

Prüfungsprotokoll

Prüfer: Professor Dr. Thomas Seidl

Fächer:

- Einführung in Datenbanken (Jarke), aber abgesprochen Buch Kemper Datenbanksysteme, Kapitel 1–7,9, 13, 14, 15
- Einführung in die Künstliche Intelligenz (Lakemeyer, Norvig/Russel), dazu habe ich ihm 5 Wochen vor der Prüfung eine Themenliste gegeben, was wir in der Vorlesung gemacht haben
- Modelle der Datenexploration (Seidl)

Datum: 20.02.2002

Note: 1.0

Dauer: genau 45 Minuten

S: Guten Tag... Womit wollen sie denn anfangen?

T: Mir egal.

S: Mmh. Da weiß ich jetzt ja gar nicht weiter, das ist ja auch meine erste Diplomprüfung. Ach, fangen wir einfach mit der Datenexploration an.

Fach 1: Datenexploration

S: Was hatte ich denn eigentlich letzte Woche so gefragt? Das habe ich schon alles wieder vergessen. Ach, egal. Welche Ähnlichkeitsmodelle für Formen gibt es denn?

T: Formhistogramme, Kantenzüge, damit geht auch partielle Ähnlichkeit, Rechtecksüberdeckungen, Algebraische Moment Invarianten und in der Vorlesung haben wir auch noch über das Morphologische Modell gesprochen, das wurde da auf Bilder angewandt, ist aber auch für allgemeinere Formen geeignet.

S: Gut. Was ist denn die Grundidee der algebraischen Moment Invarianten?

T: Da werden statistische Kennwerte der Objekte berechnet, wie z.B. der Schwerpunkt und die Hauptachsen. Das ist so ähnlich wie die Varianz.

S: Das ist die Varianz.

T: Nein, die Varianz kann man eigentlich nicht für mehrdimensionale Daten berechnen. Eigentlich berechnet man die Kovarianz und betrachtet nur einen Teil davon.

S: Stimmt eigentlich. Gut, das reicht mir dazu (mit Grinsen). Aber etwas genauere Informationen hätte ich gerne über das morphologische Modell.

T: Erklärt, und direkt auf den Filter hingewiesen. Dabei war ihm sehr wichtig, dass man “symmetrische Flächendifferenz” sagt, und dass man prägnant sagen kann, warum das aufwändig ist, und wie man das mit der granulometrischen Distanz filtern kann.

S: (zur granulometrischen Distanz): Ja, und welche Eigenschaft muss da gelten?

- T:** Lower Bounding Property, Formel dazu aufgeschrieben und erklärt, dass die Filterergebnisse vollständig sein müssen.
- S:** Gut, nun wenn man so Objekte aus einer Datenbank sucht, wie kann man das denn dann machen. Wie ist denn da der sequentielle Scan?
- T:** Nicht so gut. Der Aufwand für die Suche wächst linear mit der Anzahl der Objekte. 100 mal so viele Objekte ergibt 100 mal so lange Suche. Man sollte da besser Indexstrukturen verwenden. In normalen Datenbanken nimmt man z.B. B-Bäume, in Multimedia-Datenbanken R-Bäume oder auf hochdimensionale Daten spezialisierte Strukturen wie X-Trees.
- S:** Gut, welche Eigenschaften gibt es denn da?
- T:** Logarithmische Suche mit Basis Branchingfaktor.
- S:** Ist die garantiert?
- T:** Nein. Nur bei exakten Anfragen und überlappungsfreien Seiten. (Beispiel dazu erklärt mit Bereichsanfrage).
- S:** Gut, Wie macht man denn die Bereichsanfrage.
- T:** (Baum aufgemalt, Daran erklärt. Dazu auch Bild der Bounding Boxen im 2D-Raum)
- S:** Gut. Welche Anfragen gibt es denn noch?
- T:** Angefangen aufzuzählen, Nearest Neighbour, das ist, wenn.....
- S:** Unterbricht mich. Gut, wie macht man eine Nearest Neighbour Anfrage, und kann man dabei einen Algorithmus finden, der optimal ist?
- T:** Ja, der Prioritätsalgorithmus ist optimal in Bezug auf die gelesenen Seiten. Dazu verwendet er eine Active Page List. (und den Algorithmus erklärt)
- S:** Gut, das reicht mir dazu. (Guckt auf die Uhr, es war 10:15:30)... wir haben ja auch schon überzogen.
- S:** Was denn jetzt? Datenbanken?
- T:** Ja, da gibt es ja auch Parallelen.
- S:** Mmh, die gibt es bei KI auch, da habe ich mir auch schon was zu ausgedacht.
- T:** Stimmt, da gibt es auch Ähnlichkeiten.
- S:** An welche hatten Sie denn gedacht?
- T:** Heuristik bei A* muss optimistisch sein und die Kostenfunktionen der Filter müssen konservativ sein.
- S:** Ja, und die APL bei der Prioritätssuche ist so ähnlich wie die Prioritätsliste bei A*.

Datenbanken

- S:** Gut, aber genug davon. Stellen Sie sich vor, sie kommen zu einer Uni, und sollen da eine Vorlesungs- und Studentendatenbank machen. Wie gehen Sie da vor?
- T:** Erstmal Anforderungsanalyse. Dazu überlegen, wen man fragen muss, was man die Leute fragen muss, dann die Befragungen durchführen. Dabei entstehen verschiedene Sichtweisen. Diese müssen in der Konsolidation zu einem globalen, konzeptuellen Datenbankschema vereinigt werden. Dies geht z.B. mit dem ER-Modell. Dann übersetzt man dieses in ein Relationales Schema.
- S:** Gut. Und das ER-Modell ist ja schon sehr alt. (kleine Anekdote über den Erfinder davon). Was macht man denn heute so?
- T:** UML, das kommt aus der Objektorientierten Softwarekonstruktion.
- S:** Und was ist dabei das besondere?
- T:** Dabei wird das Verhalten der Objekte mitberücksichtigt (mehr geraten als gewusst).
- S:** Richtig. Gut, dann machen sie doch mal ein ER-Diagramm für die Vorlesungen/Studenten-Situation.
- T:** Habe mal eins aufgemalt. Mit N:M-Beziehung.
- S:** Gut. Jetzt ist der Uni ihr Service zu teuer, darum gibt die ihnen folgendes Relationale Schema vor: { VorNr, Titel, MatrNr, Name }. Was halten Sie davon?
- T:** Ja, super Schema. Nein, im Ernst. Dieses Schema ist nicht gut. Dadurch, dass da 2 Konzepte in einem Schema enthalten sind, gibt es die Anomalien...
- S:** (dazwischen fragend)... welche denn? Was machen die...
- T:** Update-, Insert- und Löschanomalie (und diese jeweils erklärt mit konkreten Beispielen dazu)
- S:** Gut. Wie kann man das denn jetzt verbessern?
- T:** Mit der relationalen Entwurfstheorie kann man das Schema in Normalformen bringen, die dazu da sind, diese Anomalien zu umgehen. Dabei muss man erstmal die funktionalen Abhängigkeiten betrachten (kurz FDs erklärt) und anhand derer dann die Relationen zerlegen.
- S:** Welche Normalformen gibt es denn, und welche strebt man an?
- T:** 1NF, 2NF, 3NF, BCNF, kurz erklärt, was die ausmacht und Vollständigkeit sowie Abhängigkeitserhaltung erklärt und erläutert wie sich das dabei verhält.
- S:** Gut, wie bringt man denn etwas in die 3NF?
- T:** Mit dem Synthesealgorithmus. Dabei muss man ... (den Synthesealgorithmus erläutert).
- S:** Gut. Dann bringen sie mal dieses Schema da in einen besseren Zustand.
- T:** Das Schema zerlegt in 3 Relationen (ist wohl klar welche, oder?) Und dabei erläutert, dass dies wenn man den Entwurf ordentlich macht eigentlich nicht nötig wäre.
- S:** Gut, was ist eine Transaktion?
- T:** Das Prinzip erklärt, ACID, und was es bedeutet. BOT, COMMIT, ABORT etc. auch.

- S:** Woher weiß das System denn, wann eine Transaktion anfängt?
- T:** Das gibt der Benutzer an, bzw. in SQL passiert das implizit.
- S:** Gut. Die Zeit ist ja auch fast schon um, was gibt es denn für Relationale Anfragesprachen?
- S:** Welche Operationen gibt es denn in der rel. Algebra?
- T:** Die 6 Basisoperationen aufgezählt und dann noch die JOINS genannt.
- T:** Relationale Algebra, Tupelkalkül und Domänenkalkül.
- S:** Wo ist denn der Zusammenhang zwischen SQL und der Rel. Algebra.
- T:** Parallelen: SELECT = Projektion, FROM= Woher (Kreuzprodukt) und WHERE=Selektion. (Dabei mehrmals wirt ausgedrückt, wegen SELECT und Selektion, das fand er aber nicht schlimm.)
- S:** Gut, und wie macht man z.B. die Vereinigung in SQL?
- T:** Zwei SELECTS in Klammern und dazwischen UNION.
- S:** Gut, das wissen nicht viele, dann wollen wir mal zum letzten Wechseln.

Fach 3: Künstliche Intelligenz

- S:** Gut, sie haben sich da mit Suchverfahren beschäftigt. Was gibt es denn da für uninformierte Suchverfahren?
- T:** Aufgezählt und erklärt. (Immer mit Bäumen und dabei war ihm eine präzise Ausdrucksweise sehr wichtig. direkt auch mit den Komplexitäten und Vor-/Nachteilen. Dabei hat er mich im wesentlichen frei referieren lassen, genickt und ab und an Kleinigkeiten nachgehakt, wichtig war ihm ID)
- S:** Gut, wie ist dass bei der Bidirektionalen Suche.
- T:** Die geht nicht immer, und erklärt warum.
- S:** Ja, und wie ist das bei den informierten Suchverfahren.
- T:** (Diese aufgezählt, erklärt und erläutert, wie man zu A* kommt)
- S:** Ja, und was muss für die Heuristik gelten? Ach ne, dass haben sie mir gerade schon gesagt. Mmh, ach sagen sie es nochmal.
- T:** Nochmal erklärt, dass die Heuristik optimistisch sein muss, damit A* optimal und vollständig ist, und erklärt, wie man mehrere Heuristiken zu einer verbinden kann, und wie man feststellt ob eine Heuristik gut ist.
- S:** Gut, dann haben sie sich weiter mit Spielen beschäftigt.
- T:** Spiele sind Suchprobleme wo man den Gegner mitmodellieren muss. Baum aufgemalt, MIN und MAX drangeschreiben, minimax-Algorithmus erklärt, erklärt dass normale Spiele so nicht lösbar sind, weil die Bäume zu groß sind und dass man dann den Baum abschneiden muss, und dass das immer noch nicht reicht und man deswegen α - β -Pruning verwendet und das man dabei noch aufpassen muss, dass man die Suche an einer stabilen Stelle abschneidet wegen des Horizont-Problems.

S: Ja, gut, damit haben sie mir meine weiteren Fragen dazu schon vorweggenommen und die Zeit ist auch um. Schade, jetzt habe ich gar nichts zu Logik gefragt.

T: Ja, dazu kann ich ihnen aber nochwas erzählen.

S: Das glaube ich ihnen, aber was haben sie denn da so gemacht.

T: Überblick über den Bereich “Intelligent Reasoning Agents” gegeben.

S: Gut. Dann gehen sie doch mal bitte kurz raus.

Etwas später durften ich wieder reinkommen. Und er hat gesagt, dass ich alle Fragen sehr gut beantwortet hätte, selbständig in die Tiefe gegangen bin, teilweise ohne dass er danach gefragt habe und dass ich eine 1.0 habe.

Fazit

Euch wünsche ich viel Erfolg, kann dazu sagen, dass die Prüfung super angenehm war. Sehr gut war, dass ich eine Woche vorher bei Prof. Seidl eine mündliche Scheinprüfung über Datenexploration gemacht hatte (Prüfungsprotokoll dazu gibt es auch).

Ich hatte die Prüfung 6 Wochen bevor sie war mit Prof. Seidl abgesprochen und dann auch angefangen zu lernen. Das hat offensichtlich gereicht.