ДИСКРЕТНАЯ МАТЕМАТИКА

ИУ5 - 4 семестр, 2015 г.

Семинар 13. ЯЗЫКИ И КОНЕЧНЫЕ АВТОМАТЫ

13.1. Алфавит, слово, язык

ОПРЕДЕЛЕНИЕ **Алфавит** — это произвольное **непустое конечное множество** $V = \{a_1, \ldots, a_n\}$, **элементы** которого называют **буквами**, или **символами**.

ОПРЕДЕЛЕНИЕ. Словом, или **цепочкой**, в алфавите V называют произвольный кортеж из множества V^k (k -ой **декартовой степени** алфавита V) для различных $k = 0, 1, 2, \ldots$

Например, если $V = \{a, b, c\}$, то (a), (b), (c), (a, b), (a, b, c), (c, b, a, a, c) и т. д. есть слова в V.

При $k = 0$ получаем **пустой кортеж**, называемый в данном контексте **пустым словом**, или **пустой цепочкой** и обозначаемый λ. Множество всех слов в алфавите V обозначают V^*, а множество всех непустых слов в V — V^+.

Пустое слово λ — это слово, не имеющее символов.

Длину слова w можно понимать как число составляющих это слово букв.

ОПРЕДЕЛЕНИЕ. Языком в алфавите V называется произвольное подмножество множества V^*.

К языкам применимы теоретико-множественные операции объединения, пересечения, разности, симметрической разности, дополнения и т.д.

Соединением (конкатенацией) языков L_1 и L_2 называют язык L_1L_2, состоящий из всех возможных соединений слов xy, в которых слово x принадлежит первому, а слово y — второму языку, т.е.

$$L_1L_2 = \{xy|x \in L_1 \text{ и } y \in L_2\}.$$

Операция соединения языков позволяет определить операцию возведения языка в произвольную натуральную степень: для любого $L \subseteq V^*$ $L^0 = \{\lambda\}$, а для любого $n > 0$ $L^n = L^{n-1}L$.

Итерацией языка L называют объединение всех его степеней:

$$L^* = \bigcup_{n=0}^{\infty} L^n.$$

Рассматривая объединение всех степеней языка L, начиная с первой, получим **позитивную итерацию**

$$L^+ = \bigcup_{n=1}^{\infty} L^n.$$

Теорема 1. Алгебра $\mathcal{L}(V) = (2^{V^*}, \cup, \cdot, \varnothing, \{\lambda\})$ есть **замкнутое полукольцо**.

В **замкнутом полукольце** $\mathcal{L}(V)$ всех языков в алфавите V рассмотрим **подалгебру**, порожденную множеством, состоящим из **пустого языка**, **языка** $\{\lambda\}$ и всех **однобуквенных языков** $\{a\}$, $a \in V$, и **замкнутую относительно итерации**.
Эта подалгебра является **полукольцом с итерацией**, обозначается $\mathcal{R}(V)$, .
Элементы полукольца $\mathcal{R}(V)$ называются **регулярными множествами**, или **регулярными языками**.

ОПРЕДЕЛЕНИЕ. Пусть фиксирован некоторый алфавит V. Тогда:
1) Пустое множество $\varnothing$, множество $\{\lambda\}$ (состоящее из одной пустой цепочки) и множество $\{a\}$ для каждого $a \in V$ является регулярным языком (множеством) в алфавите V.
2) Если P и Q — регулярные языки в алфавите V, то **объединение** $P \cup Q$ и **соединение** PQ — регулярные языки в алфавите V.
3) Если P — регулярный язык в алфавите V, то итерация P^* — регулярный язык в алфавите V.
4) Никаких других регулярных языков, кроме определенных в пп. (1) — (3), не существует.

Алгебраические операции над регулярными множествами удобно представлять с помощью так называемых **регулярных выражений**.

Каждое регулярное выражение представляет (или обозначает) некоторое однозначно определяемое регулярное множество, причем:

1) регулярные выражения $\varnothing$, λ и a обозначают регулярные множества $\varnothing$, $\{\lambda\}$ и $\{a\}$ соответственно $(a \in V)$;

2) если регулярное выражение p обозначает регулярное множество P, а q обозначает Q, то регулярные выражения $(p+q)$, (pq) и (p^*) обозначают регулярные множества $P \cup Q$, PQ и P^* соответственно.

Для регулярного выражения $\alpha\alpha^*$ или $\alpha^*\alpha$ используют обозначение α^+ и называют это выражение позитивной итерацией выражения α.

Задача 1. Доказать, что язык $L^{+k} = \bigcup\limits_{i=k>0}^{\infty} L^i$ регулярен для любого k при условии регулярности L.

Задача 2. Описать множество слов в алфавите $\{a, b, c\}$, которое задается регулярным выражением:
(а) $a^*(b + c)$;
(б) $(a + b)^* c^* (b + c)$.

13.2. Вычисление языка, допускаемого КА

Конечный автомат — это **орграф, размеченный над полукольцом** $\mathcal{R}(V)$ **регулярных языков** в **алфавите** V, с выделенной **вершиной** q_0, которая называется **начальной** и выделенным подмножеством вершин F, каждый элемент которого называется **заключительной** вершиной.

На **функцию разметки** накладываются следующие ограничения: **метка** каждой **дуги** есть либо язык $\{\lambda\}$, либо непустое подмножество **алфавита** V .

Вершины графа называют **состояниями конечного автомата**, начальную вершину — **начальным состоянием**, заключительную вершину — **заключительным состоянием конечного автомата**.

Если $e = (q, r)$ — **дуга** автомата M, и ее метка $\varphi(e)$ есть регулярное выражение λ, то в этом случае будем говорить, что в автомате M возможен **переход из состояния** q **в состояние** r **по пустой цепочке** и писать $q \rightarrow_\lambda r$. Дугу с меткой λ называют λ **-переходом** (или **пустой дугой**).

Если метка дуги e есть множество, содержащее входной символ a говорят, что в автомате M возможен **переход из состояния** q **в состояние** r **по символу** a и пишут $q \rightarrow_a r$.

Метка пути в конечном автомате есть **соединение** меток входящих в этот путь дуг (в порядке их прохождения).

Соединение в полукольце $\mathcal{R}(V)$ есть **умножение полукольца**.

Метка любого **пути конечной длины** в конечном автомате есть регулярный язык.

Метка любого **бесконечного пути** в конечном автомате есть регулярный язык, т.к. каждому контуру, включенному в путь, соответствует итерация регулярного выражения, задающего метку пути по этому контуру.

Если цепочка $x \in \varphi(W)$, где W — некоторый **путь, ведущий из вершины** q **в вершину** r конечного автомата M, то говорят, что **цепочка** x **читается на пути** W в M.
Запись: $q \Rightarrow_x^* r$.

Стоимость прохождения из состояния q в состояние r есть **объединение** меток всех путей ведущих из q в r, т.е. множество всех таких x, что $q \Rightarrow_x^* r$.

Объединение в полукольце $\mathcal{R}(V)$ есть **сложение полукольца**.

Язык $L(M)$ **конечного автомата** M есть множество всех цепочек во входном алфавите, читаемых в M на некотором пути из начального состояния в какое-либо из заключительных.

Чтобы найти язык конечного автомата, надо вычислить сумму (объединение) тех элементов матрицы стоимостей автомата, которые находятся на пересечении строки, соответствующей начальному состоянию q_0 и в столбцов, соответствующих всем заключительным состояниям $q_f \in F$.
Для этого достаточно решить одну систему линейных уравнений:

$$X^j = AX^j + \beta, \tag{13.1}$$

где A - квадратная матрица n -ого порядка, элемент a_{ij} которой является регулярным выражением, служащим меткой дуги из состояния q_i в состояние q_j , если такая дуга существует, и равный $\varnothing$, если нет дуги из q_i в q_j ;
β — столбец, все компоненты которого равны $\varnothing$, кроме компонент с номерами $t_1, \ldots, t_m,$ которые являются номерами заключительных состояний. Эти компоненты равны единице полукольца (λ).

Пример 13.1. Найдем язык конечного автомата, изображенного на рисунке.

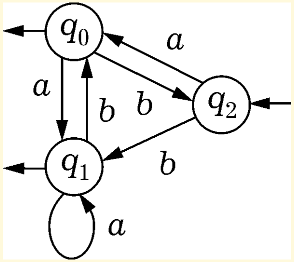

Рис. 1

Запишем для этого автомата систему уравнений:

$$\begin{cases} x_0 = ax_1 + bx_2 + \lambda, \\ x_1 = bx_0 + ax_1 + \lambda, \\ x_2 = ax_0 + bx_1, \end{cases}$$

слагаемые λ добавлено в уравнения для x_0 и x_1, так как вершины q_0 q_1 являются заключительными.

Исключая x_0, получим

$$\begin{cases} x_1 = b(ax_1 + bx_2 + \lambda) + ax_1 + \lambda, \\ x_2 = a(ax_1 + bx_2 + \lambda) + bx_1. \end{cases}$$

Раскроем скобки

$$\begin{cases} x_1 = bax_1 + bbx_2 + b\lambda + ax_1 + \lambda, \\ x_2 = aax_1 + abx_2 + a\lambda + bx_1. \end{cases}$$

$$\begin{cases} x_1 = (ba + a)x_1 + bbx_2 + b\lambda + \lambda, \\ x_2 = (aa + b)x_1 + abx_2 + a\lambda. \end{cases}$$

Отсюда

$$\begin{cases} x_1 = (ba + a)^*(b^2x_2 + b\lambda + \lambda), \\ x_2 = (a^2 + b)(ba + a)^*(b^2x_2 + b\lambda + \lambda) + abx_2 + a\lambda. \end{cases}$$

$$x_2 = (a^2 + b)(ba + a)^*b^2x_2 + abx_2 + (a^2 + b)(ba + a)^*(b\lambda + \lambda) + a\lambda.$$

Получаем регулярное выражение, обозначающее язык КА, как значение переменной x_2:

$$x_2 = ((a^2 + b)(ba + a)^*b^2 + ab)^*((a^2 + b)(ba + a)^*(b\lambda + \lambda) + a\lambda).$$

3. Найти языки, допускаемые конечными автоматами, заданными на рисунках.

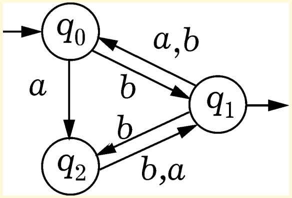

Рис. 2

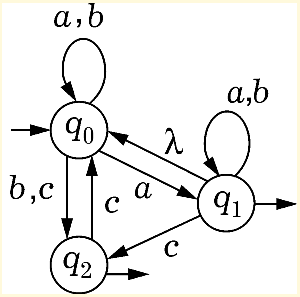

Рис. 3

4. Найти язык, допускаемый конечным автоматом:
(а) вход: q_1; выходы: q_2, q_3; дуги: (q_1, q_2, a), (q_1, q_4, a, b), (q_2, q_4, a, b), (q_3, q_4, λ), (q_4, q_3, a, b), (q_3, q_2, a, b), (q_4, q_2, b);
(б) входы: q_0, q_1; выходы: q_2, q_1; дуги: (q_0, q_2, a, b, c), (q_0, q_1, a), (q_1, q_0, a, b), (q_1, q_2, a, c), (q_2, q_0, c), (q_2, q_2, a, b).

5. Решить систему линейных уравнений с регулярными коэффициентами:

$$\begin{cases} x_1 = (01^* + 1)x_1 + x_2, \\ x_2 = 1x_1 + 00x_3 + 11, \\ x_3 = x_1 + x_2 + \lambda. \end{cases}$$

Для регулярного выражения, задающего компоненту решения x_3, построить допускающий его конечный автомат.

6. Доказать, что линейное уравнение $x = \alpha x + \beta$ с регулярными коэффициентами:

(а) имеет единственное решение при $\lambda \notin \alpha$;

(б) имеет бесконечно много решений при $\lambda \in \alpha$, причем общее решение можно записать в виде $x = \alpha^*(\beta + L)$, где L — произвольный язык.