

5. Übungsblatt

Aufgabe 18 SaM-Algorithmus

- a) Wie wird im SaM-Algorithmus der Support bestimmt? Was muß der SaM-Algorithmus für den Transaktionenarray sicherstellen, damit dies gelingt?
- b) Was wird mit dem Merge-Schritt des SaM-Algorithmus erreicht? Warum kann der Merge-Schritt der SaM-Algorithmus ein Problem darstellen? (Hinweis: Was entspricht dem Merge-Schritt im Eclat-Algorithmus?)
- c) Bestimmen Sie die Anfangsdarstellung der Transaktionsdatenbank des SaM-Algorithmus für die rechts gezeigten Transaktionen!
- | | |
|----------------|----------------|
| <i>a d e</i> | <i>a c d</i> |
| <i>b c d</i> | <i>b c</i> |
| <i>a c e</i> | <i>a c d e</i> |
| <i>a c d e</i> | <i>c b e</i> |
| <i>a e</i> | <i>a d e</i> |

Aufgabe 19 RElim-Algorithmus

- a) Wie verbessert der RElim-Algorithmus den SaM-Algorithmus? Welches Sortierprinzip wird in SaM, welches in RElim ausgenutzt?
- b) Wie können im RElim-Algorithmus äquivalente Transaktionen (oder Transaktionssuffixe) zusammengefaßt werden? Wann sollte dies geschehen?
- c) Bestimmen Sie die Anfangsdarstellung der Transaktionsdatenbank des RElim-Algorithmus für die in Aufgabe 19c gezeigten Transaktionen!

Aufgabe 20 RElim-Algorithmus

- a) Was ist eine k -Items-Maschine? Warum kann sie nur für wenige Items benutzt werden (kleines k)?
- b) Wie ist eine k -Items-Maschine mit dem RElim-Algorithmus verwandt? Worin bestehen Unterschiede und was sind die Vor- und Nachteile?
- c) Obwohl die k -Items-Maschine nur für wenige Items benutzt werden kann, ist sie für das Finden häufiger Itemmengen höchst nützlich. Warum? Wie werden Items behandelt, die über die Grenze k einer k -Items-Maschine hinausgehen?

Aufgabe 21 FP-growth-Algorithmus

- a) Was ist ein FP-Baum? Wie wird er aus einer Transaktionsdatenbank erzeugt? Wie vereinigt er eine horizontale und eine vertikale Darstellung?
- b) Wie wird mit einem FP-Baum der Support bestimmt/gezählt?
- c) Wie wird ein FP-Baum projiziert? Was liefert die Projektion?
- d) Wie erhält man im FP-growth-Algorithmus die bedingte Transaktionsdatenbank des zweiten Teilproblems?

- e) Führen Sie den FP-growth-Algorithmus für die rechts gezeigte Transaktionsdatenbank und einen minimalen Support von $s_{\min} = 3$ aus!

Sie können die Schritte der Häufigkeitsbestimmung der Items und des Umsortierens der Items in den Transaktionen nach Häufigkeit weglassen (d.h.: verwenden Sie die alphabetische Ordnung).

| | |
|----------------|----------------|
| <i>a d e</i> | <i>a c d</i> |
| <i>b c d</i> | <i>b c</i> |
| <i>a c e</i> | <i>a c d e</i> |
| <i>a c d e</i> | <i>b c e</i> |
| <i>a e</i> | <i>a d e</i> |