Рубежный контроль №2

Лиганкина Анна
Вариант 20

8 января 2026 г.

Задание 1

Дан язык SRS $bc \to cb, ac \to ccc$ над множеством базисных слов $a^n b^{n+k} c^k$. Определить, является ли язык КС.

Решение

Инварианты данной SRS:

1. количество букв b не меняется, т.е. остается $n + k$;
2. буквы c могут двигаться относительно букв b только влево;
3. количество букв a уменьшается, причем число $2|w|_a + |w|_c = 2n + k$ постоянно (количество c может только увеличиваться на 2 за счет одной буквы a).

Поэтому данный язык можно представить как:

1. В случае, если правило $ac \to ccc$ еще не было применено:

$$L_1 = \{w \mid w = a^n z \ \& \ z = (b|c)^* \ \& \ |z|_b = n + k \ \& \ |z|_c = k\}$$

2. Если правило $ac \to ccc$ было применено хотя бы раз:

$$L_2 = \{w \mid w = a^{n-i} c^{2i+1} z \ \& \ z = (b|c)^* \ \& \ |z|_b = n + k \ \& \ |z|_c = k - 1 \ \&$$

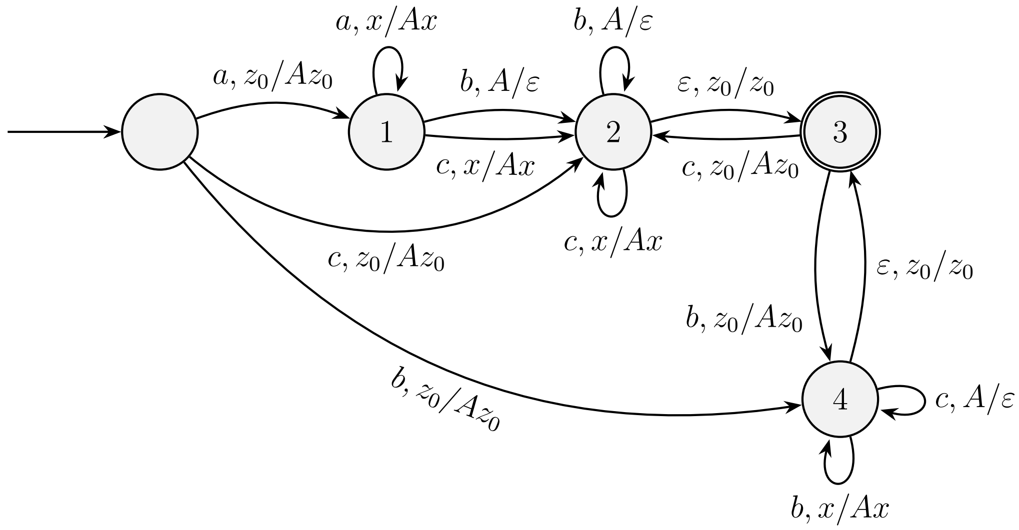
$$i, k \geq 1 \ \& \ n \geq i\}$$

Таким образом, исходный язык SRS $L = L_1 \cup L_2$. Поскольку КС языки замкнуты относительно объединения, для доказательства того, что L - КС, достаточно показать, что L_1 и L_2 - КС.

Язык L_1

$$L_1 = \{w \mid w = a^n z \ \& \ z = (b|c)^* \ \& \ |z|_b = n + k \ \& \ |z|_c = k\}$$

Для L_1 можно построить недетеминированный PDA $\Rightarrow L_1$ - КС:

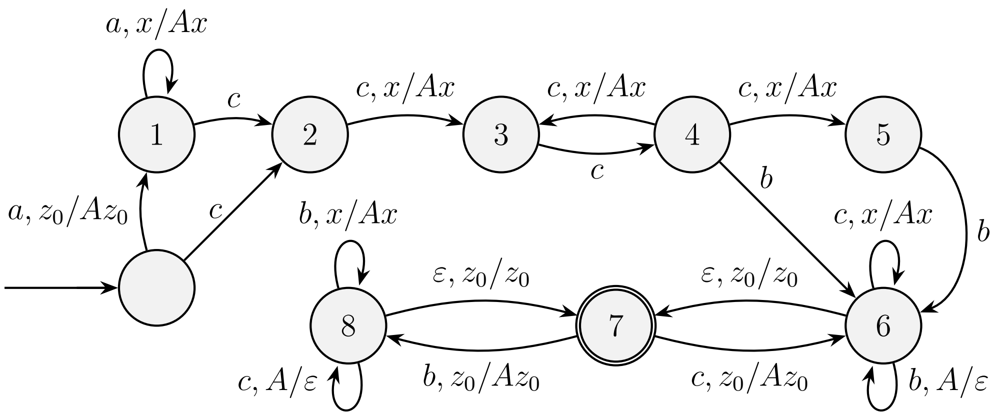


Язык L_2

Для удобства во втором случае переобозначим $n = n - i, k = k - 1$. Тогда

$$L_2 = \{w \mid w = a^n c^{2i+1} z \ \& \ z = (b|c)^* \ \& \ |z|_b = n + k + i + 1 \ \& \ |z|_c = k$$
$$\& \ i > 0 \ \& \ n, k \geq 0\}$$

Для L_2 можно построить недетеминированный PDA $\Rightarrow L_2$ - КС:



Таким образом, язык L_1 и L_2 - КС, а значит, L - КС. При этом для определения принадлежности слова языку обязателен счетчик $\Rightarrow$ язык нерегулярный.
Ответ: контекстно-свободный, нерегулярный.

Задание 2

Дан язык $\{w_1 w_2 \mid |w_1| > 1 \ \& \ w_2 = z_1 w_1^R z_2 \ \& \ |z_1| \neq |z_2|\}$. Определить, является ли он DCFL. Если да, то исследовать на LL-свойства.

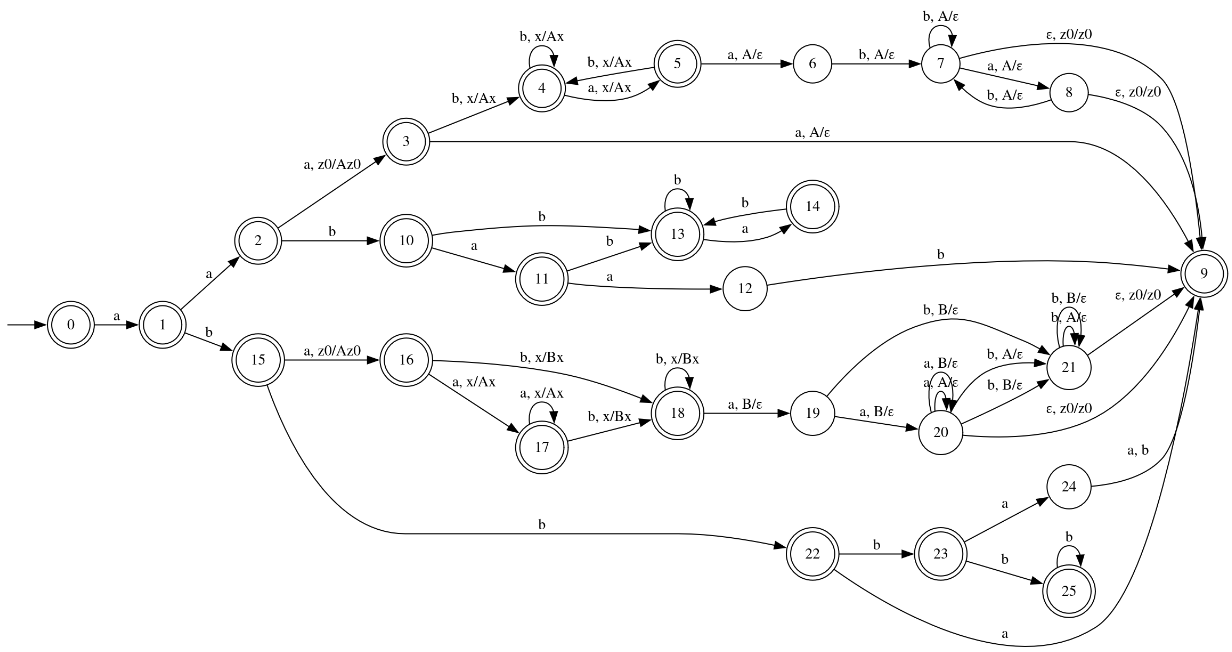
Решение

Поскольку длина $|w|$ не фиксирована, и т.к. если w^R содержится в слове, то для любой подстроки w существует как подстрока слова w^R ее реверс, можно ограничиться рассмотрением в качестве w только первых двух букв. Однако возникают следующие ситуации:

1. $\underbrace{ab}_{w}aa\underbrace{ba}_{w^R}ba$, хотя у слова есть разбиение $\underbrace{ab}_{w}aaba\underbrace{ba}_{w^R}$. Поэтому, если в слове несколько раз встречается реверс первых двух букв (причем с перекрытиями), то слово автоматически принадлежит языку (т.к. хотя бы одно из разбиений будет подходить критерию $|z_1| \neq |z_2|$).

2. $\underbrace{ab}_{w}a^n\underbrace{ba}_{w}a^n, n \geq 2$ - ba встречается лишь один раз, но существует разбиение $\underbrace{aba}_{w}a^{n-2}\underbrace{aba}_{w}a^n \Rightarrow w \in L$. Таким образом, еще одна необходимая проверка в слове - наличие реверса первых трех букв. Если это условие выполнено, то слово автоматически принадлежит языку.

Поскольку DCFL замкнуты относительно дополнения, то построим язык L' слов, не принадлежащих L. Нам подходят слова из алфавита $\{a,b\}^*$ такие, что в остатке слова либо не встречается реверс двух первых букв, либо встречается один раз, причем предшествующая им буква не совпадает с третьей, и длины отсекаемых слов равны. Половина DPDA - в файле DPDA.dot/DPDA.svg. Остальная часть - зеркальная (аналогично)



Таким образом, язык L - DCFL. Исследуем теперь язык на LL-свойства.

Язык L не является LL, что доказывается по методу неожиданного конца:
Возьмем префикс длины $> N$: $x = aab^{N+k}aa$ (N такое, что после чтения N букв b в стеке не меньше $k+3$ стековых символов). Также языку принадлежит слово с тем же префиксом x и суффиксом b^{N+k+1}. За чтение b^kaa израсходуется не более $k+2$ стековых символа, а значит, для слова с префиксом x и суффиксом ε стек окажется непустым $\Rightarrow$ язык не LL(k).
Ответ: DCFL, не LL(k).

Задание 3

Дана атрибутная грамматика:

$$\begin{array}{lcll} S & \to & TbS & ; \quad S_0.a := S_1.a + 1,\ S_1.a == T.a \\ S & \to & T & ; \quad S.a := T.a \\ T & \to & aTa & ; \quad T_0.a := T_1.a + 1 \\ T & \to & bT & ; \quad T_0.a := T_1.a \cdot 2 \\ T & \to & bbT & ; \quad T.a := 0 \\ T & \to & aTa & ; \quad T_0.a := T_1.a + 2 \\ T & \to & \varepsilon & ; \quad T.a := 0 \end{array}$$

Определить, является ли язык КС.

Решение

Докажем, что пересечение языка L, порожденного данной атрибутной грамматикой, с регулярным языком

$$R = aaaa\ (aa)^* b^+\ aa$$

или по-другому

$$R = \{a^{2p} b^q aa, p \geq 2, q \geq 1\}$$

не является контекстно-свободным, а значит, и L не является КС.
$L_R = L \cap R$ имеет вид:

$$\begin{array}{lcll} S & \to & TbT & ; \quad T_1.a == T_2.a \\ T & \to & aTa & ; \quad T_0.a := T_1.a + 1 \\ T & \to & bT & ; \quad T_0.a := T_1.a \cdot 2 \\ T & \to & bbT & ; \quad T.a := 0 \\ T & \to & aTa & ; \quad T_0.a := T_1.a + 2 \\ T & \to & \varepsilon & ; \quad T.a := 0 \end{array}$$

Обоснование. (От противного). Пусть $w \in L_R$ порождено правилом $S \to T$. Тогда, т.к. в начале стоит минимум 4 буквы a, то применяется правило $T \to aTa$ минимум 4 раза, что невозможно, поскольку в конце стоит только 2 буквы a подряд. Поэтому имеем четкое разбиение w по правилу $S \to TbS$, где $T = aaaa\ a^*$, а $S = b^+aa$.

При этом последняя S может быть раскрыта только по правилу $S \to T$, потому что если предположить, что $S \to TbS$, то T содержит только b^*, а значит, атрибут a для этого T равен 0. Поэтому (при любом представлении $S = b^*aa$) вес этого слова должен быть равен 0. Таким образом, w не принадлежит языку, поскольку вес левой части T точно больше 0.

По этой же причине нужно рассматривать блок b^* в качестве применения правила $T \to bT$, которое умножает атрибут a (иначе опять уйдем в 0). Получается, что значения атрибута a оценивается так (с учетом правила $S \to T_1bT_2$):

- для $T_1 : 2 \leq p \leq T_1.a \leq 2p$;
- для $T_2 : 2 \leq 2^{q-1} \leq T_2.a \leq 2^{q-1} \cdot 2$

Таким образом, $T_1.a = T_2.a \Leftrightarrow$

$$\begin{cases} 2^{q-1} \leq 2p & p \geq 2^{q-2} \\ p \leq 2^{q-1} \cdot 2 & p \leq 2^q \end{cases} \Leftrightarrow 2^{q-1} \leq 2p \leq 2^{q+1}$$

Т.е. $L_R = \left\{a^{2p}b^q aa \;\middle|\; p \geq 2 \;\&\; 2^{q-1} \leq 2p \leq 2^{q+1}\right\}$. Теперь воспользуемся леммой о накачке:
Возьмем слово $w = a^{2^{q-1}}b^q aa$, где $q > N$ - длина накачки. Рассмотрим разбиение $w = uvxyz$, где $|vxy| < N$:

- $vxy \subset a^{2p}$: при отрицательной накачке ($k = 0$) выходим из языка, т.к. получаем слово $a^{2p-t}b^q aa$, где $2p = 2^{q-1}, t > 0$;
- $vxy \subset b^q$: при любой положительной накачке ($k > 0$) выходим из языка, т.к. нарушается условие $2^{q-1} = 2p \geq 2^{q+t-1}$;
- vxy попадает на стык a^{2p} и b^q: число накачиваемых букв a меньше N, а значит новое число букв $a < N \cdot k$, а в части b^q накачивается хотя бы 1 буква, поэтому новое число букв $b \geq q + 1$. Получается, новое число букв a

  $$2^{q-1} + N \cdot k < 2^{q-1} + q \cdot k \;\text{ т.к. } (q > N) < 2^{q-1} + 2^{q-1} \cdot k \leq 2^q$$

  и мы выходим из языка;
- vxy не может задеть блок aa, потому что это автоматически выводит из языка.

Таким образом, согласно лемме о накачке, язык L_R не является КС, а значит, и исходный язык не является КС.
Ответ: не контекстно-свободный.