МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
высшего образования

**«Дальневосточный федеральный университет»**

**ШКОЛА ЕСТЕСТВЕННЫХ НАУК**

**Кафедра прикладной математики, механики, управления и программного обеспечения**

РАЗРАБОТКА ТРАНСЛЯТОРА, ПЕРЕВОДЯЩЕГО ПОДМНОЖЕСТВО ЯЗЫКА JAVA В ЭКВИВАЛЕНТНОЕ ПОДМНОЖЕСТВО ЯЗЫКА GO

**КУРСОВОЙ ПРОЕКТ**

по дисциплине «Теория вычислительных процессов и структур»

по образовательной программе подготовки бакалавров по направлению 02.03.03 Математическое обеспечение и администрирование информационных систем

Выполнили студенты гр. Б8117-02.03.03

\_\_\_\_\_\_\_\_\_\_\_ Дейко А.О.

\_\_\_\_\_\_\_\_\_\_\_ Проскурин Д.А.

\_\_\_\_\_\_\_\_\_\_\_ Кокорин И.В..

Руководитель: зав. кафедрой ПММУиПО, д.т.н, профессор И. Л. Артемьева

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.

(подпись)

Защищён с оценкой \_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_

(подпись) И.О. Фамилия

«\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2021 г.

г. Владивосток

2021 г.

**Оглавление**

[Введение 3](#_Toc61983585)

[1 Неформальная постановка задачи 4](#_Toc61983586)

[2 Синтаксис входного языка 9](#_Toc61983587)

[3 Контекстные условия 15](#_Toc61983588)

[4 Таблица соответствия 17](#_Toc61983589)

[5 Проект лексического анализатора 21](#_Toc61983590)

[5.1 Таблица ключевых слов 21](#_Toc61983591)

[5.2 Типы лексем 23](#_Toc61983592)

[5.3 Ошибки лексического анализатора 24](#_Toc61983593)

[5.4 Конечный автомат лексического анализатора 25](#_Toc61983594)

[6 Проект синтаксического анализатора 31](#_Toc61983595)

[Заключение 41](#_Toc61983596)

[Список литературы 42](#_Toc61983597)

# Введение

Программист, написавший программу на каком-либо языке программирования, не сможет сразу запустить свою программу. Одним из способов сделать это является перевод её на язык машины, на которой эта программа будет запущена, иначе говоря, эту программу нужно скомпилировать.

Таким образом компилятор – это программа, которая считывает текст программы, написанной на одном языке (исходном), и транслирует (переводит) его в эквивалентный текст на другом языке (целевом).

***Цель курсового проекта***: разработать проект транслятора, который преобразует программу, содержащую подмножество языка Java в программу, содержащую подмножество языка Go генерируя эквивалентный исходный код. Язык задаётся множеством цепочек, принадлежащих ему;

***Задачи:***

* разработать грамматику подмножества языка Java;
* описать контекстные условия входного языка;
* описать соответствие конструкций входного и выходного языков;
* разработать проект лексического анализатора;
* разработать проект синтаксического анализатора.
* реализовать проект лексического анализатора;
* реализовать проект синтаксического анализатора.

1. Неформальная постановка задачи

Разработать проект транслятора из подмножества языка «Java» в эквивалентное подмножество языка «Go».

Подмножество языка «Java» включает:

1. Типы данных:***byte, short, int, long, float, double, char, boolean.***
2. Операторы:

***Арифметические операторы (***Таблица 1***).***

|  |  |
| --- | --- |
| ***Оператор*** | ***Описание*** |
| + | Складывает значения по обе стороны от оператора |
| - | Вычитает правый операнд из левого операнда |
| \* | Умножает значения по обе стороны от оператора |
| / | Оператор деления делит левый операнд на правый операнд |
| % | Делит левый операнд на правый операнд и возвращает остаток |
| ++ | Инкремент - увеличивает значение операнда на 1 |
| -- | Декремент - уменьшает значение операнда на 1 |

Таблица -Арифметические операторы

***Операторы сравнения (Таблица 2).***

|  |  |
| --- | --- |
| ***Оператор*** | ***Описание*** |
| == | Проверяет, равны или нет значения двух операндов, если да, то условие становится истинным |
| != | Проверяет, равны или нет значения двух операндов, если значения не равны, то условие становится истинным |
| > | Проверяет, является ли значение левого операнда больше, чем значение правого операнда, если да, то условие становится истинным |
| < | Проверяет, является ли значение левого операнда меньше, чем значение правого операнда, если да, то условие становится истинным |
| >= | Проверяет, является ли значение левого операнда больше или равно значению правого операнда, если да, то условие становится истинным |
| <= | Проверяет, если значение левого операнда меньше или равно значению правого операнда, если да, то условие становится истинным |

Таблица -Операторы сравнения

***Побитовые операторы (Таблица 3).***

|  |  |
| --- | --- |
| ***Оператор*** | ***Описание*** |
| & (побитовое и) | Бинарный оператор AND копирует бит в результат, если он существует в обоих операндах. |
| | (побитовое или) | Бинарный оператор OR копирует бит, если он существует в любом из операндов. |
| ^ (побитовое логическое или) | Бинарный оператор XOR копирует бит, если он установлен в одном операнде, но не в обоих. |
| ~ (побитовое дополнение) | Бинарный оператор дополнения и имеет эффект «отражения» бит. |
| << (сдвиг влево) | Бинарный оператор сдвига влево. Значение левых операндов перемещается влево на количество бит, заданных правым операндом. |
| >> (сдвиг вправо) | Бинарный оператор сдвига вправо. Значение правых операндов перемещается вправо на количество бит, заданных левых операндом. |
| >>> (нулевой сдвиг вправо) | Нулевой оператор сдвига вправо. Значение левых операндов перемещается вправо на количество бит, заданных правым операндом, а сдвинутые значения заполняются нулями. |

Таблица -Побитовые операторы

***Логическое операторы (Таблица 4).***

|  |  |
| --- | --- |
| ***Оператор*** | ***Описание*** |
| && | Называется логический оператор «И». Если оба операнда являются не равны нулю, то условие становится истинным |
| || | Называется логический оператор «ИЛИ». Если любой из двух операндов не равен нулю, то условие становится истинным |
| ! | Называется логический оператор «НЕ». Использование меняет логическое состояние своего операнда. Если условие имеет значение true, то оператор логического «НЕ» будет делать false |

Таблица -Логические операторы

***Операторы присваивания (Таблица 5).***

|  |  |
| --- | --- |
| ***Оператор*** | ***Описание*** |
| = | Простой оператор присваивания, присваивает значения из правой стороны операндов к левому операнду |
| += | Оператор присваивания «Добавления», он присваивает левому операнду значения правого |
| -= | Оператор присваивания «Вычитания», он вычитает из правого операнда левый операнд |
| \*= | Оператор присваивания «Умножение», он умножает правый операнд на левый операнд |
| /= | Оператор присваивания «Деление», он делит левый операнд на правый операнд |
| %= | Оператор присваивания «Модуль», он принимает модуль, с помощью двух операндов и присваивает его результат левому операнду |

Таблица -Операторы присваивания

1. Операторы циклов: цикл for, цикл do, цикл while.
2. Операторы ветвления: if else
3. Стандартные функции:

double min( a, b ) – возвращает минимум из a и b

double log( a ) – возвращает натуральный логарифм аргумента a

double pow( a, b ) – возвращает значение a, возведенное в степень b

double sqrt( a ) – возвращает квадратный корень из a

double time() – возвращает текущее значение модельного времени (в единицах модельного времени)

System.out.println(Строка/переменная)– функция вывода в стандартный поток вывода.

1. Синтаксис входного языка

G=(N,,P,<Программа>)

N={<Программа>,<Имя класса>,<Главная функция>,<Буква>,<Цифра>,<Знак>,<Строка>,<Предложение>,<Функция вывода>,<Арифметическое выражение>,<Цикл>,<Оператор ветвления>,<Объявить переменную>,<Целый тип>,<Вещественный тип>,<Символьный тип>,<Логический тип>,<Переменные целые>,<Вещественные переменные>,<Символьные переменные>,<Логические переменные>,<Идентификатор>,<ЦифрИд>,<Целое число>,<Вещественное число>,<Знак>,<Символ>,<Булево значение>,<Стандартные функции>,<Формула>,<Функция вывода>,<Вывод>,<Присваивание>,<Операторы присваивания>,<Унарные операторы>,<Часть формулы>,<Арифметические операторы>,<Цикл>,<Цикл for>,<Цикл while>,<Цикл do>,<Инициализация счётчика>,<Булево выражение>,<Условие>,<Изменение счётчика>,<Счётчик>,<Выражение счётчика>,<Логические операторы>,<Операторы сравнения>,<Оператор ветвления>}

={A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z,a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z,1,2,3,4,5,6,7,8,9,0,+,\_,|,-,[,],’,;,-,%,/,\*,~,<,>,&,^,|,!,.,=,(,),{,}}

P:

<Программа> public class <Идентификатор>{<Главная функция>}

<Буква>A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z

<Цифра>→1|2|3|4|5|6|7|8|9|0

<Знак>→+|-|

<Строка>→<Цифра>|<Знак>|<Буква>|<Цифра><Строка>|<Знак><Строка>|<Буква><Строка>

<Главная функция>public static void main(String[] args){<Список предложений>}|  
public static void main(String[] args){}

<Предложение><Объявить переменную >| <Функция вывода> |<Присваивание> | <Цикл> | <Оператор ветвления> | <Оператор останова>

<Список предложений><Предложение>; | <Предложение> ;<Список предложений>

<Объявить переменную> <Имя типа><Объявить список переменных>

<Объявить список переменных> <Объявить одну переменную> | <Объявить список переменных>, <Объявить переменную>

<Объявить одну переменную> <Идентификатор | <Имя типа><Идентификатор> = <Выражение>

<Имя типа> byte|short|int|long|float|double|char|boolean|String

<Число>→<Вещественное число>|<Целое число>

<Целое число>→<Цифра>|<Цифра><Целое число>

<Вещественное число>→<Целое число>.<Целое число>

<Булево значение>→true|false

<Идентификатор>→\_<Символы ид-ра>|<Буква>< Символы ид-ра >|<Буква>

<Символы ид-ра> → <Цифра> | <Буква> | \_ |<Цифра><Символы ид-ра> | <Буква><Символы ид-ра> | \_ <Символы ид-ра>

<Стандартные функции>→Math.log(<Аргументы>)| Math.pow(<Аргументы>)| Math.sqrt(<Аргументы>)

<Аргументы>→<Выражение>,<Аргументы>|<Выражение>

<Функция вывода>→System.out.print(<Символьное выражение>);

<Присваивание>→<Идентификатор>=<Выражение>;|  
<Выражение><Унарные операторы>;

<Унарные операторы>→++|--

<Выражение>→<Арифметическое выражение>|<Логическое выражение>|<Символьное выражение>

<Знак операции символьного сложения>→+

<Знак операции символьного умножения>→\*

<Символьное выражение> →  
<Символьное слагаемое><Знак операции символьного сложения><Символьное выражение>|  
<Символьное слагаемое>

<Символьное слагаемое> →   
<Символьный множитель><Знак операции символьного умножения><Символьное слагаемое>|  
<Символьный множитель>

<Символьный множитель> → (<Символьное выражение >)|<Идентификатор>|‘<Строка>’ | “<Строка>”

<Знак операции типа сложение>→+|-

<Знак операции типа умножение>→\*|/|%

<Арифметическое выражение> →  
<Арифметическое слагаемое><Знак операции типа сложение><Арифметический выражение>|  
<Арифметическое слагаемое>

<Арифметическое слагаемое> →   
<Арифметический множитель><Знак операции типа умножение><Арифметическое слагаемое>|  
<Арифметический множитель>

<Арифметический множитель> → (<Арифметическое выражение >)|<Идентификатор>|<Число>|<Стандартные функции>

<Знак операции логического сложения> → ||

<Знак операции логического умножения> → &&

<Знак операции сравнения >→==|!=|>|<|>=|<=

<Логическое выражение> →  
<Логическое слагаемое><Знак операции логического сложения><Логическое выражение>|  
<Логическое слагаемое>

<Логическое слагаемое> →   
<Логический множитель><Знак операции логического умножения><Логическое слагаемое>|  
<Логический множитель>

<Логический множитель> →   
<Арифметическое выражение><Знак операции сравнения>< Арифметическое выражение>|  
<Логическое выражение><Знак операции сравнения>< Логическое выражение>|  
(<Логическое выражение>)|  
<Идентификатор>|  
<Булево значение>

<Цикл>→<Цикл for>|<Цикл while>|<Цикл do>

<Цикл for>→for (<Инициализация счётчика>;<Условие>;<Изменение счётчика>){<Список предложений>}|for (<Инициализация счётчика>;<Условие>;<Изменение счётчика>)< Предложение >

<Инициализация счётчика>→|<Объявить переменную>|<Присваивание>

<Изменение счётчика>→|<Присваивание>

<Условие>→<Логическое выражение>

<Цикл while>→While (<Условие>){< Список предложений >}

<Цикл do>→do {< Список предложений >} while (<Условие>);

<Оператор ветвления>→if (<Условие>) {< Список предложений >}|if (<Условие>){< Список предложений >}else{< Список предложений >}|if (<Условие>){< Список предложений >}else <Предложение>|if(<Условие>)<Предложение>|if(<Условие>)<Предложение> else {< Список предложений>}|if(<Условие>)<Предложение> else <Предложение>}

<Операторы останова>→break|continue

1. Контекстные условия

***Контекстные условия о правилах описания идентификаторов в программах:***

1. Все используемые в программах идентификаторы должны быть описаны до их использования в программе.
2. Каждый из идентификаторов, используемых в программе, должен быть описан один раз.

***Контекстные условия о правилах использования идентификаторов в своей области действия:***

Две основные области видимости в Java определяются классом и методом, хотя такое их разделение несколько искусственно.

1. Все используемые в программах идентификаторы должны иметь тот же регистр что и при объявлении.
2. Все используемые в программах идентификаторы должны быть объявлены в этой области видимости, либо в более глобальной, включающую эту область видимости.
3. Во внутреннем блоке кода нельзя объявлять переменные с тем же именем, что и во внешней области действия.

***Контекстные условия, определяющие правила соответствия видов величин, входящих в синтаксические конструкции программ:***

1. Соответствие фактических и формальных параметров идёт в порядке перечисления (т.е. первый фактический параметр соответствует первому формальному, второй фактический- второму формальному и т.д).
2. Фактические параметры должны быть совместимы с формальными или быть приводимыми (автоматически или с помощью операторов «кастования»).
3. В операторе присваивания типы переменной и выражения должны совпадать, либо быть приводимыми (автоматически или с помощью операторов «кастования»).
4. Размер массива обязательно указывается при создании. Размерность массива может быть указана при помощи числа или числовой константы.
5. При инициализировании массива собственными значениями, тип элементов из списка значений должен соответствовать объявленному типу массива.
6. При операции присваивания и т.д типы данных должны соответствовать.

***Контекстные условия, задающие различные количественные ограничения:***

1. При достаточном размере выделенной памяти, массив может иметь длину *Integer.MAX\_VALUE - 8* значений
2. Идентификатор может быть любой длины.
3. Количество идентификаторов не более 1000.
4. Количество параметров в функциях, процедурах не более 1000.
5. Вложенность скобок в выражении не более 5 уровней.
6. Таблица соответствия

В данной главе представлена таблица соответствия (Таблица 6).

|  |  |
| --- | --- |
| **Java** | **Go** |
| ***Структура программы*** | |
| public class <Имя Класса> {  public static void main(String[] args) {  [Объявление переменных]  [Тело программы]  } | func main() {  [Объявление переменных]  [Тело программы]  } |
| ***Циклы*** | |
| for ([инициализация счетчика]; [условие]; [изменение счетчика])  {  // действия  }  do{  // действия  }  while ([условие]);  while ([условие]){  // действия  } | for [инициализация счетчика]; [условие]; [изменение счетчика]{  // действия  }  for {; [условие];{  // действия  } |
| ***Оператор ветвления*** | |
| If([условие]){[блок]}  If([условие]){[блок]}else{[блок]} | If [условие] {[блок]}  If [условие] {[блок]}else{[блок]} |

Таблица -Таблица соответствия Java-Go

Продолжение таблицы 6 соответствия Java-Go

|  |  |  |
| --- | --- | --- |
| **Java** | **Go** | |
| ***Типы данных*** | | |
| Byte  Short  Int  Long  Double  Float  Char  Boolean | | Int8  Int16  Int32  Int64  Float32  Float64  String  Bool |
| ***Логический тип данных*** | | |
| True  False | | True  False |
| ***Логические операции*** | | |
| !  &&  || | | !  &&  || |
| ***Математические операции*** | | |
| +  -  \*  /  %  ++  -- | | +  -  \*  /  %  ++  -- |

Продолжение таблицы 6 соответствия Java-Go

|  |  |  |
| --- | --- | --- |
| **Java** | **Go** | |
| ***Поразрядные операции*** | | |
| &  |  ^  ~  <<  >>  >>> | | &  |  ^  &^  <<  >>  Нет аналога |
| ***Условные выражения*** | | |
| ==  !=  <  >  <=  >= | | ==  !=  <  >  <=  >= |
| ***Операторы присваивания*** | | |
| =  +=  -=  \*=  /=  %= | | =  [Идентификатор]=[Идентификатор] +[Выражение]  [Идентификатор]=[Идентификатор] -[Выражение]  [Идентификатор]=[Идентификатор] \*[Выражение]  [Идентификатор]=[Идентификатор] /[Выражение]  [Идентификатор]=[Идентификатор] %[Выражение] |

Окончание таблицы соответствия Java-Go

|  |  |  |
| --- | --- | --- |
| **Java** | **Go** | |
| ***Встроенные функции языка*** | | |
| min( a, b )  log( a )  pow( a, b )  sqrt( a )  time()  System.out.println([Выражение]) | | Нет аналога  log( a )  pow( a, b )  sqrt( a )  time.Now()  fmt.Println([Выражение]) |

1. Проект лексического анализатора

# Таблица ключевых слов

Ниже представлена таблица ключевых слов (Таблица 7).

|  |  |
| --- | --- |
| ***Ключевое слово*** | ***Ключ*** |
| break | К1 |
| class | К2 |
| const | К3 |
| continue | К4 |
| do | К5 |
| else | К6 |
| for | К7 |
| goto; | К8 |
| if | К9 |
| static | К10 |
| void | К11 |
| while | К12 |
| public | К13 |
| return | К14 |

Таблица -Таблица ключевых слов

Ниже представлена таблица зарезервированных имён (Таблица 8).

|  |  |
| --- | --- |
| ***Зарезервированное имя*** | ***Значение*** |
| boolean | R1 (имя типа данных) |
| byte — | R2 (имя типа данных) |
| char | R3 (имя типа данных) |
| double | R4 (имя типа данных) |
| float | R5 (имя типа данных) |

Таблица -Таблица зарезервированных имён

Окончание таблицы 8 зарезервированных имён

|  |  |
| --- | --- |
| ***Зарезервированное имя*** | ***Значение*** |
| int | R6 (целочисленный тип;) |
| long | R7 (целочисленный тип;) |
| short | R8 (целочисленный тип;) |
| min | R9 (встроенная функция) |
| log | R10 (встроенная функция) |
| pow | R11 (встроенная функция) |
| sqrt | R12 (встроенная функция) |
| time | R13 (встроенная функция) |
| System.out.println | R14 (встроенная функция) |
| True | R15 (логический тип) |
| False | R16 (логический тип) |

# Типы лексем

Ниже представлена таблица типов лексем (Таблица 9).

|  |  |
| --- | --- |
| Идентификатор | Ключевое слово, зарезервированное имя |
| Оператор | +(О1) |
| -(О2) |
| \*(О3) |
| /(О4) |
| %(О5) |
| ++(О6) |
| --(О7) |
| >(О8) |
| <(О9) |
| >=(О10) |
| <=(О11) |
| ==(О12) |
| !=(О13) |
| &&(О14) |
| ||(О15) |
| !(О16) |
| ~(О17) |
| =(О18) |
| +=(О19) |
| -=(О20) |
| \*=(О21) |
| /=(О22) |

Таблица -Типы лексем

Окончание таблицы 9 Типы лексем

|  |  |
| --- | --- |
|  | %=(О23) |
| <<(О24) |
| >>(О25) |
| &(О26) |
| |(О27) |
| ^(О28) |
| >>>(О29) |
| Разделитель | . (D1) |
| , (D2) |
| ; (D3) |
| { (D4) |
| } (D5) |
| ( (D6) |
| ) (D7) |
| Число | Целое, вещественное |
| Строка | Символ (‘ ‘) |

# Ошибки лексического анализатора

В данном проекте лексический анализатор распознает следующие синтаксические ошибки:

1. Ошибка в образовании вещественного числа;
2. Неправильное формирование символа;
3. Неправильное формирование оператора;
4. Неверное образование имен переменных;
5. Недопустимый грамматикой символ;

# Конечный автомат лексического анализатора

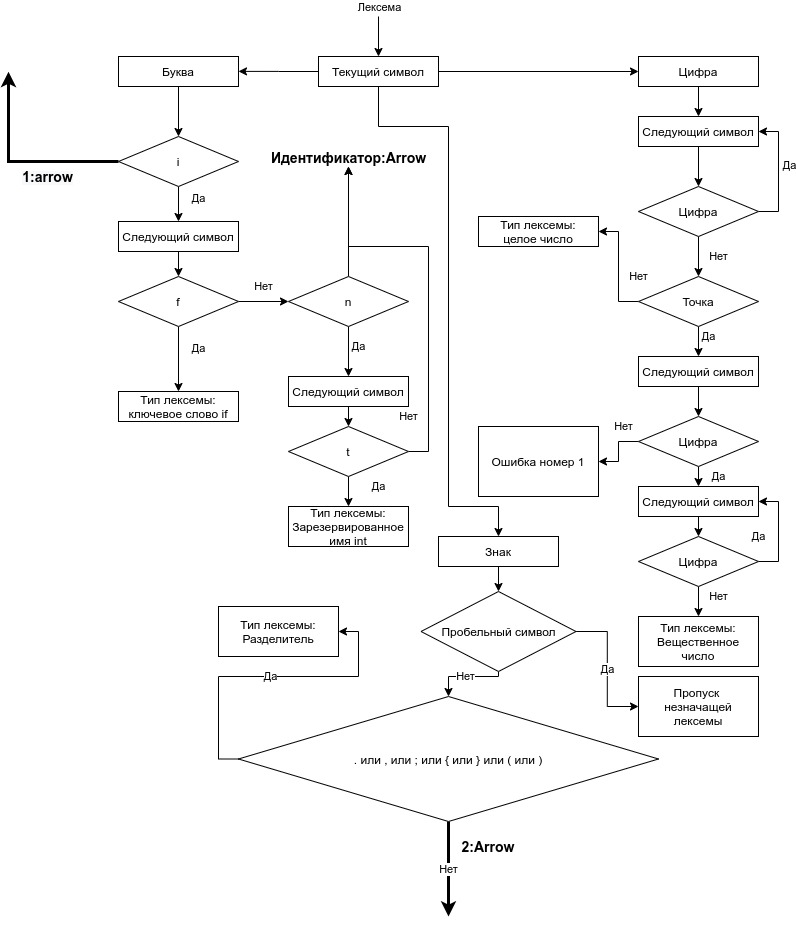


Рисунок - Конечный автомат 1 часть

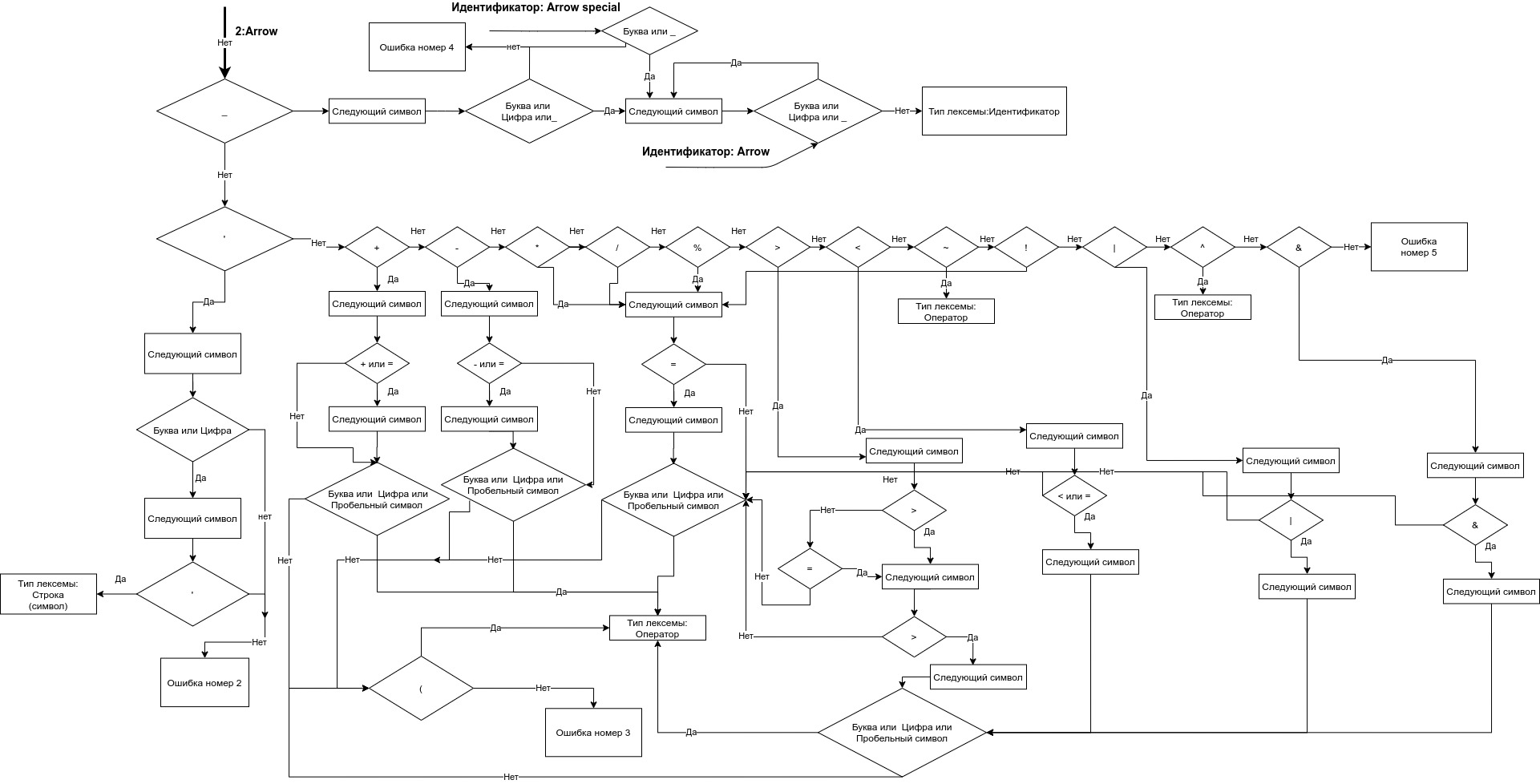


Рисунок -Конечный автомат 2 часть

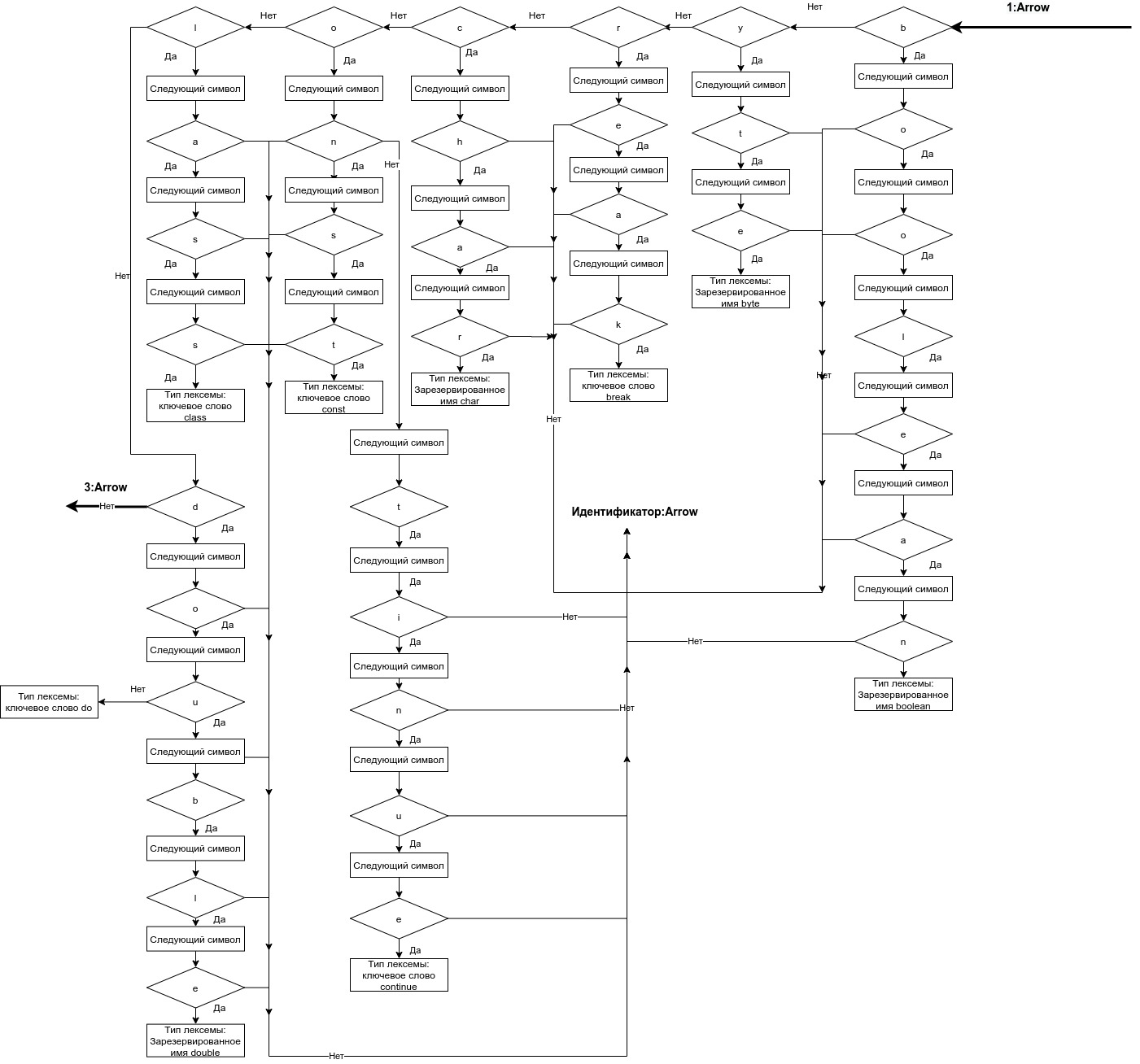


Рисунок -Конечный автомат 3 часть

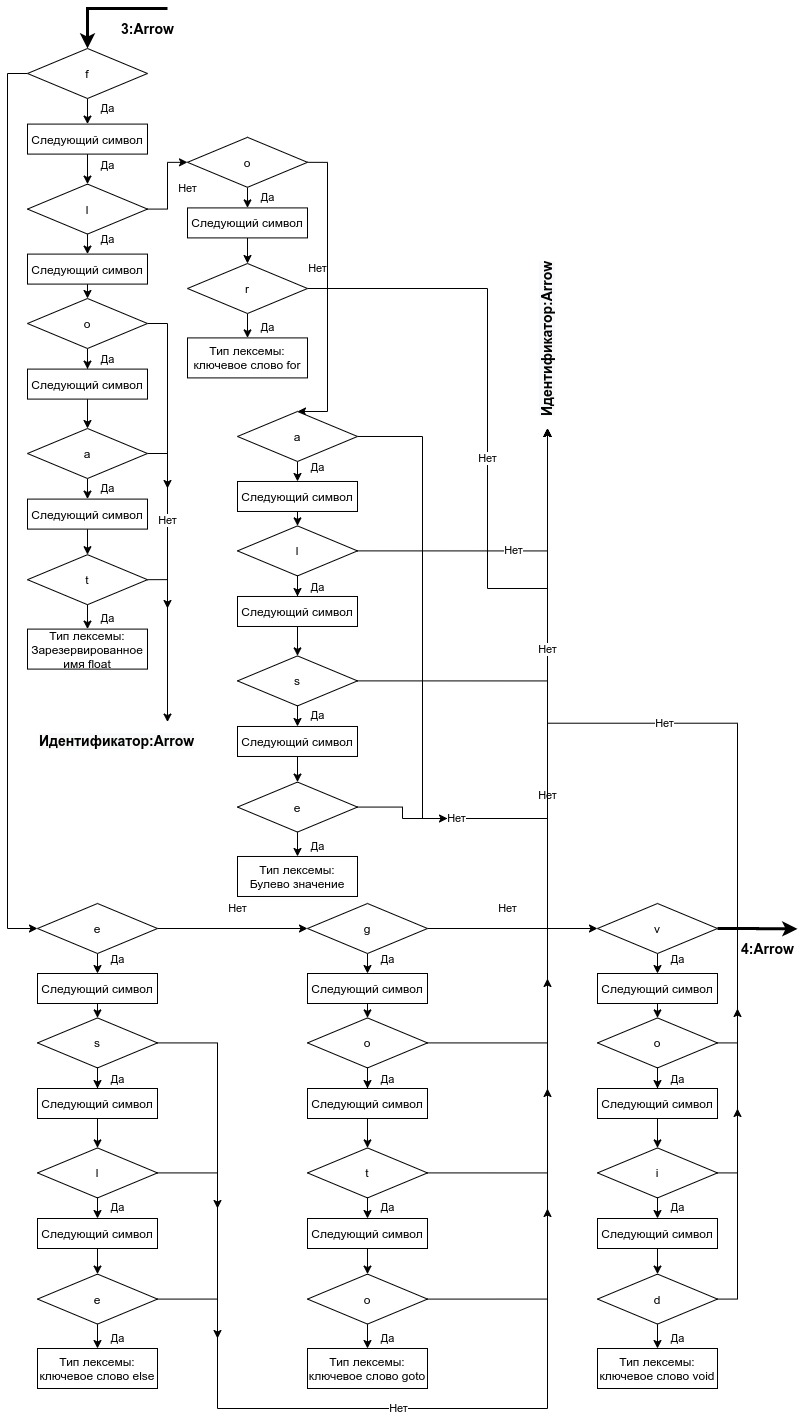


Рисунок -Конечный автомат 4 часть

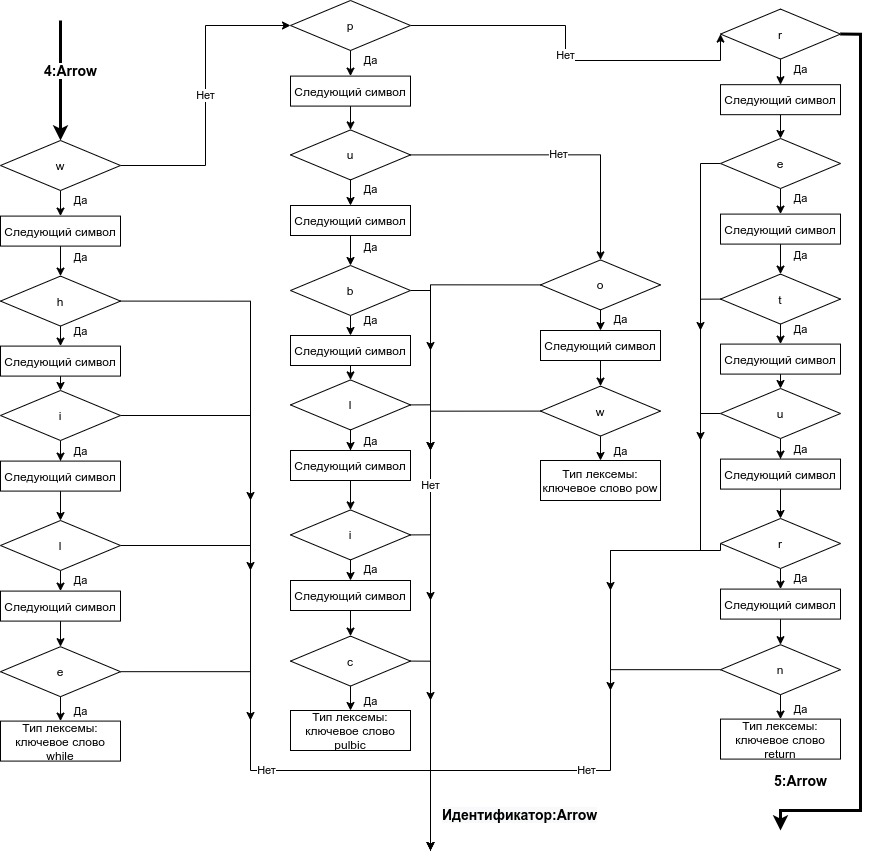


Рисунок -Конечный автомат часть 5

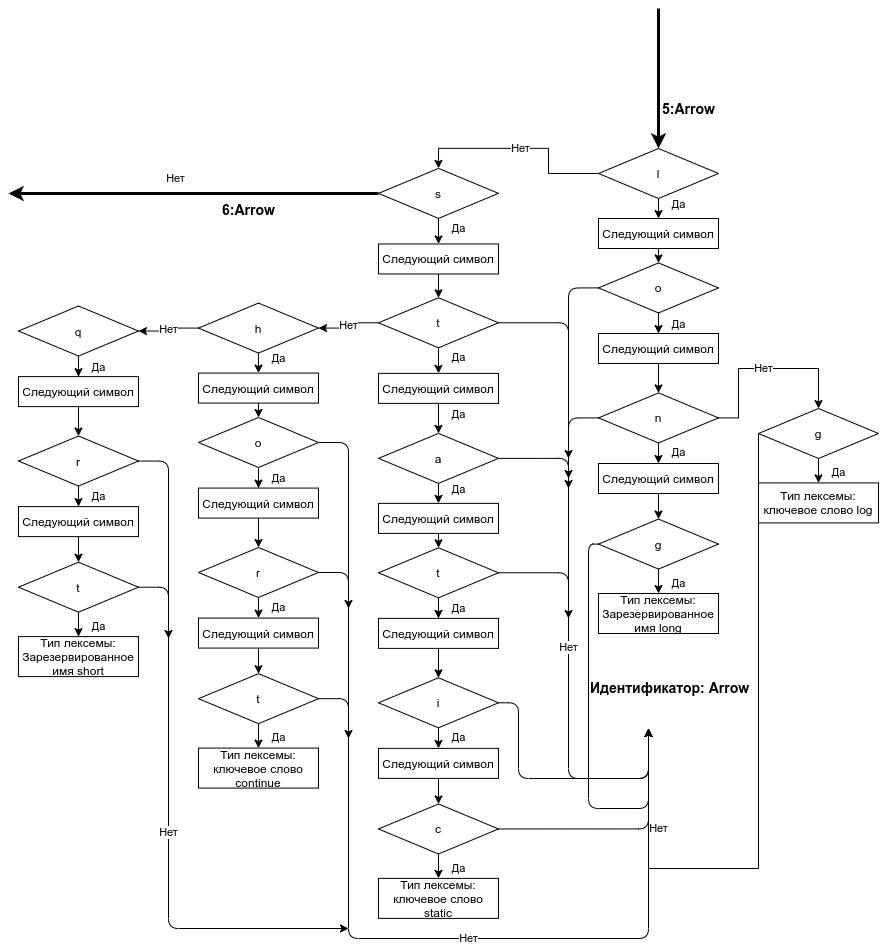


Рисунок -Конечный автомат 6 часть

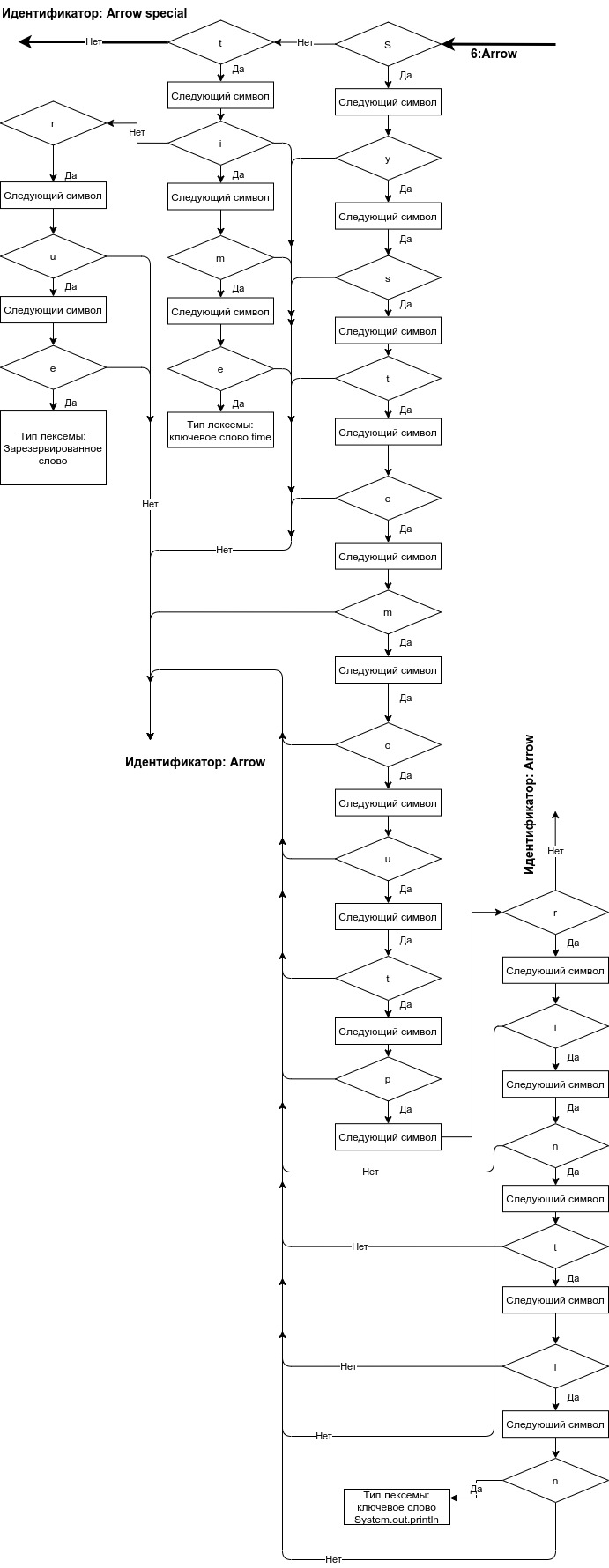


Рисунок -Конечный автомат 7 часть

1. Проект синтаксического анализатора

Задача синтаксического анализа – проверка правильности структуры программы, описываемой контекстно-свободной грамматикой. Класс контекстно-свободных грамматик шире класса автоматных грамматик, поэтому и синтаксический анализатор сложнее лексического.

Для синтаксического анализатора был выбран алгоритм Эрли восходящего разбора.

**Модель данных**

Язык реализации – C++. Синтаксический анализатор реализован в классе SyntaxAnalyzer. По умолчанию задана грамматика класса Grammar и лексический анализатор класса LexicalAnalyzer. Алгоритм разбора реализован в классе EarleyAlgorithm. Основной метод класса – разбор цепочки – получает на вход цепочку символов, результатом работы метода является последовательность номеров правил – правый вывод входной цепочки.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рисунок – Диаграмма потока данных

**Класс FileReader**

read(filepath, pathdir, namepattern) // метод читает файлы с программами и директории и возвращает их в виде массива строк

**Класс LexicalAnalyzer**

check(str) // метод проверяет строку на наличие недопустимых символов

deletecomments(instr) // метод удаляет комментарии из текста программы

skan(indata) // метод возвращает множество токенов входной цепочки.

**Класс Env**

insertnewscope(ptr) // добавляет новую таблицу имен в текущее окружение

searchelem(ptr, id) // поиск иденификатора в текущем окружении

**Класс SyntaxAnalyzer**

earley(rule, text) // метод возвращает список разбора цепочки text, где rule – аксиома грамматики.

right\_parsing(table) // метод возвращает список правил – правый разбор цепочки, для которой составлен список разбора table.

**Класс Grammar**

getTerminal() // метод возвращает множество терминальных символов грамматики.

getNonterminal() // метод возвращает множество нетерминальных символов грамматики.

getAxiom() // метод возвращает начальный символ грамматики.

getRules() // метод возвращает множество правил вывода грамматики.

setTerminal(newTerminal) // метод устанавливает множество терминальных символов грамматики.

setNonterminal(newNonterminal) // метод устанавливает нетерминальных символов грамматики.

setAxiom(newAxiom) // метод устанавливает начальный символ грамматики.

setRules(newRules) // метод устанавливает множество правил вывода грамматики.

getRuleNumber(rule) // метод возвращает номер правила.

**Класс State**

completed() // возвращает true, если точка на конце, иначе false.

next\_term(): // сдвигает точку на одну позицию вправо.

getindexXk(xk):// возвращает индекс символа xk в состоянии, если такой символ отсутствует, то -1.

**Класс Rule**

add(productions) // добавляет к правилу правые части productions.

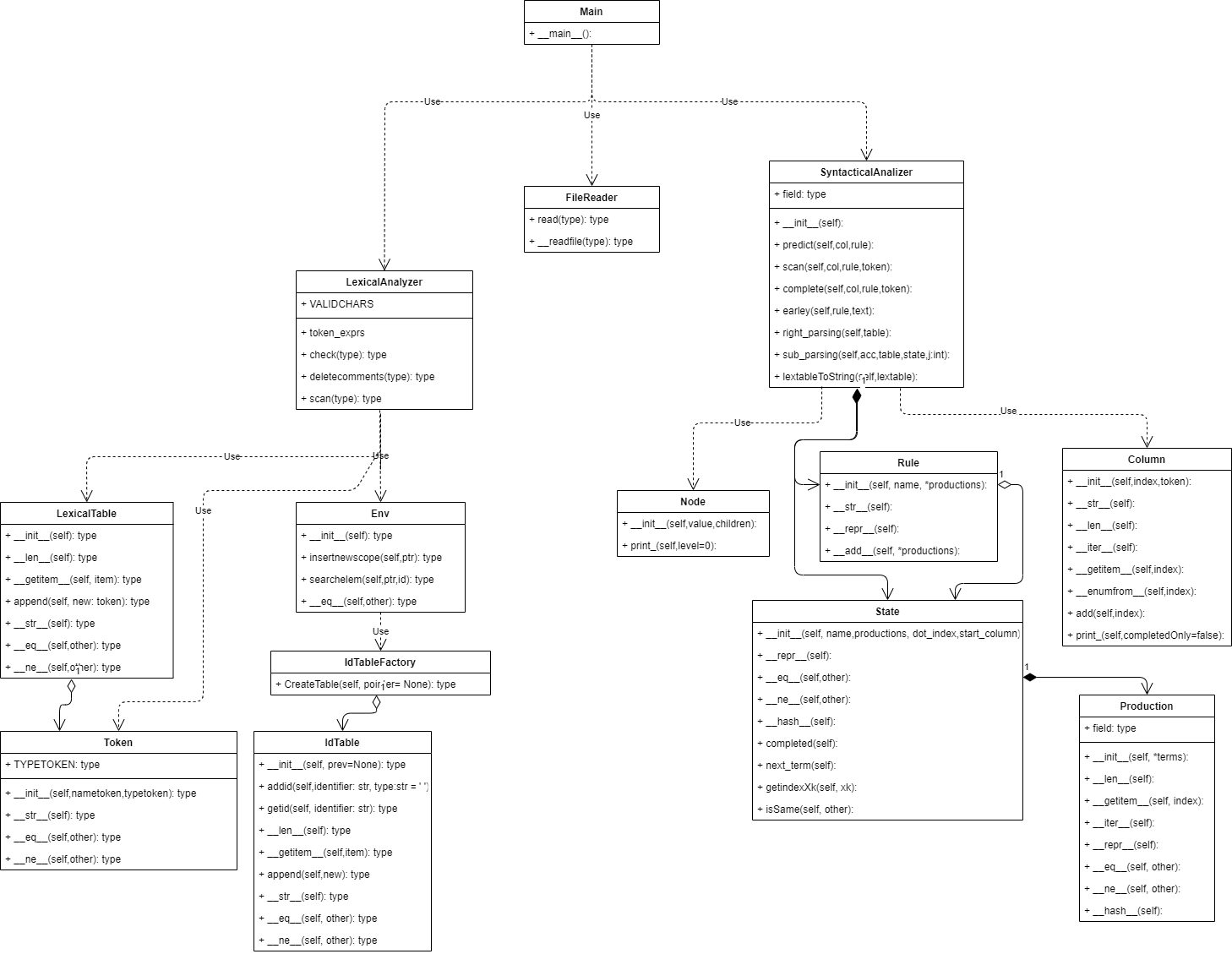


Рисунок – Диаграмма классов синтаксического анализатора

**Проект алгоритма Эрли**

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рисунок – Проект алгоритма разбора

Изображение выглядит как текст

Автоматически созданное описание

Рисунок – Проект алгоритма составления списка разбора

Изображение выглядит как текст, карта

Автоматически созданное описание

Рисунок – Проект алгоритма построения разбора

Изображение выглядит как текст, карта

Автоматически созданное описание

Рисунок – Проект процедуры R(правый разбор)

**Пример дерева разбора для следующей входной цепочки**

public class Main{

public static void main(String[] args){

System.out.print(“Hello”);

}

}

Пример дерева её разбора приведён на рисунке 7:

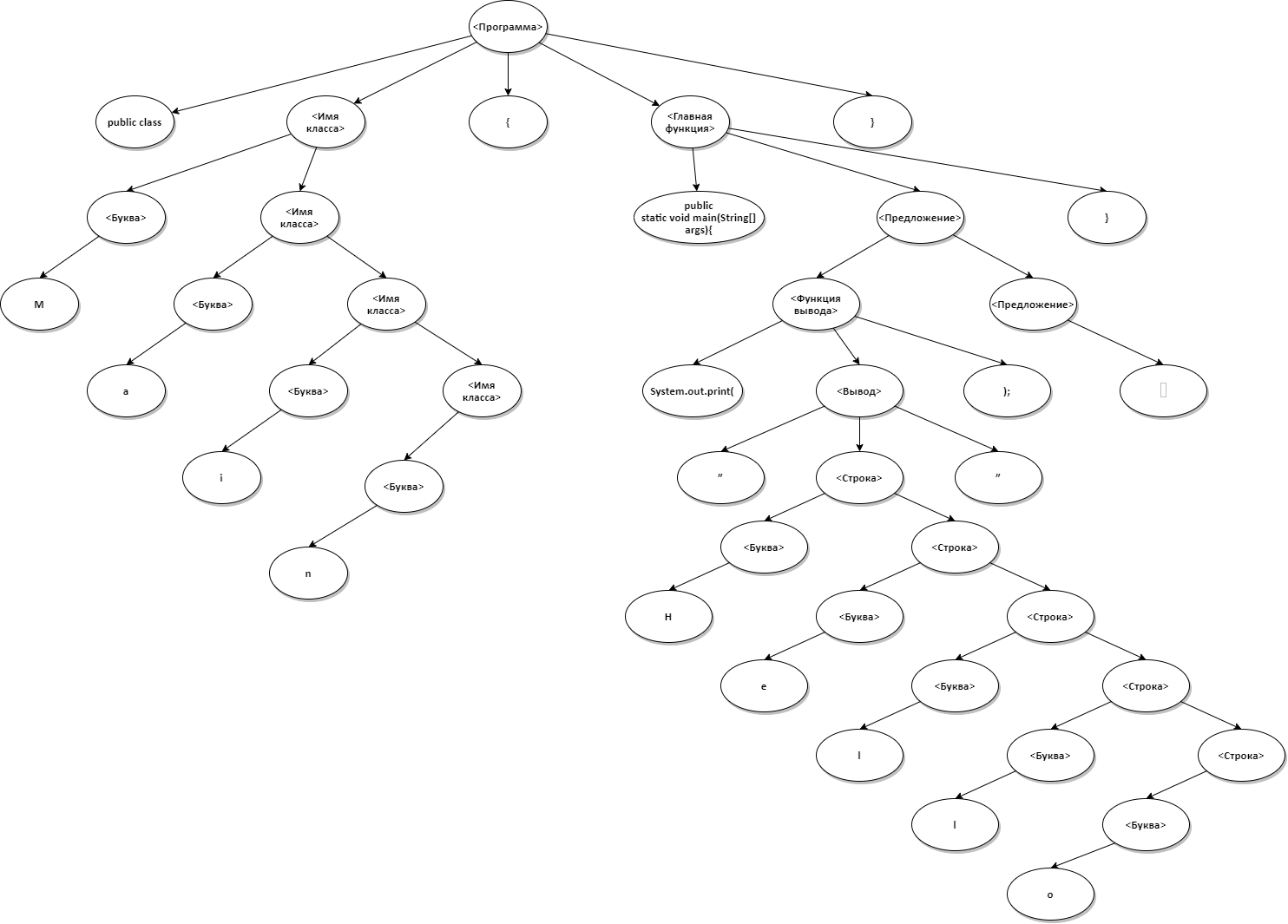


Рисунок -Дерево разбора

**Ошибки синтаксичекского анализатора**

В данном проекте синтаксический анализатор распознает следующие синтаксические ошибки:

1. Разбор цепочки, не принадлежащей языку исходной грамматики.
2. Фрагменты кода

lexicalanalyzer.py

import re  
import sys  
  
from .env import Env  
from .lexicaltable import LexicalTable  
from .tok import Token  
  
  
class LexicalAnalyzer:  
 *'''returns table of lexems'''* VALIDCHARS = "ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz1234567890+\_-[]'\",;-%/\*~<>&^|!.=(){} \n"  
 '''Strict order'''  
 token\_exprs = [  
 (r"[\n\s]+", None),  
 (r'\(', Token.TYPETOKEN[6]),  
 (r'\)', Token.TYPETOKEN[6]),  
 (r'\}', Token.TYPETOKEN[6]),  
 (r'\{', Token.TYPETOKEN[6]),  
 (r'\;', Token.TYPETOKEN[6]),  
 (r'\'[A-Za-z][A-Za-z0-9\_ ]\*\'', Token.TYPETOKEN[5]),  
 (r'\'', Token.TYPETOKEN[6]),  
 (r'\,', Token.TYPETOKEN[6]),  
 (r'\"', Token.TYPETOKEN[6]),  
 (r'\:', Token.TYPETOKEN[6]),  
 (r'\+\+', Token.TYPETOKEN[4]),  
 (r'\+', Token.TYPETOKEN[4]),  
 (r'\-\-', Token.TYPETOKEN[4]),  
 (r'\-', Token.TYPETOKEN[4]),  
 (r'\\*', Token.TYPETOKEN[4]),  
 (r'\/', Token.TYPETOKEN[4]),  
 (r'\>\=', Token.TYPETOKEN[4]),  
 (r'\>', Token.TYPETOKEN[4]),  
 (r'\<\=', Token.TYPETOKEN[4]),  
 (r'\<', Token.TYPETOKEN[4]),  
 (r'\!\= ', Token.TYPETOKEN[4]),  
 (r'\!', Token.TYPETOKEN[4]),  
 (r'\==', Token.TYPETOKEN[4]),  
 (r'\=', Token.TYPETOKEN[4]),  
 (r'\%', Token.TYPETOKEN[4]),  
 (r'\&\&', Token.TYPETOKEN[4]),  
 (r'\|\|', Token.TYPETOKEN[4]),  
 (r'break', Token.TYPETOKEN[2]),  
 (r'class', Token.TYPETOKEN[2]),  
 (r'continue', Token.TYPETOKEN[2]),  
 (r'double', Token.TYPETOKEN[1]),  
 (r'do', Token.TYPETOKEN[2]),  
 (r'for', Token.TYPETOKEN[2]),  
 (r'static', Token.TYPETOKEN[2]),  
 (r'void', Token.TYPETOKEN[2]),  
 (r'while', Token.TYPETOKEN[2]),  
 (r'main', Token.TYPETOKEN[2]),  
 (r'public', Token.TYPETOKEN[2]),  
 (r'String\[\]', Token.TYPETOKEN[2]),  
 (r'String', Token.TYPETOKEN[2]),  
 (r'args', Token.TYPETOKEN[2]),  
 (r'if', Token.TYPETOKEN[2]),  
 (r'else', Token.TYPETOKEN[2]),  
 (r'boolean', Token.TYPETOKEN[1]),  
 (r'byte', Token.TYPETOKEN[1]),  
 (r'char', Token.TYPETOKEN[1]),  
 (r'float', Token.TYPETOKEN[1]),  
 (r'int', Token.TYPETOKEN[1]),  
 (r'long', Token.TYPETOKEN[1]),  
 (r'short', Token.TYPETOKEN[1]),  
 (r'Math\.log', Token.TYPETOKEN[1]),  
 (r'Math\.pow', Token.TYPETOKEN[1]),  
 (r'Math\.sqrt', Token.TYPETOKEN[1]),  
 (r'System\.out\.println', Token.TYPETOKEN[1]),  
 (r'System\.out\.print', Token.TYPETOKEN[1]),  
 (r'true', Token.TYPETOKEN[1]),  
 (r'false', Token.TYPETOKEN[1]),  
 (r'(\d+(.\d+)?)', Token.TYPETOKEN[3]),  
 (r'[A-Za-z\_][A-Za-z0-9\_]\*', Token.TYPETOKEN[0]),  
 ]  
  
 def check(self, str):  
 *'''checks the entire string for the validity of characters'''* checklist = []  
 for i in str:  
 if i not in self.VALIDCHARS:  
 checklist.append(i)  
 return checklist  
  
 def deletecomments(self, instr):  
 *'''delete all comments from string '''* res = re.sub(r'\s+//[\w\s]+\n', '', instr)  
 res = re.sub(r'\s+/\\*[\w\s]+\\*/\n', '', res)  
 return res  
  
 def skan(self, indata):  
 *'''Parses text into tokens, forms a tree of name tables and returns a table of tokens'''* lextable = LexicalTable()  
 env = Env()  
 namefile = indata[0]  
 filedata = indata[1]  
 checkdata = self.check(filedata)  
 ptr = env.root  
 if checkdata:  
 return print("Invalid chars {0} in program {1}".format(checkdata, namefile))  
 else:  
 characters = self.deletecomments(filedata)  
 pos = 0  
 while pos < len(characters):  
 match = None  
 for token\_expr in self.token\_exprs:  
 pattern, tag = token\_expr  
 regex = re.compile(pattern)  
 match = regex.match(str(characters), pos)  
 if match:  
 text = match.group(0)  
 if tag:  
 if text == '{':  
 ptr = env.insertnewscope(ptr)  
 if text == '}':  
 ptr = ptr.prev  
 token = Token(text, tag)  
 lextable.append(token)  
 if tag == Token.TYPETOKEN[0] and ptr is not None:  
 ptr.addid(token.name, token.type)  
 break  
 if not match:  
 sys.stderr.write('Illegal character \'{}\',{}: %s\n'.format(namefile, pos) % characters[pos])  
 sys.exit(1)  
 else:  
 pos = match.end(0)  
 return [lextable, env]

syntaxanalyzer.py

from .column import Column  
from .node import Node  
from .production import Production  
from .rule import Rule  
from .state import State  
from translator.lexicalanalyzer.tok import Token  
from .syntacticaltree import SyntacticalTree  
import copy  
  
  
class SyntacticalAnalyzer:  
  
 def \_\_init\_\_(self):  
 self.LETTER = Rule("LETTER", Production("A"), Production("B"), Production("C"), Production("D"),  
 Production("E"),  
 Production("F"), Production("G"), Production("H"), Production("I"), Production("J"),  
 Production("K"),  
 Production("L"), Production("M"), Production("N"), Production("O"), Production("P"),  
 Production("Q"),  
 Production("R"), Production("S"), Production("T"), Production("U"), Production("V"),  
 Production("W"),  
 Production("X"), Production("Y"), Production("Z"),  
 Production("a"), Production("b"), Production("c"), Production("d"), Production("e"),  
 Production("f"),  
 Production("g"), Production("h"), Production("i"), Production("j"), Production("k"),  
 Production("l"),  
 Production("m"), Production("n"), Production("o"), Production("p"), Production("q"),  
 Production("r"),  
 Production("s"), Production("t"), Production("u"), Production("v"), Production("w"),  
 Production("x"),  
 Production("y"), Production("z"))  
  
 self.DIGIT = Rule("DIGIT", Production("1"), Production("2"), Production("3"), Production("4"), Production("5"),  
 Production("6"), Production("7"), Production("8"), Production("9"), Production("0"))  
 self.SIGN = Rule("SIGN", Production("+"), Production("-"), Production(""))  
 self.BOOL\_VAL = Rule("BOOL VALUE", Production("true"), Production("false"))  
 self.TYPE\_NAME = Rule("TYPE NAME", Production("byte"), Production("short"), Production("int"),  
 Production("long"),  
 Production("float"), Production("double"), Production("char"), Production("boolean"),  
 Production("String"))  
 self.ID\_SYMBOLS = Rule("IDENTIFICATOR SYMBOLS", Production(self.DIGIT), Production(self.LETTER),  
 Production("\_"))  
 self.ID\_SYMBOLS.add(Production(self.DIGIT, self.ID\_SYMBOLS), Production(self.LETTER, self.ID\_SYMBOLS),  
 Production("\_", self.ID\_SYMBOLS))  
 self.IDENTIFICATOR = Rule("IDENTIFICATOR", Production("\_", self.ID\_SYMBOLS),  
 Production(self.LETTER, self.ID\_SYMBOLS),  
 Production(self.LETTER))  
  
 self.STRING = Rule("STRING", Production(self.DIGIT), Production(self.SIGN), Production(self.LETTER))  
 self.STRING.add(Production(self.DIGIT, self.STRING), Production(self.SIGN, self.STRING),  
 Production(self.LETTER, self.STRING))  
 self.STR\_SGN\_SUM = Rule("STRING SUM SIGN", Production("+"))  
 self.STR\_SGN\_MUL = Rule("STRING MUL SIGN", Production("\*"))  
 self.STR\_MUL = Rule("STRING MULTIPLIER", Production(self.IDENTIFICATOR), Production("'", self.STRING, "'"),  
 Production("\"", self.STRING, "\""))  
 self.STR\_SUM = Rule("STRING ADDENDUM", Production(self.STR\_MUL))  
 self.STR\_SUM.add(Production(self.STR\_MUL, self.STR\_SGN\_MUL, self.STR\_SUM))  
 self.STR\_EXPR = Rule("STRING EXPRESSION", Production(self.STR\_SUM))  
 self.STR\_EXPR.add(Production(self.STR\_SUM, self.STR\_SGN\_SUM, self.STR\_EXPR))  
 self.STR\_MUL.add(Production("(", self.STR\_EXPR, ")"))  
  
 self.INT\_NUM = Rule("INTEGER NUMBER", Production(self.DIGIT))  
 self.INT\_NUM.add(Production(self.DIGIT, self.INT\_NUM))  
 self.REAL\_NUM = Rule("REAL NUMBER", Production(self.INT\_NUM, ".", self.INT\_NUM))  
 self.NUM = Rule("NUMBER", Production(self.INT\_NUM), Production(self.REAL\_NUM),  
 Production(self.SIGN, self.INT\_NUM), Production(self.SIGN, self.REAL\_NUM))  
 self.ARITH\_SGN\_SUM = Rule("ARITHMETIC SUM SIGN", Production("+"), Production("-"))  
 self.ARITH\_SGN\_MUL = Rule("ARITHMETIC MUL SIGN", Production("\*"), Production("/"), Production("%"))  
 self.ARITH\_MUL = Rule("ARITHMETIC MULTIPLIER", Production(self.IDENTIFICATOR), Production(self.NUM))  
 self.ARITH\_SUM = Rule("ARITHMETIC ADDENUM", Production(self.ARITH\_MUL))  
 self.ARITH\_SUM.add(Production(self.ARITH\_MUL, self.ARITH\_SGN\_MUL, self.ARITH\_SUM))  
 self.ARITH\_EXPR = Rule("ARITHMETIC EXPRESSION", Production(self.ARITH\_SUM))  
 self.ARITH\_EXPR.add(Production(self.ARITH\_SUM, self.ARITH\_SGN\_SUM, self.ARITH\_EXPR))  
 self.ARITH\_MUL.add(Production("(", self.ARITH\_EXPR, ")"))  
  
 self.LOG\_SGN\_SUM = Rule("LOGIC SUM SIGN", Production("||"))  
 self.LOG\_SGN\_MUL = Rule("LOGIC MUL SIGN", Production("&&"))  
 self.LOG\_SGN\_CMP = Rule("LOGIC COMPARE SIGN", Production("=="), Production("!="), Production(">"),  
 Production("<"), Production(">="), Production("<="))  
 self.LOG\_MUL = Rule("LOGIC MULTIPLIER", Production(self.IDENTIFICATOR),  
 Production(self.BOOL\_VAL), Production(self.ARITH\_EXPR, self.LOG\_SGN\_CMP, self.ARITH\_EXPR))  
 self.LOG\_SUM = Rule("LOGIC ADDENUM", Production(self.LOG\_MUL))  
 self.LOG\_SUM.add(Production(self.LOG\_MUL, self.LOG\_SGN\_MUL, self.LOG\_SUM))  
 self.LOG\_EXPR = Rule("LOGIC EXPRESSION", Production(self.LOG\_SUM))  
 self.LOG\_EXPR.add(Production(self.LOG\_SUM, self.LOG\_SGN\_SUM, self.LOG\_EXPR))  
 self.LOG\_MUL.add(Production(self.LOG\_EXPR, self.LOG\_SGN\_CMP, self.LOG\_EXPR),  
 Production("(", self.LOG\_EXPR, ")"))  
  
 self.EXPR = Rule("EXPRESSION", Production(self.STR\_EXPR), Production(self.ARITH\_EXPR),  
 Production(self.LOG\_EXPR))  
  
 self.ARGS = Rule("ARGUMENTS", Production(self.EXPR))  
 self.ARGS.add(Production(self.EXPR, ",", self.ARGS))  
 self.STD\_FUNCS = Rule("STANDARD FUNCTIONS", Production("Math.log", "(", self.ARGS, ")"),  
 Production("Math.pow", "(", self.ARGS, ")"),  
 Production("Math.sqrt", "(", self.ARGS, ")"))  
 self.ARITH\_MUL.add(Production(self.STD\_FUNCS))  
  
 self.UNARY\_OPERS = Rule("UNARY OPERATORS", Production("++"), Production("--"))  
 self.ASSGN = Rule("ASSIGNMENT", Production(self.IDENTIFICATOR, "=", self.EXPR),  
 Production(self.EXPR, self.UNARY\_OPERS))  
 self.OUTPUT\_FUNC = Rule("OUTPUT FUNCTION", Production("System.out.print", "(", self.STR\_EXPR, ")"))  
  
 '''   
 '''  
 self.DECLARE\_VAR\_ONE = Rule("DECLARE A ONE VARIABLE", Production(self.IDENTIFICATOR),  
 Production(self.IDENTIFICATOR, "=", self.EXPR))  
 self.DECLARE\_VAR\_LIST = Rule("DECLARE A VARIABLE LIST", Production(self.DECLARE\_VAR\_ONE))  
 self.DECLARE\_VAR\_LIST.add(Production(self.DECLARE\_VAR\_ONE, ",", self.DECLARE\_VAR\_LIST))  
 self.DECLARE\_VAR = Rule("DECLARE A VARIABLES", Production(self.TYPE\_NAME, self.DECLARE\_VAR\_LIST))  
 '''  
 self.DECLARE\_VAR\_ONE = Rule("DECLARE A ONE VARIABLE", Production(self.TYPE\_NAME, self.IDENTIFICATOR),  
 Production(self.TYPE\_NAME, self.IDENTIFICATOR, "=", self.EXPR))  
 self.DECLARE\_VAR = Rule("DECLARE A VARIABLE", Production(self.DECLARE\_VAR\_ONE))  
 self.DECLARE\_VAR.add(Production(self.DECLARE\_VAR\_ONE, ",", self.DECLARE\_VAR))  
 '''  
  
 self.SUGGESTION = Rule("SUGGESTION", Production(self.DECLARE\_VAR), Production(self.OUTPUT\_FUNC),  
 Production(self.ASSGN))  
 self.SUGGESTION\_LIST = Rule("SUGGESTION\_LIST", Production(self.SUGGESTION, ";"))  
 self.SUGGESTION\_LIST.add(Production(self.SUGGESTION, ";", self.SUGGESTION\_LIST))  
  
 self.MAIN\_FUNC = Rule("MAIN FUNCTION", Production("public", "static", "void", "main",  
 "(", "String[]", "args", ")", "{", self.SUGGESTION\_LIST, "}"),  
 Production("public", "static", "void", "main",  
 "(", "String[]", "args", ")", "{", "}")  
 )  
  
 self.PROGRAMM = Rule("PROGRAMM", Production("public", "class", self.IDENTIFICATOR, "{", self.MAIN\_FUNC, "}"))  
  
 self.BREAK\_OPERS = Rule("BREAK OPEARTORS", Production("break"), Production("continue"))  
 self.INIT\_COUNT = Rule("INIT COUNTER", Production(""), Production(self.DECLARE\_VAR), Production(self.ASSGN))  
 self.MOD\_COUNT = Rule("MODIFY COUNTER", Production(""), Production(self.ASSGN))  
 self.COND = Rule("CONDITION", Production(self.LOG\_EXPR))  
 self.CYCLE\_FOR = Rule("CYCLE FOR", Production("for", "(", self.INIT\_COUNT, ";",  
 self.COND, ";", self.MOD\_COUNT, ")", "{",  
 self.SUGGESTION\_LIST, "}"),  
 Production("for", "(", self.INIT\_COUNT, ";", self.COND, ";", self.MOD\_COUNT,  
 ")", self.SUGGESTION))  
 self.CYCLE\_WHILE = Rule("CYCLE WHILE", Production("while", "(", self.COND, ")", "{",  
 self.SUGGESTION\_LIST, "}"))  
 self.CYCLE\_DO = Rule("CYCLE DO", Production("do", "{", self.SUGGESTION\_LIST, "}",  
 "while", "(", self.COND, ")", ))  
 self.CYCLE = Rule("CYCLE", Production(self.CYCLE\_FOR), Production(self.CYCLE\_DO), Production(self.CYCLE\_WHILE))  
  
 self.IF\_OPER = Rule("IF OPERATOR", Production("if", "(", self.COND, ")", "{",  
 self.SUGGESTION\_LIST, "}"),  
 Production("if", "(", self.COND, ")", "{", self.SUGGESTION\_LIST, "}",  
 "else", "{", self.SUGGESTION\_LIST, "}"),  
 '''  
 Production("if", "(", self.COND, ")", "{", self.SUGGESTION\_LIST, "}",  
 "else", self.SUGGESTION),  
 Production("if", "(", self.COND, ")", self.SUGGESTION,  
 "else", "{", self.SUGGESTION\_LIST, "}"),  
 Production("if", "(", self.COND, ")", self.SUGGESTION)  
 '''  
 )  
 self.SUGGESTION.add(Production(self.IF\_OPER), Production(self.CYCLE))  
  
 self.GAMMA\_RULE = u"GAMMA"  
  
 def predict(self, col, rule):  
 for prod in rule.productions:  
 col.add(State(rule.name, prod, 0, col))  
  
 def scan(self, col, state, token):  
 if token != col.token:  
 return  
 col.add(State(state.name, state.production, state.dot\_index + 1, state.start\_column))  
  
 def complete(self, col, state):  
 if not state.completed():  
 return  
 for st in state.start\_column:  
 term = st.next\_term()  
 if not isinstance(term, Rule):  
 continue  
 if term.name == state.name:  
 col.add(State(st.name, st.production, st.dot\_index + 1, st.start\_column))  
  
 def earley(self, rule, text):  
 table = [Column(i, tok) for i, tok in enumerate([None] + text.split())]  
 # axiom = rule.name  
 # predict(table[0], rule)  
 table[0].add(State(self.GAMMA\_RULE, Production(rule), 0, table[0]))  
  
 for i, col in enumerate(table):  
 for state in col:  
 if state.completed():  
 self.complete(col, state)  
 else:  
 term = state.next\_term()  
 if isinstance(term, Rule):  
 self.predict(col, term)  
 elif i + 1 < len(table):  
 self.scan(table[i + 1], state, term)  
 for st in table[-1]:  
 if st.name == self.GAMMA\_RULE and st.completed():  
 return table  
 else:  
 raise ValueError("Parsing failed")  
  
 def right\_parsing(self, table):  
 state = None  
 for st in table[-1]:  
 if st.name == self.GAMMA\_RULE and st.completed():  
 state = st  
 return self.sub\_parsing([], table, state, state.end\_column.index)  
  
 def sub\_parsing(self, acc, table, state: State, j: int):  
 acc.append(Rule(state.name, state.production))  
 k = len(state.production)  
 c = j  
 breaked = False  
 while k != 0:  
 breaked = False  
 Xk = state.production[k - 1]  
 if not isinstance(Xk, Rule):  
 k -= 1  
 c -= 1  
 else:  
 Ic = table[c]  
 # founding the state for Nonterminal Xk  
 for st in Ic:  
 if breaked:  
 break  
 if st.completed() and st.name == Xk.name:  
 r = st.start\_column.index  
 Ir = table[r]  
 expectedst = copy.copy(state)  
 expectedst.dot\_index = k - 1  
 # founding the previous state of Nonterminal Xk  
 for prevst in Ir:  
 # exit from cycle  
 if breaked:  
 break  
 if not prevst.completed() and prevst.isSame(expectedst):  
 self.sub\_parsing(acc, table, st, c)  
 k -= 1  
 c = r  
 breaked = True  
 return acc  
  
 def lextableToString(self, lextable):  
 *'''return a special format string for earley alghoritm from lexical table'''* res = ""  
 for i in lextable:  
 if i.type == Token.TYPETOKEN[0] or i.type == Token.TYPETOKEN[3] or i.type == Token.TYPETOKEN[5]:  
 for j in i.name:  
 res += ' ' + j  
 else:  
 res += ' ' + i.name  
 return res  
  
 def toTree(self, list):  
 return SyntacticalTree(list)

Полный код проекта представлен по ссылке: https://github.com/DieNice/JGT/

# Заключение

В ходе курсового проекта были решены следующие задачи:

1. Сформулирована неформальная постановка задачи.
2. Разработан синтаксис входного языка.
3. Описаны контекстные условия.
4. Создан проект лексического анализатора, включающий в себя таблицы соответствий, ключевых и зарезервированных слов, типов лексем, конечный автомат.
5. Создан проект синтаксического анализатора, включающий в себя модель данных, архитектуру классов синтаксического анализатора, проект алгоритма синтаксического анализа
6. Реализован лексический анализатор на языке Python.
7. Реализован синтаксический анализатор на языке Python.

# Список литературы

1. Хорстманн, К.С. Java. Библиотека профессионала [Текст]: в 2-х т. / К.С. Хорстманн. – М.: Вильямс, 2014. – 864 с. – 1 т.
2. Донован, А.А. Язык программирования Go. [Текст] / А.А. Донован, Б.У. Керниган. – М.: Вильямс, 2016. – 432 с.
3. Ахо, А. Теория синтаксического анализа, перевода и компиляции [Текст]: в 2-х т. / А. Ахо, Д. Ульман. – М.: Мир, 1978 – 384 с. – 1 т.
4. Серебряков, В.А. Основы конструирования компиляторов [Текст] / В.А. Серебряков, М.П. Галочкин. – М.: УРСС, 2001 – 192 с.
5. Грис, Д. Конструирование компиляторов для цифровых вычислительных машин [Текст] / Д. Грис. - М. : Мир, 1975 – 575 с.