

Universität Dortmund SoSe 2004  
 Übungen zu *Wissensentdeckung in Datenbanken*

Blatt 12. Abgabe bis Montag, den 19.7.2004

**Aufgabe 1** Bestimmen Sie die 1-freien Mengen, die 2-freien Mengen und die closed item sets in der folgenden Transaktionstabelle.

| tid | A | B | C | D |
|-----|---|---|---|---|
| 1   | 0 | 1 | 1 | 0 |
| 2   | 1 | 1 | 0 | 0 |
| 3   | 1 | 0 | 0 | 1 |
| 4   | 0 | 0 | 1 | 1 |
| 5   | 0 | 1 | 1 | 0 |
| 6   | 0 | 1 | 1 | 1 |

| tid | A | B | C | D |
|-----|---|---|---|---|
| 7   | 1 | 1 | 0 | 1 |
| 8   | 0 | 0 | 1 | 0 |
| 9   | 0 | 1 | 0 | 0 |
| 10  | 1 | 1 | 0 | 0 |
| 11  | 0 | 1 | 0 | 1 |
| 12  | 0 | 0 | 1 | 1 |

**Aufgabe 2** An einem Tag im April wird zu jeder vollen Stunde das Wetter beobachtet; die Ergebnisse finden sich in der folgenden Tabelle.

| Zeit | Wind      | Niederschlag |
|------|-----------|--------------|
| 0h   | windstill | kein         |
| 1h   | windstill | kein         |
| 2h   | windstill | kein         |
| 3h   | windstill | kein         |
| 4h   | windstill | kein         |
| 5h   | Brise     | Regen        |
| 6h   | windstill | kein         |
| 7h   | windstill | kein         |
| 8h   | windstill | Schnee       |
| 9h   | windstill | Schnee       |
| 10h  | windstill | kein         |
| 11h  | frisch    | kein         |

| Zeit | Wind      | Niederschlag |
|------|-----------|--------------|
| 12h  | frisch    | kein         |
| 13h  | frisch    | kein         |
| 14h  | frisch    | Regen        |
| 15h  | frisch    | Regen        |
| 16h  | windstill | kein         |
| 17h  | windstill | kein         |
| 18h  | frisch    | kein         |
| 19h  | frisch    | kein         |
| 20h  | frisch    | Regen        |
| 21h  | windstill | kein         |
| 22h  | windstill | kein         |
| 23h  | windstill | Schnee       |

- Überführen Sie die Tabelle in die Repräsentation zeitlicher Daten, die in Mannilas WINEPI-Ansatz verwendet wird (Folien Teil 13 ab Nr. 86), d.h. geben Sie die Beobachtungssequenz  $s$  an. Als Ereignis gilt dabei das **Eintreten** eines neuen Wind- **oder** Niederschlagsverhältnisses, also können null bis zwei Ereignisse zum sel-

ben Zeitpunkt stattfinden. Formalisieren Sie die Episoden “Auf keinen Niederschlag folgt frischer Wind, danach Regen” und “Schnee fällt bei Windstille”. Geben Sie die diesem Ansatz entsprechende Fensterung von  $s$  an bei einer Fensterbreite von 6 Stunden, mit einer Schrittweite von 2 Stunden. Bestimmen Sie die Häufigkeit der beiden Episoden anhand dieser Fensterung. Bestimmen Sie zusätzlich die Konfidenz der Regel “Wenn es windstill ist, fällt Schnee”.

- Überführen Sie die Tabelle in die Repräsentation zeitlicher Daten, die in Höppners Ansatz verwendet wird (Folien Teil 13 ab Nr. 112), d.h. geben Sie die Zustandsintervalle an. Schieben Sie dann ein 6 Stunden breites Fenster mit Schrittweite 4 Stunden über die Zustandssequenz und bestimmen Sie für jedes Fenster die Intervallrelationen, die zwischen den Windverhältnissen einerseits und den Niederschlagsverhältnissen andererseits bestehen. Die möglichen Intervallrelationen sind die folgenden [Auswahl aus Allens Ansatz]:

| Relation     | Gilt wenn   |
|--------------|---|
| A before B   | A endet, bevor B beginnt  |
| A meets B    | A endet genau dort, wo B beginnt                                    |
| A overlaps B | A beginnt eher als B und endet, nachdem B beginnt und bevor B endet |
| A starts B   | A und B beginnen gleichzeitig, aber A endet eher                    |
| A finishes B | A und B enden gleichzeitig, aber B beginnt eher                     |
| A contains B | A beginnt eher als B und endet, nachdem B endet                     |
| A equals B   | A und B beginnen und enden gleichzeitig                             |

Bestimmen Sie den Support der Muster “Regen fällt bei Wind (frischer Wind oder Brise)” und “Schnee fällt bei Windstille”.

- Aufgabe 3** In dieser Aufgabe sollen Sie ein Datenbankschema zur zentralen Wettererfassung und Auswertung entwerfen. In einer endlichen und festen Anzahl von Städten befinden sich Wetterstationen, die jeden Tag die minimale, maximale und durchschnittliche Temperatur festhalten, sowie den Niederschlag an diesem Tag und die durchschnittliche Windstärke. Entwerfen Sie ein Datenbankschema, mit dem sich folgende Fragen beantworten lassen:

1. Wie waren durchschnittlich Temperatur, Wind oder Niederschlag in einer Stadt an einem gegebenen Tag, in einer gegebenen Woche, einem gegebenen Monat oder einem gegebenen Jahr?
2. Dasselbe für einen Monat über den gesamten Zeitraum (d.h. z.B. die Durchschnittstemperatur über alle Januare).
3. Wie waren die durchschnittlichen Wetterwerte gemittelt über alle Wetterstationen in einem bestimmten Monat?
4. War der Niederschlag an Arbeitstagen größer als an Tagen, die frei sind? Wie sieht es mit der Durchschnittstemperatur aus?

Hinweise:

- Sie können alle Zeitgranularitäten als Integer modellieren (wie auf den Folien angegeben).
- Sie müssen die Tabellen, die Sie anlegen, nicht mit Daten füllen (d.h. sowohl Wetter- als auch Datuminformationen müssen nicht eingetragen werden). Wenn Sie Ihren Ansatz aber testen wollen, können Sie einfach zufällige Werte eintragen.
- Geben Sie in Ihrer Lösung die SQL-Anweisungen an, mit denen Sie die Tabellen erstellt und die Anfragen gestartet haben.