

1 Практика

1.1 Конечные автоматы

1. $\{w \in \{0, 1\}^* \mid w \text{ содержит подстроку } 01\}$. (Разобрано на занятии 1)

2. Два DFA:

- $\{w \in \{0, 1\}^* \mid w \text{ начинается на } 01\}$.
- $\{w \in \{0, 1\}^* \mid w \text{ оканчивается на } 01\}$.

(Разобрано на занятии 1)

3. $\{w \in \{0, 1\}^* \mid \text{если в строке более одного символа, предпоследний равен } 1\}$. (Разобрано на занятии 3)

4. $\{w \in \{0, 1, 2\}^* \mid \text{сумма цифр делится на } 3\}$. (Разобрано на занятии 1)

5. Покажите, что $\{a^n b^n \mid n \geq 0\}$ не распознаётся никаким DFA. (Разобрано на занятии 1)

6. $\{w \in \{0, 1\}^* \mid \text{на каждой чётной позиции стоит } 1\}$. Индексация с единицы. Например, распознаётся строка $0\bar{1}1\bar{1}0\bar{1}0\bar{1}$. (Разобрано на занятии 2)

7. Язык, состоящий из слов чётной длины, начинающихся с 1, и слов нечётной длины, начинающихся с 0. (Разобрано на занятии 3)

8. Язык, состоящий из слов $\{0, 1\}^*$, в последних пяти символах которых хотя бы два раза встречался 0. (Разобрано на занятии 3)

9. Сначала несколько определений:

Определение (Отделимые слова). Будем говорить, что язык L отделяет слова s_1 и s_2 , если существует такое слово t , что одно из слов $s_1 t$ и $s_2 t$ принадлежит L , а другое — нет. Для неотделимых L слов будем писать, что $s_1 \equiv_L s_2$.

Определение (Индекс языка). Индексом L назовём супремум размеров множеств, в которых все строки попарно отделимы языком L (если все такие множества конечные, то индекс — это размер наибольшего; иначе индекс бесконечный).

Само задание:

- Найдите для языка из первой задачи какое-нибудь наибольшего размера множество попарно разделимых им строк.
- Найдите индекс языка из первой задачи.

(Разобрано на занятии 2)

10. Найдите индекс языка из задачи №4 из домашнего задания. (Разобрано на занятии 2)

11. Регулярный ли язык $P(L) = \{w \mid \text{все префиксы } w \text{ входят в язык } L\}$, если L регулярный? (Разобрано на занятии 3)

12. Регулярный ли язык $L^2 = \{ww \mid w \in L\}$, если L регулярный? (Разобрано на занятии 3)

13. Если L регулярный, то регулярный ли язык $\{w \mid \text{число префиксов } w, \text{ входящих в } L, \text{ чётно}\}$? (Разобрано на занятии 3)

14. Важные определения:

Определение (Недетерминированный конечный автомат). Недетерминированный конечный автомат (сокращённо “NFA” или “НКА”) — это набор вида $(\Sigma, Q, Q_0, \Delta, T)$ такой, что

- Σ, Q, T определены как в DFA;
- $Q_0 \subseteq Q$ — подмножество состояний, называемое начальными;
- $\Delta : Q \times \Sigma \rightarrow 2^Q$ — функция, которая осуществляет переход из состояния в какое-то множество состояний по символу или по пустой строке.

NFA A принимает язык L , если A заканчивает хотя бы одно из вычислений на любом входе $w \in L$ в состоянии из T . Формально:

$$L(A) = \{w = a_1 \cdots a_n \mid a_1, \dots, a_n \in \Sigma, \exists r_0, \dots, r_n \in Q : r_0 \in Q_0, r_i \in \Delta(r_{i-1}, a_i), r_n \in T\}$$

Определение (co-NFA). co-NFA — это автомат, который устроен точно так же, как NFA, но принимает слово w , только если *каждое* вычисление на входе x в состоянии из T .

Определение (ε -NFA). ε -НКА — НКА, в котором переходы могут быть не только по символам, но и по пустым строкам.

Факт. Язык, принимаемый ε -НКА, принимается и НКА. (Лекция; вкратце: когда кто-то переходит в состояние q , из которого по пустой строке можно попасть в состояния E , то переход происходит и в E).

Факт. Язык, принимаемый НКА, принимается и ДКА. (Лекция; вкратце: построим новый ДКА, множество состояний которого — 2^Q , а дальше всё само получается).

Приведите DFA и NFA для языка слов, распознаваемых регулярным выражением $a^*b^*c^*$. (Разобрано на занятии 4)

- Покажите, что язык слов над $\{a, b, c\}$, в которых последний символ уже встречался до этого, распознаётся DFA. (Разобрано на занятии 4)
- Покажите, что язык слов над $\{a, b, c\}$, в которых последний символ ещё не встречался, распознаётся DFA. (Разобрано на занятии 4)
- Постройте NFA для языка слов, описываемых регулярным выражением $((01)^* | (010)^*)$. (Разобрано на занятии 4)
- Напишите в автоматном стиле программу на Си, которая печатает первое слово из каждой поданной на вход строки. (Разобрано на занятии 5)
- Приведите регулярное выражение для языка слов над алфавитом $\{0, 1\}$, не содержащих смежных единиц. (Разобрано на занятии 6)
- Приведите регулярное выражение для языка слов над алфавитом $\{0, 1\}$, содержащих ровно одну пару смежных единиц. (Не разобрано и не будет)
- Приведите регулярное выражение для языка слов над алфавитом $\{0, 1\}$, содержащих либо ни одной пары смежных единиц, либо одну. (Не разобрано и не будет)
- Нарисуйте NFA для языка $(0+1)01$. (Не разобрано и не будет)
- Нарисуйте NFA для языка $00(0+1)^*$. (Не разобрано и не будет)
- Приведите регулярное выражение для языка троичных чисел, сумма разрядов которых делится на три. (Не разобрано и не будет)
- Приведите регулярное выражение для языка двоичных чисел, кратных пяти. (Не разобрано и не будет)

1.2 Контекстно-свободные языки

26. Определение:

Определение (Контекстно-свободная грамматика). Контекстно-свободная грамматика (КС-грамматика, context-free grammar, cf-grammar) G — это четвёрка (Σ, N, R, S) такая, что:

- Σ — конечный алфавит; его элементы назовём *терминальными символами*, или *терминалами*.
- N — некоторое конечное множество, причём $N \cap \Sigma = \emptyset$. Его элементы назовём *нетерминальными символами*, или *нетерминалами*.
- $S \in N$ — стартовый символ.
- R — конечное множество правил вывода вида $A \rightarrow \alpha$, $A \in N, \alpha \in (N \cup \Sigma)^*$; иными словами, R — конечное подмножество $2^{N \times (N \cup \Sigma)^*}$.

Говорим, что язык $L(G)$ — это язык, *распознаваемый* (*генерируемый*) грамматикой G , множество строк из Σ^* , которые можно вывести из S по правилам вывода из R .

Зададим грамматику с $\Sigma = \{a, b\}$, $N = \{S\}$ и следующим R :

- $S \rightarrow aSb$
- $S \rightarrow SS$
- $S \rightarrow \varepsilon$

(Можно записать R иначе: $\{S \rightarrow aSb \mid SS \mid \varepsilon\}$).

Объясните, что за язык эта грамматика распознаёт и почему. (*Разобрано на занятии 7*)

27. Приведите грамматику для языка всех слов нечётной длины над алфавитом $\{a, b\}$. (*Разобрано на занятии 7*)
28. Приведите грамматику для языка $\{a^i b^j \mid i \geq j\}$. (*Разобрано на занятии 7*)
29. Приведите грамматику для языка $\{a^i b^j c^k \mid i \neq j\}$. (*Разобрано на занятии 7*)
30. Приведите грамматику для языка $\{a^i b^j c^k \mid i \neq k\}$. (*Разобрано на занятии 7*)
31. Приведите грамматику для языка $\{a^i b^j c^k \mid i \neq j \vee i \neq k\}$. (*Разобрано на занятии 7*)
32. Опишите, что распознаёт грамматика $S \rightarrow aSa \mid aa \mid a$, и докажите это в обе стороны по индукции.
33. Опишите, что распознаёт грамматика $S \rightarrow aSa \mid aTb \mid bTa \mid bTb$, $T \rightarrow aT \mid bT \mid \varepsilon$, и докажите это в обе стороны по индукции.
34. Приведите грамматику для языка $0^*1(0+1)^*$. Напишите левосторонний и правосторонний вывод в ней для строк 00101, 1001, 00011.
35. Дан алфавит $\{0, 1, (,), +, *, \emptyset, \varepsilon\}$. Приведите грамматику языка регулярных выражений.
36. Приведите грамматику для языка $\{w \mid w \in \{a, b\}^*, |w_a| = 2 \cdot |w_b|\}$.
37. Покажите, что грамматика $S \rightarrow aS \mid aSbS \mid \varepsilon$ неоднозначная.
38. Докажите, что грамматика из задания (37) распознаёт ровно те строки, в каждом префиксе которых количество a не меньше количества b , то есть ровно $\text{PREFIX}(L)$, где L — язык правильных скобочных последовательностей.
39. Опишите, какие строки обладают ровно одним деревом разбора в грамматике, определённой в задании (37).
40. Является ли неоднозначной грамматика $S \rightarrow SaS \mid a \mid b$? Почему?

2 Домашнее задание

2.1 Конечные автоматы

1. $\{w \in \{0, 1\}^* \mid w \text{ не содержит подстроку } 01\}$. (1.5 б.) (Разобрано на занятии 2)
2. Покажите, что язык L распознаётся DFA тогда и только тогда, когда \bar{L} распознаётся DFA, где $\bar{L} = \Sigma^* \setminus L$. (2 б.) (Разобрано на занятии 2)
3. Докажите, что любой конечный язык (иными словами, конечное множество слов) распознаётся DFA. (2.5 б.) (Разобрано на занятии 2)
4. $\{a^m b^n \mid m, n \geq 0, m \equiv n \pmod{3}\}$. (2.5 б.) (Разобрано на занятии 2)
5. $\{w \in \{0, 1\}^* \mid w \text{ делится на } 5 \text{ как двоичное число}\}$. (3.25 б.) (Разобрано на занятии 3)
6. Правда ли, что если L и M распознаются DFA, то $L \cap M$ обязательно тоже? Обоснуйте. (3.25 б.) (Разобрано на занятии 2)
7. Правда ли, что если $L \cap M$ распознаётся DFA, то L и M обязательно тоже? Обоснуйте. (2 б.) (Разобрано на занятии 3)
8. Покажите, что невозможно распознать DFA язык $\{a^n b^m \mid n \neq m\}$. (2 б.) (Разобрано на занятии 2)
9. Покажите, что невозможно распознать DFA язык $\{a^n b^m \mid m, n \geq 0, \gcd(m, n) > 1\}$. (2.75 б.) (Разобрано на занятии 3)
10. Покажите, что класс распознаваемых DFA языков замкнут относительно операции $\sqrt{L} = \{m \mid mm \in L\}$. (3 б.) (Разобрано на занятии 3)
11. Покажите, что класс распознаваемых DFA языков замкнут относительно операции $L^R = \{m^R \mid m \in L\}$, где $(a_1 a_2 \dots a_n)^R = a_n a_{n-1} \dots a_1$. (1.75 б.) (Разобрано на занятии 5)
12. Докажите, что язык L распознаётся неким DFA тогда и только тогда, когда индекс L конечен. (2.75 б.) (Разобрано на занятии 3)
13. Докажите, что индекс языка L равен минимальному размеру распознающего L автомата. (2.75 б.) (Разобрано на занятии 3)
14. Всегда ли, когда L распознаётся DFA, распознаётся и $L/L' = \{u \mid \exists v : uv \in L, v \in L'\}$, где $L' \subseteq \Sigma^*$ — произвольный язык над тем же алфавитом? (3 б.) (Разобрано на занятии 3)
15. Всегда ли, когда L распознаётся DFA, распознаётся и $\text{PERMUTE}(L) = \{a_{k_1} a_{k_2} \dots a_{k_n} \mid n \geq 0, (k_1 k_2 \dots k_n) \text{ — некоторая перестановка, } a_i \in \Sigma, a_1 a_2 \dots a_n \in L\}$? (2.75 б.) (Разобрано на занятии 4)
16. Всегда ли, когда L распознаётся DFA, распознаётся и $\text{SUBSEQ}(L) = \{a_1 a_2 \dots a_n \mid n \geq 0, a_i \in \Sigma, \exists u_0, \dots, u_n \in \Sigma^* : u_0 a_1 u_1 a_2 u_2 \dots a_n u_n \in L\}$? (2.5 б.) (Разобрано на занятии 6)
17. Всегда ли, когда L распознаётся DFA, распознаётся и $\frac{1}{2}L = \{u \mid u \in \Sigma^*, \exists v \in \Sigma^* : |u| = |v|, uv \in L\}$? (2.5 б.) (Разобрано на занятии 4)
18. Всегда ли, когда L распознаётся DFA, распознаётся и $\text{SHIFT}(L) = \bigcup_{k \geq 0} \{a_{k+1} a_{k+2} \dots a_n a_1 a_2 \dots a_k \mid a_1 a_2 \dots a_n \in L\}$? (3 б.) (Разобрано на занятии 4)

19. Приведите такой пример множества $A \subseteq \mathbb{N}$, что язык чисел из A , представленных в двоичной системе, распознаётся DFA, а язык тех же чисел, представленных в троичной, — нет. (3.5 б.) (Разобрано на занятии 4)
20. Всегда ли, когда L распознаётся DFA, распознаётся и $h(L) = \{h(x) \mid x \in L\}$, где $L \subseteq \Sigma^*$ и h — некий гомоморфизм на Σ^* ? (3 б.) (Разобрано на занятии 5)
21. Всегда ли, когда L распознаётся DFA, распознаётся и $h^{-1}(L) = \{y \mid h(y) \in L\}$, где $L \subseteq \Sigma^*$ и h — некий гомоморфизм на Σ^* ? (3.25 б.) (Разобрано на занятии 5)
22. Докажите, что нельзя распознать DFA язык $\{0^n \mid n \text{ — полный квадрат}\}$. (2.25 б.) (Разобрано на занятии 4)
23. Покажите, что если детерминированный автомат имеет k состояний и синхронизируется какой-то строкой, то он синхронизируется и строкой длины не более k^3 .

Определение (Синхронизация строкой). Автомат синхронизируется строкой s , если $\forall q_1, q_2 \in Q : \delta(q_1, s) = \delta(q_2, s)$.

(Разобрано на занятии 6)

24. Покажите, что класс языков, распознающихся co-NFA, совпадает с классом языков, распознающихся NFA. (Разобрано на занятии 6)
25. Существует ли такое семейство языков $\{E_n\}_{n \in \mathbb{N}}$, что E_n распознаётся NFA с n состояниями, но требует DFA размером как минимум c^n для некоторого $c > 1$? (Разобрано на занятии 6)
26. Покажите, что не распознаётся DFA язык $\{w \in \{0, 1\}^* \mid w \text{ — двоичное представление простого числа}\}$. (Разобрано на занятии 7)
27. Приведите алгоритм, который по данному DFA A вычисляет количество распознаваемых им слов длины n за $O(\text{poly}(|Q_A| \cdot n))$, где $\text{poly}(x)$ — некая полиномиальная функция. (Разобрано на занятии 6)
28. Решите предыдущую задачу за $O(\text{poly}(|Q_A|) \cdot \log(n))$.
29. Будем говорить, что $L_1 \ll L_2$, если $L_1 \subset L_2$ и $|L_2 \setminus L_1| = \infty$. Докажите, что если L_1 и L_2 распознаются DFA и $L_1 \ll L_2$, то существует такой распознаваемый DFA язык L_3 , что $L_1 \ll L_3 \ll L_2$. (Разобрано на занятии 7)
30. Пусть M_1 и M_2 — DFA, имеющие k_1 и k_2 состояний соответственно, и пусть $U = L(M_1) \cup L(M_2)$, где $L(A)$ — язык, распознаваемый автоматом A . Пусть $U \neq \emptyset$ и $U \neq \Sigma^*$. Докажите, что U содержит некоторую строку s_1 длины не более $\max(k_1, k_2)$ и что существует не принадлежащая U строка s_2 длины не более $k_1 k_2$. (Разобрано на занятии 7)
31. Приведите регулярное выражение для языка слов из 0 и 1, в которых каждая пара смежных 0 находится перед парой смежных 1. (Разобрано на занятии 7)
32. Приведите регулярное выражение для языка слов из 0 и 1, в которых число символов 0 делится на 3, а число символов 1 чётно. (Разобрано на занятии 7)
33. Докажите, что замыкание Клини унарного языка (языка над алфавитом из одного символа) является регулярным языком.

2.2 Контекстно-свободные грамматики

34. Покажите, что является контекстно-свободным язык $\{a^n b^{n+m} a^m \mid n, m \geq 0\}$. (Разобрано на занятии 8)
35. Покажите, что является контекстно-свободным язык $\{w \mid w \in \{a, b\}^*, |w|_a = |w|_b\}$. (Разобрано на занятии 8)
36. Покажите, что является контекстно-свободным язык \bar{L} , где $L = \{ww \mid w \in \{a, b\}^*\}$.
37. Покажите, что класс контекстно-свободных языков замкнут относительно объединения с контекстно-свободными и пересечения с регулярными языками. (Разобрано на занятии 8)
38. Покажите, что класс контекстно-свободных языков замкнут относительно операции $\text{SUFFIX}(L) = \{v \mid \exists u : uv \in L\}$. (Разобрано на занятии 8)
39. Покажите, что класс контекстно-свободных языков замкнут относительно операции $h(L) = \{h(x) \mid x \in L\}$, где $h : \Sigma^* \rightarrow \Gamma^*$ — произвольный гомоморфизм. (Разобрано на занятии 8)
40. Покажите, что класс контекстно-свободных языков замкнут относительно операции $\text{SHIFT}(L)$.
41. Покажите, что если L_1 и L_2 — регулярные языки, то $L_3 = \{uv \mid u \in L_1, v \in L_2, |u| = |v|\}$ контекстно-свободный.
42. Пусть $G = (\Sigma, N, R, S)$ — грамматика, в которой $|R| = p$ и длина правой части каждого правила не превосходит m . Пусть из $A \in N$ выводится пустая строка. Докажите, что тогда из A возможно вывести пустую строку за не более чем $\frac{m^p - 1}{m - 1}$ шагов.

3 Дополнительные задания

3.1 Регулярные языки

3.1.1 Формат сдачи

Для тех задач, где требуется программа, ответ нужно представить в виде программы на любом языке. Программа должна следовать указанному формату.

Формат описания конечного автомата Тестирующая система выдаёт и ожидает

- На первой строке два неотрицательных числа: k , количество букв в алфавите, и n , количество состояний в данном автомате.
- На второй строке через пробел представлено k символов — допустимых букв данного алфавита.
- На третьей строке дано единственное число $[0; n)$ — индекс начального состояния автомата.
- На четвёртой строке через пробел сначала дано число z , а затем перечислено z чисел $[0; n)$ — индексов терминальных состояний автомата.
- Следующие $n \cdot k$ строк содержат единственное число $[0; n)$. То, что на строке i содержится число j , означает, что из состояния $[i/k] \in [0; n)$ по символу с индексом $i \pmod k \in [0; k)$ осуществляется переход в состояние j .

Пример представления автомата, который распознаёт слова, начинающиеся со строки 01, (комментарии представлены для наглядности, их тестирующая система писать не будет):

```
2 4 # 2 letters , 4 states
0 1 # Sigma = {0, 1}
0 # the initial state is 0
1 2 # there is one terminal state , 2
1 # 0 -0-> 1
3 # 0 -1-> 3
3 # 1 -0-> 3
2 # 1 -1-> 2
2 # 2 -0-> 2
2 # 2 -1-> 2
3 # 3 -0-> 3
3 # 3 -1-> 3
```

Формат описания недетерминированного конечного автомата

- На первой строке два неотрицательных числа: k , количество букв в алфавите, и n , количество состояний в данном автомате.
- На второй строке через пробел представлено k символов — допустимых букв данного алфавита.
- На третьей строке дано единственное число $[0; n)$ — индекс начального состояния автомата.
- На четвёртой строке через пробел сначала дано число z , а затем перечислено z чисел $[0; n)$ — индексов терминальных состояний автомата.
- Следующие $n \cdot (k+1)$ строк имеют следующий вид: сначала число a_i , а затем a_i чисел $[0; n)$ подряд. То, что на строке i записаны числа a_{ij} , означает, что из состояния $[i/(k+1)] \in [0; n)$ можно попасть во все состояния a_{ij} либо по символу r , где $r = (i \bmod (k+1))$, если $r < k$, либо по пустой строке, если $r \equiv k \bmod (k+1)$.

Формат описания регулярного выражения Регулярным выражениям соответствует два формата.

- Формат, который тестирующая система принимает на вход, — это обычные регулярные выражения. Регулярное выражение — это одна из следующих конструкций:
 - Символ алфавита.
 - Регулярное выражение | регулярное выражение.
 - Регулярное выражение, за которым тут же записано другое регулярное выражение.
 - Регулярное выражение между парой круглых скобок.
 - Регулярное выражение, за которым идёт $*$.

Пробелы в этой записи не допускаются.

Пример регулярного выражения:

```
a((ab|b|c)c)*(cde|fg)
```

- Формат, который выдаёт тестирующая система, устроен для простоты считывания. Регулярное выражение — это одна из следующих конструкций:
 - `empty()` — пустое регулярное выражение.

- `char`(, затем единственный символ рассматриваемого алфавита, затем).
- `kleene`(, затем регулярное выражение, затем).
- `or`(, затем регулярное выражение, затем) (, затем ещё одно регулярное выражение, затем).
- `append`(, затем регулярное выражение, затем) (, затем ещё одно регулярное выражение, затем).

Представление того же регулярного выражения, что было в примере к предыдущему пункту (пробелы добавлены для читаемости; в том, что выдаёт тестирующая система, гарантированно нет ни пробелов, ни каких-то ещё лишних символов):

```
append(char(a))
  (append(kleene(append(or(append(char(a))
                        (char(b)))
                    (or(char(b))
                      (char(c))))
        (char(c))))
  (or(append(char(c))
      (append(char(d))
              (char(e))))
  (append(char(f))
          (char(g))))))
```

3.1.2 Задачи

1. Напишите программу, которая принимает на вход описание недетерминированного конечного автомата и возвращает описание детерминированного конечного автомата, который распознаёт тот же язык.
Напишите также программу, которая производит преобразование в обратную сторону.
2. Напишите программу, которая принимает на вход описание детерминированного конечного автомата и возвращает регулярное выражение, которое распознаёт тот же язык.
Напишите также программу, которая производит преобразование в обратную сторону.
3. Напишите программу, которая принимает на вход регулярное выражение для языка L и возвращает регулярное выражение для языка L^R .
4. Напишите программу, которая принимает на вход регулярное выражение для языка L и возвращает регулярное выражение для языка \sqrt{L} .
5. Напишите программу, которая принимает на вход регулярное выражение для языка L и возвращает регулярное выражение для языка $\text{SHIFT}(L)$.
6. Напишите программу, которая принимает на вход регулярное выражение для языка L и возвращает регулярное выражение для языка $\text{SUBSEQ}(L)$.
7. Напишите программу, которая принимает на вход регулярное выражение для языка L и возвращает регулярное выражение для языка $\frac{1}{2}(L)$.
8. Напишите программу, которая принимает на вход регулярное выражение для языка L и гомоморфизм $h : \Sigma^* \rightarrow \Gamma^*$ и возвращает регулярное выражение для языка $h^{-1}(L)$.
Формат входа следующий:
 - Строка с единственным числом k .
 - Строка с k разделёнными пробелом символами — буквами алфавита Σ .

- Строка с k разделёнными пробелом словами, где i 'ое слово — это образ i 'ого символа под h . Обратите внимание, что гомоморфный образ может быть строкой над другим алфавитом.
- Регулярное выражение над алфавитом Γ .