



**POLITECNICO
DI TORINO**



Data Science Lab

Exercises

DataBase and Data Mining Group

Andrea Pasini, Elena Baralis

- The following transactional dataset is given

Tid	Items
1	A,B,C,E,G
2	B,C,D
3	A,B,C,G,H
4	B,D,F,G
5	A,C,G
6	C,D,G
7	A,B,D,G
8	C,D,E,G

- Q1: Considering minsup=3, list the itemsets generated by the step analyzing prefix B in FP-growth
- Q2: Considering minsup=3, which of the following itemsets are generated by the step analyzing prefix B in FP-growth?
- Example of alternative question: ... by step 3 in Apriori

Q1: Considering minsup=3, list the itemsets generated by the step analyzing prefix B in FP-growth

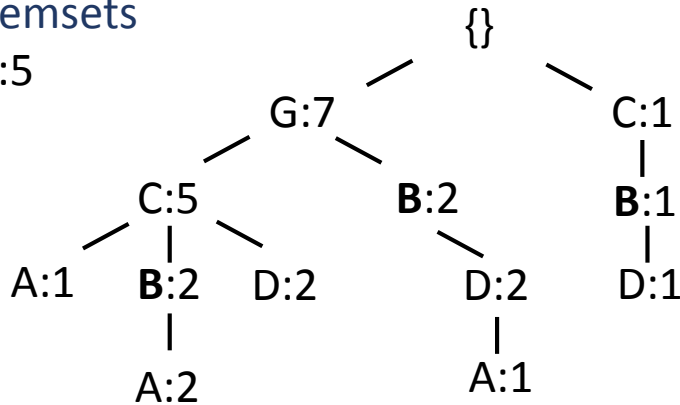
Solution

Frequent items (header table):

G:7,C:6,B:5,D:5,A:4

Itemsets

B:5



B-CPB

CG:2

G:2

C:1

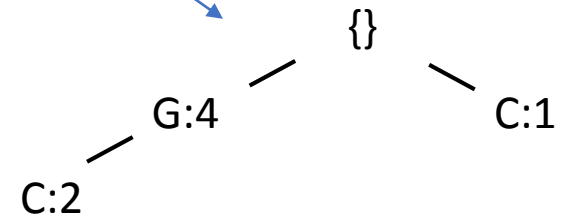
Itemsets

BG:4,

BC:3

H. Table

G:4,C:3



End: all itemsets below support

Itemsets extracted with prefix B

B:5

BC:3

BG:4

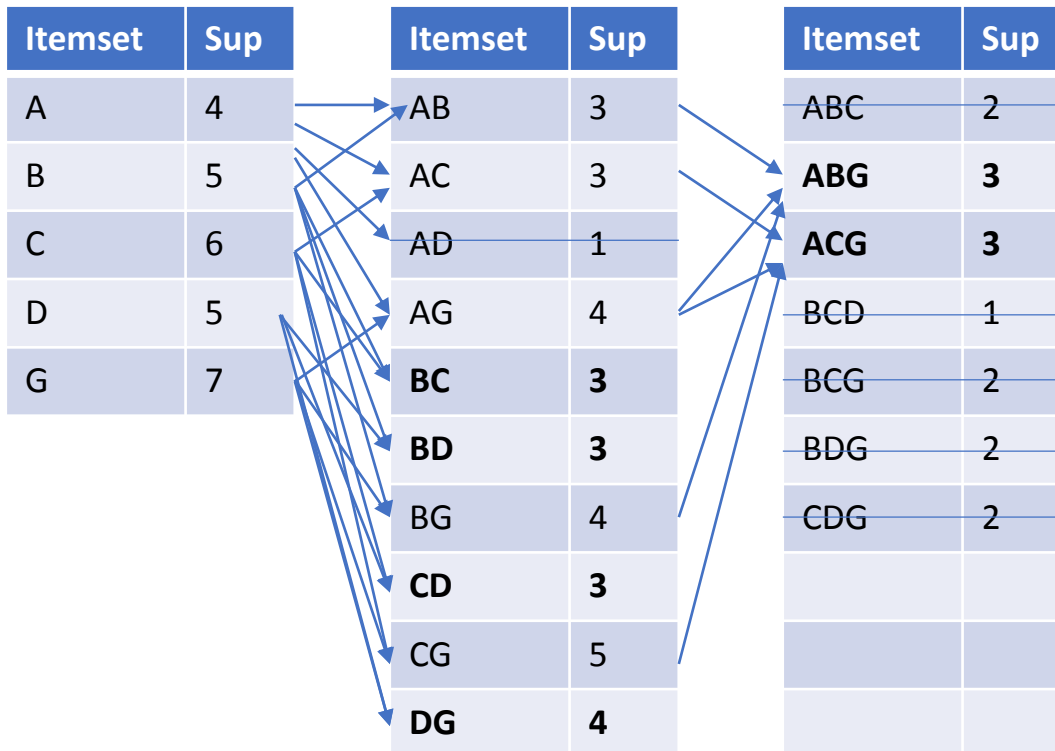
- Q2: Considering minsup=3, which of the following itemsets are generated by the step analyzing prefix B in FP-growth?
 - a) B, BD, BG
 - b) B, BE, BG
 - c) B, BC, BG
 - d) B, BC, BCG
 - e) No itemset is generated with prefix B
- Correct answer: c)

- Q3: Considering minsup=3, which of the following itemsets are maximal itemsets?

- a) ABC
- b) BG
- c) BD
- d) ACD
- e) No answer is correct

Tid	Items
1	A,B,C,E,G
2	B,C,D
3	A,B,C,G,H
4	B,D,F,G
5	A,C,G
6	C,D,G
7	A,B,D,G
8	C,D,E,G

- Solution: use apriori



Q3: Maximal itemsets (= do not have a superset which is frequent):
 ABG (3), ACG (3), BC (3), **BD** (3), CD (3), DG (4) → Correct answer: c)

- Q4: Considering $\text{minsup}=3$, which of the following itemsets are closed itemsets?
 - a) ABC
 - b) A
 - c) AB
 - d) B
 - e) All answers are correct

- Solution: use apriori

Itemset	Sup	Itemset	Sup	Itemset	Sup
A	4	AB	3	ABC	2
B	5	AC	3	ABG	3
C	6	AD	1	ACG	3
D	5	AG	4	BCD	1
G	7	BC	3	BCG	2
		BD	3	BDG	2
		BG	4	CDG	2
		CD	3		
		CG	5		
		DG	4		

Legend:

→ found a superset with same support

Q4: Closed itemsets (no superset with same support):

ABG (3), ACG (3), BC (3), BD (3), CD (3), DG (4), → Correct answer: d)
 AG (4), BG (4), CG (5), B (5), C (6), D (5), G (7)