

Exercise Sheet 5

Exercise 19 SaM Algorithm

- a) How is support counted in the SaM algorithm?
What does SaM have to ensure for the transaction array for this to work?
- b) What does the merge step of the SaM algorithm achieve?
Why can the merge step of the SaM algorithm be a problem?
(Hint: What corresponds to the merge step in the Eclat algorithm?)
- c) Construct the initial database representation for the SaM algorithm for the transaction database shown on the right!
- | | |
|----------------|----------------|
| <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> |

Exercise 20 RElim Algorithm

- a) How does the RElim algorithm improve over the SaM algorithm?
Which sorting principle is exploited in SaM, which in RElim?
- b) How can the RElim algorithm achieve merging of equivalent transactions (or transaction suffixes)? When should this be done?
- c) Construct the initial database representation for the RElim algorithm for the transactions shown in Exercise 19c!

Exercise 21 k -items Machine

- a) What is a k -items machine? Why can it be used only for few items (small k)?
- b) How is a k -items machine related to the RElim algorithm?
What are the differences and the advantages/disadvantages?
- c) Even though the k -items machine can be used only for few items, it is highly useful for frequent item set mining. Why? How are items beyond the limit k of the k -items machine handled?

Exercise 22 FP-growth Algorithm

- a) What is an FP-tree? How is it constructed from a transaction database?
Why does it combine a horizontal and a vertical representation?
- b) How is support counted with an FP-tree?
- c) How is an FP-tree projected? What does the projection yield?
- d) How do we obtain the conditional transaction database for the second subproblem in the FP-growth algorithm?

- e) Execute the FP-growth algorithm for the transaction database shown on the right and a minimum support $s_{\min} = 3!$

You may skip the steps of finding the item frequencies and resorting the items in the transactions w.r.t. the frequency order (that is: use the alphabetical order).

<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>