1 Грамматики и источники

1.1 Переход от грамматики к источнику

Пусть дана грамматика

$$G = (T, N, I, P),$$

такая что $T = \{a, b, \dots\}$, $N = \{A, B, \dots\}$, $I \in N$ и $P = \{A \to aB, A \to a, I \to b\}$.
Построим источник, порождающий тот же язык, что и грамматика. Каждой букве нетерминального алфавита ставится в соответствие вершина источника, причем аксиоме I соответствует начальная вершина q_0, а заключительная вершина q_f вводится отдельно.
Обработка правил вывода:

- Каждому правилу вида $A \to aB$ ставится в соответствие дуга с буквой a из вершины A в вершину B.
- Каждому правилу $A \to a$ – дуга с буквой a, идущая из A в заключительную вершину q_f.
- Каждому правилу вида $A \to B$ ставится в соответствие пустая дуга из A в B.
- Если правило имеет вид $A \to \Lambda$, проводится пустая дуга из вершины A в заключительную вершину q_f.

Пример.
Построим источник, порождающий тот же язык, что и следующая грамматика:

$$\begin{array}{l} I \to A \mid aB \\ A \to bB \mid a \\ B \to \Lambda \end{array}$$

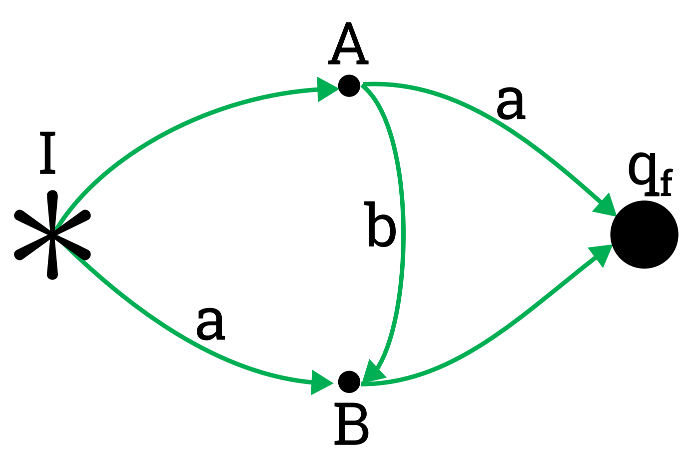

Рис. 1: Источник, порождающий тот же язык, что и грамматика

1.2 Переход от источника к грамматике

Пусть L – язык, порожденный источником **И**.
Построим грамматику, порождающую L.
Начальной вершине поставим в соответствие аксиому I, остальным вершинам $v_1, v_2, \ldots$ – различные буквы $A_1, A_2, \ldots$, которые будут составлять нетерминальный алфавит.
Каждой дуге вида $\bullet \xrightarrow{a} \bullet$ ставится в соответствие правило вывода $A_i \to aA_j$ ($A_i \to A_j$, если это пустая дуга).
Если v_j – заключительная вершина, то добавляем еще и правило $A_i \to a$ ($A_i \to \Lambda$, если ребро пустое).
Если начальная вершина является заключительной, то добавляем правило $I \to \Lambda$.

2 Детерминированные источники

Определение 2.1. *Источник называется детерминированным, если он содержит ровно одну начальную вершину, а из каждой вершины выходит ровно $|A|$ ребер, и всем этим ребрам приписаны разные буквы алфавита A.*

Для любого источника можно построить эквивалентный ему детерминированный источник, порождающий тот же язык.

2.1 Процесс детерминизации источника

Возьмем произвольный источник **И** и построим эквивалентный ему детерминированный источник **Д** таким образом, что вершинам источника **Д** будут поставлены в соответствие некоторые подмножества (будем называть их массивами) множества вершин исходного источника **И**.

1. Перенумеруем вершины **И** таким образом, чтобы множество вершин можно было отождествить с последовательностью $\{1, 2, \ldots, n\}$.

2. В качестве начальной вершины источника **Д** возьмем массив, состоящий из начальных вершин источника **И** и вершин, которые достижимы из этих начальных по путям, составленных из пустых дуг (эти пути порождают пустое слово Λ). Обозначим этот массив через 1^0 и поставим ему в соответствие точку 1^0 на плоскости.

3. Далее, для каждой буквы $a \in A$ рассмотрим всевозможные пути, выходящие из некоторой вершины массива 1^0 и порождающие слово a (если такие существуют). Множество концов таких путей обозначим через 2^0 и поставим ему в соответствие точку 2^0 на плоскости. Из вершины 1^0 проведем дугу с буквой a в вершину 2^0.

   Если же путей с указанным свойством не существует, то введем вершину (и точку на плоскости) f^0, соответствующую пустому множеству вершин источника **И**, и направим ребро с буквой a из 1^0 в f^0.

4. Повторим описанную процедуру для остальных вершин.

   Из вершины k^0 для каждой $a \in A$ проводится ребро a в вершину l^0, образованную номерами вершин, достижимых в **И** из вершин множества k^0 по путям, порождающим слово a. Если таких вершин нет, то из k^0 проводим ребро с буквой a к упомянутой вершине f^0. Если k^0 – пустое множество, то проводятся петли с каждой буквой алфавита A.

5. Заключительными вершинами объявляются массивы $\{i_1, \ldots, i_t\}$, содержащие хотя бы одну заключительную вершину источника **И**.

3 Автоматы

3.1 Определение и способы задания

Рассмотрим алфавит A – входной алфавит, его элементы – входные буквы, алфавит V – выходной алфавит, его элементы – выходные буквы.
Составленные из этих букв слова называются соответственно входными и выходными.
Также вводится Q – множество (алфавит) состояний.
В дальнейшем буквы из A, V, Q будем обозначать, соответственно, через $a, b, \ldots,$ $x, y, \ldots$ и $q_0, q_1, q_2, \ldots$.
Чтобы определить автомат, нужно задать, в каких состояниях автомат будет находиться в зависимости от поступившей последовательности входных букв, и какие выходные буквы будет выдавать.

Определение 3.1. *Конечным неинициальным автоматом называется пятерка*

$$\mathfrak{A} = (A, Q, V, \varphi, f),$$

где $\varphi : Q \times A \to Q$ – функция переходов, $f : Q \times A \to V$ – функция выходов.

Определение 3.2. *Если выделено начальное состояние q_0, автомат $\mathfrak{A} = (A, Q, V, \varphi, f, q_0)$ называется инициальным.*

Если в момент времени t на вход автомату $\mathfrak{A}$ поступает буква x_t, на выходе получается некоторая буква y_t, и q_t означает состояние автомата в момент времени t, то функционирование автомата может быть задано следующей системой уравнений:

$$\begin{cases} y_t = f(q_{t-1}, x_t), \\ q_t = \varphi(q_{t-1}, x_t) \end{cases}$$

То есть в каждый момент времени t автомат, в зависимости от текущего состояния, выдаёт выходной символ y_t и переходит в новое состояние q_t.
Пример 1.
Дан детерминированный источник. Каждому ребру, помимо буквы входного алфавита, приписана буква из алфавита V. Заключительные вершины не учитываются. Такой источник можно считать конечным инициальным автоматом. Вершины называются состояниями автомата.

Автомат можно представить в виде псевдографа, такое отображение называют диаграммой Мура.

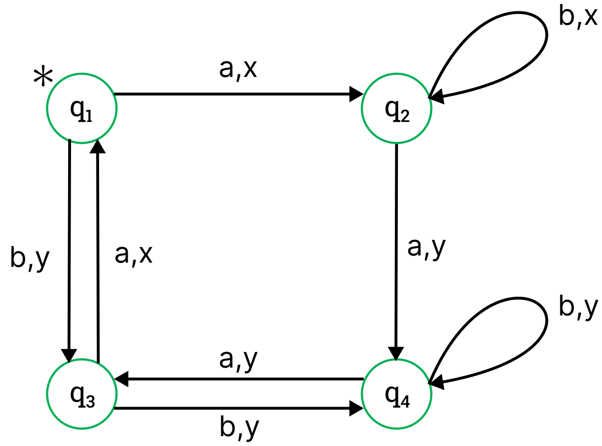

Рис. 2: Диаграмма Мура для автомата

Также распространен способ задания в виде таблицы, в которой задаются функции φ и f, при этом в строки отвечают за буквы входного алфавита, а столбцы за состояния.

	q_1	q_2	q_3	q_4
a	q_2, x	q_4, y	q_1, x	q_3, y
b	q_3, y	q_2, x	q_4, y	q_4, y

Пример 2.
Пусть

$$\mathfrak{A} = (\{0, 1\}, \{0, 1\}, \{0, 1\}, \varphi, f),$$

где $\varphi(q, a) = a$ и $f(q, a) = q$.

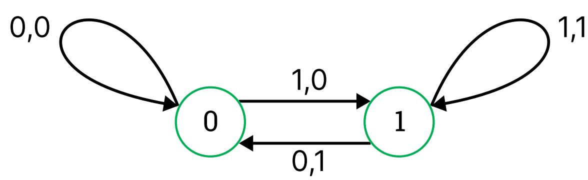

Рис. 3: Диаграмма Мура для автомата

	$q_1 = 0$	$q_2 = 1$
$a = 0$	0, 0	0, 1
$a = 1$	1, 0	1, 1

3.2 Отображение языков, задаваемое автоматом

Пусть дан инициальный автомат. Возьмем произвольное слово $\alpha \in A^*$, $\alpha = \alpha[1]\alpha[2]\dots\alpha[r]$.
Рассмотрим в диаграмме Мура путь, начинающийся в начальной вершине q_0, дальше ведущий по дуге $\alpha[1]$, потом по $\alpha[2]$ и так далее до $\alpha[r]$.
Выпишем последовательно буквы выходного алфавита, соответствующие этим дугам

$$\beta[1]\beta[2]\dots\beta[r] \in V^*.$$

Получится отображение

$$\Phi : A^* \to V^*,$$

$\Phi(\alpha[1]\dots\alpha[r]) = \beta[1]\dots\beta[r]$, в частности, $\Phi(\Lambda) = \Lambda$.
Пусть начальное состояние автомата из примера 2 $q = 0$. Подадим в автомат последовательность символов из входного алфавита $s_A = (0, 1, 0, 0, 1, 1, 0)$, тогда последовательность выходных символов примет вид $s_V = (0, 0, 1, 0, 0, 1, 1)$. Видно, что этот автомат осуществляет задержку входной буквы на один такт.

Определение 3.3. *Два автомата с одними и теми же входными и выходными алфавитами называются эквивалентными, если задаваемые ими отображения совпадают.*

3.3 Распознавание автоматом языка

Определение 3.4. *Автомат $\mathfrak{A} = (A, Q, V, \varphi, f)$ распознает язык L с помощью выделенного символа выходного алфавита $y \in V$ следующим образом*

$$\alpha \in L \Leftrightarrow \Phi(\alpha) = \beta y,$$

то есть последний символ выходного слова есть y.

4 Автомат Мура

Для распознавания языков удобно использовать автомат следующего вида: для каждого состояния $q_i \in Q$ на всех ребрах, входящих в q_i, написана одна и та же буква выходного алфавита V (V-метка).
Диаграмму в этом случае можно изображать проще: состояния изображаются в виде круга, внутрь которого вписана V-метка. На ребрах выходные метки не отображаются.

Теорема 4.1. *Для любого автомата можно построить эквивалентный ему муровский автомат.*

Доказательство. Рассмотрим состояние q_i и ребра, в него входящие.
Возможны два случая:

1. На всех входящих ребрах V-метки одинаковые.

   Назовем такое состояние простым. Тогда стираем метки на ребрах и пишем это выходное значение в состоянии. Таким образом, простому состоянию обычного автомата соответствует одно состояние муровского автомата.

2. V-метки разные. Назовем такое состояние сложным.

   Пусть число различных V-меток, написанных на ребрах, входящих в вершину q_i, равно k.

   Тогда в муровском автомате состоянию q_i будет соответствовать ровно k состояний. В каждом из k состояний своя V-метка, и в него ведут ребра, входившие в q_i в исходном автомате с этой V-меткой.

   При этом q_i стираем, а ребра, выходившие из него, выводим из каждой новой вершины.

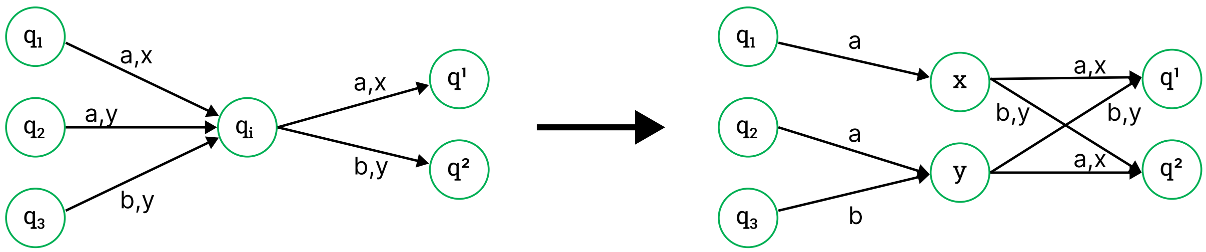

Рис. 4: Пример расщепления

Если в сложном состоянии есть петля, расщепление происходит так:

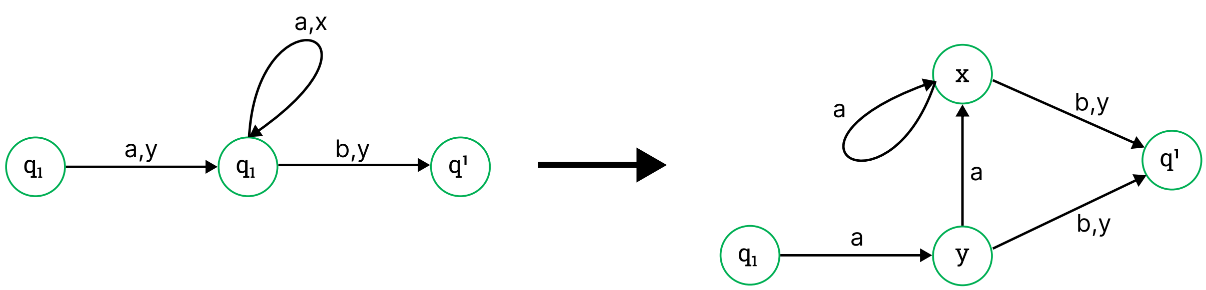

Рис. 5: Пример расщепления с петлёй

Эта процедура проделывается для каждого состояния. В итоге получается муровский автомат, эквивалентный данному. □