

Практические задания №3. Григорьев И.С. 6304

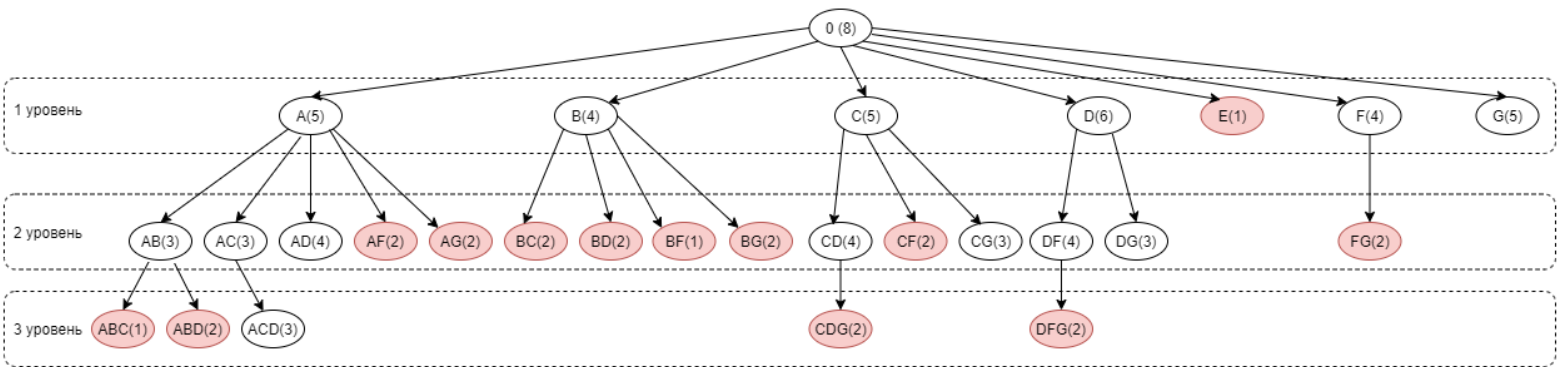
Задание №1

Дан набор данных:

| tid | itemset |
|-------|--------------|
| t_1 | <i>ABCD</i> |
| t_2 | <i>ACDF</i> |
| t_3 | <i>ACDEG</i> |
| t_4 | <i>ABDF</i> |
| t_5 | <i>BCG</i> |
| t_6 | <i>DFG</i> |
| t_7 | <i>ABG</i> |
| t_8 | <i>CDFG</i> |

1. Предположив, что минимальный уровень поддержки равен 3 / 8.

Продemonстрируйте, как алгоритм Apriori перебирает данный набор данных.

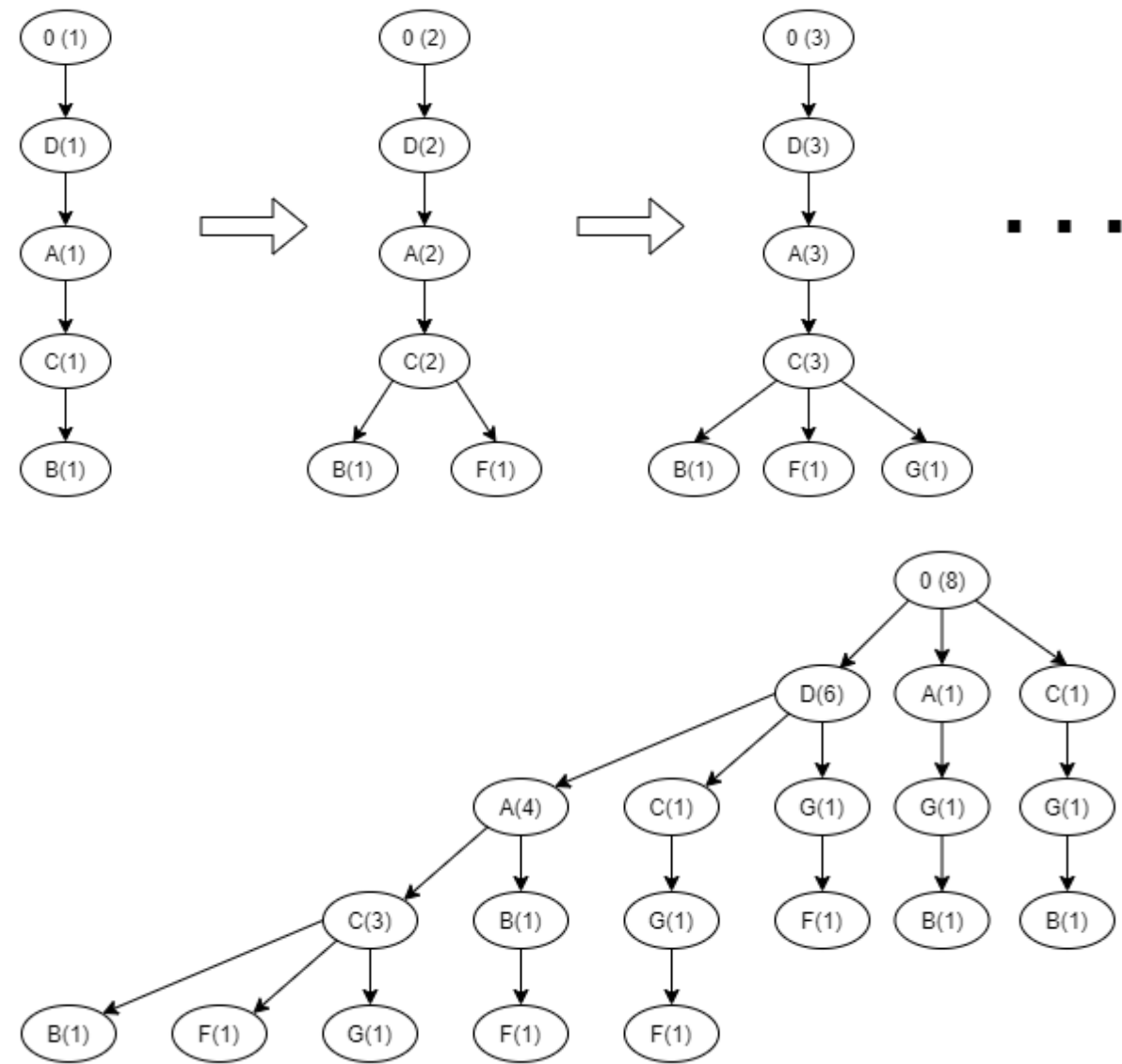


2. Предположив, что минимальный уровень поддержки равен 2 / 8.

Продemonстрируйте, как алгоритм FPGrowth перебирает данный набор данных.

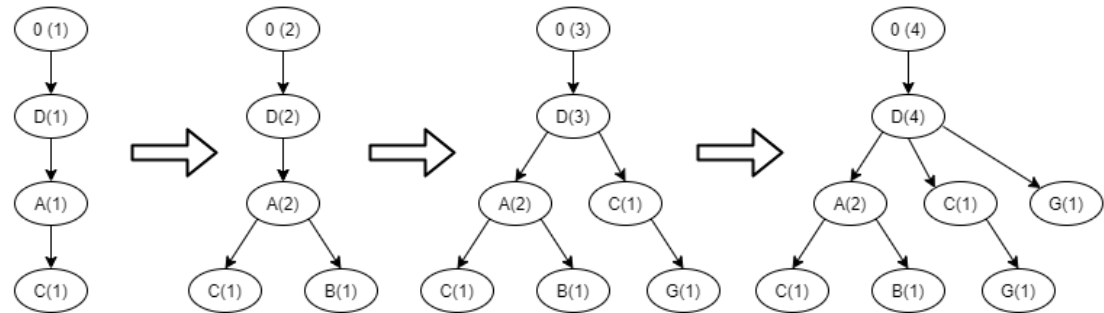
D(6) A(5) C(5) G(5) B(4) F(4)

FP-дерево (подается на вход алгоритму):



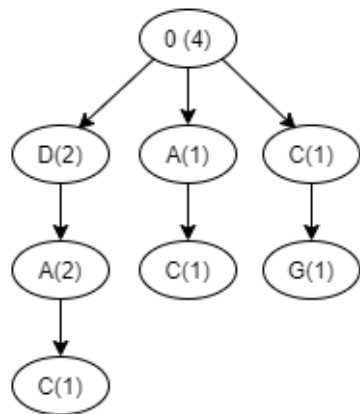
Спроецированные FP-деревья для каждого объекта:

F:



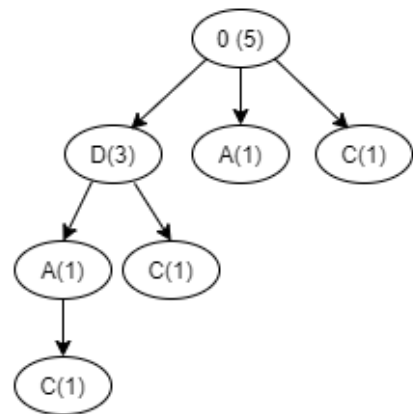
FD(4) FA(2) FC(2) FG(2)
FDA(2) FDC(2) FDG(2)

B:



BD(2) BA(3) BC(2) BG(2)
BDA(2)

G:



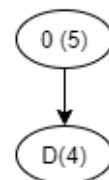
GD(3) GA(2) GC(3)
GDC(2)

C:



CD(4) CA(3) CDA(3)

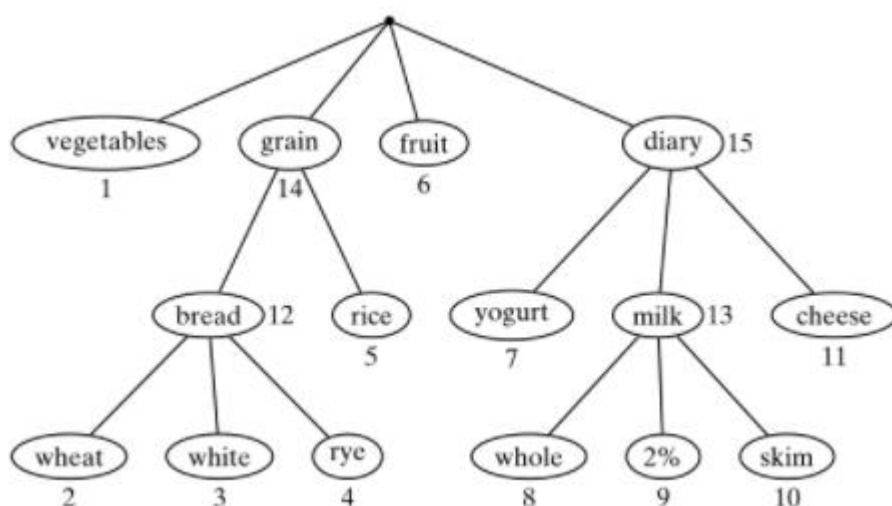
A:



AD(4)

Задание №2

На рисунке представлена классификация различных продуктов. Каждый лист дерева – это конкретный продукт, внутренний узел дерева представляет категорию продукта более верхнего уровня.



Был получен следующий набор данных:

| tid | itemset |
|-----|-------------|
| 1 | 2 3 6 7 |
| 2 | 1 3 4 8 11 |
| 3 | 3 9 11 |
| 4 | 1 5 6 7 |
| 5 | 1 3 8 10 11 |
| 6 | 3 5 7 9 11 |
| 7 | 4 6 8 10 11 |
| 8 | 1 3 5 8 11 |

1. Каков размер области поиска наборов элементов, если ограничиваться только наборами, состоящими из простых элементов?

$$2^{11} - 1 = 2047$$

2. Предположив, что минимальный уровень поддержки = 7/8. Найдите все часто встречающиеся наборы элементов, состоящие только из элементов высокого уровня в таксономии. Имейте в виду, что, если в транзакции появляется простой элемент, предполагается, что все его предки высокого уровня также присутствуют в транзакции.

| | |
|----------|-----|
| (12) | 7/8 |
| (14) | 8/8 |
| (15) | 8/8 |
| (12, 14) | 7/8 |
| (12, 15) | 7/8 |
| (14, 15) | 8/8 |