

1. Найдите все минимальные генераторы для минимального уровня поддержки = 1.

Support = 1 — уровень поддержки

Набор данных:

tid	itemset
1	ACD
2	BCD
3	AC
4	ABD
5	ABCD
6	BCD

F — frequent set

	support	itemsets
0	4.0	(A)
1	4.0	(B)
2	5.0	(C)
3	5.0	(D)
4	2.0	(B, A)
5	3.0	(C, A)
6	3.0	(D, A)
7	3.0	(C, B)
8	4.0	(B, D)
9	4.0	(C, D)
10	1.0	(C, B, A)
11	2.0	(B, A, D)
12	2.0	(C, D, A)
13	3.0	(C, B, D)
14	1.0	(C, B, A, D)

Сгруппируем F по поддержке:

5 D, C
 4 A, B, BD, CD
 3 CA, DA, CB, CBD
 2 BAD, CDA, BA
 1 CBA, CBAD

Получим closed frequent set **C** и minimal generators **G**

tidset	C	G
12456	D	D
12356	C	C
1345	A	A
2456	BD	B
1256	CD	CD
135	CA	CA
145	DA	DA
256	CBD	CB
45	BAD	BA
15	CDA	CDA
5	CBAD	CBA

Пример получения значений:

Для C:

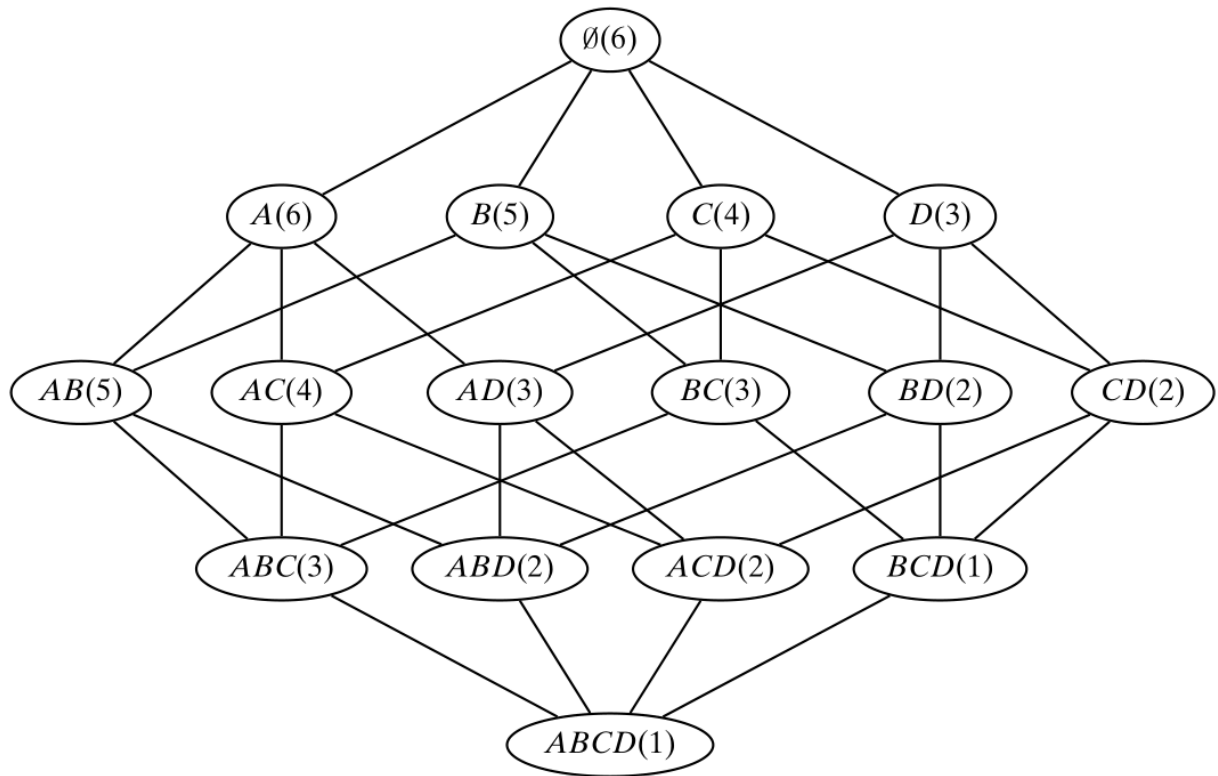
$c(BA) = i(t(BA)) = i(45) = BAD$ — BA не входит в C, тк $X \neq C(X)$

$c(BAD) = i(t(BAD)) = i(45) = BAD$ — BAD входит в C, тк $X = C(X)$

Для G:

тк BAD — закрытый сет для BA и только BA и BAD делят $tidset=45$, BA — минимальный генератор

2.



Выполните следующие задание:

1. Выпишите список всех закрытых наборов (closed itemsets)
2. Является ли набор BCD выводимым? Является ли набор ABCD выводимым? Какие границы их поддержки?.

Сгруппируем по поддержке:

6 — A
5 — B, AB,
4 — C, AC
3 — D, AD, BC, ABC
2 — BD, CD, ABD, ACD
1 — BCD, ABCD

Тогда, closed frequent set **C**:

C

A

AB

AC

AD

ABC

ABD

ACD

ABCD

выводимы ли BCD и ABCD (derived)? Какие границы поддержки?

$$X = BCD$$

UB – UPPER BORDER

LB – LOWER BORDER

$$Y \subseteq W \subset X$$

$$Y = BCD; |X/Y| = 0 - \text{even}; W - \text{not exist}; \sup(X) \geq 0 - LB$$

$$Y = BC; |X/Y| = 1 - \text{odd}; W = \{BC\}; \sup(X) \leq (-1)^2 * \sup(BC) = 3 - UB$$

$$Y = BD; |X/Y| = 1 - \text{odd}; W = \{BD\}; \sup(X) \leq (-1)^2 * \sup(BD) = 2 - UB$$

$$Y = CD; |X/Y| = 1 - \text{odd}; W = \{CD\}; \sup(X) \leq (-1)^2 * \sup(CD) = 2 - UB$$

$$Y = B; |X/Y| = 2 - \text{even}; W = \{B, BC, BD\};$$

$$\sup(X) \geq (-1)^3 * \sup(B) + (-1)^2 * \sup(BC) + (-1)^2 * \sup(BD) = -5 + 3 + 2 = 0 - LB$$

$$Y = C; |X/Y| = 2 - \text{even}; W = \{C, BC, CD\};$$

$$\sup(X) \geq (-1)^3 * \sup(C) + (-1)^2 * \sup(BC) + (-1)^2 * \sup(CD) = -4 + 3 + 2 = 1 - LB$$

$$Y = D; |X/Y| = 2 - \text{even}; W = \{D, BD, CD\};$$

$$\sup(X) \geq (-1)^3 * \sup(D) + (-1)^2 * \sup(BD) + (-1)^2 * \sup(CD) = -3 + 2 + 2 = 1 - LB$$

$$Y = \emptyset; |X/Y| = 3 - \text{odd}; W = \{B, C, D, BC, BD, CD, \emptyset\};$$

$$\sup(X) \leq (-1)^3 * \sup(B) + (-1)^3 * \sup(C) + (-1)^3 * \sup(D)$$

$$+ (-1)^2 * \sup(BD) + (-1)^2 * \sup(BC) + (-1)^2 * \sup(CD) + (-1)^4 * \sup(\emptyset)$$

$$= -5 - 4 - 3 + 3 + 2 + 2 + 6 = 1 - UB$$

$$\sup(X) \in [\max\{LB(X)\}, \min\{UB(X)\}] = [1, 1] - \text{borders}$$

if $\max\{LB(X)\} = \min\{UB(X)\}$, then X – derived $\rightarrow 1 = 1$, then **BCD – derived**

$$X = ABCD$$

$$Y = ABCD; |X/Y| = 0 - \text{even}; \sup(x) \geq 0$$

$$Y = ABC; |X/Y| = 1 - \text{odd}; \sup(x) \leq 3$$

$$Y = ACD; |X/Y| = 1 - \text{odd}; \sup(x) \leq 2$$

$$Y = ABD; |X/Y| = 1 - \text{odd}; \sup(x) \leq 2$$

$$Y = BCD; |X/Y| = 1 - \text{odd}; \sup(x) \leq 1$$

$$Y = AB; |X/Y| = 2 - \text{even}; \sup(x) \geq 0$$

$$Y = AC; |X/Y| = 2 - \text{even}; \sup(x) \geq 1$$

$$Y = AD; |X/Y| = 2 - \text{even}; \sup(x) \geq 1$$

$$Y = BC; |X/Y| = 2 - \text{even}; \sup(x) \geq 1$$

$$Y = BD; |X/Y| = 2 - \text{even}; \sup(x) \geq 1$$

$$Y = CD; |X/Y| = 2 - \text{even}; \sup(x) \geq 1$$

$$Y = A; |X/Y| = 3 - \text{odd}; \sup(x) \leq 1$$

$$Y = B; |X/Y| = 3 - \text{odd}; \sup(x) \leq 1$$

$$Y = C; |X/Y| = 3 - \text{odd}; \sup(x) \leq 1$$

$$Y = D; |X/Y| = 3 - \text{odd}; \sup(x) \leq 1$$

$$Y = \emptyset; |X/Y| = 4 - \text{even}; \sup(x) \geq 1$$

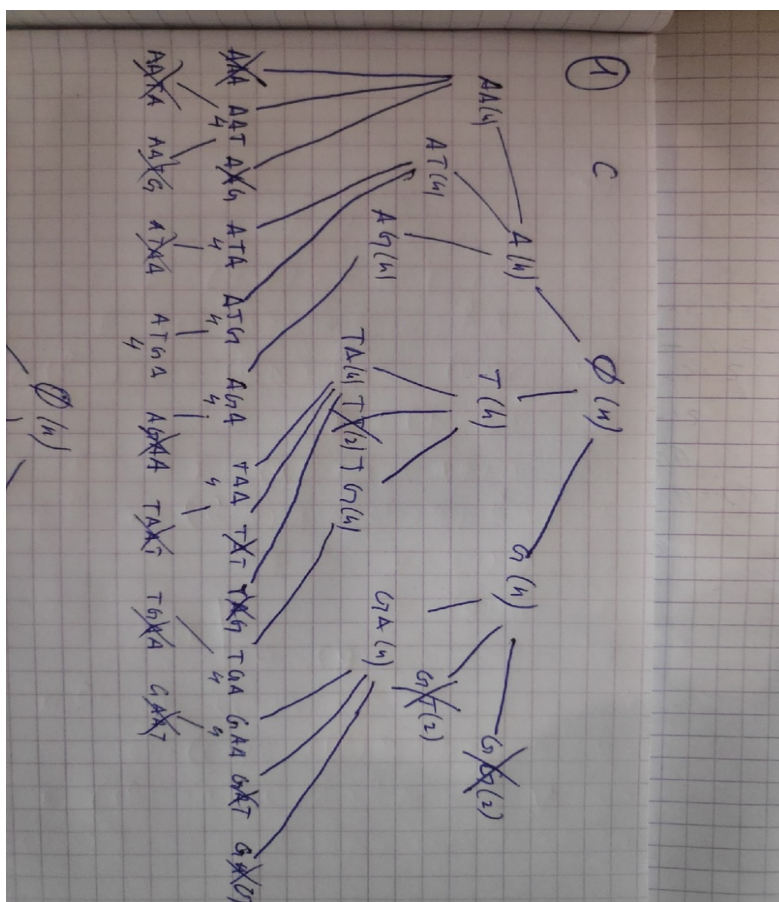
$$\sup(X) \in [\max\{LB(X)\}, \min\{UB(X)\}] = [1, 1] - \text{borders}$$

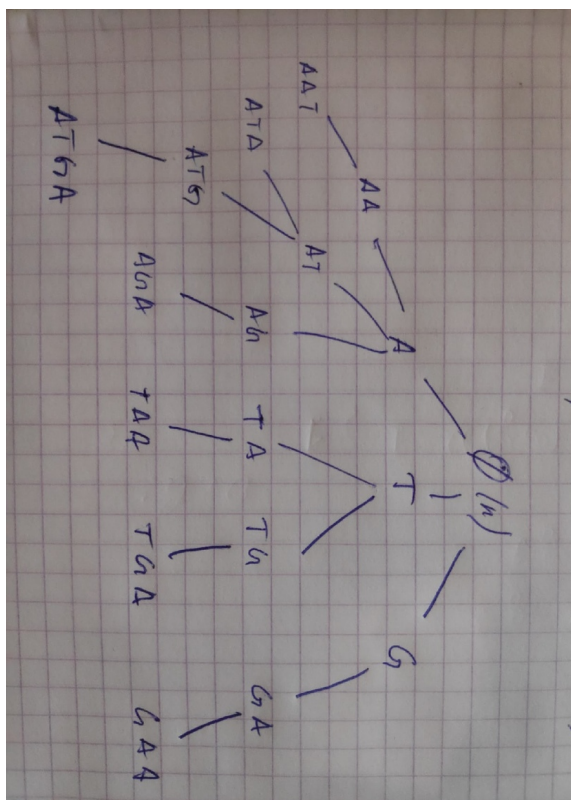
$$\text{if } \max\{LB(X)\} = \min\{UB(X)\}, \text{ then } X - \text{derived} \rightarrow 1 = 1, \text{ then } ABCD - \text{derived}$$

3. Найдите все подпоследовательности в минимальном уровне поддержки = 4
Даны последовательности:

Id	Sequence
s_1	<i>AATACAAGAAC</i>
s_2	<i>GTATGGTGAT</i>
s_3	<i>AACATGGCCAA</i>
s_4	<i>AAGCGTGGTCAA</i>

Как проходило построение дерева подпоследовательностей:



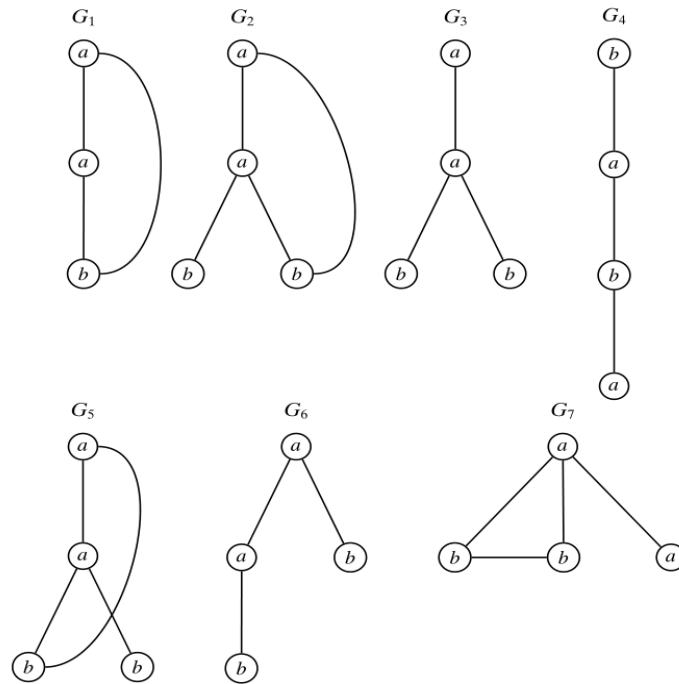


Итоговое дерево подпоследовательностей без лишних узлов

Для алфавита {A,C,G,T} посчитайте, сколько всего может быть разных последовательностей длины k

$$N = |\{A,C,G,T\}|^k$$

4. Дан набор графов. Разделите их на изоморфные группы



G_2, G_5 — изоморфные графы

G_1, G_3, G_4, G_6 — их изоморфные подграфы

G_3 — изоморфный подграф G_7