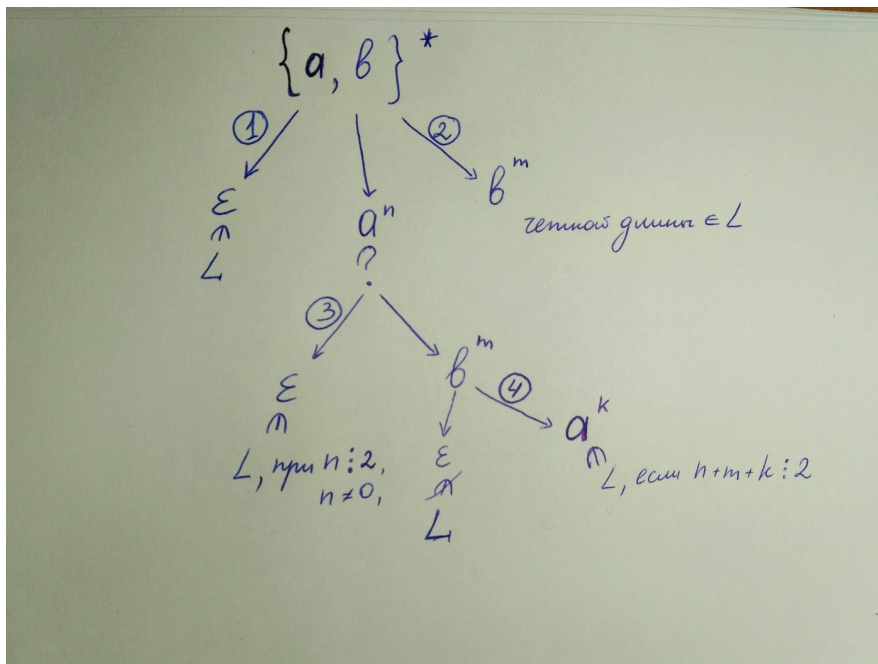


Домашнее задание 2

- (1) Постройте грамматику, порождающую все слова языка $\{a, b\}^* \setminus \{a^n b^m | n \geq 0, m \geq 1\}$ четной длины и только их.

Рассмотрим возможные варианты:



$$S_1 \rightarrow \epsilon;$$

$$S_2 \rightarrow baB|b^2B; B \rightarrow abB|bbB|baB|aaB|\epsilon$$

$$S_3 \rightarrow a^2A; A \rightarrow a^2A|\epsilon$$

$$S_4 \rightarrow S_{41}|S_{42}; S_{41} \rightarrow a^2C; C \rightarrow a^2C|b^2C|abC|baD; D \rightarrow \epsilon|B$$

$$S_{42} \rightarrow aS_{41}a|aS_{41}b$$

$$S \rightarrow S_1|S_2|S_3|S_4$$

- (2) Докажите по индукции, что грамматика $\{\{S, B, C\}, \{a, b, c\}, P, S\}$, где $P = \{S \rightarrow aSBC|aBC, CB \rightarrow BC, aB \rightarrow ab, bB \rightarrow bb, bC \rightarrow bc, cC \rightarrow cc\}$ порождает язык $\{a^n b^n c^n | n \geq 1\}$

Скажем, что: 0) $S \rightarrow aSBC|1)aBC$,

$$2) CB \rightarrow BC,$$

$$3) aB \rightarrow ab,$$

$$4) bB \rightarrow bb,$$

$$5) bC \rightarrow bc,$$

$$6) cC \rightarrow cc$$

Рассмотрим несколько первых шагов: $S \vdash_1 aBC \vdash_3 abC \vdash_5 abc$,

$$S \vdash_0 aSBC \vdash_1 a^2(BC)^2 \vdash_2 a^2BBCC \vdash_{3,4} aabbCC \vdash_{5,6} aabbcc.$$

Под знаком \vdash указываем, какое правило применяем, а над знаком - количество раз.

Проведем доказательство в 2 стороны: сначала покажем по индукции, что цепочка вида $a^n b^n c^n$ порождается данной грамматикой, затем покажем, что данная грамматика не порождает никакие другие цепочки.

1) Выше было показано, что цепочка abc выводима. Предположим, что при $n = k$ цепочка $a^k b^k c^k$ поражается грамматикой, докажем, что данной грамматикой поражается цепочка $a^{k+1} b^{k+1} c^{k+1}$, $n = k + 1$. Очевидно, что цепочка $a^k b^k c^k$ может быть выводима только из цепочки вида $a^{k-1} S(BC)^{k-1}$. Применим к этой цепочке правила грамматики: $a^{k-1} S(BC)^{k-1} \vdash_0 a^k S(BC)^k \vdash_1 a^{k+1} (BC)^{k+1}$, теперь применяем правило 2 столько раз, чтобы $(BC)^{k+1} \rightarrow B^{k+1} C^{k+1}$ (можно заметить что это правило надо применять $1 + 2 + \dots + n = n(n-1)/2$ раз ($n = k+1$), так как при добавлении новой пары BC к уже "отредактированной" последовательности, то правило 2 применяется к добавленной букве B столько раз, сколько букв C стоит перед ней). Получаем $a^{k+1} (BC)^{k+1} \vdash_2^{n(n-1)/2} a^{k+1} B^{k+1} C^{k+1} \vdash_3 a^{k+1} b B^k C^{k+1} \vdash_4^{n-1} a^{k+1} b^{k+1} C^{k+1} \vdash_{5,6}^n a^{k+1} b^{k+1} c^{k+1}$, доказано.

2) Теперь докажем, что других цепочек грамматика не поражает. Действительно, обозначим $N(a)$ - количество символов a . Легко заметить, что на любом шаге $N(a) = N(b) + N(B) = N(c) + N(C)$. На любом шаге любое a стоит левее любого b и любого B , любое b стоит левее любого c и любого C . Также после любого нетерминального символа идет нетерминальный. То есть любое a стоит левее любого b, B, c, C . Таким образом, применяя различные правила к цепочке вида $a^n S(BC)^n$ мы получаем ту же цепочку $a^n b^n c^n$ или тупиковые цепочки.

(3) **Докажите, что язык $\{a^n b^n | n > 0\}$ нерегулярен.**

Язык регулярен, тогда и только тогда, когда существует ДКА, принимающий этот язык. Предположим, что существует ДКА, принимающий данный язык: $Auto = (Q, a, b, D, Start, Final)$. Автомат конечный, следовательно он имеет конечное число состояний. Пусть $|Q| = k$. Так как ДКА принимает любое слово из языка, следовательно принимает и слово $a^{k+1} b^{k+1}$.

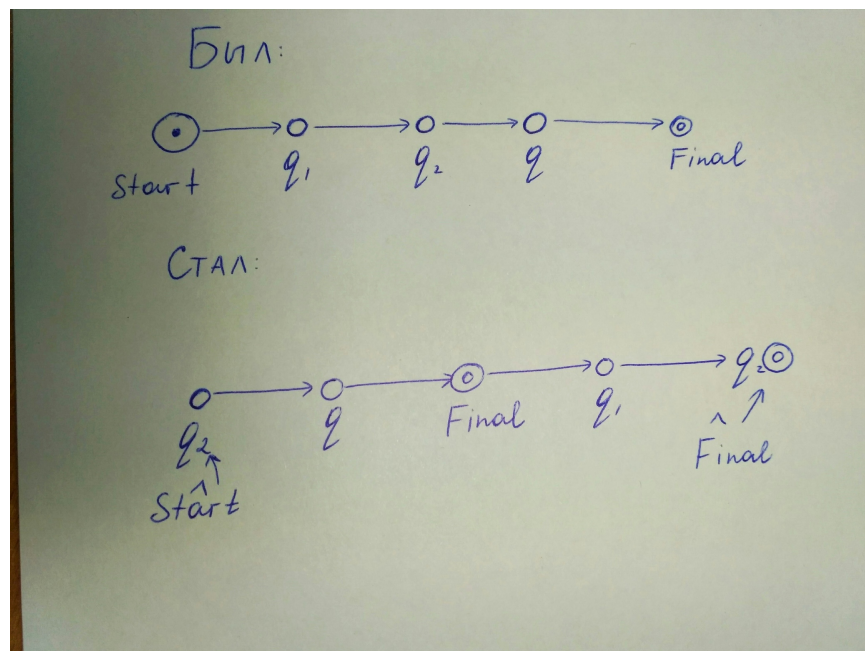
$(Start, a^{k+1} b^{k+1}) \vdash (q_1, a^k b^{k+1}) \vdash \dots \vdash (q_k, a b^{k+1}) \vdash (Final, \epsilon)$. Чтобы $k+1$ раз считать букву k автомату нужно вернуться в одно из пройденных состояний, потому что "петель" в автомате быть не может, иначе мы не сможем контролировать количество считанных букв a , и автомат будет принимать так же слова не из языка. Следовательно, автомат не является ДКА, получаем противоречие. То есть для данного языка не существует ДКА, который его принимает, следовательно, язык не является регулярным.

(4) **Пусть язык $L \subseteq T^*$ - регулярен. Докажите, что язык $L' = \{w \in T^* | w = yx, xy \in L, |x| = 2\}$ тоже будет регулярным.**

L - регулярен $\Rightarrow \exists$ ДКА, принимающий этот язык: $A = (Q, T, D, Start, Final)$

Чтобы доказать регулярность языка L' построим НКА, принимающий этот язык: $\hat{A} = (\hat{Q}, T, \hat{D}, \hat{Start}, \hat{Final})$

Если представить автомат A в виде графа, то, чтобы построить автомат \hat{A} , мы скопируем первые ребра исходного автомата, и создав 2 новые вершины, добавим эти скопированные ребра в конец исходного автомата, а далее добавляем 2 скопированных ребра. Делаем такие автоматы для всех путей длины 2, получаем много таких автоматов, далее заводим новую вершину и по ϵ -переходам соединяем стартовые вершины полученных автоматов. В виде графа, автомат для конкретного выбранного пути длины 2 выглядит следующим образом:



Опишем формально: $\hat{A} = (\hat{Q}, T, \hat{D}, \hat{Start}, \hat{Final})$

$$\hat{Q} = Q \times Q \times Q;$$

$$\hat{Start} = (q_{1i}; q_{2j}; q_{2j});$$

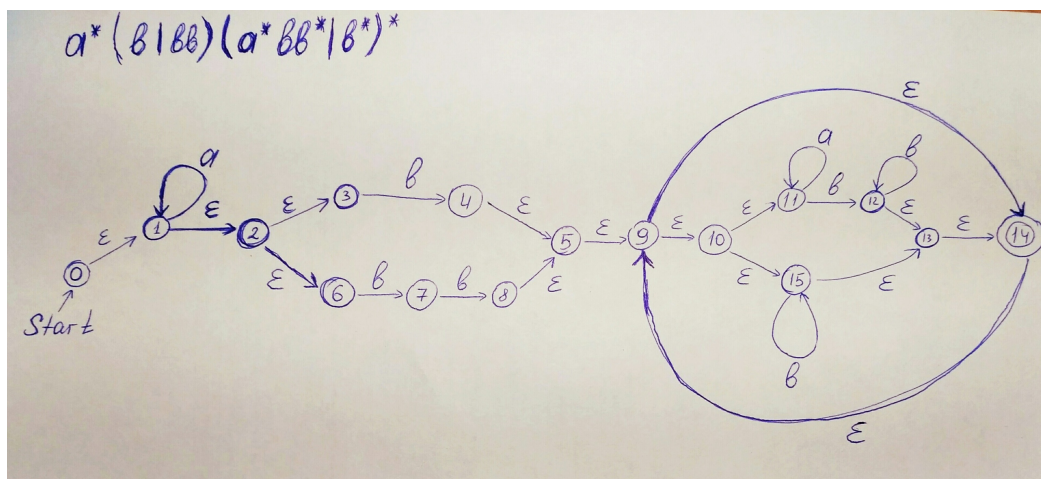
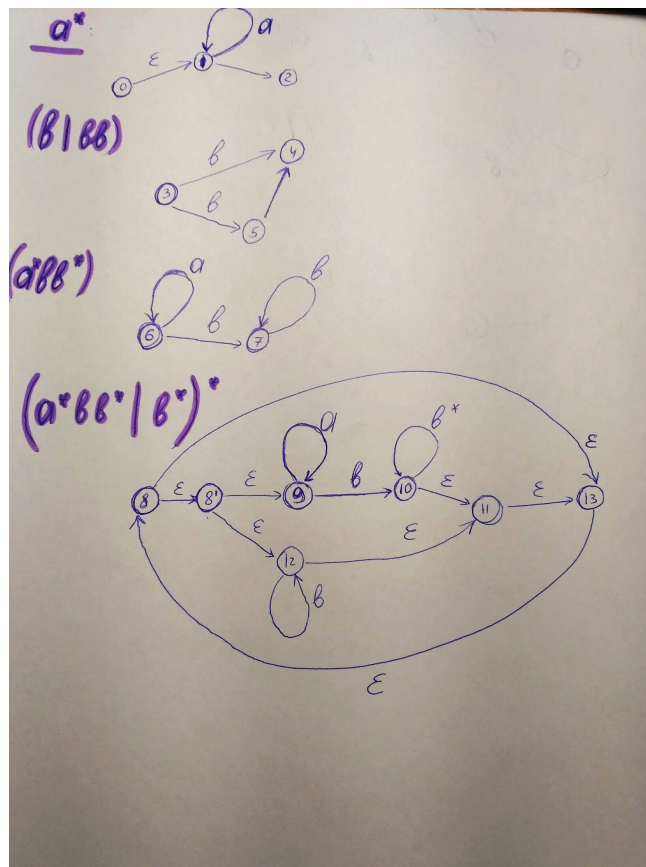
$$\hat{Final} = (q_{1i}, q_{2j}, q_{2j});$$

$$\hat{D}(q_{1i}, q_{2j}, q, a) = (q_{1i}, q_{2j}, D(q, a)) = (q_{1i}, q_{2j}, q'); \hat{D}(q_{1i}, q_{2j}, Final, a) = (q_{1i}, q_{2j}, q_{1i}) \forall a \in T,$$

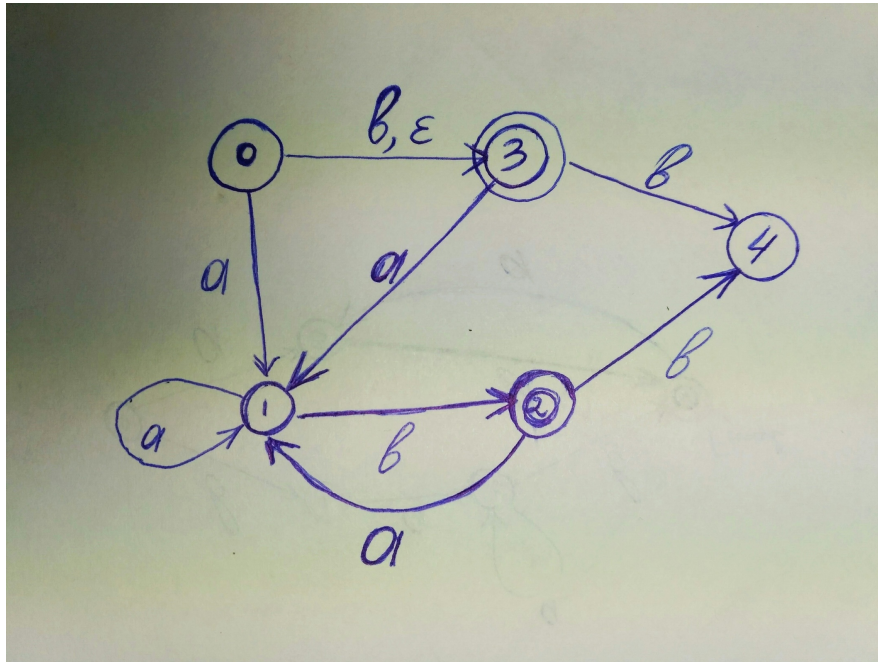
где $\exists a \in T : D(q_{1i}, a) = q_{2j}$, то есть существует такие 2 ребра $(start, q_{1i})$ и (q_{1i}, q_{2j}) в исходном автомате.

Докажем корректность построенного автомата: 1) Очевидно, что слова вида $yx, xy \in L, |x| = 2$ автомат принимает. 2) Докажем, что автомат не принимает никакие другие слова. Пусть автомат принял слово $w \notin L'$, по построению автомата получаем, что тогда $w = yx$, где $|x| = 2, xy \in L$. Но по предположению, $w \notin L'$. Получаем противоречие. Построили НКА, принимающий только слова из языка L' и никакие другие.

- (5) Постройте для РВ $a^*(b|bb)(a^*bb^*|b^*)^*$ над алфавитом $\{a, b\}$ недетерминированный конечный автомат.



- 4 (6) (4) Пусть язык $L \subseteq T^*$ – регулярный. Докажите, что язык $\sqrt{L} = \{w \in T^* | ww \in L\}$ тоже будет регулярным.
- Пусть автомат A принимает язык L . $A = (Q, T, D, Start, Final)$; построим автомат, который будет принимать язык L' : $\hat{A} = (\hat{Q}, T, \hat{D}, \hat{Start}, \hat{Final})$, $\hat{Q} = Q \times Q \times Q$, $\hat{Start} = (q_0, q, q_0^*)$, $\hat{Final} = (q, q, q_{final})$, $\hat{D}((q_1, q_2, q_3), x) = (D(q_1, x), q_2, D(q_3, x))$. Будем фиксировать состояние q и запускать машину "параллельно то есть выбрав фиксированное состояние q , делим путь на 2 части и прочитываем слово параллельно (одновременно по обоим частям). q_0 - начало первой половины, q_0^* - начало второй. Таким образом будем перебирать все возможные состояния q . Если для какого-то из фиксированных состояний первая половина слова и вторая считались одновременно, то слово принадлежит языку \sqrt{L} .
- Докажем корректность: 1) Допустим, $w : ww \in L$ не принимает построенный автомат. Это значит, что $\forall q \in Q$ слово w не может быть считано построенным выше образом, но это значит, что $ww \notin L$. Очевидно, что слова $w : ww \in L$ принимаются автоматом \hat{A} .
- 2) Теперь докажем, что автомат не может принимать никакие другие слова. Пусть \hat{A} принял слово $u : u \notin \sqrt{L} \Rightarrow u^2 \notin L$. Если автомат принял это слово, значит $\exists q \in Q : ((q_0, q, q), u) \vdash \dots \vdash ((q, q, q_{final}), \epsilon)$, но это значит, что слово $u^2 \in L$, получаем противоречие. Доказали, что построенный автомат принимает только цепочки из языка \sqrt{L} и никакие другие.
- (7) (3) Постройте КА для языка всех слов над алфавитом $\{a, b\}$, в который после каждой буквы a далее встречается (не обязательно следующим символом) буква b и нет двух b подряд.
- Построим регулярное выражение для слов языка: $b(aa^*b)^*(aa^*b)^*$



Очевидно, что автомат принимает слова, оканчивающиеся на b , в которых нет 2-х b подряд и пустое слово. Докажем по индукции, что построенный автомат принимает язык (натуральный параметр - число шагов, то есть длина входного слова).

1) База: $n = 1 : (q_0, b) \vdash (q_2, \epsilon)$ - принимается языком, $(q_0, a) \vdash (q_1, \epsilon)$ - не принимается, $(q_0, \epsilon) \vdash (q_2, \epsilon)$ - принимается языком.

2) Индукционный переход: $n = k$. Пусть автомат принимает все слова длины $\leq k$, которые оканчиваются на b и в которых нет 2-х букв b подряд и только эти слова. Докажем, что это утверждение справедливо для $n = k + 1$. Мы знаем, в каком состоянии находится автомат после считывания первых k символов: если он находится в финальном состоянии, из этого следует, что слово оканчивается на b и нет 2-х b подряд, $k+1$ буква - $a|b$, но очевидно, что такое слово автоматом не принимается, потому что ни из одного финального состояния нет перехода в финальное состояние. Если первые k символов не образуют цепочку, которая принимается автоматом, значит: 1) либо в них есть 2 b подряд, тогда слово из k таких символов и одной любой буквы, очевидно, тоже автоматом не принимается, 2) либо цепочка оканчивается на a и в ней нет двух букв b подряд. Тогда цепочка, состоящая из k этих символов и буквы b , очевидно, принимается автоматом, а состоящая из k этих символов и буквы a нет. Таким образом, мы рассмотрели все возможные варианты. Доказали, что автомат принимает только слова из языка и никакие другие.