

2. Übungsblatt

Exercise 5 Suchen nach häufigen Mustern

- Wie groß ist der Suchraum der Suche nach häufigen Itemmengen?
Wie wächst die Größe des Suchraums mit der Anzahl der Items?
- Warum ist eine erschöpfende Suche über alle Itemmengen nicht ratsam?
Was ist die (asymptotische) Zeitkomplexität einer erschöpfenden Suche?
Ist die gestutze/reduzierte Suche grundsätzlich effizienter?
Wenn ja, in welcher Weise? Wenn nein, warum könnte sie trotzdem hilfreich sein?

Exercise 6 Suchen nach häufigen Mustern

- Wie viele Teilmengen mit $k - 1$ Items hat eine Itemmenge mit k Items?
Warum stellt dies ein Problem für die Kandidatengenerierung dar?
- Zeigen Sie, warum jede häufige Itemmenge mit $k + 1$ Items auf $j = (k + 1)k/2$ Weisen mit der Originalversion des Apriori-Algorithmus erzeugt werden kann!
- Auf wie viele Arten kann eine *nicht* häufige Itemmenge mit $k + 1$ Items erzeugt werden? Unterscheidet sich die Anzahl von der für häufige Itemmengen? Warum?

Exercise 7 Suchen nach häufigen Mustern

- Welche Struktur des Suchraums wird für die Suche ausgenutzt?
Warum ist diese Struktur vorteilhaft, gegeben die Eigenschaften des Support?
- Was ist ein Hasse-Diagramm? Was stellt es dar?
Wofür verwenden wir Hasse-Diagramme?
- Wie wird das Hasse-Diagramm verändert/transformiert, um die Suche zu verbessern? Was ist der Zweck dieser Transformation?
- Was erreichen wir durch die Zuordnung eindeutiger Elter-Itemmengen?
Könnten wir auch eindeutige Kind-Itemmengen zuordnen?
- Gibt es Itemmengen, die immer eine eindeutige Elter-Itemmenge haben,
auch im unveränderten Hasse-Diagramm?

Aufgabe 8 Suchen nach häufigen Mustern

- Wie können wir eindeutige Elter-Itemmengen definieren (konkrete Methode)?
Gibt es Alternativen? Wie helfen eindeutige Eltern bei der Suche?
- Wenn eindeutige Elter-Itemmengen zugeordnet sind: Auf wie viele Weisen können wir eine Itemmenge der Größe $k + 1$ aus Itemmengen der Größe k erzeugen?
- Wie sieht das allgemeine Schema der Suche mit eindeutigen Eltern aus?

Aufgabe 9 Kanonische Form von Itemmengen

- a) Was ist eine kanonische Form einer Itemmenge? Wie kann man sie definieren? Braucht man dieses Konzept tatsächlich für die Suche nach häufigen Itemmengen? Warum betrachten wir kanonische Formen für Itemmengen trotzdem?
- b) Was ist die Präfixeigenschaft? Wie wird sie ausgedrückt (zwei Varianten)? Welche Vorteile folgen aus der Präfixeigenschaft?
- c) Beschreiben Sie, wie die rekursive Suche nach häufigen Mustern vereinfacht werden kann, wenn die gewählte kanonische Form die Präfixeigenschaft hat!
- d) Die Präfixeigenschaft ist eine notwendige Bedingung um sicherzustellen, daß die Suche durch Anhängen von Items an kanonische Kodewörter erschöpfend ist. Was braucht man zusätzlich, um eine hinreichende Bedingung zu erhalten?
- e) Was ist eine kanonische Erweiterungsregel? Gibt es eine kanonische Erweiterungsregel für die Suche nach häufigen Itemmengen?