Конечные автоматы

Конечным автоматом называется пятерка вида $\mathfrak{A} = \langle \Sigma, Q, s, F, \Delta \rangle$, где Σ — конечное множество букв, называемое *входным алфавитом*, Q — конечное множество, называемое *множеством состояний*, s — элемент Q, называемый *начальным состоянием*, F — подмножество Q, называемое *множеством принимающих состояний* и Δ — подмножество $\Sigma \times Q \times Q$, называемое *множеством допустимых переходов*.

Автомату $\mathfrak{A}$ подаются слова в алфавите Σ и он производит над ними *процесс вычисления*. Автомат начинает свою работу в начале слова в состояние s. Далее он производит следующие шаги вычисления до тех пор пока он не дойдет до конца слова: пусть очередная буква слова $a \in \Sigma$ и автомат находится в состояние $q \in Q$, тогда автомат может перейти к следующей букве (или концу слова, если текущая буква последняя) в любое из состояний q' таких, что $\langle a, q, q' \rangle \in \Delta$.

1. Приведите примеры автомата $\mathfrak{A}$ со входным алфавитом $\{a\}$ и слова α в этом алфавите такого, что:
а) есть более, чем один процесс вычисления $\mathfrak{A}$ на слове α, доходящий до конца слова.
б) нет ни одного процесса вычисления $\mathfrak{A}$ на слове α, доходящего до конца слова.

Мы говорим, что автомат $\mathfrak{A}$ *принимает слово* α в алфавите Σ, если найдется доходящий до конца α процесс вычисления, заканчивающийся принимающим состоянием $q \in F$. Мы говорим, что автомат $\mathfrak{A}$ *распознает* множество слов $\mathbf{L}$ в алфавите Σ, если $\mathbf{L}$ — это множество всех слов в алфавите Σ, принимаемых автоматом $\mathfrak{A}$. Множество слов, распознаваемое каким-либо автоматом называется *автоматоматным*.

2. Докажите, что существуют автоматы, распознающие следующие множества слов в алфавите $\{a, b\}$:
а) $\{ab\}$;
б) $\{\varepsilon\}$ (мы обозначаем через ε пустое слово);
в) $\{a^n = \underbrace{a \dots a}_{n \text{ раз}} \mid n \in \mathbb{N}\}$;
г) $\{\alpha \mid$ в α входит по крайней мере 3 буквы $b\}$;
д) $\{\alpha \mid$ в α входит нечетное число букв $a\}$;
е) $\{\alpha \mid \alpha$ имеет длину кратную $3\}$;
ж) $\{\alpha \mid$ в α входит подслово $abba\}$;
з) $\{\alpha \mid$ в α входит подслово $ababa$ или α имеет нечетную длину$\}$;
и) $\{\alpha \mid$ в α входит подслово $ababa$ и α имеет нечетную длину$\}$.

3. **а)** Какова мощность множества всех автоматных множеств слов? **б)** Докажите, что существует множество слов в алфавите $\{a, b\}$, которое не является автоматным.

4. *(Лемма о накачке).* Пусть дан автомат $\mathfrak{A} = \langle \Sigma, Q, s, F, \Delta \rangle$. Обозначим через $\mathbf{L}$ множество слов, распознаваемое автоматом $\mathfrak{A}$ и через c количество состояний автомата $\mathfrak{A}$. **а)** Докажите, что каждое слово $\alpha \in \mathbf{L}$ длины $|\alpha| \geq c$ можно разбить на три части $\alpha = \beta\gamma\delta$ таким образом, что $1 \leq |\gamma| \leq c$ и всякое слово из множества $\{\beta\gamma^n\delta \mid n \in \mathbb{N}\}$ лежит в $\mathbf{L}$. Докажите, что дополнительно в пункте а) можно наложить условие **б)** $|\beta\gamma| < c$ или наложить условие **в)** $|\gamma\delta| < c$.

5. Автоматны ли следующие множества:
а) множество $\{a^{2n+5} \mid n \in \mathbb{N}\}$;
б) множество $\{a^{n^2} \mid n \in \mathbb{N}\}$;
в) множество $\{a^n b^m \mid n, m \in \mathbb{N}\}$;
г) множество $\{a^n b^{m+n} \mid n, m \in \mathbb{N}\}$;
д) множество правильных скобочных последовательностей в алфавите $\{[,]\}$;
е) множество правильных скобочных последовательностей глубины вложенности 3 в алфавите $\{[,]\}$?

6. Какие из пунктов задачи 4 могут выполняться для следующих $\mathbf{L}$ при подходящем выборе c:
а) $\mathbf{L}$ состоит из всех слов α в алфавите $\{a, b\}$ таких, что в α букв a меньше, чем букв b;

б) $\mathbf{L} = \{a^{n+1}b^{m^2} \mid n, m \in \mathbb{N}\}$?

7. Пусть множества слов $\mathbf{L}$ и $\mathbf{K}$ в алфавите Σ автоматны. Докажите, что следующие множества слов также автоматны:
а) $\mathbf{L} \cup \mathbf{K}$;
б) $\mathbf{L} \cap \mathbf{K}$;
в) $\mathbf{L} \cdot \mathbf{K} = \{\alpha\beta \mid \alpha \in \mathbf{L}, \beta \in \mathbf{K}\}$;
г) $\mathbf{L}^\star = \{\alpha_1\alpha_2\ldots\alpha_n \mid n \in \mathbb{N}$ и $\alpha_1, \ldots, \alpha_n \in \mathbf{L}\}$ (операция $(\cdot)^\star$ известна как звезда Клини).

Автомат $\mathfrak{A} = \langle \Sigma, Q, s, F, \Delta \rangle$ называется *детерминированным*, если для любого $q \in Q$ и $a \in \Sigma$ найдется единственное $q' \in Q$ такое, что $\langle a, q, q' \rangle \in \Delta$. Автомат называется *недетерминированным*, если он не является детерминированным.

8. Докажите, что детерминированными автоматами распознаются все множества слов из задачи 2.

9. Пусть множества слов $\mathbf{L}, \mathbf{K}$ в алфавите Σ распознаются детерминированными автоматами. Покажите, что детерминированными автоматами распознаются множества слов:
а) $\mathbf{L} \cup \mathbf{K}$;
б) $\mathbf{L} \cap \mathbf{K}$;
в) $\mathbf{L} \cdot \mathbf{K}$;
г) $\overline{\mathbf{L}}$ — множество всех слов в алфавите Σ, не лежащих в $\mathbf{L}$.

10. Пусть дан автомат $\mathfrak{A} = \langle \Sigma, Q, s, F, \Delta \rangle$.

а) Рассмотрим алфавит $\Theta_{\mathfrak{A}} = \Sigma \times Q$. Покажите, что некоторым конечным автоматом распознается множество слов $(a_1, q_1)(a_2, q_2)\ldots(a_n, q_n)$ в алфавите $\Theta_{\mathfrak{A}}$ таких, что некоторый процесс выполнения автомата на слове $a_1a_2\ldots a_n$ соответствует последовательности состояний $s, q_1, q_2, \ldots, q_n$.

б) Рассмотрим алфавит $\Theta_{\mathfrak{A}}^+ = \Sigma \times \mathcal{P}(Q)$ (напомним, что $\mathcal{P}(Q)$ — это множество всех подмножеств множества Q). Покажите, что существует детерминированный автомат $\mathfrak{B}$, который распознает слово $(a_1, R_1)(a_2, R_2)\ldots(a_n, R_n)$ в алфавите $\Theta_{\mathfrak{A}}^+$ тогда и только тогда, когда для каждого $i \leq n$ множество R_i — это множество всех q таких, что существует процесс выполнения автомата $\mathfrak{A}$ на слове $a_1a_2\ldots a_i$, завершающийся состоянием q.

11. Докажите, что по каждому данному недетерминированному автомату $\mathfrak{A}$, можно построить детерминированный автомат $\mathfrak{B}$ такой, что $\mathfrak{A}$ и $\mathfrak{B}$ распознают одно и тоже множество слов.

Регулярное выражение в алфавите Σ — это запись одного из следующих видов:
1) $\emptyset$;
2) ε;
3) символ a из алфавита Σ;
4) $R_1 \cdot R_2$, где R_1 и R_2 — это регулярные выражения в алфавите Σ;
5) $R_1 \cup R_2$, где R_1 и R_2 — это регулярные выражения в алфавите Σ;
6) $R^\star$, где R — это регулярное выражение в алфавите Σ.

Естественным образом, индукцией по построению, регулярным выражениям ставятся в соответствие определяемые ими множества слов:

1) регулярное выражение $\emptyset$ определяет пустое множество слов;

2) регулярное выражение ε определяет множество слов $\{\varepsilon\}$;

3) регулярное выражение вида $a \in \Sigma$ определяет множество слов $\{a\}$;

4) регулярное выражение вида $R_1 \cdot R_2$ определяет множество слов $\mathbf{L}_1 \cdot \mathbf{L}_2$, где $\mathbf{L}_1$ и $\mathbf{L}_2$ — это множества слов определенные выражениями R_1 и R_2, соответственно;

5) регулярное выражение вида $R_1 \cup R_2$ определяет множество слов $\mathbf{L}_1 \cup \mathbf{L}_2$, где $\mathbf{L}_1$ и $\mathbf{L}_2$ — это множества слов определяемые выражениями R_1 и R_2, соответственно;

6) регулярное выражение вида $R^\star$ определяет множество слов $\mathbf{L}^\star$, где $\mathbf{L}$ — это множество слов определяемое выражением R.

12. Постройте регулярные выражения для всех множеств слов из задачи 2.

13. Для всякого n постройте регулярное выражение для множества всех правильных скобочных последовательностей из [и] глубины вложенности не более n.

14. Рассмотрим алфавит состоящий из $\{$a, b, ..., z, «␣», $<, >, /\}$. Постройте регулярное выражение определяющее множество всех корректно составленных документов в следующем (сильно урезанном) варианте языка разметки `HTML`. Документ записан в одну строку. Он должен начинаться с подслова `<html>` и заканчиваться подсловом `</html>`. Далее описываются условия на содержание между первым подсловом `<html>` и последним подсловом `</html>`:

i) должно иметься ровно по одному вхождению следующих подслов: `<head>`, `</head>`, `<title>`, `</title>`, `<body>` и `</body>`;

ii) никаких других вхождений символов $<$ и $>$ быть не должно;

iii) вхождения `</head>`, `</title>` и `</body>` находятся после вхождений `<head>`, `<title>` и `<body>`, соответственно;

iv) вхождения `<title>` и `</title>` находятся между вхождениями `<head>` и `</head>`;

v) вхождение `<body>` находится после вхождениями `</head>`.

Пример:

```
<html> <head><title> example </title> </head><body> //example// </body></html>
```

15. *(Теорема Клини)* Докажите, что **а)** всякое множество слов, определяемое некоторым регулярным выражением является автоматным и **б)*** всякое автоматное множество слов определяется некоторым регулярным выражением.

Указание. В пункте б) докажите, что для каждого автомата $\mathfrak{A} = \langle \Sigma, Q, s, F, \Delta \rangle$ найдется регулярное выражение, соответствующее тому же множеству слов. Для этого, для всяких $q_1, q_2 \in Q$ и $T \subset Q$ покажите, что регулярными выражениями определяются множества слов $\mathbf{L}_{q_1,T,q_2}$, состоящие из всех слов α таких, что для автомата $\mathfrak{A}_{q_1} = \langle \Sigma, Q, q_1, F, \Delta \rangle$ есть процесс выполнения на α, заканчивающийся состоянием q_2 и использующий лишь состояния из множества T во внутренней части процесса выполнения.

Определим *теорию* WS1S. Формулы WS1S используют переменные двух сортов — переменные по натуральным числам $x, y, z, \ldots$ и переменные по конечным множествам натуральных чисел $X, Y, Z, \ldots$. Атомарными формулами WS1S являются формулы вида $Succ(x, y)$ и $x \in Y$, означающие, что y равен $x + 1$ и что натуральное число x лежит в множестве Y. Формулы WS1S строятся из атомарных при помощи булевых связок $\wedge, \vee, \neg, \to$ и кванторов $\forall x$, $\exists x$, $\forall X$, $\exists X$. Истинность формул определяется исходя из интерпретации переменных первого сорта натуральными числами, а переменных второго сорта множествами натуральных чисел.

16. Покажите, что следующие отношения выразимы формулами WS1S:

а) $x < y$;

б) x — это первый элемент Y;

в) $X \subseteq Y$;

г) x четно.

WS1S формула $\varphi(x_1, \ldots, x_n, X_1, \ldots, X_m)$ *задает* множество слов $\mathbf{L}$ в алфавите $\{0,1\}^{n+m}$, состоящее из всех слов $\langle a_1^1, a_2^1, \ldots, a_n^1, b_1^1, \ldots, b_m^1 \rangle \ldots \langle a_1^k, a_2^k, \ldots, a_n^k, b_1^k, \ldots, b_m^k \rangle$ таких, что

i) для каждого i от 1 до n существует единственное натуральное число u_i такое, что $a_i^{u_i} = 1$;

ii) истинна формула $\varphi(u_1, \ldots, u_n, K_1, \ldots, K_m)$, где K_i — это множество всех таких $j \le k$, что $b_i^j = 1$.

17. Пусть множество слов $\mathbf{L}$ в алфавите $\{0,1\}^n$ автоматно и таково, что для всякого слова $\alpha \in \mathbf{L}$ в $\mathbf{L}$ также лежит всякое слово вида $\alpha \langle 0, 0, \ldots, 0 \rangle \ldots \langle 0, 0, \ldots, 0 \rangle$. Покажите, что по $\mathfrak{A}$ можно построить WS1S формулу $\varphi(X_1, \ldots, X_n)$, которая задает $\mathbf{L}$.

18. * Пусть WS1S формула $\varphi(x_1, \ldots, x_n, X_1, \ldots, X_m)$ задает множество слов $\mathbf{L}$ в алфавите $\{0,1\}^{n+m}$. Покажите, что по φ можно построить автомат $\mathfrak{A}$, распознающий множество $\mathbf{L}$.

19. Придумайте метод проверки того, что данный автомат принимает по крайней мере одно слово.

20. * Придумайте метод проверки истинности WS1S формул, в которых все переменные находятся под кванторами.