Лабораторная работа №3

Недетерминированные магазинные автоматы.

Теоретическая часть

Модель магазинного автомата (рис.1) состоит из

- входной ленты,
- устройства управления и
- вспомогательной ленты, называемой магазином или стеком.

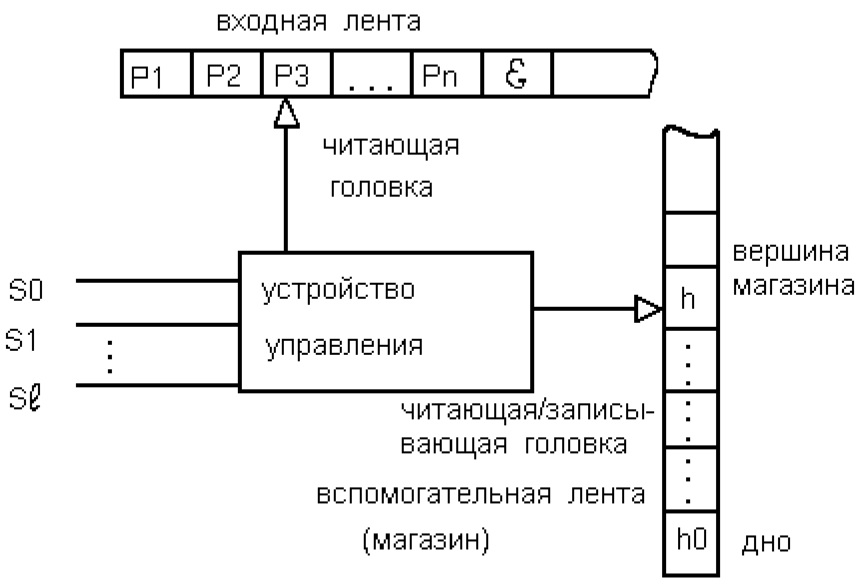


Рис.1. Модель магазинного автомата

Входная лента разделяется на клетки (позиции), в каждой из которых может быть записан символ входного алфавита. При этом предполагается, что в неиспользуемых клетках входной ленты расположены пустые символы λ.

Вспомогательная лента также разделена на клетки, в которых могут располагаться символы магазинного алфавита. Начало вспомогательной ленты называется дном магазина. Связь устройства управления с лентами осуществляется двумя головками, которые могут перемещаться вдоль лент.

Головка входной ленты может перемещаться только в одну сторону - вправо или оставаться на месте. Она может выполнять только чтение. Головка вспомогательной ленты способна выполнять как чтение, так и запись, но эти операции связаны с перемещением головки определенным образом:

- при записи головка предварительно сдвигается на одну позицию вверх, а затем символ заносится на ленту,
- при чтении символ, находящийся под головкой считывается с ленты, а затем головка сдвигается на одну позицию вниз, т.о. головка всегда установлена против последнего записанного символа. Позицию, находящуюся в рассматриваемый момент времени под головкой, называют вершиной магазина.

Определение. Магазинный автомат **M** определяется следующей совокупностью семи объектов: **$M=\{S, P, Z, \delta, s_0, h_0, F\}$,** где

S - алфавит состояний,

P - входной алфавит,

Z - алфавит магазинных символов, записываемых на вспомогательную ленту,

δ - функция переходов, $\delta : \{S \times P \cup \{\lambda\} \times H\} \cup \square S \times H^*$, если М-автомат - детерминированный, и $\delta : \{S \times P \cup \{\lambda\} \times H\} \cup \square 2^{S \times H^*}$, если М-автомат - недетерминированный.

s_0 - начальное состояние, $s_0 \in S$.

h_0- маркер дна, он всегда записывается на дно магазина , $h_0 \in H$.

F - множество конечных состояний. **F** является подмножеством **S**.

Функция **δ** отображает тройки (p_i , s_j , h_k) в пары (s_r , υ) , где $\upsilon \in H^*$ и h_k - символ в вершине магазина, для детерминированного автомата или в множество таких пар для недетерминированного автомата.

Эта функция описывает изменение состояния магазинного автомата, происходящее при чтении символа с входной ленты и перемещении входной головки.

В дальнейшем при построении магазинных автоматов потребуются две разновидности функций переходов:

1) функция переходов с пустым символом в качестве входного символа:

$\delta^0(s, \lambda, h) = (s', \beta)$, которая, независимо от того, какой символ находится под читающей головкой входной ленты, предписывает прочитать символ h из вершины магазина, изменить состояние автомата на s' и записать цепочку β □в магазин.

2) функция переходов с определенным входным символом:

$\delta (s, a, \square\square h) = (s', \beta)$, которая предписывает прочитать входной символ a, изменить состояние автомата на s’ и заменить верхний символ магазина h цепочкой β.

Работа магазинного автомата

Чтобы описать, как работает автомат, введем понятие конфигурации.

Определение. Конфигурацией автомата **M** называют тройку $(s, \alpha, \gamma) \in \mathbf{S \times P^* \times H^*}$, где

s - текущее состояние управляющего устройства,

α - неиспользованная часть входной цепочки $\alpha \in P^*$, самый левый символ этой цепочки находится под головкой. Если $a = \lambda$, то считается, что вся входная цепочка прочитана.

γ - цепочка, записанная в магазине, $\gamma \in H^*$, самый правый символ цепочки считается вершиной магазина. Если $\gamma=\lambda$, то магазин пуст.

Работа автомата может быть представлена как смена конфигураций. Один такт работы автомата заключается в определении новой конфигурации по заданной. Это записывается так:

$(s, a\alpha, \gamma h) \vdash (s', \alpha, \gamma\beta)$

Такая смена конфигураций возможна, если функция **δ(s, a, h)** (или **δ(s, λ, h)**) определена и имеет значение **(s', β).** При этом предполагается, что автомат

- читает символ a, находящийся под головкой. Или не читает ничего, в случае входного символа λ.
- определяет новое состояние s'
- читает символ h, находящийся в вершине магазина и
- записывает цепочку β в магазин вместо символа h. Если $\beta = \lambda$, то верхний символ оказывается удаленным из магазина.

Таким образом, могут быть три случая при работе автомата:

- **δ(s, a, h)** определена и выполняется такт работы,
- **δ(s, a, h)** не определена, но определена функция **δ(s, λ, h)** и выполняется пустой такт (без чтения входной информации).
- функции **δ(s, a, h)** и **δ(s, λ, h)** не определены, в этом случае дальнейшая работа автомата невозможна.

Начальной конфигурацией называется конфигурация (s_0, α, h_0), где s_0 - начальное состояние, α - исходная цепочка, h_0 - маркер дна магазина.
Заключительная конфигурация – (s, λ, λ), где s принадлежит множеству заключительных состояний F.

Для обозначения последовательности сменяющих друг друга конфигураций условимся использовать знак $\vdash^*$. Таким образом последовательность

$$(s_1, \alpha_1, \gamma_1) \vdash (s_2, \alpha_2, \gamma_2) \vdash \ldots \vdash (s_n, \alpha_n, \gamma_n)$$

записывается в сокращенном виде как:

$$(s_1, \alpha_1, \gamma_1) \vdash^* (s_n, \alpha_n, \gamma_n).$$

Язык, допускаемый магазинным автоматом

Определение. Цепочка α называется **допустимой** для автомата M, если существует последовательность конфигураций, в которой первая конфигурация является начальной с цепочкой α, а последняя – заключительной. (s_0, α, h_0) $\vdash^*$ (s_1, λ , λ) , где $s_1 \in F$.
Определение. Множество цепочек, допускаемых автоматом M, называется языком, **допускаемым** или определяемым автоматом M, и обозначается **L(M)**.

$$\mathbf{L(M)= \{\alpha \mid (s_0, \alpha, h_0) \vdash^* (s', \lambda, \lambda) \text{ и } s' \in F \}}$$

Чтобы лучше представить себе работу магазинного автомата, рассмотрим два примера. Пусть задан магазинный автомат M_1 в следующем виде:

$$M_1: P = \{a , b\}; S = \{s_0 , s_1 , s_2\}; Z = \{h_0 , a\}; F = \{s_0\};$$

$$\delta (s_0 , a , h_0) = (s1 , h_0a),$$
$$\delta (s_1 , a , a) = (s_1 , aa),$$
$$\delta (s_1 , b , a) = (s_2 , \lambda),$$
$$\delta (s_2 , b , a) = (s_2 , \lambda),$$
$$\delta (s_2 , \lambda , h_0) = (s_0 , \lambda).$$

Этот автомат является **детерминированным**, поскольку каждому набору аргументов соответствует единственное значение функции. Работу автомата при распознавании входной цепочки aabb можно представить в виде последовательности конфигураций:

$(s_0,aabb,h_0) \vdash (s_1,abb,h_0a) \vdash (s_1,bb,h_0aa) \vdash (s_2,b,h_0a) \vdash (s_2,\lambda,h_0) \vdash (s_0,\lambda,\lambda)$.

Нетрудно проверить, что при задании входной цепочки aabbb автомат не сможет закончить работу. Следовательно, эта цепочка не принадлежит языку, допускаемому автоматом M_1.

Магазинный автомат M_2, заданный следующим описанием:

$$M_2: P = \{a , b\}; S = \{s_0, s_1 , s_2\}; Z = \{h_0, a , b\}; F = \{s_2\};$$

(1) $\delta (s_0 , a , h_0) = (s_0, h_0a)$,
(2) $\delta (s_0 , b , h_0) = (s_0, h_0b)$,
(3) $\delta (s_0 , a , a) = \{(s_0,aa) , (s_1 , \lambda)\}$,
(4) $\delta (s_0 , b , a) = (s_0,ab)$,
(5) $\delta (s_0 , a , b) = (s0 , ba)$,
(6) $\delta (s_0 , b , b) = \{(s_0 , bb) , (s_1 , \lambda)\}$,
(7) $\delta (s_1 , a , a) = (s_1, \lambda)$,
(8) $\delta (s_1 , b , b) = (s_1, \lambda)$,
(9) $\delta (s_2 , \varepsilon , h_0) = (s_2 , \lambda)$,

является **недетерминированным** автоматом, поскольку одному и тому же набору аргументов, например (s_o , a, a), соответствуют два значения функции. Работу автомата рассмотрим для входной цепочки abba. Если использовать последовательность команд (1),(4),(6.1),(5), то получим последовательность конфигураций:

$(s_0,abba,h_0)$
$\vdash (s_0,bba,h_0a),$ (1)
$\vdash (s_0,ba,h_0ab),$ (4)
$\vdash (s_0,a,h_0abb),$ (6.1)
$\vdash (s_0,\lambda,h_0abba).$ (5)

которая показывает, что дальнейшая работа невозможна, т.к. входная цепочка прочитана и переход (s_0,λ,h_0abba) не определен. Если же использовать последовательность команд (1),(4),(6.2),(3),(9), то получим заключительную конфигурацию:

$(s_0,abba,h_0)$
$\vdash (s_0,bba,h_0a),$ (1)
$\vdash (s_0,ba,h_0ab),$ (4)
$\vdash (s1,a,h_0a),$ (6.2)
$\vdash (s_1,\lambda,h_0),$ (3)
$\vdash (s2,\lambda,\lambda)$. (9).

Т.о. можно сделать вывод о том, что цепочка abba допускается автоматом M_2.

Построение магазинного автомата

Пусть задана грамматика G = {VN, VT, I, P}. Определим компоненты автомата M следующим образом:

$$\mathbf{S = \{s_0\}, P = VT, Z = VN \cup VT \cup \{h_0\} , F = \{s_0\},}$$

в качестве начального состояния автомата примем s_0 и построим функцию переходов так:

1. Для всех **A ∈ VN** таких, что встречаются в левой части правил $\mathbf{A \rightarrow \alpha}$, построим команды вида:

   $\boldsymbol{\delta^0(s_0, \lambda, A) = (s_0, \alpha^R)},$

   где α^R – реверс цепочки α .
2. Для всех **a ∈VT** построим команды вида

   $\boldsymbol{\delta (s_0, a, a) = (s_0, \lambda)}$
3. Для перехода в конечное состояние построим команду

   $\boldsymbol{\delta (s_0, \lambda, h_0) = (s_0, \lambda)}$
4. Начальную конфигурацию автомата определим в виде:

   $\boldsymbol{(s_0, \omega, h_0 I)},$

   где ω – исходная цепочка, записанная на входной ленте.

Автомат, построенный по приведенным выше правилам, работает следующим образом. Если в вершине магазина находится терминал, и символ, читаемый с входной ленты, совпадает с ним, то по команде типа (2) терминал удаляется из магазина, а входная головка сдвигается. Если же в вершине магазина находится нетерминал, то выполняется команда типа (1), которая вместо терминала записывает в магазин цепочку, представляющую собой правую часть правила грамматики. Следовательно, автомат, последовательно заменяя нетерминалы, появляющиеся в вершине магазина, строит в магазине левый вывод входной цепочки, удаляя полученные терминальные символы, совпадающие с символами входной цепочки. Это означает, что каждая цепочка, которая может быть получена с помощью левого вывода в грамматике G, допускается построенным автоматом M.

Пример построения автомата

Процедуру построения автомата рассмотрим на примере грамматики с правилами:

$$\mathbf{E \rightarrow E + T \mid T}$$
$$\mathbf{T \rightarrow T * F \mid F}$$
$$\mathbf{F \rightarrow (E) \mid a}$$

Для искомого автомата имеем:

$P = \{a, +, *,), (\}, Z = \{E, T, F, a, +, *) , h_0, (\}, S = \{s_0 \}, F = \{s_0\}$

Для всех правил грамматики строим команды типа (1):

(1) $\delta^0 (s_0 , \varepsilon , E) = \{(s_0 , T{+}E) ; (s_0 , T)\},$

(2) $\delta^0 (s_0 , \lambda , T) = \{(s_0 , F{*}T) ; (s_0 , F)\},$

(3) $\delta^0 (s_0 , \lambda , F) = \{(s_0 , (E)) ; (s_0 , a)\},$

Для всех терминальных символов строим команды типа (2):

(4) $\delta (s_0, a, a) = (s_0, \lambda),$

(5) $\delta (s_0 , + , +) = (s_0 , \lambda),$

(6) $\delta (s_0 , * , *) = (s_0 , \lambda),$

(7) $\delta (s_0 , (, ( ) = (s_0 , \lambda),$

(8) $\delta (s_0 ,) ,)) = (s_0 , \lambda),$

Для перехода в конечное состояние построим команду:

$$(9)\ \delta(s_0, \lambda, h_0) = (s_0, \lambda).$$

Построенный автомат является недетерминированным.

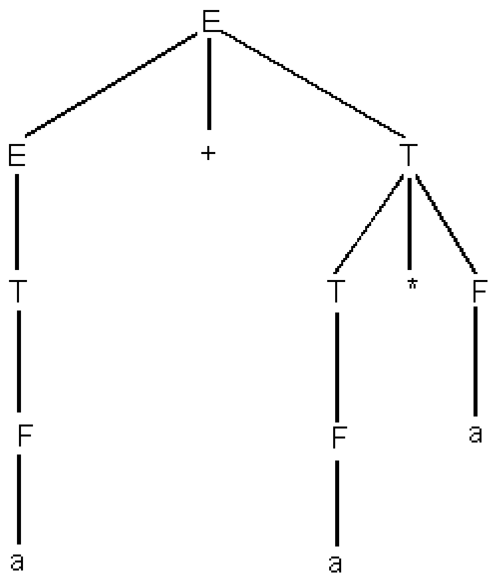
Начальную конфигурацию с цепочкой a + a*a запишем так: $(s_0, a+a*a, h_0E)$.
Последовательность тактов работы построенного автомата, показывающая, что заданная цепочка допустима, имеет вид:
$(s_0, a+a*a, h_0E) \vdash (s_0, a+a*a, h_0T+E) \vdash (s_0, a+a*a, h_0T+T) \vdash (s_0, a+a*a, h_0T+F) \vdash$
$(s_0, a+a*a, h_0T+a) \vdash (s_0, +a*a, h_0T+) \vdash (s_0, a*a, h_0T) \vdash (s_0, a*a, h_0F*T) \vdash$
$(s_0, a*a, h_0F*F) \vdash (s_0, a*a, h_0F*a) \vdash (s_0, *a, h_0F*) \vdash (s_0, a, h_0F) \vdash$
$(s_0, a, h_0a) \vdash (s_0, \lambda, h_0) \vdash (s_0, \lambda, \lambda)$.

Отметим, что последовательность правил, используемая построенным автоматом, соответствует левому выводу входной цепочки:

$E \Rightarrow E+T \Rightarrow T+T \Rightarrow F+T \Rightarrow a+T \Rightarrow a+T*F \Rightarrow a+F*F \Rightarrow a+a*F \Rightarrow$ □a+a*a.

Если по такому выводу строить дерево, то построение будет происходить сверху вниз, т.е. от корня дерева к листьям.

Такой способ построения дерева по заданной цепочке называется **нисходящим**.



Магазинные автоматы называют часто распознавателями, поскольку они определяют, является ли цепочка, подаваемая на вход автомата, допустимой или нет, и следовательно, отвечают на вопрос, принадлежит ли эта цепочка языку, пораждаемому грамматикой, использованной для построения автомата.

Учитывая характер построения вывода в магазине, автоматы рассмотренного типа называют **нисходящими распознавателями**.

Задание на лабораторную работу

Написать программу, реализующую работу недетерминированного магазинного автомата.

Входные данные:

1. Текстовый файл с описанием грамматики, для которой строится магазинный автомат.

   Каждая строка в файле может задавать несколько правил грамматики с одинаковой левой частью. В этом случае правила отделяются друг от друга символом ‘|’.

   Для отделения левой части правила от правой используется символ ‘>’. В одной строке может быть несколько символов ‘>’. В этом случае первый из них трактуется, как символ, отделяющий правую и левую части продукции, все последующие – как терминальный символ.

   Все нетерминалы задаются с помощью прописных букв латинского алфавита.

   Все символы, не описанные выше, являются терминальными символами.

   Правил в грамматике (а значит и во входном файле) – неограниченное количество.

   Рекомендуется считать пробельные символы в файле незначащими, а для терминала «пробел» использовать какой-нибудь другой символ (например, ~, либо заключать пробел в апострофы).

   <u>Пример входного файла</u>

   E > □E + T | T
   T > □T * F | F
   F > (E) | a | ~

2. Строка символов, которую нужно проанализировать с помощью построенного автомата и дать заключение о допустимости (или недопустимости) автоматом данной цепочки символов.

Выходные данные

1. На оценку «удовлетворительно»: заключение о допустимости (или недопустимости) автоматом цепочки символов.
2. На оценку «отлично»: значение множеств P, Z и т.д.; список команд (1) – (4) (см раздел «Построение магазинного автомата»); цепочка конфигураций магазинного автомата, полученная в процессе его работы; заключение о допустимости (или недопустимости) автоматом цепочки символов.

Литература

При подготовка данной лабораторной работы были использованы материалы сайта http://mf.grsu.by/.