchen Zusammenhang gibt es zwischen Relationen Algebra und SQL?**

SQL basiert auf RA, allerdings kann SQL mehr

**Welche Grundoperationen gibt es in RA?**

6 aufzählen

**Eine Anfrage in RA machen. Die gleiche in Domänenkalkül.**

**Wie werden JOINS zusammengestellt, welche Methoden gibt es ?**

Nested-Loop. Hashing, Merge-Sort, Indexstruktur

**Was passiert mit einer Anfrage?**

Anfragebearbeitung (Lexikalische Prüfung, Semantische Prüfung, Integritäts- und Zugriffsberechtigungsprüfung, Anfragegraph aufbauen, Anfrage normalisieren, optimieren, Code generieren).

**Wie sieht ein Anfragegraph aus? (Er meinte den Operatorgraphen, ich den Anfragegraphen, haben dann ins Buch geschaut um zu gucken wer was meint) Dann zu einem SQL-Statement den Operatorgraphen aufzeichnen.**

**Wie wird die Anfrage optimiert ?**

algebraisch und nicht-algebraisch, Kostenmodell und Selektivitätsfaktor erklären

(ab hier waren wir schon längst über der Zeit, die letzten Fragen kamen also entsprechend kurz und schnell)

**Was ist eine Transaktion?**

...

**Was sind die ACID-Eigenschaften?**

alle 4 aufzählen und erklären

**Was sind optimistische Synchronisationsverfahren?**

gehen davon aus, daß Blockierungen bei Transaktionen selten auftreten...

**Wie arbeiten sie?**

BOCC und FOCC erklären.

**2PL, striktes 2PL und Preclaiming erklären.**