

## 8. Übungsblatt

### Aufgabe 28 Geschlossene und maximale Itemmengen

- Nehmen Sie an, wir versuchten, all geschlossenen (oder alle maximalen) (häufigen) Itemmengen dadurch zu finden, daß wir erst alle häufigen Itemmengen finden und dann das Ergebnis gemäß der definierenden Bedingungen für geschlossene (oder maximale) Itemmengen filtern. Ist dies eine gute Idee? Begründen Sie Ihre Antwort!
- Was bedeuten die Begriffe *Kopf* (“*head*”) und *Schwanz* (“*tail*”) eines Suchbaumknotens (oder eines Teilproblems)? Besteht eine Beziehung zu dem Begriff des *Präfix* eines Teilproblems?
- Gegeben eine globale Itemordnung, gibt es eine (feste) Beziehung zwischen den Items im Kopf (head) und denen im Schwanz (tail)?
- Gilt die Beziehung  $head \cup tail = B$  für alle Suchbaumknoten? Falls nein, gibt es bestimmte Suchbaumknoten oder bestimmte Bedingungen, die erfüllt sein müssen, damit dies gilt?

### Aufgabe 29 Ausgeschlossene Items

- Was ist die Menge der *ausgeschlossenen Items* (“*excluded items*”) (relativ zu einem Suchbaumknoten oder einem Teilproblem)? Wie kann man diese Menge mit Hilfe der Zerlegungen in Teilprobleme charakterisieren, die wir im Teile-und-Herrsche-Algorithmus ausführen? Was könnte man analog die Menge der *eingeschlossenen Items* nennen?
- Nehmen Sie an, das Teile-und-Herrsche-Schema zum Finden häufiger Itemmengen werde mit dem Stutzen mit Hilfe perfekter Erweiterungen verbessert, so daß Teilprobleme durch Tripel  $S = (T, P, X)$  beschrieben werden. Ist  $P \cup X$  immer eine geschlossene Itemmenge? Ist es immer eine maximale Itemmenge? Begründen Sie Ihre Antwort!
- Ist die folgende Aussage wahr? Wenn der Schwanz (tail) eines Suchbaumknotens nicht leer ist, dann ist sein Kopf eine maximale Itemmenge. Begründen Sie Ihre Antwort!
- Ist die folgende Aussage wahr? Wenn der Schwanz (tail) eines Suchbaumknotens kein Item enthält, das den gleichen Support hat wie der Kopf (head), dann ist der Kopf eine geschlossene Itemmenge. Begründen Sie Ihre Antwort!

### Aufgabe 30 Geschlossene und maximale Itemmengen

- Wie kann man die definierende Bedingung geschlossener Itemmengen mit Hilfe einer *horizontalen* Transaktionsdarstellung prüfen?
- Wie kann man die definierende Bedingung geschlossener Itemmengen mit Hilfe einer *vertikalen* Transaktionsdarstellung prüfen?

- c) Worin besteht eine Alternative zum Prüfen der definierenden Bedingung? Wie wird die Prüfung in diesem alternativen Ansatz ausgeführt? Welche Möglichkeiten gibt es (z.B. an Datenstrukturen), um diese Alternative zu implementieren?

**Aufgabe 31** Geschlossene und maximale Itemmengen: Stutzen

- a) In den vorangehenden Aufgaben haben wir betrachtet, wie man herausfinden kann, ob eine Itemmenge geschlossen (oder maximal) ist und damit, wie man die Ausgabe nicht geschlossener (oder nicht maximaler) Itemmengen vermeiden kann. Ist dieses einfache Filtern der einzige Vorteil? Oder kann man auch die Suche selbst mit Hilfe dieser Methoden vereinfachen/stutzen? Wenn ja, wie?
- b) Wie kann ein Stutzen mit perfekten Erweiterungen bei der Suche nach geschlossenen Itemmengen ausgenutzt werden? Warum ist es sicher, daß wir mit diesem Stutzen keine geschlossenen Itemmengen verlieren?
- c) Was versteht man unter Kopf-vereinigt-Schwanz-Stutzen (“head union tail pruning”) für die Suche nach maximalen Itemmengen (zwei Varianten)?

**Zusatzaufgabe** Alternativen zum Support

- a) Finden Sie ein vom Support verschiedenes Maß (oder mehrere Maße) auf der teilgeordneten Menge  $(2^B, \subseteq)$  das/die anti-monoton (oder nach unten abgeschlossen) sind!
- b) Wir definieren den *Träger*  $L_T(I)$  einer Itemmenge  $I$  bzgl. einer Transaktionsdatenbank  $T$  als

$$\begin{aligned} L_T(I) &= \{k \in \{1, \dots, n\} \mid I \cap t_k \neq \emptyset\} \\ &= \{k \in \{1, \dots, n\} \mid \exists i \in I: i \in t_k\} \\ &= \bigcup_{i \in I} K_T(\{i\}). \end{aligned}$$

Die *Ausdehnung*  $r_T(I)$  einer Itemmenge  $I$  bzgl. einer Transaktionsdatenbank  $T$  ist die Größe ihres Trägers, d.h.,  $r_T(I) = |L_T(I)|$ . Ist die Ausdehnung anti-monoton (oder nach unten abgeschlossen)? Wenn nein, welches Verhalten zeigt sie?

- c) Könnte man Support und Ausdehnung kombinieren, um ein besseres Maß für die Assoziation von Items zu erhalten? (Hinweis: Schauen Sie sich das Maß *Jaccard-Index* an und überlegen Sie, wie es verallgemeinert werden könnte.)