



Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe14

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://www-db.in.tum.de/teaching/ss14/impldb/>

Blatt Nr. 7

Bitte führen Sie alle anzugebenden Datalog-Anfragen tatsächlich aus!
Beispielsweise hier:

<http://dbkemper4-vm10.informatik.tu-muenchen.de:1998/index.html>

Aufgabe 1

Entwerfen Sie einen Algorithmus, um den Klassifikationsbaum, wie er in Abbildung 1 dargestellt ist automatisch zu ermitteln.

Schadenshöhe			
wiealt	Geschlecht	Autotyp	Schäden
45	w	Van	gering
18	w	Coupé	gering
22	w	Van	gering
38	w	Coupé	gering
19	m	Coupé	hoch
24	m	Van	hoch
40	m	Coupé	hoch
40	m	Van	gering
⋮	⋮	⋮	⋮

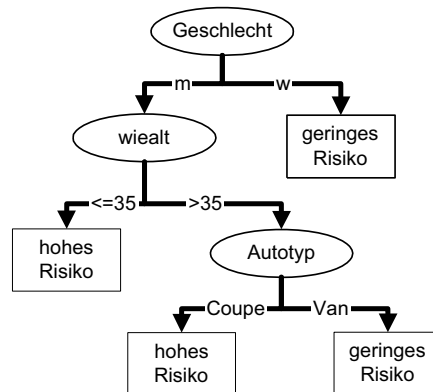


Abbildung 1: Klassifikationsschema für Haftpflicht-Risikoabschätzung.

Aufgabe 2

Gegeben eine Datenbank mit dem Schema aus Abbildung 2. Bestimmen Sie die Wahlbeteiligung bei der Bundestagswahl 2005. Die Wahlbeteiligung soll sowohl für ganz Deutschland, wie auch aufgeschlüsselt nach den Bundesländern, nach Wahlkreisen und sogar nach Wahllokalen abrufbar sein. Geben Sie SQL unter Verwendung des CUBE Operators an.

Aufgabe 3

Die in Abbildung 3 dargestellten Relationen Mietspiegel und Kindergarten dienen der Bewertung von Wohngebieten im Großraum München. Für eine junge Familie ist ausschlaggebend, wie hoch die Lebenshaltungskosten gemessen an zu zahlender Miete und zu entrichtender Gebühr für den Kindergarten im jeweiligen Wohnort ausfallen. Illustrieren Sie die Ausführung einer Top-1-Berechnung (zur Bestimmung des günstigsten Wohnorts) für eine junge Familie mit zwei Kindern. Zeigen Sie die phasenweise Berechnung des Ergebnisses jeweils mit dem Threshold- und dem NRA-Algorithmus.

Wahlbezirke			
Nr	Wahlberechtigte	Wahllokal	Wahlkreis
21967	4500	Rathaus	26
21921	6700	Bürgerhaus	153
28424	3400	Gymnasium	213

Wahlkreise		
Nr	Bezeichnung	Bundesland
26	Neuholm	NRW
153	Oberbach	Bayern
213	Berch	Hessen

Direktkandidaten			
SozialVNr	Name	Partei	Wahlkreis
2005-DK01	Meier	SPD	26
2005-DK02	Müller	CDU	26
2005-DK03	Schmidt	FDP	26
2005-DK04	Huber	null	26

Erststimmen			
Wahlbezirk	Jahr	Kandidat	Stimmen
21967	2005	2005-DK01	450
21967	2005	2005-DK02	750
21967	2005	2005-DK03	600

Bundesländer	
Name	Einwohner
Bayern	12.000.000
NRW	18.000.000
Hessen	6.100.000

Zweitstimmen			
Wahlbezirk	Jahr	Partei	Stimmen
21967	2005	CDU	535
21967	2005	SPD	252
21967	2005	FDP	363
21967	2005	B'90/Grüne	377
21921	2005	CSU	439

Parteien	
Name	Mitglieder
CDU	580.000
CSU	170.000
SPD	600.000
FDP	67.000
B'90/Grüne	44.000

Abbildung 2: Schema des Wahlsystems

Mietspiegel		Kindergarten		WohnLage	
Ort	Miete	Ort	Beitrag	Ort	Lage
Garching	800	Grünwald	-100	Grünwald	München-Süd
Ismaning	900	Unterföhring	0	Unterföhring	München-Nord
Unterföhring	1000	Bogenhausen	100	Ismaning	München-Nord
Nymphenburg	1500	Ismaning	200	Garching	München-Nord
Bogenhausen	1600	Garching	250	Bogenhausen	München-City
Grünwald	1700	Nymphenburg	300	Nymphenburg	München-City

Abbildung 3: Münchner Wohnlagen zur Berechnung der monatlichen Kosten für eine Familie.

Aufgabe 4

Schreiben Sie eine SQL-Anfrage, die basierend auf dem Schema aus Abbildung 4 einen dreidimensionalen Quader berechnet, der es unserem Handelsunternehmen erlaubt, entlang der folgenden Dimensionen drill-down/roll-up Anfragen zu stellen:

- Produkttyp,
- Bezirk,
- Alter.

Das Handelsunternehmen ist dabei nur an Daten aus Deutschland interessiert, die in die Hochsommersaison fallen. Verwenden Sie den **cube**-Operator.

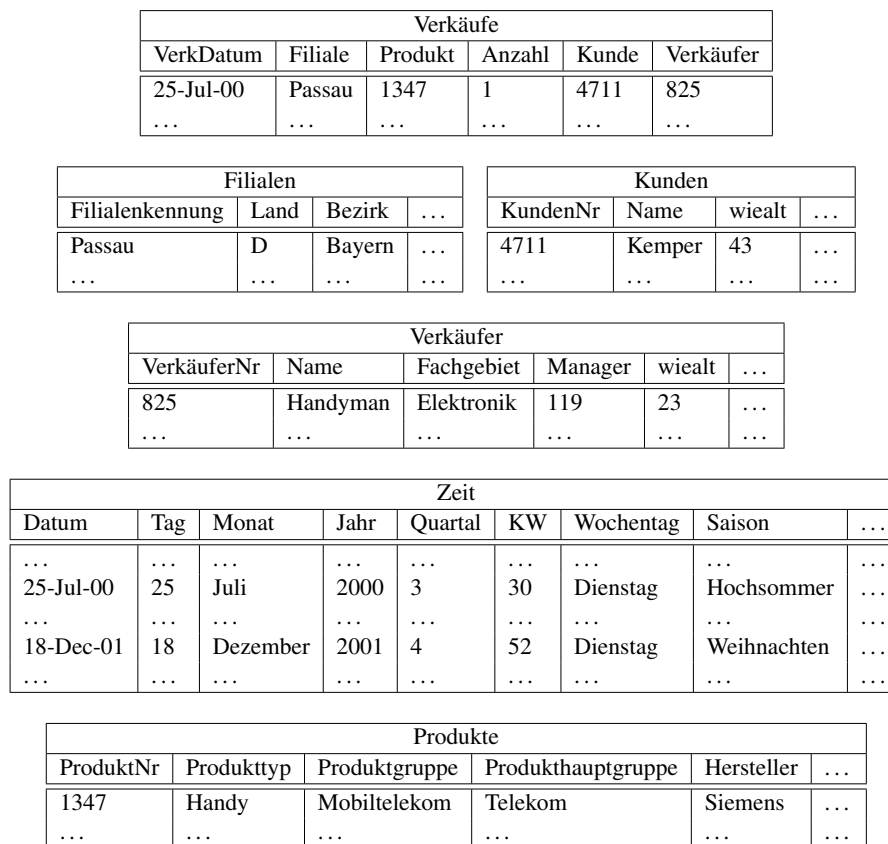


Abbildung 4: Schema des Handelsunternehmens.

Aufgabe 5

Zeigen Sie die weiteren Phasen des à priori-Algorithmus für unser Beispiel in Abbildung 5 (hier ist lediglich bis inkl. 2. Phase dargestellt). Damit eine Menge von Produkten ein frequent itemset ist, muss sie in mindestens $3/5$ aller Verkäufe enthalten sein, d.h. $minsupp = 3/5$.

VerkaufsTransaktionen	
TransID	Produkt
111	Drucker
111	Papier
111	PC
111	Toner
222	PC
222	Scanner
333	Drucker
333	Papier
333	Toner
444	Drucker
444	PC
555	Drucker
555	Papier
555	PC
555	Scanner
555	Toner

Zwischenergebnisse	
FI-Kandidat	Anzahl
{Drucker}	4
{Papier}	3
{PC}	4
{Scanner}	2
{Toner}	3
{Drucker, Papier}	3
{Drucker, PC}	3
{Drucker, Scanner}	3
{Drucker, Toner}	3
{Papier, PC}	2
{Papier, Scanner}	3
{Papier, Toner}	3
{PC, Scanner}	2
{PC, Toner}	2
{Scanner, Toner}	2

Abbildung 5: Ausgangssituation für den à priori-Algorithmus