

7. Übungsblatt

Aufgabe 25 Carpenter-Algorithmus

- a) Wie nutzt der Carpenter-Algorithmus die alternative Charakterisierung geschlossener Item-Mengen aus (d.h., diejenige, die auf einem Abschlußoperator basiert)? Was ist das Kernproblem dieser Charakterisierung? (Hinweis: minimaler Support)
- b) Welche Datenstruktur benutzt der Carpenter-Algorithmus?
Könnte er auch eine horizontale Transaktionsdarstellung nutzen? *a d e*
Woran könnte es liegen, das er so meist *nicht* implementiert wird? *b c d*
- c) Führen Sie den Carpenter-Algorithmus für die rechts gezeigte Transaktionsdatenbank aus! Verwenden Sie eine einfache Menge zum Speichern der Item-Mengen, nicht, wie in der Vorlesung, einen Präfixbaum! *a c e*
b c d
a e
- d) Wie kann man die Carpenter-Suche mit dem minimalen Support stützen?

Aufgabe 26 Carpenter-Algorithmus

- a) Kann Stützen mit perfekten Erweiterungen im Carpenter-Algorithmus verwendet werden? Wenn ja, wie? Wie ist dies analog zum Stützen mit perfekten Erweiterungen in Item-Mengen-Aufzählungsalgorithmen (wie Eclat, FP-Growth etc.)?
- b) Kann man den Carpenter-Algorithmus benutzen, um die maximalen (häufigen) Item-Mengen zu finden?
- c) Könnte man den Carpenter-Algorithmus verwenden, um *alle* häufigen Item-Mengen zu finden?
- d) Warum ist der Carpenter-Algorithmus für die Warenkorbanalyse nicht empfehlenswert? Für welche Art von Daten ist er empfehlenswert? Warum?
(Hinweis: Zeitkomplexitäten in den Parametern der Datenbank.)

Aufgabe 27 IsTa-Algorithmus

- a) Welche rekursive Beziehung nutzt der IsTa-Algorithmus aus, um geschlossene Item-Mengen zu finden? Wo liegt die Komplexität?
- b) Nehmen wir an, der IsTa-Algorithmus hat Arbeit beendet und der wir haben daher den fertigen Präfixbaum der gefundenen geschlossenen häufigen Item-Mengen vorliegen. Wie werden diese geschlossenen häufigen Item-Mengen aus diesem Präfixbaum ausgegeben? Welche Bedingungen müssen beim Durchlaufen des Präfixbaumes geprüft werden?
- c) Wie könnte man aus diesem Präfixbaum heraus die *maximalen* Itemmengen ausgeben? Welche Bedingungen müssen in diesem Fall geprüft werden?
- d) Warum ist der IsTa-Algorithmus für die Warenkorbanalyse nicht empfehlenswert? Für welche Art von Daten ist er empfehlenswert? Warum?
(Hinweis: Zeitkomplexitäten in den Parametern der Datenbank.)