# **Введение**

Основной целью данной курсовой работы является разработка транслятора для языка программирования DVV-2018. Главной задачей для транслятора будет сделать язык DVV-2018 понятным для вычислительной машины. В данном курсовом проекте будет происходить трансляция кода на языке DVV-2018 в код на языке ассемблер.

Для данного курсового проекта были определены следующие задачи:

* разработка спецификации языка программирования;
* разработка структуры транслятора;
* разработка лексического анализатора;
* разработка синтаксического анализатора;
* разработка семантического анализатора;
* обработка выражений;
* генерация кода в Assembler;
* тестирование транслятора.

Решение для каждой из поставленных задач будут отображены в следующих главах курсового проекта:

* спецификация языка программирования;
* структура транслятора;
* лексический анализатор;
* синтаксический анализатор;
* семантический анализатор;
* преобразование выражений;
* генерация кода;
* тестирование транслятора.

Язык программирования DVV-2018 предназначается для выполнения простейших арифметический и логических действий , а также для операций над строками.

**Глава 1 Спецификация языка программирования**

**1.1 Характеристика языка программирования**

Язык DVV-2018 – это процедурный, универсальный, строготипизированный, компилируемый язык. Так же язык не является объектно-ориентированным.

* Универсальный — язык, система терминов, определенных строго и однозначно, а потому допускающих над собой чисто формальные операции.
* Процедурный – язык при программировании на котором последовательно выполняемые операторы можно собрать в подпрограммы, то есть более крупные целостные единицы кода, с помощью механизмов самого языка.
* Сильной(строгой) типизацией в языке является отсуствия приведения типов
* Компилируемый – язык который конвертируется в машинный код.

**1.2 Алфавит языка**

Алфавит языка DVV-2018 основан на кодировке Windows-1251, представленной на рисунке 1.1. Исходный код DVV-2018 может содержать символы латинского алфавита, цифры десятичной системы счисления от 0 до 9, символы русского алфавита можно использовать только в строковых литералах. Код, написанный на языке DVV-2018, может содержать символы : +, -, \*, /, {, }, (, ), ', >, <, =, : .

**1.3 Символы сепараторы**

Символы, которые являются сепараторами в языке DVV-2018 представлены в таблице 1.1.

Таблица 1.1 – Сепараторы

|  |  |
| --- | --- |
| Сепаратор | Назначение |
| (space)  (\n)  .  , | разделение инструкций |
| {  } | программный блок |
| (  ) | параметры/приоритетность операций (в выражениях) |

**1.4 Применяемы кодировки**

Для написания исходного кода на языке программирования DVV-2018 используется кодировка Windows-1251. Кодировка Windows-1251 представлена на рисунке 1.1



Рисунок 1.1 – Алфавит входных символов

**1.5 Типы данных**

В языке DVV-2018 есть 3 типа данных: целочисленный integer, строковый string и логический bool. Описание типов данных, предусмотренных данным языком представлены в таблице 1.2.

Таблица 1.2 – Типы данных языка DVV-2018

|  |  |
| --- | --- |
| Тип данных | Описание типа данных |
| integer | Фундаментальный тип данных. Предусмотрен для объявления целочисленных положительных данных.  Автоматически инициализируется нулевым значением.  Возможные операции:  + – суммирование;  - – вычитание;  \* – умножение;  / – деление;  = – присваивание значения; |
| String | Фундаментальный тип данных. Предусмотрен для объявления строк. Автоматическая инициализируется строкой нулевой длины. Максимальное количество символов в строке – 255. |
| Bool | Фундаментальный тип данных. Предусмотрен для использования в конструкции if() else. Может быть инициализирован 2 значениями truth или false. Автоматическая инициализация не предусмотрена.  Возможные операции:  == , < , > – логические операторы сравнения. |

**1.6 Преобразование типов данных**

Так как язык DVV-2018 является строготипизированным , то по определению

сильной(строгой) типизации преобразование(приведение) типов в языке отсутствует.

**1.7 Идентификаторы**

В языке DVV-2018 для идентификатора можно использовать строчные буквы латинского алфавита. Максимальная длина идентификатора 4096 символов.

**1.8Литералы**

Для языка DVV-2018 характерно присутствие литералов 3 типов: литералы целого типа, строковые литералы, а также литералы типа bool. Описание данных литералов можно прочитать ниже в таблице 1.3.

Таблица 1.3 – Описание литералов

|  |  |
| --- | --- |
| Тип литерала | Описание |
| Литералы целого типа | Целочисленные неотрицательные литералы, инициализируются 0. Литералы только могут быть только rvalue. Диапазон изменения от 0 до 4294967296. |
| Строковые литералы | Символы, заключённые в " " (двойные кавычки), инициализируются пустой строкой, строковые переменные. Литералы могут быть только rvalue. |
| Логические литералы | Логические литералы могут быть инициализированы только двумя значениями truth или false. |

**1.9 Область видимости идентификаторов**

Область видимости «сверху вниз». В языке DVV-2018 требуется обязательное объявление переменной перед её использованием. Все переменные должны располагаться после точки входа в приложение. Возможность объявления переменных c одинаковым идентификаторами отсутствует.

**1.10 Инициализация данных**

Таблица 1.4 – Инициализация данных

|  |  |
| --- | --- |
| Вид инициализации | Примечание |
| <тип данных><идентификатор>; | Автоматическая инициализация: переменные типа integer инициализируются нулём, переменные типа string – пустой строкой, переменные типа bool не имеют автоматической инициализации. |
| <идентификатор> : <значение>; | Присваивание переменной значения. |

**1.11 Инструкции языка**

Все возможные инструкции языка программирования DVV-2018 представлены в общем виде в таблице 1.5.

Таблица 1.5 – Инструкции языка программирования DVV-2018

|  |  |
| --- | --- |
| Инструкция | Запись на языке DVV-2018 |
| Объявление переменной | <тип данных> <идентификатор>; |
| Присваивание | <идентификатор> : <значение>/<идентификатор>; |
| Блок инструкций | [… ] |
| Вывод данных | print (<идентификатор> / <литерал>); |

**1.12 Операции языка**

Язык программирования DVV-2018 может выполнять арифметические и логические операции, представленные в таблице 1.5.

Таблица 1.6 – Приоритетности операций языка программирования DVV-2018

|  |  |
| --- | --- |
| Операция | Приоритетность операции |
| (  ) | 0 |
| , | 1 |
| \*  / | 3 |
| +  - | 2 |
| = , > , < | Данные операции не имеют приоритета |

Максимальным значением приоритетности является “3”, минимальным “0” соответственно.

**1.13 Выражения и их вычисления**

Выражения могут содержать вызов библиотечных функций. Допускаются выражения которые вычисляют значение из операндов. Вычисление в выражения происходит по приоритетности операций языка.

**1.14 Программные конструкции языка**

Ключевые программные конструкции языка программирования DVV-2018 представлены в таблице 1.7.

Таблица 1.7 – Программные конструкции языка DVV-2018

|  |  |
| --- | --- |
| Точка входа в приложение | dvv  [  …  ] |
| Блок if() else | if(проверяемое условие) […]  else […] |

**1.15 Область видимости**

В языке DVV-2018 переменные находятся внутри одной области видимости расположенной внутри блока dvv. Объявление пользовательских областей видимости не предусмотрено.

**1.16 Семантические проверки**

Таблица с перечнем семантических проверок, предусмотренных языком, приведена в таблице 1.8.

Таблица 1.8 – Семантические проверки

|  |  |
| --- | --- |
| Номер | Правило |
| 1 | Тип данных передаваемых значений в функцию стандартной библиотеки должен соответствовать заявленному. |
| 2 | Объявление идентификатора должно происходить до его использования. |
| 3 | В арифметических выражениях операнды должны иметь один тип данных. |

## 

**1.17 Распределение оперативной памяти на этапе выполнения**

Размещение всех переменных происходит в куче. Распределение оперативной памяти происходит на этапе генерации. Таблица лексем и таблица идентификаторов сохраняются в структуры с выделенной динамической памятью, которая очищается по окончанию работы транслятора.

**1.18 Стандартная библиотека и её состав**

Функции стандартной библиотеки с описанием представлены в таблице 1.9. Стандартная библиотека написана на языке программирования C++.

Таблица 1.9 – Состав стандартной библиотеки

|  |  |  |  |
| --- | --- | --- | --- |
| Имя функции | Возвращаемое значение | Принимаемые параметры | Описание |
| Strlen | Integer | str s - строка | Возвращает длину строки s |
| Substr | String | string x – строка,  string y – строка | Возвращает значение либо throw либо false в зависимости от того является ли строка 2 подстрокой строки 1. |

**1.19 Ввод и вывод данных**

Ввод данных не предусмотрен в языке DVV-2018.А за вывод данных отвечает функция: print (<идентификатор или литерал>);

**1.20 Точка входа**

Точкой входа в языке DVV-2018 является – ключевое слово dvv.

**1.21 Препроцессор**

Препроцессор для языка DVV-2018 не предусмотрен.

**1.22 Соглашение о вызовах**

В языке вызов функций происходит по соглашению о вызовах stdcall. Выбор пал на это соглашение о вызовах так как это stdcall является одним из самых популярных соглашениях о вызовах.

Особенности stdcall:

* параметры передаются через стек;
* запись в стек идет справа налево
* очистку стека производит вызываемый код

**1.23 Объектный код**

Код на языке DVV-2018 транслируется в код на языке ассемблера.

**1.24 Классификация сообщений транслятора**

При возникновении ошибки в коде программы на языке DVV-2018 и выявления её транслятором в файл протокола выводится сообщение. Классификация сообщений приведена в таблице 1.10.

Таблица 1.10 – Классификация сообщений транслятора

|  |  |
| --- | --- |
| Интервал | Описание ошибок |
| 100,104 | Ошибки параметров |
| 107-112 | Ошибки открытия и чтения файлов |
| 113-121 | Ошибки лексического анализа |
| 600-607 | Ошибки синтаксического анализа |
| 700-799 | Ошибки семантического анализа |

## **1.25 Контрольный пример**

Пример программы на языке DVV-2018 представлен ниже:

dvv

[

integer x : 20.

integer y : 10.

integer l,z,q,w,m.

string str : 'mystring'.

string mystr : 'mynewstring'.

bool mb.

mb : false.

if (x > y)

[

integer p.

p : x + y.

print (p).

]

else

[

l : strlen(str).

]

print (mb).

z : (x + y) \* (x - y).

bool nb : substr(mystr,str).

print (mystr).

return z.

]

# **Глава 2 Структура транслятора**

## **2.1 Компоненты транслятора их назначение и принципы взаимодействия**

Транслятор – обслуживающая программа, преобразующая исходную программу, предоставленную на входном языке программирования, в рабочую программу, представленную на объектном языке.

Компоненты транслятора:

* Лексический анализатор

1. Выполняет аналитический разбор входной последовательности символов на распознанные группы — лексемы.
2. Входным параметром является исходный код.
3. Выходными параметрами являются таблица лексем и таблица идентификаторов.

* Синтаксический анализатор

1. Выполняет сопоставление линейной последовательности лексем естественного или формального языка с его формальной грамматикой.
2. Входным параметром является таблица лексем.
3. Выходным параметром является дерево разбора или иначе говоря синтаксическое дерево

* Семантический анализатор

1. Выполняет последовательности действий алгоритма автоматического понимания текстов, заключающегося в выделении семантических отношений, формировании семантического представления текстов.
2. Входными параметрами являются таблица идентификаторов и дерево разбора.
3. Выходными параметрами является семантическая структура

* Генератор объектного кода

1. Занимается построением для программы на входном языке эквивалентной машинной программы.
2. Входным параметром является промежуточное представление программы.
3. Выходным параметром является объектный код.

Подробное описание компонентов транслятора можно найти в главах курсового проекта под названием:

* Лексический анализатор
* Синтаксический анализатор
* Семантический анализатор
* Генерация кода.

## **2.2 Перечень входных параметров транслятора**

Перечень входных параметров для языка DVV-2018 представлен в таблице 2.1

Таблица 2.1

|  |  |
| --- | --- |
| Имя входного параметра | Выполняемая функция |
| -in | Указывает на файл из которого транслятор будет брать код |
| -log | Указывает на файл в который будет выводиться протокол работы. |
| -id | Указывает на файл в который будет выводиться таблицу идентификаторов |
| -lx | Указывает на файл в который будет выводиться таблица лексем |
| -rl | Указывает на файл в который будет выводиться дерево разбора или иначе говоря синтаксическое дерево |

## **2.3 Протоколы, формируемые транслятором**

Для языка DVV-2018 предусмотрено формирование трех протоколов представленных в таблице 2.2

Таблица 2.2

|  |  |
| --- | --- |
| Имя входного протокола | Информация которая содержится в протоколе |
| Имяфайла.txt.log | Ошибки, если они есть, иначе будет выводится, что тест выполнен без ошибок. Дата создания протокола.  Перечень входных параметров языка DVV-2018.  Исходные данные: Количество символов, Количество строк. |
| Имяфайла.txt.id | Таблица идентификаторов. Дата создания протокола. |
| Имяфайла.txt.lx | Таблица лексем. Дата создания протокола. |
| Имяфайла.txt.rl | Дерево разбора(Синтаксическое дерево)ю Дата создания протокола. |

# **Глава 3 Лексический анализатор**

## **3.1 Структура лексического анализатора**

Лексический анализатор представляет собой программу, которая транслирует исходную программу в набор символ, которая формирует таблицу лексем и таблицу идентификаторов, записывая в таблицы полученную информацию об идентификаторах и литералах.

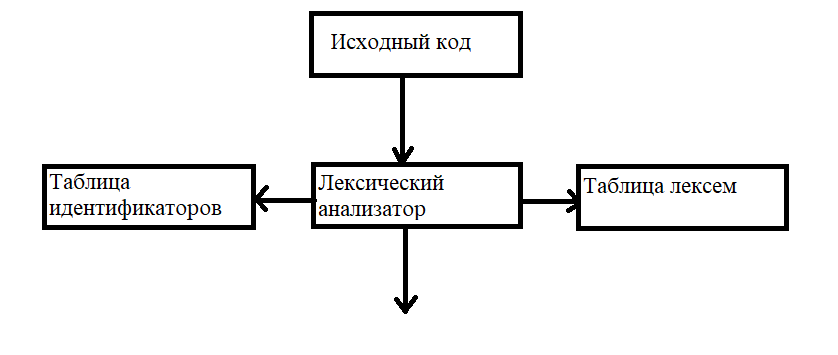


Рис. 3.1 Структура лексического анализатора

## **3.2 Контроль входных символов**

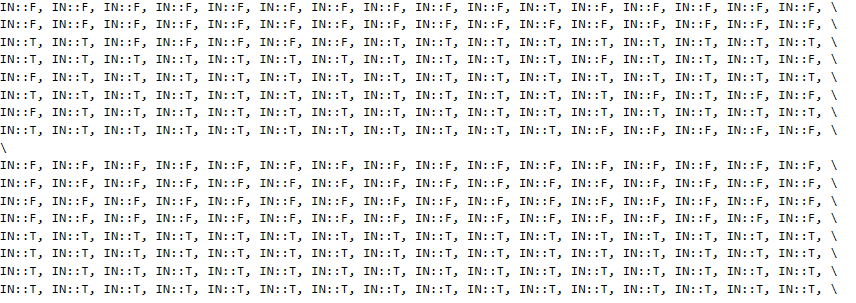


Рис 3.2 Входная таблица символов языка DVV-2018

Таблица контроля входных символов представлена на рис. 3.2. Принцип работы таблицы состоит в следующем: каждому элементу в таблице соответствует значение в 16 с/с такой же, как и таблица ASCII. Допустимые символы на рисунке обозначены буквой - T, запрещенные – F.

**3.3 Удаление избыточных символов**

Алгоритм удаления избыточных символов(пробел и переход на новую строку) происходит путем их игнорирования то есть если при считывании из файла в буфер попадает только избыточный символ то процесс сравнивания содержимого буфера с конечными автоматами пропускается.

## **3.4 Перечень ключевых слов**

Перечень ключевых слов, сепараторов, конечных автоматов языка DVV-2018 представлен в приложении А.

**3.5 Основные структуры данных**

Основными структурами данных лексического анализатора являются таблица лексем и таблица идентификаторов.

Структур таблицы лексем:

struct Entry//строка таблицы лексем

{

char lexema; // лексема

int sn;//номер строки в тексте

int indxTI;//индекс в таблице идентификаторов или LT\_TI\_NULLIDX

};

struct LexTable//экземпяр таблицы лексем

{

int maxsize;//емкость таблицы лексем( < LT\_MAXSIZE)

int size;//текущий размер таблицы лексем ( < maxsize)

Entry\* table;//массив строк таблицы лексем

};

Структуры таблицы идентификаторов:

struct Entry

{

short idxfirstLE;//индекс первой строки в таблице лексем

char id[ID\_MAXSIZE \* 2];//идентификатор

IDDATATYPE iddatatype;//тип данных

IDTYPE idtype;//тип идентификатора

union

{

char vbool[6];//значение bool

int vint;//значение инт

struct

{

int len;//длина строки стр

char str[TI\_STR\_MAXSIZE - 1];//строка

}vstr[TI\_STR\_MAXSIZE];//значение стр

}value;//значение идентификатора

};

struct IdTable//экземпляр таблицы идентификаторов

{

int maxsize;//макс емкость таблицы идентификаторов( < TI\_MAXSIZE)

int size;//текущий размер таблицы идентификаторов ( < maxsize)

Entry\* table;//массив строк таблицы идентификаторов

};

## **3.6. Структура и перечень сообщений лексического анализатора**

Для лексического анализатора языка DVV-2018 в таблице ошибок зарезервированы ошибки с 113 по 121. Структура ошибок одинакова для всех частей транслятора при возникновении ошибки в консоль выводится:

* Номер ошибки
* Сообщение об ошибке
* Строка где возникла ошибка
* Позиция в строке, где возникла ошибка

В таблице 3.1 представлены номера и текст ошибок лексического анализатора.

Таблица 3.1

|  |  |
| --- | --- |
| Номер ошибки | Выводимое сообщение |
| 113 | Превышен максимальный размер таблицы лексем |
| 114 | Превышен максимальный размер таблицы идентификаторов |
| 115 | Недопустимы номер возвращаемой строки таблицы лексем |
| 117 | Отсутствует таблица лексем |
| 118 | Отсутствует таблица идентификаторов |
| 119 | Недопустимый номер возвращаемой строки таблицы идентификаторов |
| 120 | Размер длины строки превышен(256) |
| 121 | Превышено максимальное значение целочисленного литерала |

## **3.7. Принцип обработки ошибок**

При обнаружении в коде ошибки лексический анализатор прекращает свою работу и выводит в консоль одну из ошибок представленных в таблице 3.1. В языке задан лимит на количество ошибок для лексического анализа их количество равно 8.

## **3.8 Параметры лексического анализатора**

Транслятор DVV-2018P допускает использование параметров для управления работой лексического анализатора. Принцип их использования описан в табл. 2.1.

**3.9. Алгоритм лексического анализа**

Алгоритм лексического анализа представлен ниже:

* Из исходно файла подается текст пока в тексте не будет встречен один из символов сепараторов.
* При достижении символа сепаратора буфер усекается на нем.
* Если это был сепаратор который не является пробелом или переходом на другую строку то он добавляется во второй буфер.
* Дальше идет проверка текста записанного в первый буфер на то совпадает он с конечным автоматом.
* При совпадении происходит запись лексемы соответствующей совпавшему автомату в таблицу лексем
* Если совпавший автомат был идентификатором или лексемой он также заносится в таблицу лексем
* После этого производится проверка на совпадение записанного во второй буфер.
* При совпадении происходит запись лексемы соответствующей совпавшему автомату в таблицу лексем.

**3.10 Контрольный пример**

Результат работы Лексического Анализа на основе контрольного примера представленного в пункте 1.25 ниже на Рис. 3.1 и Рис 3.2

Рис 3.2 Таблица Идентификаторов

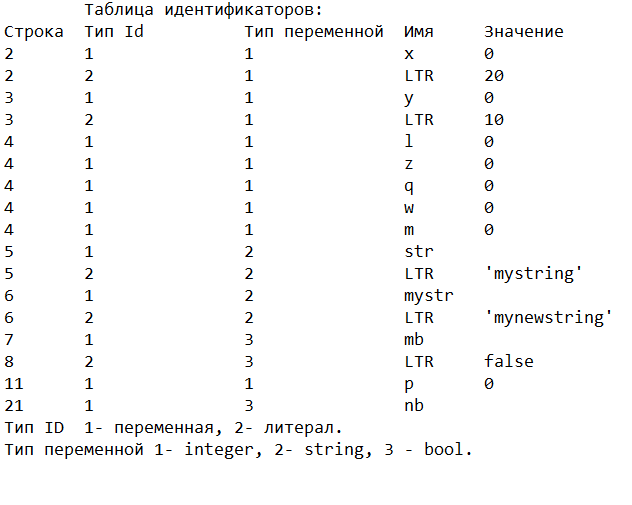
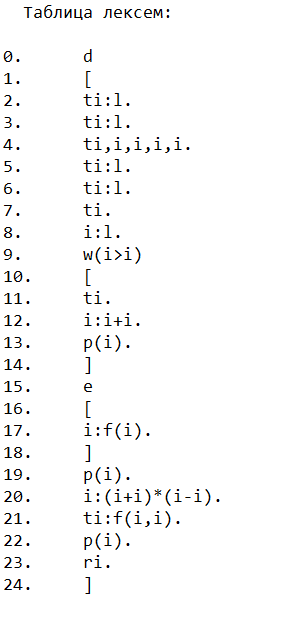


Рис 3.3 Таблица Лексем



# **Глава 4 Синтаксический анализатор**

## **4.1 Структура синтаксического анализатора**

Синтаксический анализатор представляет собой программу, проверяющую таблицу лексем по определенным правилам грамматики Грейбах. При удачной проверке таблицы лексем – формирует протокол работы, представляющий из дерево разбора или синтаксическое дерево.

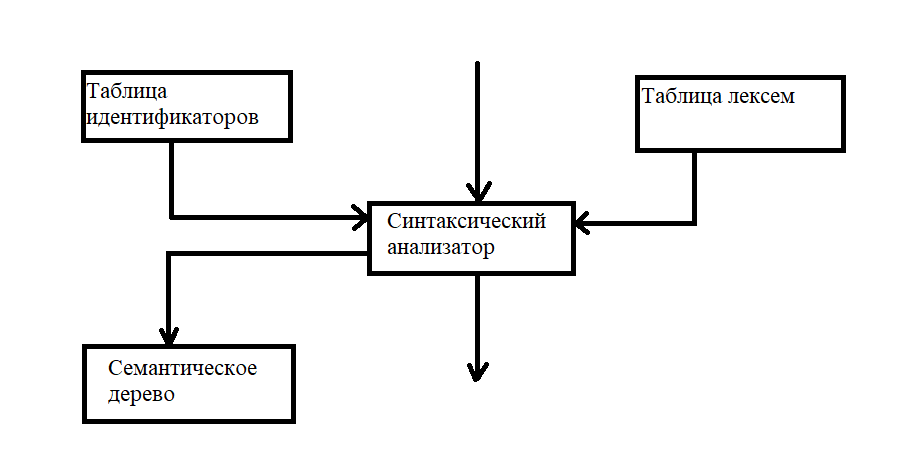


Рис. 4.1 Структура синтаксического анализатора

## **4.2. Контекстно-свободная грамматика, описывающая синтаксис языка**

Грамматика для синтаксического разбора языка DVV-2018 представлена

в таблице 4.1.

Таблица 4.1 – Таблица правил перехода нетерминальных автоматов

|  |  |
| --- | --- |
| Нетерминал | Нетерминал |
| S→ | d[N] |
| N→ | ti:E.N  ti.N  tE.N  ti:P.N  i:P.N  i:E.N  w(K)[C]e[C]N  w(K)[C]N  p(E).N  p(E).  rE.  ri:P. |
| E→ | i  l  i,E  f(E)  f(E)E |
| P→ | i  l  iMP  lMP  (P)  (P)MP  (P)P |
| C→ | i:E.  i:E.C  i:P.  i:P.C  ti:E.C  ti:E.  ti:P.C  ti:P.  ti.C  ti.  p(E).C  p(E). |
| M→ | +  -  \*  / |
| B→ | >  <  = |
| K→ | iBi  lBl  f(E)Bf(E)  f(E)  l |

## **4.3 Построение конечного формального автомата**

Формальное описание конечного магазинного автомата для грамматики представлено на Рис 4.2.

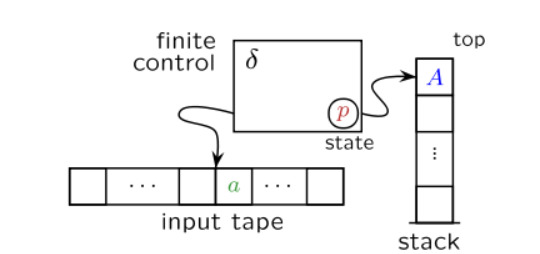


Рис 4.2. Формальное описание конечного магазинного автомата для грамматики

**4.4Основные структуры данных**

Основные структуры данных для синтаксического анализа в языке DVV-2018 будут представлены ниже.