

Universität Dortmund SoSe 2004  
Übungen zu *Wissensentdeckung in Datenbanken*

Blatt 11. Abgabe bis Montag, den 12.7.2004

**Aufgabe 1** Für die folgenden Fragen geben Sie bitte wieder die jeweilige SQL-Abfrage ab, die sie beantwortet (nicht die Antwort auf die Frage abgeben). Die Fragen beziehen sich auf die Daten vom vorigen Übungsblatt. Versuchen Sie, möglichst wenige Abfragen pro Frage zu verwenden; notwendig ist jeweils nur eine.

1. Wer hat im Schnitt länger telefoniert, Frauen oder Männer?
2. Wieviele verschiedene Nummern wählten die Kunden im Durchschnitt?
3. Wieviele verschiedene Nummern wählten diejenigen Kunden, die weniger als 100 Anrufe tätigten, im Durchschnitt?

Legen Sie außerdem eine Sicht an, die für jede **vorkommende** Kombination von Werten der Attribute `callclass` und `callunits` die Summe der Anruflängen (`calllength`) für diese Wertekombination enthält, sowie die Summe der Anruflängen pro jeweiligem Wert eines jeden Attributes.

**Aufgabe 2** Sie haben in der Vorlesung den Algorithmus Apriori zur Bestimmung von Assoziationsregeln kennengelernt (Folien Teil 13, Nr. 31 bis 37). Im folgenden sollen Assoziationsregeln über das Fernsehverhalten von Testpersonen gefunden werden. In der unten angegebenen Tabelle sehen Sie, welche Zuschauer welche Sendungen verfolgt haben.

Sendung	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$	$z_{11}$	$z_{12}$
Wer will Millionen?	x	x					x		x	x		
Quizzy			x		x	x		x			x	x
Quiz Express		x			x		x			x		
Birkenallee	x							x			x	x
Gut Josepha			x		x	x		x			x	x
Gutes Bier, Schlechtes Bier	x	x	x				x		x	x	x	
Bettina Schläfer				x	x					x		
Nora am Abend		x		x	x		x	x		x		
Olli-G.-Show		x		x	x			x		x		

1. Bestimmen Sie alle Mengen von Fernsehsendungen mit einem support größer gleich 0.4, sowie mit support größer gleich 0.25. Geben Sie dabei für jeden Schritt die Kandidatenmenge sowie die Menge der large item sets (d.h. derjenigen Mengen, die tatsächlich den Mindestsupport erfüllen) an.
2. Geben Sie alle Regeln mit minimalem support 0.25/0.4 und minimaler confidence 0.8 an. Berechnen Sie für die Regeln mit minimalem support 0.4 außerdem jeweils:
  - (a) Konfidenz
  - (b) Sicherheitsmaß
  - (c) Signifikanz

**Aufgabe 3** Zeigen Sie für folgende Aussagen, ob sie gelten oder nicht. Geben Sie dabei immer einen Beweis oder ein Gegenbeispiel an:

1.  $(conf(A \rightarrow B) = \alpha) \wedge (conf(B \rightarrow C) = \beta) \Rightarrow conf(A \rightarrow C) = \alpha\beta$
2.  $conf(A \rightarrow B) = conf(B \rightarrow A) \Rightarrow (h(A) = h(B))$  (wobei  $h(A) > 0$  und  $h(B) > 0$  jeweils die Anzahl der Transaktionen sind, in denen  $A$  bzw.  $B$  vorkommen).
3.  $s(X \rightarrow Y) \geq s(X \rightarrow \emptyset)s(Y \rightarrow \emptyset)$