

Informationssysteme II

Übungsblatt 5

Christine Pries, Mareike Wagner, Felix Oppermann (Gruppe 18)

15. Juni 2005

Aufgabe 1 - Verteilte Datenbanken/Fragmentierung (50 Punkte)

1.1

Fragmentierung: Der Ausdruck Fragmentierung bezeichnet allgemein die Aufteilung eines Größeren in Teile. Im Zusammenhang mit der verteilten Datenhaltung bei Datenbanksystemen bezeichnet die Fragmentierung die Aufteilung einer Relation mit dem Ziel die einzelnen Teile auf unterschiedlichen Systemen abzulegen. Man unterscheidet hierbei zwischen horizontaler und vertikaler Fragmentierung. Bei horizontaler Fragmentierung unterteilt man die Tabellen zeilenweise, alle Attribute bleiben in allen Tabellen erhalten, jedoch enthalten die einzelnen Tabellen nur einen Teil der Tupelmenge. Bei der vertikalen Fragmentierung unterteilt man die Tabellen spaltenweise. Die einzelnen Tabellen enthalten nur einen Teil der Attribute, jedoch aus allen ursprünglich vorhanden Tupeln. Alle Fragmente enthalten den Primärschlüssel. Zusätzlich gibt es noch die gemischte Fragmentierung, welche eine Kombination der horizontalen und vertikalen Fragmentierung darstellt. Außerdem gibt es die abgeleitete Fragmentierung, es handelt sich hierbei um eine Fragmentierung einer Tabelle in Abhängigkeit von der Fragmentierung einer referenzierten Tabelle. So lässt sich also folgende Tabelle darstellen:

ID	Stadt	Land
1	Berlin	D
2	Oldenburg	D
3	Amsterdam	NL
4	Groningen	NL
5	London	GB

Diese Tabelle könnte bei der horizontalen Fragmentierung z.B. nach dem Attribut Land fragmentiert werden. In diesem Fall ergibt sich eine Aufteilung in folgende Tabellen:

ID	Stadt	Land
1	Berlin	D
2	Oldenburg	D

ID	Stadt	Land
3	Amsterdam	NL
4	Groningen	NL

ID	Stadt	Land
5	London	GB

Für die vertikale Fragmentierung ergeben sich folgende Tabellen:

ID	Stadt
1	Berlin
2	Oldenburg
3	Amsterdam
4	Groningen
5	London

ID	Stadt
1	D
2	D
3	NL
4	NL
5	GB

Ein Beispiel für abgeleitete Fragmentierung kann man unter Aufgabe 1.2 (b) finden.

Allokation: Allokation legt fest, auf welchem Rechner ein Fragment gehalten wird. Hierbei kann eine replizierte Speicherung eines Fragments auf mehreren Rechner vorgenommen werden. In unserem Beispiel der horizontalen Fragmentierung könnten beispielsweise die ersten beiden Fragmente auf einem und das letzte Fragment auf einem anderen Rechner abgelegt werden.

(1.2a)

Professoren1				
PersNr	Name	Rang	Raum	Fakultät
2125	Sokrates	C4	226	Philosophie
2126	Russel	C4	232	Philosophie

Professoren2				
PersNr	Name	Rang	Raum	Fakultät
2137	Kant	C4	7	Philosophie

Professoren3				
PersNr	Name	Rang	Raum	Fakultät
2133	Kopernikus	C3	310	Physik

Professoren4				
PersNr	Name	Rang	Raum	Fakultät
2134	Augustinus	C3	309	Theologie

Professoren5				
PersNr	Name	Rang	Raum	Fakultät
2136	Curie	C4	36	Physik

(1.2b)

Vorlesungen1			
VorlNr	Titel	Umfang	gelesenVon
5001	Grundlagen	4	2137
4630	Die 3 Kritiken	4	2137

Vorlesungen2			
VorlNr	Titel	Umfang	gelesenVon
5041	Ethik	4	2125
5049	Mäeutik	2	2125
5043	Erkenntnistheorie	3	2126
5052	Wissenschaftstheorie	3	2126
5216	Bioethik	2	2126

Aufgabe 2 - Vertraut machen mit WEKA (50 Punkte)

(a)

Programm wurde erfolgreich installiert

(b)

@relation uebung05_2b

@attribute Marktstimmung normal, aufgeregt, verhagelt

@attribute Aktienkruf sehr_niedrig, niedrig, normal, hoch, sehr_hoch

@attribute Kreditlinie sehr_niedrig, niedrig, normal, hoch, sehr_hoch

@attribute Bewertung erfolgreich, nicht_erfolgreich

@attribute Kreditwuerdig JA, NEIN

@data

aufgeregt,normal,hoch,erfolgreich,NEIN

aufgeregt,normal,hoch,erfolgreich,NEIN

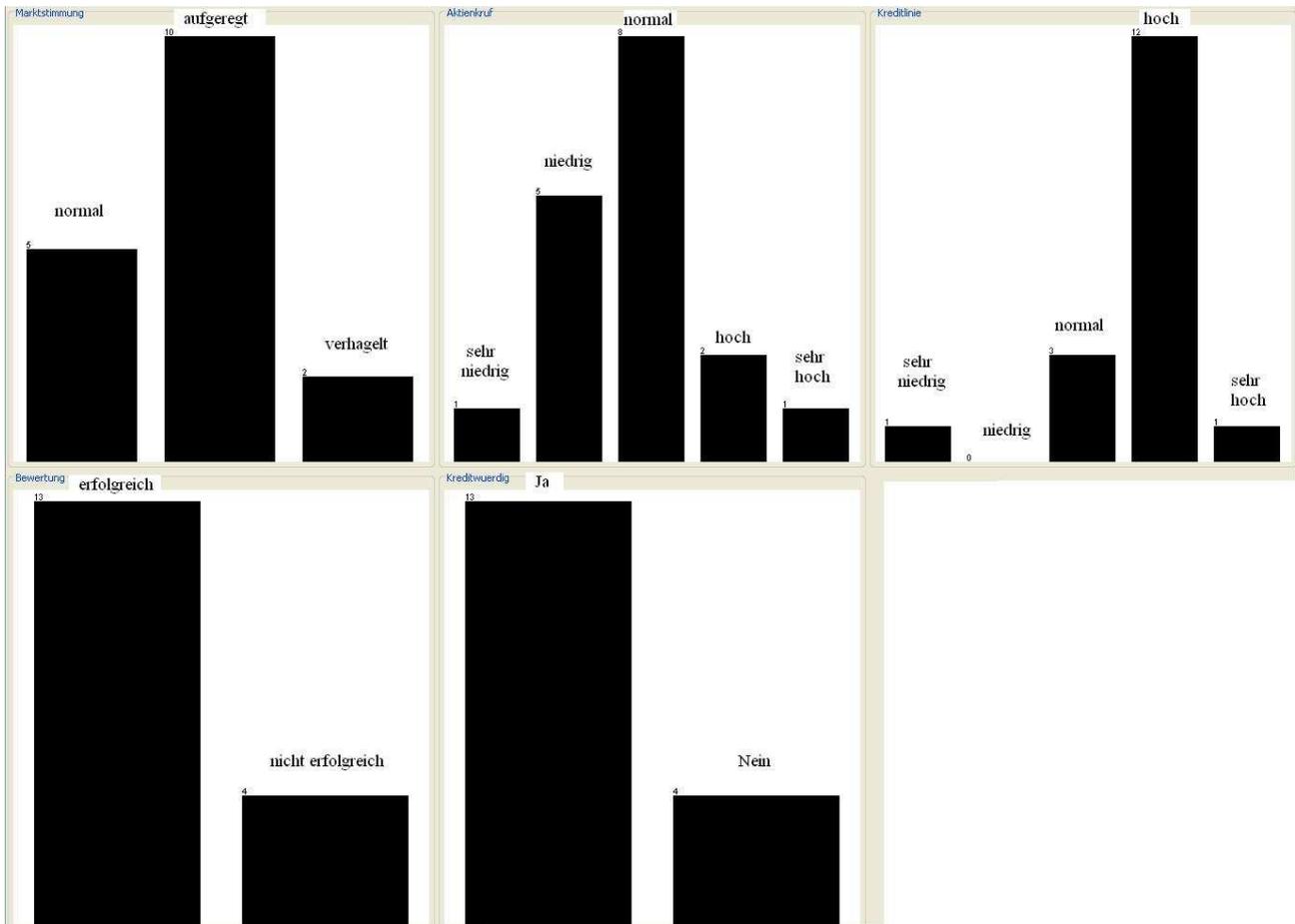
aufgeregt,sehr_niedrig,hoch,erfolgreich,JA

aufgeregt,normal,hoch,erfolgreich,JA

aufgeregt,normal,hoch,nicht_erfolgreich,NEIN

normal,normal,hoch,erfolgreich,NEIN
 verhagelt,niedrig,normal,nicht_erfolgreich,JA
 normal,niedrig,normal,nicht_erfolgreich,JA
 normal,niedrig,normal,nicht_erfolgreich,JA
 aufgeregt,normal,hoch,erfolgreich,JA
 aufgeregt,normal,hoch,erfolgreich,JA
 aufgeregt,normal,hoch,erfolgreich,JA
 aufgeregt,hoch,hoch,erfolgreich,JA
 aufgeregt,niedrig,hoch,erfolgreich,JA
 normal,niedrig,hoch,erfolgreich,JA
 verhagelt,hoch,sehr_hoch,erfolgreich,JA
 normal,sehr_hoch,sehr_niedrig,erfolgreich,JA

(c)



(d)

```
=== Run information ===

Scheme:      weka.associations.Apriori -N 2 -T 0 -C 0.0 -D 0.01 -U 1.0 -M 0.0 -S -1.0
Relation:    uebung05_2b
Instances:   17
Attributes:  5
              Marktstimmung
              Aktienkruf
              Kreditlinie
              Bewertung
              Kreditwuerdig
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.67
Minimum metric <confidence>: 0
Number of cycles performed: 33

Generated sets of large itemsets:

Size of set of large itemsets L(1): 3

Size of set of large itemsets L(2): 1

Best rules found:

1. Kreditlinie=hoch 12 ==> Bewertung=erfolgreich11   conf:(0.92)
2. Bewertung=erfolgreich 13 ==> Kreditlinie=hoch 11   conf:(0.85)
```

Dies ist der Screenshot, der Ergebnisausgabe.

Hieraus ergeben sich die zwei Regeln mit dem höchsten Support:

- Kreditlinie=hoch 12 → Bewertung=erfolgreich 11 conf:(0.92)
- Bewertung=erfolgreich 13 → Kreditlinie=hoch 11 conf:(0.85)

Das heißt, dass wenn die Kreditlinie hoch ist, ist auch die Bewertung erfolgreich (mit einer confidence von 0.92) und umgekehrt genauso (mit einer confidence von 0.85).