Міністерство освіти і науки України

Національний університет “Львівська політехніка”



**Курсовий проект**

З дисципліни «Системне програмування»

на тему: "Розробка системних програмних модулів

та компонент систем програмування.

Розробка транслятора з вхідної мови програмування"

**Варіант №18**

**Виконала:** ст. гр. КІ-308

Продан К.О.

**Перевірив:**

Козак Н. Б.

Львів-2024

Анотація

Цей курсовий проект приводить до розробки транслятора, який здатен конвертувати вхідну мову, визначену відповідно до варіанту, у мову C. Процес трансляції включає в себе лексичний аналіз, синтаксичний аналіз та генерацію коду.

Лексичний аналіз розбиває вхідну послідовність символів на лексеми, які записуються у відповідну таблицю лексем. Кожній лексемі присвоюється числове значення для полегшення порівнянь, а також зберігається додаткова інформація, така як номер рядка, значення (якщо тип лексеми є числом) та інші деталі.

Синтаксичний аналіз: використовується висхідний метод аналізу без повернення. Призначений для побудови дерева розбору, послідовно рухаючись від листків вгору до кореня дерева розбору.

Генерація коду включає повторне прочитання таблиці лексем та створення відповідного коду на мові С для кожного блоку лексем. Отриманий код записується у результуючий файл, готовий для виконання.

Зміст

[Анотація 2](#_Toc187251040)

[Завдання до курсового проекту 4](#_Toc187251041)

[Вступ 5](#_Toc187251042)

[1. Огляд методів та способів проектування трансляторів 6](#_Toc187251043)

[2. Формальний опис вхідної мови програмування 9](#_Toc187251044)

[2.1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура 9](#_Toc187251045)

[2.2. Опис термінальних символів та ключових слів 11](#_Toc187251046)

[3. Розробка транслятора вхідної мови програмування 13](#_Toc187251047)

[3.1. Вибір технології програмування 13](#_Toc187251048)

[3.2. Проектування таблиць транслятора 14](#_Toc187251049)

[3.3. Розробка лексичного аналізатора 16](#_Toc187251050)

[3.3.1. Розробка блок-схеми алгоритму 18](#_Toc187251051)

[3.3.2. Опис програми реалізації лексичного аналізатора 19](#_Toc187251052)

[3.4. Розробка синтаксичного та семантичного аналізатора 22](#_Toc187251053)

[3.4.1. Розробка алгоритму роботи синтаксичного і семантичного аналізатора 23](#_Toc187251054)

[3.4.2. Опис програми реалізації синтаксичного та семантичного аналізатора 24](#_Toc187251055)

[3.4.3. Розробка граф-схеми алгоритму 27](#_Toc187251056)

[3.5. Розробка генератора коду 27](#_Toc187251057)

[3.5.1. Розробка граф-схеми алгоритму 29](#_Toc187251058)

[3.5.2. Опис програми реалізації генератора коду 30](#_Toc187251059)

[4. Опис програми 31](#_Toc187251060)

[4.1. Опис інтерфейсу та інструкція користувачеві 34](#_Toc187251061)

[5. Відлагодження та тестування програми 35](#_Toc187251062)

[5.1. Виявлення лексичних та синтаксичних помилок 35](#_Toc187251063)

[5.2. Виявлення семантичних помилок 36](#_Toc187251064)

[5.3. Загальна перевірка коректності роботи транслятора 36](#_Toc187251065)

[5.4. Тестова програма №1 37](#_Toc187251066)

[5.5. Тестова програма №2 37](#_Toc187251067)

[5.6. Тестова програма №3 39](#_Toc187251068)

[Висновки 41](#_Toc187251069)

[Список використаної літератури 42](#_Toc187251070)

[Додатки 43](#_Toc187251071)

[Додаток А (таблиці лексем для тесових програм) 43](#_Toc187251072)

[Додаток Б (Код на мові C) 57](#_Toc187251073)

[Додаток В (Абстрактне синтаксичне дерево для тестових прикладів) 60](#_Toc187251074)

[Додаток Г (Документований текст програмних модулів (лістинги)) 68](#_Toc187251075)

[Додаток Д (тестові програми) 120](#_Toc187251076)

[Додаток Е (Алгоритм транслятора В03) 124](#_Toc187251077)

Завдання до курсового проекту

**Варіант 18**

Завдання на курсовий проект

1. Цільова мова транслятора – мова програмування С.

2. Для отримання виконавчого файлу на виході розробленого транслятора скористатися середовищем Microsoft Visual Studio або будь-яким іншим.

3. Мова розробки транслятора: C++.

4. Реалізувати оболонку або інтерфейс з командного рядка.

5. На вхід розробленого транслятора має подаватися текстовий файл, написаний на заданій мові програмування.

6. На виході розробленого транслятора мають створюватись такі файли:

* *файл з лексемами;*
* *файл з повідомленнями про помилки (або про їх відсутність);*
* *файл на мові C;*
* *об’єктний файл;*
* *виконавчий файл.*

7. Назва вхідної мови програмування утворюється від першої букви у прізвищі студента та останніх двох цифр номера його варіанту. Саме таке розширення повинні мати текстові файли, написані на цій мові програмування.

В моєму випадку це .p18

Опис вхідної мови програмування:

* Тип даних: Int32
* Блок тіла програми: StartProgram StartBlok Variable…; EndBlok
* Оператор вводу: Scan
* Оператор виводу: Print
* Оператори: If Else (C)

Goto (C)

For-To-Do (Паскаль)

For-DownTo-Do (Паскаль)

While (Бейсік)

Repeat-Until (Паскаль)

* Регістр ключових слів: Up-Low перший символ Up
* Регістр ідентифікаторів: Low-Up6 перший символ \_
* Операції арифметичні: ++, --, \*\*, Div, Mod
* Операції порівняння: Eq, Neq, Ls, Gr
* Операції логічні: Not, And, Or
* Коментар: { … }
* Ідентифікатори змінних, числові константи
* Оператор присвоєння: ==>

Вступ

Термін "транслятор" визначає програму, яка виконує переклад (трансляцію) початкової програми, написаної на вхідній мові, у еквівалентну їй об'єктну програму. У випадку, коли мова високого рівня є вхідною, а мова асемблера або машинна – вихідною, такий транслятор отримує назву компілятора.

Транслятори можуть бути розділені на два основних типи: компілятори та інтерпретатори. Процес компіляції включає дві основні фази: аналіз та синтез. Під час аналізу вхідну програму розбивають на окремі елементи (лексеми), перевіряють її відповідність граматичним правилам і створюють проміжне представлення програми. На етапі синтезу з проміжного представлення формується програма в машинних кодах, яку називають об'єктною програмою. Останню можна виконати на комп'ютері без додаткової трансляції.

У відміну від компіляторів, інтерпретатор не створює нову програму; він лише виконує – інтерпретує – кожну інструкцію вхідної мови програмування. Подібно компілятору, інтерпретатор аналізує вхідну програму, створює проміжне представлення, але не формує об'єктну програму, а негайно виконує команди, передбачені вхідною програмою.

Компілятор виконує переклад програми з однієї мови програмування в іншу. На вхід компілятора надходить ланцюг символів, який представляє вхідну програму на певній мові програмування. На виході компілятора (об'єктна програма) також представляє собою ланцюг символів, що вже відповідає іншій мові програмування, наприклад, машинній мові конкретного комп'ютера. При цьому сам компілятор може бути написаний на третій мові.

1. Огляд методів та способів проектування трансляторів

Термін "транслятор" визначає обслуговуючу програму, що проводить трансляцію вихідної програми, представленої на вхідній мові програмування, у робочу програму, яка відображена на іншій цільовій мові програмування, такій як C. Наведене визначення застосовне до різноманітних транслюючих програм. Однак кожна з таких програм може виявляти свої особливості в організації процесу трансляції. В сучасному контексті транслятори поділяються на три основні групи: асемблери, компілятори та інтерпретатори.

Компілятор - обслуговуюча програма, яка виконує трансляцію програми, написаної мовою оригіналу програмування, в іншу мову, наприклад, мову C. Схоже до асемблера, компілятор виконує перетворення програми з однієї мови в іншу, часто генеруючи код, який може бути виконаний іншими компіляторами.

Інтерпретатор - це програма чи пристрій, що виконує пооператорну трансляцію та виконання вихідної програми. Відмінно від компілятора, інтерпретатор не створює на виході нову програму мовою C. Розпізнавши команду вихідної мови, він негайно її виконує, забезпечуючи більшу гнучкість у процесі розробки та налагодження програм.

Процес трансляції включає фази лексичного аналізу, синтаксичного та семантичного аналізу, оптимізації коду та генерації коду. Лексичний аналіз розбиває вхідну програму на лексеми, що представляють слова відповідно до визначень мови. Синтаксичний аналіз визначає структуру програми, створюючи синтаксичне дерево. Семантичний аналіз виявляє залежності між частинами програми, недосяжні контекстно-вільним синтаксисом. Оптимізація коду та генерація коду спрямовані на створення коду мовою C, з урахуванням ефективності його виконання.

Зазначені фази можуть об'єднуватися або відсутні у трансляторах залежно від їхньої реалізації. Наприклад, у простих однопрохідних трансляторах може бути відсутня фаза генерації проміжного представлення та оптимізації, а інші фази можуть об'єднуватися.

Під час процесу виділення лексем лексичний аналізатор може виконувати дві основні функції: автоматичну побудову таблиць об'єктів (таких як ідентифікатори, рядки, числа тощо) і видачу значень для кожної лексеми при кожному новому зверненні до нього. У цьому контексті таблиці об'єктів формуються в подальших етапах, наприклад, під час синтаксичного аналізу.

На етапі лексичного аналізу виявляються деякі прості помилки, такі як неприпустимі символи або невірний формат чисел та ідентифікаторів.

Основним завданням синтаксичного аналізу є розбір структури програми. Зазвичай під структурою розуміється дерево, яке відповідає розбору в контекстно-вільній граматиці мови програмування. У сучасній практиці найчастіше використовуються методи аналізу, такі як LL(1) або LR(1) та їхні варіанти (рекурсивний спуск для LL(1) або LR(1), LR(0), SLR(1), LALR(1) та інші для LR(1)). Рекурсивний спуск застосовується частіше при ручному програмуванні синтаксичного аналізатора, тоді як LR(1) використовується при автоматичній генерації синтаксичних аналізаторів.

Результатом синтаксичного аналізу є синтаксичне дерево з посиланнями на таблиці об'єктів. Під час синтаксичного аналізу також виявляються помилки, пов'язані зі структурою програми.

На етапі контекстного аналізу виявляються взаємозалежності між різними частинами програми, які не можуть бути адекватно описані за допомогою контекстно-вільної граматики. Ці взаємозалежності, зокрема, включають аналіз типів об'єктів, областей видимості, відповідності параметрів, міток та інших аспектів "опис-використання". У ході контекстного аналізу таблиці об'єктів доповнюються інформацією, пов'язаною з описами (властивостями) об'єктів.

В основі контекстного аналізу лежить апарат атрибутних граматик. Результатом цього аналізу є створення атрибутованого дерева програми, де інформація про об'єкти може бути розсіяна в самому дереві чи сконцентрована в окремих таблицях об'єктів. Під час контекстного аналізу також можуть бути виявлені помилки, пов'язані з неправильним використанням об'єктів.

Після завершення контекстного аналізу програма може бути перетворена у внутрішнє представлення. Це здійснюється з метою оптимізації та/або для полегшення генерації коду мовою C. Крім того, перетворення програми у внутрішнє представлення може бути використано для створення переносимого транслятора. У цьому випадку, тільки остання фаза (генерація коду) є залежною від конкретної архітектури. В якості внутрішнього представлення може використовуватися орієнтований граф, трійки, четвірки та інші формати.

Фаза оптимізації транслятора може включати декілька етапів, які спрямовані на покращення якості та ефективності згенерованого коду. Ці оптимізації часто розподіляються за двома головними критеріями: машинно-залежні та машинно-незалежні, а також локальні та глобальні.

Машинно-незалежні оптимізації орієнтовані на спрощення коду або видалення надлишкових обчислень, тоді як машинно-залежні оптимізації проводяться на етапі генерації коду.

Фінальна фаза трансляції - генерація коду мовою C. На цьому етапі можуть застосовуватися деякі локальні оптимізації для полегшення генерації ефективного та читабельного коду.

Важливо відзначити, що фази транслятора можуть бути відсутніми або об'єднаними залежно від конкретної реалізації. У простіших випадках, таких як однопрохідні транслятори, може бути відсутній окремий етап генерації проміжного представлення та оптимізації, а інші фази можуть бути об'єднані в одну, без створення явно побудованого синтаксичного дерева.

1. Формальний опис вхідної мови програмування
   1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура

Однією з перших задач, що виникають при побудові компілятора, є визначення вхідної мови програмування. Для цього використовують різні способи формального опису, серед яких я застосував розширену нотацію Бекуса-Наура (extended Backus/Naur Form - EBNF).

**<програма>** ::= ‘StartProgram’ ‘;’ ‘Variable’ <оголошення змінних> ‘;’ <тіло програми> ‘EndBlok’ ‘;’

**<оголошення змінних>** ::= [<тип даних> <список змінних>]

**<тип даних>** ::= ‘Int32’

**<список змінних>** ::= <ідентифікатор> [‘==>’ <арифметичний вираз>] { ‘,’ **<ідентифікатор>** [‘==>’ <арифметичний вираз>] }

**<ідентифікатор>** ::= '\_' <буква> { <буква або цифра> }

**<буква або цифра>** ::= <буква> | <цифра>

**<буква>** ::= <велика буква> | <мала буква>

**<велика буква>** ::= ‘A’ | ‘B’ | ‘C’ | … | ‘Z’

**<мала буква>** ::= ‘a’ | ‘b’ | ‘c’ | … | ‘z’

**<цифра>** ::= ‘0’ | ‘1’ | ‘2’ | ‘3’ | ‘4’ | ‘5’ | ‘6’ | ‘7’ | ‘8’ | ‘9’

**<арифметичний вираз>** ::= <доданок> { ‘++’ <доданок> | ‘--’ <доданок> }

**<доданок>** ::= <множник> { ‘\*\*’ <множник> | ‘Div’ <множник> | ‘Mod’ <множник> }

**<множник>** ::= <ідентифікатор> | <число> | ‘(’ <арифметичний вираз> ‘)’

**<число>** ::= [‘-’ | ‘+’] <цифра> { <цифра> }

**<тіло програми>** ::= { <оператор> }

**<оператор>** ::= <мітка> | <присвоєння> | <ввід> | <вивід> | <умовний оператор> | <безумовний перехід> | <оператор циклу forTo> | <оператор циклу forDownTo> | <оператор циклу while> | <оператор циклу repeat> | <складений оператор>

**<мітка>** ::= <ідентифікатор> ‘:’

**<присвоєння>** ::= <ідентифікатор> ‘==>’ <арифметичний вираз> ‘;’

**<ввід>** ::= ‘Scan’ <ідентифікатор> ‘;’

**<вивід>** ::= ‘Print’ <арифметичний вираз> ‘;’

**<умовний оператор>** ::= ‘If’ <логічний вираз> ‘Goto’ <мітка>

**<логічний вираз>** ::= <вираз І> { ‘Or’ <вираз І> }

**<вираз І>** ::= <порівняння> { ‘And’ <порівняння> }

**<порівняння>** ::= <операція порівняння> | ‘Not’ ‘(’ <логічний вираз> ‘)’ | ‘(’ <логічний вираз> ‘)’

**<операція порівняння>** ::= <арифметичний вираз> <менше-більше> <арифметичний вираз>

**<менше-більше>** ::= ‘Eq’ | ‘Neq’ | ‘Ls’ | ‘Gr’

**<оператор циклу forTo>** ::= ‘For’ <ідентифікатор> ‘==>’ <арифметичний вираз> ‘To’ <арифметичний вираз> ‘Do’ <оператор>

**<оператор циклу forDownTo>** ::= ‘For’ <ідентифікатор> ‘==>’ <арифметичний вираз> ‘DownTo’ <арифметичний вираз> ‘Do’ <оператор>

**<оператор циклу while>** ::= ‘While’ ‘(’ <логічний вираз> ‘)’ ‘Do’ <оператор> ‘EndWhile’

**<оператор циклу repeat>** ::= ‘Repeat’ <оператор> ‘Until’ ‘(’ <логічний вираз> ‘)’ ‘;’

**<безумовний перехід>** ::= ‘Goto’ <мітка> ‘;’

**<складений оператор>** ::= ‘StartBlok’ <тіло програми> ‘EndBlok’ ‘;’

* 1. Опис термінальних символів та ключових слів

Визначимо окремі термінальні символи та нерозривні набори термінальних символів (ключові слова):

|  |  |
| --- | --- |
| Термінальний символ або ключове слово | Значення |
| StartProgram | Початок програми |
| StartBlok | Початок тексту програми |
| Variable | Початок блоку опису змінних |
| EndBlok | Кінець розділу операторів |
| Scan | Оператор вводу змінних |
| Print | Оператор виводу (змінних або рядкових констант) |
| ==> | Оператор присвоєння |
| If | Оператор умови |
| Else | Оператор умови |
| Goto | Оператор переходу |
| Label | Мітка переходу |
| For | Оператор циклу |
| To | Інкремент циклу |
| DownTo | Декремент циклу |
| Do | Початок тіла циклу |
| While | Оператор циклу |
| Repeat | Початок тіла циклу |
| Until | Оператор циклу |
| ++ | Оператор додавання |
| -- | Оператор віднімання |
| \*\* | Оператор множення |
| Div | Оператор ділення |
| Mod | Оператор знаходження залишку від ділення |
| Eq | Оператор перевірки на рівність |
| Neq | Оператор перевірки на нерівність |
| Ls | Оператор перевірки чи менше |
| Gr | Оператор перевірки чи більше |
| Not | Оператор логічного заперечення |
| And | Оператор кон’юнкції |
| Or | Оператор диз’юнкції |
| Int32 | 16-ти розрядні знакові цілі |
| {…} | Коментар |
| , | Розділювач |
| ; | Ознака кінця оператора |
| ( | Відкриваюча дужка |
| ) | Закриваюча дужка |

До термінальних символів віднесемо також усі цифри (0-9), латинські букви (a-z, A-Z), символи табуляції, символ переходу на нову стрічку, пробілу.

1. Розробка транслятора вхідної мови програмування
   1. Вибір технології програмування

Для ефективної роботи створюваної програми важливу роль відіграє попереднє складення алгоритму роботи програми, алгоритму написання програми і вибір технології програмування.

Тому при складанні транслятора треба брати до уваги швидкість компіляції, якість об’єктної програми. Проект повинен давати можливість просто вносити зміни.

В реалізації мов високого рівня часто використовується специфічний тільки для компіляції засіб “розкрутки”. З кожним транслятором завжди зв`язані три мови програмування: Х – початкова, Y – об`єктна та Z – інструментальна. Транслятор перекладає програми мовою Х в програми, складені мовою Y, при цьому сам транслятор є програмою написаною мовою Z.

При розробці даного курсового проекту був використаний висхідний метод синтаксичного аналізу.

Також був обраний прямий метод лексичного аналізу. Характерною ознакою цього методу є те, що його реалізація відбувається без повернення назад. Його можна сприймати, як один спільний скінченний автомат. Такий автомат на кожному кроці читає один вхідний символ і переходить у наступний стан, що наближає його до розпізнавання поточної лексеми чи формування інформації про помилки. Для лексем, що мають однакові підланцюжки, автомат має спільні фрагменти, що реалізують єдину множину станів. Частини, що відрізняються, реалізуються своїми фрагментами

* 1. Проектування таблиць транслятора

Використання таблиць значно полегшує створення трансляторів, тому у даному випадку використовуються наступне:

1. Таблиця лексем з елементами, які мають таку структуру:

struct Token

{

char name[16]; // ім'я лексеми

int value; // значення лексеми (для цілих констант)

int line; // номер рядка

TypeOfTokens type; // тип лексеми

};

1. Таблиця лексичних класів

enum TypeOfTokens

{

Mainprogram,

StartProgram,

Variable,

Type,

EndProgram,

Input,

Output,

If,

Else,

Goto,

Label,

For,

To,

DownTo,

Do,

While,

Repeat,

Until,

Identifier,

Number,

Assign,

Add,

Sub,

Mul,

Div,

Mod,

Equality,

NotEquality,

Greate,

Less,

Not,

And,

Or,

LBraket,

RBraket,

Semicolon,

Colon,

Comma,

Unknown

};

Якщо у стовпці «Значення» відсутня інформація про токен, то це означає що його значення визначається користувачем під час написання коду на створеній мові програмування.

Таблиця 2 Опис термінальних символі та ключових слів

|  |  |
| --- | --- |
| **Токен** | **Значення** |
| StartProgram | StartProgram |
| StartBody | StartBlok |
| Variable | Variable |
| End | EndBlok |
| Type | Int32 |
| Input | Scan |
| Output | Print |
| Assign | ==> |
| If | If |
| Else | Else |
| Goto | Goto |
| Label |  |
| For | For |
| To | To |
| DownTo | DownTo |
| Do | Do |
| While | While |
| Repeat | Repeat |
| Until | Until |
| Add | ++ |
| Sub | -- |
| Mul | \*\* |
| Div | Div |
| Mod | Mod |
| Equality | Eg |
| NotEquality | Ne |
| Less | Ls |
| Greate | Gr |
| Not | Not |
| And | And |
| Or | Or |
| Identifier |  |
| Number |  |
| Unknown |  |
| Comma | , |
| Semicolon | ; |
| LBraket | ( |
| RBraket | ) |

* 1. Розробка лексичного аналізатора

На фазі лексичного аналізу вхідна програма, що представляє собою потік літер, розбивається на лексеми - слова у відповідності з визначеннями мови. Лексичний аналізатор може працювати в двох основних режимах: або як підпрограма, що викликається синтаксичним аналізатором для отримання чергової лексеми, або як повний прохід, результатом якого є файл лексем.

Для нашої програми виберемо другий варіант. Тобто, спочатку буде виконуватись фаза лексичного аналізу. Результатом цієї фази буде файл з списком лексем. Але лексеми записуються у файл не як послідовність символів. Кожній лексемі присвоюється певний символ, тип, значення та рядок. Ці дані далі записуються у файл. Такий підхід дозволяє спростити роботу синтаксичного аналізатора.

Також на етапі лексичного аналізу виявляються деякі (найпростіші) помилки (неприпустимі символи, неправильний запис чисел, ідентифікаторів та ін.)

На вхід лексичного аналізатора надходить текст вихідної програми, а вихідна інформація передається для подальшої обробки компілятором на етапі синтаксичного аналізу.

Існує кілька причин, з яких до складу практично всіх компіляторів включають лексичний аналіз:

* застосування лексичного аналізатора спрощує роботу з текстом вихідної програми на етапі синтаксичного розбору;
* для виділення в тексті та розбору лексем можливо застосовувати просту, ефективну і теоретично добре пророблену техніку аналізу;

За варіантом розділено лексеми на типи або лексичні класи:

* Ключові слова (0-StartProgram, 1-Variable, 2-StartBlok, 3-EndBlok, 4-Scan, 5-Print, 6-Int32, 7-If, 8-Else, 9-Goto, 10-For, 11-To, 12-Do, 13-Downto, 14-While, 15-Repeat,16-Until, 17-Exit)
* Ідентифікатори (18-максимум 6 великих літер)
* Числові константи (19-ціле число)
* Оператор присвоєння (20- ==>)
* Знаки операції (21- ++, 22- --, 23- \*\*, 24- Div, 25- Mod, 26- Ls, 27- Gr, 28- Eq, 29- Neq, 30- Not, 31- And, 32- Or)
* Розділювачі(33- ;, 34- ,)
* Дужки (35- (, 36- ))
* Невідома лексема (37- символи і ланцюжки символів, які не підпадають під вище описані правила)
  + 1. Розробка блок-схеми алгоритму

Лексичний аналізатор виконує обробку тексту з вхідного файлу та визначає типи лексем за наступними етапами. Процес роботи можна описати наступним чином::

* **Початок роботи:**  
  алгоритм починається зі зчитування чергового слова або символу з файлу (state).
* **Перевірка на кінець файлу:**  
  якщо зчитане значення state дорівнює EndOfFile, тип лексеми визначається як EndOfFile, після чого завершується обробка.
* **Ключове слово або символ:**  
  якщо зчитане значення є символом, виконується обробка лексеми, і тип визначається як KeyWords.
* **Розпізнавання слова (ідентифікатора):**  
  якщо значення є підкремленням, проводиться обробка ідентифікатора, а тип встановлюється як Identifiers.
* **Розпізнавання числа:**  
  якщо значення є цифрою, тип лексеми визначається як Digit.
* **Невідомий тип:**  
  якщо жодна з вищезазначених умов не виконана, тип лексеми визначається як Another.
* **Завершення роботи:**  
  після визначення типу лексеми алгоритм повертається до зчитування наступного слова або символу, доки не буде досягнуто стану EndOfFile.

Алгоритм працює на основі автомату з наступними станами:

* **Start** – початковий стан, початок виділення чергової лексеми.
* **Finish** – кінець виділення чергової лексеми.
* **EndOfFile** – кінцевий стан, завершення обробки файлу.
* **KeyWords** – обробка зарезервованих слів або символів.
* **Identifiers** – обробка лексем, що є ідентифікаторами.
* **Separators** – видалення пробілів, символів табуляції і переходу на новий рядок.
* **Digit** – обробка числових значень.
* **Another** – обробка зарезервованих символів.
* **Comment** – видалення коментаря.

Ця модель дозволяє покроково аналізувати вхідний файл, виділяючи ключові слова, ідентифікатори, числа та інші елементи, забезпечуючи коректну обробку кожної лексеми.

Алгоритм роботи лексичного аналізатора можна зобразити у вигляді блок-схеми.



Рис. 3.1 Блок-схема роботи лексичного аналізатора

* + 1. Опис програми реалізації лексичного аналізатора

Основна задача лексичного аналізу – розбити вихідний текст, що складається з послідовності одиночних символів, на послідовність слів, або лексем, тобто виділити ці слова з безперервної послідовності символів. Всі символи вхідної послідовності з цієї точки зору розділяються на символи, що належать яким-небудь лексемам, і символи, що розділяють лексеми. В цьому випадку використовуються звичайні засоби обробки рядків. Вхiдна програма проглядається послідовно з початку до кінця. Базовi елементи, або лексичнi одиницi, роздiляються пробілами, знаками операцiй i спецiальними символами (новий рядок, знак табуляції), i таким чином видiляються та розпізнаються iдентифiкатори, лiтерали i термiнальнi символи (операцiї, ключові слова).

Програма аналізує файл поки не досягне його кінця. Для вхідного файлу викликається функція Parser (). Вона зчитує з файлу його вміст та кожну лексему порівнює з зарезервованою словами якщо є співпадіння то присвоює лексемі відповідний тип або значення, якщо це числова константа.

При виділенні лексеми вона розпізнається та записується у таблиця за допомогою відповідного типу лексеми, що є унікальним для кожної лексеми із усього можливого їх набору. Це дає можливість наступним фазам компiляції звертатись до лексеми не як до послідовності символів, а як до унікального типу лексеми, що значно спрощує роботу синтаксичного аналізатора: легко перевіряти належність лексеми до відповідної синтаксичної конструкції та є можливість легкого перегляду програми, як вгору, так і вниз, від поточної позиції аналізу. Також в таблиці лексем ведуться записи, щодо рядка відповідної лексеми – для місця помилки – та додаткова інформація.

При лексичному аналiзі виявляються i вiдзначаються лексичнi помилки (наприклад, недопустимi символи i неправильнi iдентифiкатори). Лексична фаза вiдкидає також коментарi, оскiльки вони не мають нiякого впливу на виконання програми, отже й на синтаксичний розбір та генерацію коду.

В даному курсовому проекті реалізовано прямий лексичний аналізатор, який виділяє з вхідного тексту програми окремі лексеми і на основі цього формує таблицю.

*Створимо структуру даних для зберігання стану аналізатора:*

// перерахування, яке описує стани лексичного аналізатора

enum States

{

Start, // початок виділення чергової лексеми

Finish, // кінець виділення чергової лексеми

Letter, // опрацювання слів (ключові слова і ідентифікатори)

Digit, // опрацювання цифри

Separators, // видалення пробілів, символів табуляції і переходу на новий рядок

Another, // опрацювання інших символів

EndOfFile, // кінець файлу

SComment, // початок коментаря

Comment // видалення коментаря

};

*Напишемо функцію, яка реалізує лексичний аналіз:*

// функція отримує лексеми з вхідного файлу F і записує їх у таблицю лексем TokenTable

// результат функції - кількість лексем

unsigned int GetTokens(FILE\* F, Token TokenTable[], FILE\* errFile)

*І функції, які друкують список лексем:*

// функція друкує таблицю лексем на екран

void PrintTokens(Token TokenTable[], unsigned int TokensNum);

// функція друкує таблицю лексем у файл

void PrintTokensToFile(char\* FileName, Token TokenTable[], unsigned int TokensNum);

* 1. Розробка синтаксичного та семантичного аналізатора

Синтаксичний аналізатор - частина компілятора, яка відповідає за виявлення основних синтаксичних конструкцій вхідної мови. У завдання синтаксичного аналізатора входить: знайти і виділити основні синтаксичні конструкції в тексті вхідної програми, встановити тип і перевірити правильність кожної синтаксичної конструкції у вигляді, зручному для подальшої генерації тексту результуючої програми.

В основі синтаксичного аналізатора лежить розпізнавач тексту вхідної програми на основі граматики вхідної мови. Як правило, синтаксичні конструкції мов програмування можуть бути описані за допомогою КС-граматик, рідше зустрічаються мови, які можуть бути описані за допомогою регулярних граматик. Найчастіше регулярні граматики застосовні до мов асемблера, а мови високого рівня побудовані на основі КС-мов.

Синтаксичний розбір - це основна частина компіляції на етапі аналізу. Без виконання синтаксичного розбору робота компілятора безглузда, у той час як лексичний аналізатор є зовсім необов'язковим. Усі завдання з перевірки лексики вхідного мови можуть бути вирішені на етапі синтаксичного розбору. Сканер тільки дозволяє позбавити складний за структурою лексичний аналізатор від рішення примітивних завдань з виявлення та запам'ятовування лексем вхідний програми.

В даному курсовому проекті синтаксичний аналіз можна виконувати лише після виконання лексичного аналізу, він являється окремим етапом трансляції.

На вході даного аналізатора є файл лексем, який є результатом виконання лексичного аналізу, на базі цього файлу синтаксичний аналізатор формує таблицю ідентифікаторів та змінних.

* + 1. Розробка алгоритму роботи синтаксичного і семантичного аналізатора

Одним із найбільш простих і популярних методів низхідного синтаксичного аналізу є метод рекурсивного спуску (recursive descent method). Цей метод ґрунтується на тому, що синтаксичний аналізатор складається з набору рекурсивних процедур для кожного правила граматики. У даній курсовій роботі метод рекурсивного спуску реалізовано відповідно до граматики мови p18.

Назви процедур, що відповідають нетерміналам граматики, визначено наступним чином:

// Набір функцій для рекурсивного спуску

// Для кожного правила – окрема функція

void program(); // <program> ::= "StartProgram" ";" <variable\_declaration> ";" <program\_body> "EndBlok" ";"

void variable\_declaration(); // <variable\_declaration> ::= "Int32" <variable\_list>

void variable\_list(); // <variable\_list> ::= <identifier> [ '=>' <arithmetic\_expression> ] { ',' <identifier> [ '=>' <arithmetic\_expression> ] }

void program\_body(); // <program\_body> ::= { <statement> }

void statement(); // <statement> ::= <assignment> | <input> | <output> | <conditional> | <goto\_statement> | <label> | <loop> | <compound\_statement>

void assignment(); // <assignment> ::= <identifier> '=>' <arithmetic\_expression> ';'

void arithmetic\_expression(); // <arithmetic\_expression> ::= <term> { ('++' | '--') <term> }

void term(); // <term> ::= <factor> { ('\*\*' | 'Div' | 'Mod') <factor> }

void factor(); // <factor> ::= <identifier> | <number> | '(' <arithmetic\_expression> ')'

void input(); // <input> ::= 'Scan' <identifier> ';'

void output(); // <output> ::= 'Print' <arithmetic\_expression> ';'

void conditional(); // <conditional> ::= 'If' '(' <logical\_expression> ')' <statement> ['Else' <statement>]

void goto\_statement(); // <goto\_statement> ::= 'Goto' <label> ';'

void label(); // <label> ::= <identifier> ':'

void loop(); // <loop> ::= <for\_loop> | <while\_loop> | <repeat\_loop>

void for\_loop(); // <for\_loop> ::= 'For' <identifier> '=>' <arithmetic\_expression> ('To' | 'DownTo') <arithmetic\_expression> 'Do' <statement>

void while\_loop(); // <while\_loop> ::= 'While' '(' <logical\_expression> ')' 'Do' <statement>

void repeat\_loop(); // <repeat\_loop> ::= 'Repeat' <statement> 'Until' '(' <logical\_expression> ')' ';'

void logical\_expression(); // <logical\_expression> ::= <comparison> { 'And' <comparison> | 'Or' <comparison> }

void comparison(); // <comparison> ::= <arithmetic\_expression> ('Eq' | 'Neq' | 'Ls' | 'Gr') <arithmetic\_expression>

void compound\_statement(); // <compound\_statement> ::= 'StartBlok' <program\_body> 'EndBlok' ';'

* + 1. Опис програми реалізації синтаксичного та семантичного аналізатора

На вхід синтаксичного аналізатора подіється таблиця лексем створена на етапі лексичного аналізу. Аналізатор проходить по ній і перевіряє чи набір лексем відповідає раніше описаним формам нотації Бекуса-Наура. І разі не відповідності у файл з помилками виводиться інформація про помилку і про рядок на якій вона знаходиться.

При знаходженні оператора присвоєння або математичних виразів здійснюється перевірка балансу дужок(кількість відкриваючих дужок має дорівнювати кількості закриваючих). Також здійснюється перевірка чи не йдуть підряд декілька лексем одного типу

Результатом синтаксичного аналізу є синтаксичне дерево з посиланнями на таблиці об'єктів. У процесі синтаксичного аналізу також виявляються помилки, пов'язані зі структурою програми.

В основі синтаксичного аналізатора лежить розпізнавач тексту вхідної програми на основі граматики вхідної мови.

Аналізатор працює за принципом рекурсивного спуску, де кожне правило граматики реалізується окремою функцією.

Основні етапи роботи аналізатора:

1. **Ініціалізація**: Виклик функції Parser(), яка починає аналіз програми.
2. **Аналіз програми**: Функція program() аналізує основну структуру програми, включаючи оголошення змінних та тіло програми.
3. **Аналіз операторів**: Функція statement() визначає тип оператора (ввід, вивід, умовний оператор, присвоєння тощо) та викликає відповідну функцію для його аналізу.
4. **Аналіз виразів**: Функції arithmetic\_expression(), term(), factor() аналізують арифметичні вирази, включаючи операції додавання, віднімання, множення та ділення.
5. **Аналіз умов**: Функції logical\_expression(), and\_expression(), comparison() аналізують логічні вирази та операції порівняння.

Основні функції

* **program()**: Аналізує основну структуру програми.
* **variable\_declaration()**: Аналізує оголошення змінних.
* **variable\_list()**: Аналізує список змінних.
* **program\_body()**: Аналізує тіло програми.
* **statement()**: Визначає тип оператора та викликає відповідну функцію для його аналізу.
* **assignment()**: Аналізує оператор присвоєння.
* **arithmetic\_expression()**: Аналізує арифметичний вираз.
* **term()**: Аналізує доданок у виразі.
* **factor()**: Аналізує множник у виразі.
* **input()**: Аналізує оператор вводу.
* **output()**: Аналізує оператор виводу.
* **conditional()**: Аналізує умовний оператор.
* **goto\_statement()**: Аналізує оператор переходу.
* **label\_statement()**: Аналізує мітку.
* **for\_to\_do()**: Аналізує цикл for з інкрементом.
* **for\_downto\_do()**: Аналізує цикл for з декрементом.
* **while\_statement()**: Аналізує цикл while.
* **repeat\_until()**: Аналізує цикл repeat until.
* **logical\_expression()**: Аналізує логічний вираз.
* **and\_expression()**: Аналізує логічний вираз з операцією AND.
* **comparison()**: Аналізує операції порівняння.
* **compound\_statement()**: Аналізує складений оператор.

Цей аналізатор забезпечує перевірку синтаксичної коректності програми та виявлення синтаксичних помилок. Якщо виявляється помилка, аналізатор виводить повідомлення про помилку та завершує роботу.

Структура синтаксичного аналізатора буде такою:

FILE\* errFile;

if (fopen\_s(&errFile, ErrFile, "w") != 0)

{

printf("Error: Cannot open file for writing: %s\n", ErrFile);

return 1;

}

// таблиця лексем

extern Token\* TokenTable;

// кількість лексем

extern unsigned int TokensNum;

void Parser(FILE\* errFile)

{

program(errFile);

fprintf(errFile, "\nNo errors found.\n");

}

Синтаксичний аналізатор працює за методом рекурсивного спуску, а отже функція parser() викликає функцію program(), яка в свою чергу викликає інші функції.

Семантичний аналіз у нашому випадку буде реалізований у функції, яка опрацьовує оголошення і використання ідентифікаторів:

// функція записує оголошені ідентифікатори в таблицю ідентифікаторів IdTable

// повертає кількість ідентифікаторів

// перевіряє чи усі використані ідентифікатори оголошені

unsigned int IdIdentification(Id IdTable[], Token TokenTable[], unsigned int tokenCount, FILE\* errFile);

* + 1. Розробка граф-схеми алгоритму



Рис. 3.2 Граф-схема роботи синтаксичного аналізатора

* 1. Розробка генератора коду

Синтаксичне дерево в чистому вигляді несе тільки інформацію про структуру програми. Насправді в процесі генерації коду потрібна також інформація про змінні, операції, мітки і т.д. Для представлення цієї інформації можливі різні рішення. Найбільш поширені два:

* інформація зберігається у таблицях генератора коду;
* інформація зберігається у відповідних вершинах дерева.

Розглянемо, наприклад, структуру таблиць, які можуть бути використані в поєднанні з Лідер-представленням. Оскільки Лідер-представлення не містить інформації про адреси змінних, значить, цю інформацію потрібно формувати в процесі обробки оголошень і зберігати в таблицях. Це стосується і описів масивів, записів і т.д. Крім того, в таблицях також повинна міститися інформація про операції.

Генерація коду – це машинно-залежний етап компіляції, під час якого відбувається побудова машинного еквівалента вхідної програми. 3азвичай входом для генератора коду служить проміжна форма представлення програми, а на виході може з’являтися об’єктний код або модуль завантаження.

Генератор C коду приймає масив лексем без помилок. Якщо на двох попередніх етапах виявлено помилки, то ця фаза не виконується.

В даному курсовому проекті генерація коду реалізується як окремий етап. Можливість його виконання є лише за умови, що попередньо успішно виконався етап синтаксичного аналізу. І використовує результат виконання попереднього аналізу, тобто два файли: перший містить згенерований C код відповідно операторам які були в програмі, другий файл містить таблицю змінних.

* + 1. Розробка граф-схеми алгоритму



Рис. 3.3 Блок схема генератора коду

* + 1. Опис програми реалізації генератора коду

У компілятора, реалізованого в даному курсовому проекті, вихідна мова - програма на мові С. Ця програма записується у файл, що має таку ж саму назву, як і файл з вхідним текстом, але розширення “.c”. Генерація коду відбувається одразу ж після синтаксичного аналізу.

В даному трансляторі генератор коду послідовно викликає окремі функції, які записують у вихідний файл частини коду.

Першим кроком генерації коду записується заголовки, необхідні для програми на С, та визначається основна функція main(). Далі виконується аналіз коду та визначаються змінні, які використовуються.

Проаналізувавши змінні, які є у програмі, генератор формує секцію оголошення змінних для програми на С. Для цього з таблиці лексем вибирається ім’я змінної (типи змінних відповідають типам у С, наприклад int), та записується її початкове значення, якщо воно задано.

Аналіз наявних операторів необхідний у зв’язку з тим, що введення/виведення, виконання арифметичних та логічних операцій виконуються як окремі конструкції, і у випадку їх відсутності немає сенсу записувати у вихідний файл зайву інформацію.

Після цього зчитується лексема з таблиці лексем. Також відбувається перевірка, чи це не остання лексема. Якщо це остання лексема, то функція завершується.

Наступним кроком є аналіз таблиці лексем та безпосередня генерація коду у відповідності до вхідної програми.

Генератор коду зчитує лексему та генерує відповідний код, який записується у файл. Наприклад, якщо це лексема виведення, то у основну програму записується виклик функції printf, яка формує вихідний текст. Якщо це арифметична операція, то у вихідний файл записується вираз, що відповідає правилам С, із врахуванням пріоритетів операцій.

Генератор закінчує свою роботу, коли зчитує лексему, що відповідає кінцю файлу.

В кінці своєї роботи генератор формує завершення програми на С, додаючи повернення значення 0 з основної функції.

1. Опис програми

Дана програма написана мовою С++ з використанням визначень нових типів та перечислень:

// cтруктура генератора коду

// таблиця лексем

extern Token\* TokenTable;

// кількість лексем

extern unsigned int TokensNum;

// таблиця ідентифікаторів

extern Id\* IdTable;

// кількість ідентифікаторів

extern unsigned int IdNum;

static int pos = 3;

void generateCCode(FILE\* outFile)

{

fprintf(outFile, "#include <stdio.h>\n#include <stdlib.h>\n\n");

fprintf(outFile, "int main() {\n");

offset++;

variable\_declaration(outFile);

program\_body(outFile);

fprintf(outFile, "return 0;\n");

fprintf(outFile, "}\n");

}

enum TypeOfTokens

{

StartProgram,

StartBody,

Variable,

Type,

End,

Input,

Output,

If,

Else,

Goto,

Label,

For,

To,

DownTo,

Do,

While,

Repeat,

Until,

Identifier,

Number,

Assign,

Add,

Sub,

Mul,

Div,

Mod,

Equality,

NotEquality,

Greate,

Less,

Not,

And,

Or,

LBraket,

RBraket,

Semicolon,

Colon,

Comma,

Unknown

};

// структура для зберігання інформації про лексему

struct Token

{

char name[16]; // ім'я лексеми

int value; // значення лексеми

int line; // номер рядка

TypeOfTokens type; // тип лексеми

};

// структура для зберігання інформації про ідентифікатор

struct Id

{

char name[16];

};

// перерахування, яке описує стани лексичного аналізатора

enum States

{

Start, // початок виділення чергової лексеми

Finish, // кінець виділення чергової лексеми

Letter, // опрацювання слів (ключові слова і ідентифікатори)

Digit, // опрацювання цифри

Separators, // видалення пробілів, символів табуляції і переходу на новий рядок

Another, // опрацювання інших символів

EndOfFile, // кінець файлу

SComment, // початок коментаря

Comment // видалення коментаря

};

Спочатку вхідна програма за допомогою функції unsigned int GetTokens(FILE\* F, Token TokenTable[]) розбивається на відповідні токени для запису у таблицю та подальше їх використання в процесі синтаксичного аналізу та генерації коду.

Далі відбувається синтаксичний аналіз вхідної програми за допомогою функції void Parser(). Всі правила запису як різноманітних операцій так і програми в цілому відбувається за нотатками Бекуса-Наура, за допомогою яких можна легко описати синтаксис всіх операцій.

Нище наведено опис структури програми за допомогою нотаток Бекуса-Наура.

void program()

{

match(StartProgram);

match(StartBody);

match(Variable);

variable\_declaration();

match(Semicolon);

program\_body();

match(End);

}

Наступним етапом є генерація С коду. Алгоритм генерації працює за принципом синтаксичного аналізу але при вибірці певної лексеми або операції генерує відповідний С код який записується у вихідний файл.

Нище наведено генерацію С коду на прикладі операції присвоєння.

void assignment(FILE\* outFile)

{

fprintf(outFile, " ");

fprintf(outFile, TokenTable[pos++].name);

fprintf(outFile, " = ");

pos++;

arithmetic\_expression(outFile);

pos++;

fprintf(outFile, ";\n");

}

Така структура програми дозволяє без проблем аналізувати великі програми, написані на вхідній мові програмування. Також використання правил Бекуса-Наура дозволяє ефективно анадізувати програми великого обсягу.

* 1. Опис інтерфейсу та інструкція користувачеві

Вхідним файлом для даної програми є звичайний текстовий файл з розширенням p18. У цьому файлі необхідно набрати бажану для трансляції програму та зберегти її. Синтаксис повинен відповідати вхідній мові.

Створений транслятор є консольною програмою, що запускається з командної стрічки з параметром: "p18translator.exe <ім’я програми>.p18"

Якщо обидва файли мають місце на диску та правильно сформовані, програма буде запущена на виконання.

Початковою фазою обробки є лексичний аналіз (розбиття на окремі лексеми). Результатом цього етапу є файл <ім’я програми>.token, який містить таблицю лексем. Вміст цього файлу складається з 4 полів – 1 – безпосередньо сама лексема; 2 – тип лексеми; 3 – значення лексеми (необхідне для чисел і ідентифікаторів); 4 – рядок, у якому лексема знаходиться. Наступним етапом є перевірка на правильність написання програми (вхідної). Інформацію про наявність чи відсутність помилок можна переглянути у файлі Errors.txt. Якщо граматичний розбір виконаний успішно, файл буде містити відповідне повідомлення. Інакше, у файлі будуть зазначені помилки з їх описом та вказанням їх місця у тексті програми.

Останнім етапом є генерація коду. Транслятор переходить до цього етапу, лише у випадку, коли відсутні граматичні помилки у вхідній програмі. Згенерований код записується у файлу <ім’я програми>.с.

1. Відлагодження та тестування програми

Тестування програмного забезпечення є важливим етапом розробки продукту. На цьому етапі знаходяться помилки допущені на попередніх етапах. Цей етап дозволяє покращити певні характеристики продукту, наприклад – інтерфейс. Дає можливість знайти та вподальшому виправити слабкі сторони, якщо вони є.

Відлагодження даної програми здійснюється за допомогою набору кількох програм, які відповідають заданій граматиці. Та перевірці коректності коду, що генерується, коректності знаходження помилок та розбивки на лексеми.

* 1. Виявлення лексичних та синтаксичних помилок

Помилки у вхідній програмі виявляються на етапі лексичного аналізу.

Наприклад, у програмі зробимо лексичну помилку – у 13-му рядку неправильно вкажемо оператор вводу:

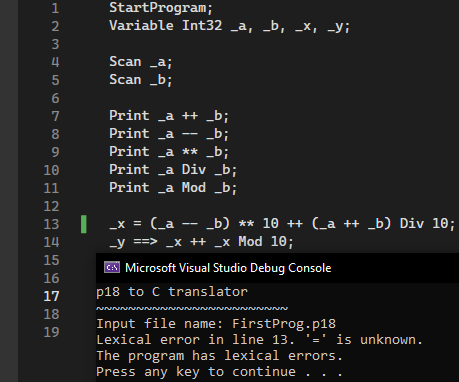


Рис. 5.1.1. Вивід інформації про лексичну помилку.

Зробимо синтаксичну помилку – у 7-му рядку, пропущу крапку з комою:

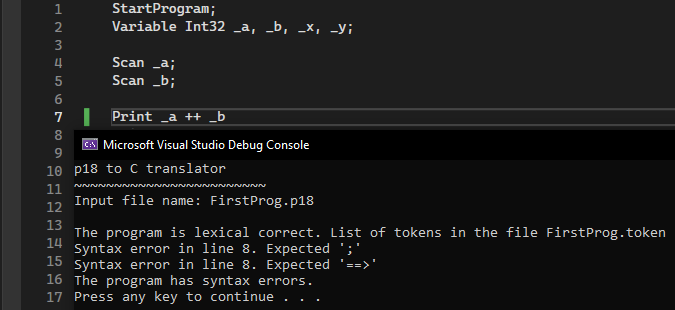


Рис. 5.1.2. Вивід інформації про синтаксичну помилку.

* 1. Виявлення семантичних помилок

Суттю виявлення семантичних помилок є перевірка числових констант на відповідність типу Int32, тобто знаковому цілому числу з відповідним діапазоном значень і перевірку на коректність використання змінних Int32 у цілочисельних і логічних виразах.

Зробимо семантичну помилку – не оголосимо змінну “x”:

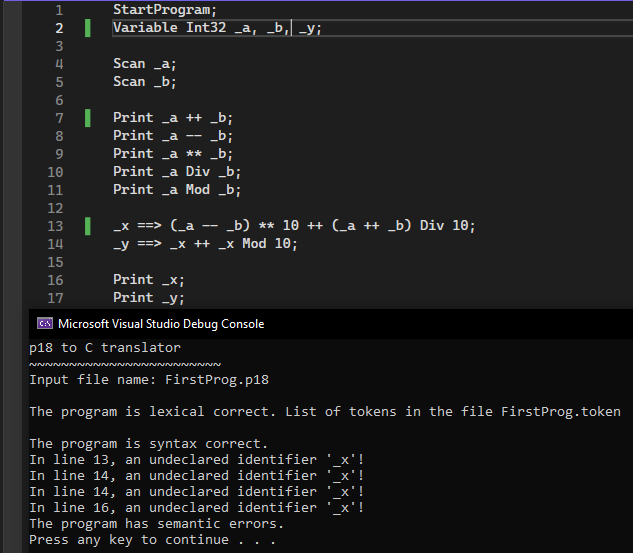


Рис. 5.2. Вивід інформації про семантичну помилку.

* 1. Загальна перевірка коректності роботи транслятора

Будь-яке програмне забезпечення необхідно протестувати і налагодити. Після опрацювання синтаксичних і семантичних помилок необхідно переконатися, що розроблене програмне забезпечення функціонує так, як очікувалось.

Для перевірки коректності роботи розробленого транслятора необхідно буде написати тестові задачі на вхідній мові програмування, отримати код на мові програмування С і переконатись, що він працює правильно.

Розроблений транслятор має простий консольний інтерфейс. При запуску програми необхідно ввести ім’я файлу з текстом програми на вхідній мові програмування:

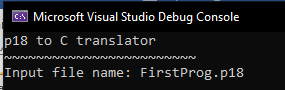


Рис. 5.3.1. Інтерфейс розробленого транслятора.

Після вводу імені програми отримуємо результати роботи транслятора:

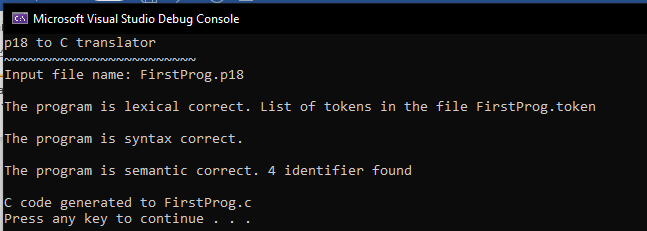


Рис. 5.3.2. Результати роботи розробленого транслятора.

* 1. Тестова програма №1

**Тестова програма «*Лінійний алгоритм*»**

1. Ввести два числа А і В (імена змінних можуть бути іншими і мають відповідати правилам запису ідентифікаторів згідно індивідуального завдання).

2. Вивести на екран:

А + В (результат операції додавання);

А - В (результат операції віднімання);

А \* В (результат операції множення);

А / В (результат операції ділення);

А % В (результат операції отримання залишку від ділення).

3. Обрахувати значення виразів

Х = (А - В) \* 10 + (А + В) / 10

У = Х + Х % 10

4. Вивести значення Х і У на екран

Напишемо програму на вхідній мові програмування(p18):

StartProgram;

Variable Int32 \_a, \_b, \_x, \_y;

Scan \_a;

Scan \_b;

Print \_a ++ \_b;

Print \_a -- \_b;

Print \_a \*\* \_b;

Print \_a Div \_b;

Print \_a Mod \_b;

\_x ==> (\_a -- \_b) \*\* 10 ++ (\_a ++ \_b) Div 10;

\_y ==> \_x ++ \_x Mod 10;

Print \_x;

Print \_y;

EndBlok;

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_x = 0, \_y = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

printf("%d\n", \_a + \_b);

printf("%d\n", \_a - \_b);

printf("%d\n", \_a \* \_b);

printf("%d\n", \_a / (\_b == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : \_b));

printf("%d\n", \_a % (\_b == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : \_b));

\_x = (\_a - \_b) \* 10 + (\_a + \_b) / (10 == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : 10);

\_y = \_x + \_x % (10 == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : 10);

printf("%d\n", \_x);

printf("%d\n", \_y);

return 0;

}

Отриманий код на мові С протестуємо у новому проєкті у середовищі Visual

Studio 2022 і отримаємо такі результати:

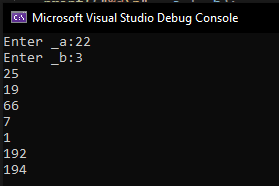


Рис. 5.4. Результати виконання тестової задачі 1.

* 1. Тестова програма №2

**Тестова програма «*Алгоритм з розгалуженням*»**

1. Ввести три числа А, В, С (імена змінних можуть бути іншими і мають відповідати правилам запису ідентифікаторів згідно індивідуального завдання).

2. Знайти найбільше з них і вивести його на екран.

3. Вивести на екран число 1, якщо усі числа однакові (логічний вираз в умовному операторі має виглядати так: «(А=В) і (А=С) і (В=С)»), інакше вивести 0.

4. Вивести на екран число -1, якщо хоча б одне з чисел від’ємне (логічний вираз в умовному операторі має виглядати так: «(А<0) або (В<0) або (С<0)»), інакше вивести 0.

5. Вивести на екран число 10, якщо число А більше за суму чисел В і С (логічний вираз в умовному операторі має виглядати так: «!(А<(В+С))»), інакше вивести 0.

Напишемо програму на вхідній мові програмування(p18):

StartProgram;

Variable Int32 \_a, \_b, \_c;

Scan \_a;

Scan \_b;

Scan \_c;

If(\_a Gr \_b)

If(\_a Gr \_b)

Print \_a;

Else

Print \_c;

Else

If(\_b Gr \_c)

Print \_b;

Else

Print \_c;

If ((\_a Eq \_b) And (\_a Eq \_c) And (\_b Eq \_c))

Print 1;

Else

Print 0;

If ((\_a Ls 0) Or (\_b Ls 0) Or (\_c Ls 0))

Print -1;

Else

Print 0;

If (Not (\_a Gr (\_b ++ \_c)))

Print 10;

Else

Print 0;

EndBlok;

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_c = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

printf("Enter \_c:");

scanf\_s("%d", &\_c);

if (\_a > \_b) {

if (\_a > \_b) {

printf("%d\n", \_a);

}

else {

printf("%d\n", \_c);

}

}

else {

if (\_b > \_c) {

printf("%d\n", \_b);

}

else {

printf("%d\n", \_c);

}

}

if ((\_a == \_b) && (\_a == \_c) && (\_b == \_c)) {

printf("%d\n", 1);

}

else {

printf("%d\n", 0);

}

if ((\_a < 0) || (\_b < 0) || (\_c < 0)) {

printf("%d\n", -1);

}

else {

printf("%d\n", 0);

}

if (!(\_a < (\_b + \_c))) {

printf("%d\n", 10);

}

else {

printf("%d\n", 0);

}

return 0;

}

Отриманий код на мові С протестуємо у новому проєкті у середовищі Visual Studio 2022 і отримаємо такі результати:

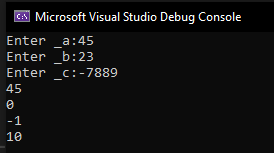


Рис. 5.5. Результати виконання тестової задачі 2.

* 1. Тестова програма №3

**Тестова програма «*Циклічний алгоритм*»**

1. Ввести два числа А і В, причому А<В (імена змінних можуть бути іншими і мають відповідати правилам запису ідентифікаторів згідно індивідуального завдання).

2. Вивести на екран квадрати чисел від А до В включно.

3. Обрахувати Х=А\*В за наступним алгоритмом:

Х = 0

Цикл від 1 до А з кроком 1

Цикл від 1 до В з кроком 1

Х = Х + 1

4. Вивести значення Х на екран.

Напишемо програму на вхідній мові програмування(p18):

StartProgram;

Variable Int32 \_a, \_b, \_x, \_i, \_j;

Scan \_a;

Scan \_b;

For \_i ==> \_a To \_b Do

Print \_i \*\* \_i;

\_x ==> 0;

For \_i ==> 1 To \_a Do

For \_j ==> 1 To \_b Do

\_x ==> \_x ++ 1;

Print \_x;

EndBlok;

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_x = 0, \_i = 0, \_j = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

for (int \_i = \_a; \_i < \_b + 1; \_i++) {

printf("%d\n", \_i \* \_i);

}

\_x = 0;

for (int \_i = 1; \_i < \_a; \_i++) {

for (int \_j = 1; \_j < \_b; \_j++) {

\_x = \_x + 1;

}}

printf("%d\n", \_x);

return 0;

}

Отриманий код на мові С протестуємо у новому проєкті у середовищі Visual Studio 2022 і отримаємо такі результати:

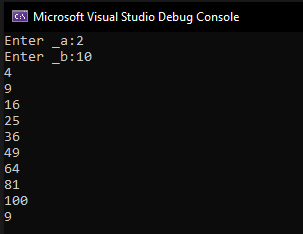


Рис. 5.6. Результати виконання тестової задачі 3.

Висновки

В процесі виконання курсового проєкту було виконано наступне:

1. Складено формальний опис мови програмування p18, в термінах розширеної нотації Бекуса-Наура, виділено усі термінальні символи та ключові слова.
2. Створено транслятор мови програмування p18, а саме:
   1. Розроблено прямий лексичний аналізатор, орієнтований на розпізнавання лексем, що є заявлені в формальному описі мови програмування.
   2. Розроблено синтаксичний аналізатор на основі низхідного методу. Складено деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура
   3. Розроблено генератор коду, відповідні процедури якого викликаються після перевірки синтаксичним аналізатором коректності запису лексем, мови програмування p18. Вихідним кодом генератора є програма на мові C.
3. Проведене тестування транслятора на тестових програмах за наступними пунктами:
   1. На виявлення лексичних помилок.
   2. На виявлення синтаксичних помилок.
   3. Загальна перевірка роботи транслятора.

Тестування не виявило помилок в роботі транслятора, і всі помилки в тестових програмах на мові p18 були успішно виявлені і відповідно оброблені.

В результаті виконання даної курсового проєкту було засвоєно методи розробки та реалізації компонент систем програмування.

Список використаної літератури

1. Основи проектування трансляторів: Конспект лекцій : [Електронний ресурс] : навч. посіб. для студ. спеціальності 123 – «Комп’ютерна інженерія» / О. І. Марченко ; КПІ ім. Ігоря Сікорського. – Київ: КПІ ім. Ігоря Сікорського, 2021. – 108 с.

2. Формальні мови, граматики та автомати: Навчальний посібник / Гавриленко С.Ю. –Харків: НТУ «ХПІ», 2021. – 133 с.

3. Сопронюк Т.М. Системне програмування. Частина І. Елементи теорії формальних мов: Навчальний посібник у двох частинах. – Чернівці: ЧНУ, 2008. – 84 c.

4. Сопронюк Т.М. Системне програмування. Частина ІІ. Елементи теорії компіляції: Навчальний посібник у двох частинах. – Чернівці: ЧНУ, 2008. – 84 c.

5. Alfred V. Aho, Monica S. Lam, Ravi Seth, Jeffrey D. Ullma. Compilers, principles, techniques, and tools, Second Edition, New York, 2007. – 1038 c.

6. Системне програмування (курсовий проект) [Електронний ресурс] – Режим доступу до ресурсу: <https://vns.lpnu.ua/course/view.php?id=11685>.

7. MIT OpenCourseWare. Computer Language Engineering [Електронний ресурс] – Режим доступу до ресурсу: https://ocw.mit.edu/courses/6-035-computer-language-engineering-spring2010.

8. Рисований О.М. Системне програмування: підручник для студентів напрямку

“Комп’ютерна інженерія” вищих навчальних закладів в 2-х томах. Том 1. – Видання четверте: виправлено та доповнено – Х.: “Слово”, 2015. – 576 с.

9. Рисований О.М. Системне програмування: підручник для студентів напрямку

“Комп’ютерна інженерія” вищих навчальних закладів в 2-х томах. Том 2. – Видання четверте: виправлено та доповнено – Х.: “Слово”, 2015. – 378 с.

Додатки

Додаток А (таблиці лексем для тесових програм)

Тестова програма «Лінійний алгоритм»

**---------------------------------------------------------------------------**

**| TOKEN TABLE |**

**---------------------------------------------------------------------------**

**| line number | token | value | token code | type of token |**

**---------------------------------------------------------------------------**

**| 1 | StartProgram | 0 | 0 | StartProgram |**

**---------------------------------------------------------------------------**

**| 1 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 2 | Variable | 0 | 2 | Variable |**

**---------------------------------------------------------------------------**

**| 2 | Int32 | 0 | 3 | Integer |**

**---------------------------------------------------------------------------**

**| 2 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_y | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 4 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 4 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 4 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 5 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 5 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 5 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 7 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 7 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 7 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 7 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 7 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 8 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 8 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | -- | 0 | 22 | Sub |**

**---------------------------------------------------------------------------**

**| 8 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 9 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 9 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 9 | \*\* | 0 | 23 | Mul |**

**---------------------------------------------------------------------------**

**| 9 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 9 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 10 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 10 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 10 | Div | 0 | 24 | Div |**

**---------------------------------------------------------------------------**

**| 10 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 10 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 11 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 11 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 11 | Mod | 0 | 25 | Mod |**

**---------------------------------------------------------------------------**

**| 11 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 11 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 13 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 13 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 13 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | -- | 0 | 22 | Sub |**

**---------------------------------------------------------------------------**

**| 13 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 13 | \*\* | 0 | 23 | Mul |**

**---------------------------------------------------------------------------**

**| 13 | 10 | 10 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 13 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 13 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 13 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 13 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 13 | Div | 0 | 24 | Div |**

**---------------------------------------------------------------------------**

**| 13 | 10 | 10 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 13 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 14 | \_y | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 14 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 14 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 14 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 14 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 14 | Mod | 0 | 25 | Mod |**

**---------------------------------------------------------------------------**

**| 14 | 10 | 10 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 14 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 16 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 16 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 16 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 17 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 17 | \_y | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 17 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 19 | EndBlok | 0 | 4 | End |**

**---------------------------------------------------------------------------**

**| 19 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

Тестова програма «Алгоритм з розгалуженням»

**---------------------------------------------------------------------------**

**| TOKEN TABLE |**

**---------------------------------------------------------------------------**

**| line number | token | value | token code | type of token |**

**---------------------------------------------------------------------------**

**| 1 | StartProgram | 0 | 0 | StartProgram |**

**---------------------------------------------------------------------------**

**| 1 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 2 | Variable | 0 | 2 | Variable |**

**---------------------------------------------------------------------------**

**| 2 | Int32 | 0 | 3 | Integer |**

**---------------------------------------------------------------------------**

**| 2 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 4 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 4 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 4 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 5 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 5 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 5 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 6 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 6 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 6 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 8 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 8 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 8 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | Gr | 0 | 28 | Greate |**

**---------------------------------------------------------------------------**

**| 8 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 9 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 9 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 9 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 9 | Gr | 0 | 28 | Greate |**

**---------------------------------------------------------------------------**

**| 9 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 9 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 10 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 10 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 10 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 11 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 12 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 12 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 12 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 13 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 14 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 14 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 14 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 14 | Gr | 0 | 28 | Greate |**

**---------------------------------------------------------------------------**

**| 14 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 14 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 15 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 15 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 15 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 16 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 17 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 17 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 17 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 19 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 19 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 19 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 19 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | Eq | 0 | 26 | Equality |**

**---------------------------------------------------------------------------**

**| 19 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 19 | And | 0 | 31 | And |**

**---------------------------------------------------------------------------**

**| 19 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 19 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | Eq | 0 | 26 | Equality |**

**---------------------------------------------------------------------------**

**| 19 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 19 | And | 0 | 31 | And |**

**---------------------------------------------------------------------------**

**| 19 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 19 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | Eq | 0 | 26 | Equality |**

**---------------------------------------------------------------------------**

**| 19 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 19 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 19 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 20 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 20 | 1 | 1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 20 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 21 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 22 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 22 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 22 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 24 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 24 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 24 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 24 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 24 | Ls | 0 | 29 | Less |**

**---------------------------------------------------------------------------**

**| 24 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 24 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 24 | Or | 0 | 32 | Or |**

**---------------------------------------------------------------------------**

**| 24 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 24 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 24 | Ls | 0 | 29 | Less |**

**---------------------------------------------------------------------------**

**| 24 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 24 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 24 | Or | 0 | 32 | Or |**

**---------------------------------------------------------------------------**

**| 24 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 24 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 24 | Ls | 0 | 29 | Less |**

**---------------------------------------------------------------------------**

**| 24 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 24 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 24 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 25 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 25 | -1 | -1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 25 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 26 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 27 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 27 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 27 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 29 | If | 0 | 7 | If |**

**---------------------------------------------------------------------------**

**| 29 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 29 | Not | 0 | 30 | Not |**

**---------------------------------------------------------------------------**

**| 29 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 29 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 29 | Ls | 0 | 29 | Less |**

**---------------------------------------------------------------------------**

**| 29 | ( | 0 | 33 | LBraket |**

**---------------------------------------------------------------------------**

**| 29 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 29 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 29 | \_c | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 29 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 29 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 29 | ) | 0 | 34 | RBraket |**

**---------------------------------------------------------------------------**

**| 30 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 30 | 10 | 10 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 30 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 31 | Else | 0 | 8 | Else |**

**---------------------------------------------------------------------------**

**| 32 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 32 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 32 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 34 | EndBlok | 0 | 4 | End |**

**---------------------------------------------------------------------------**

**| 34 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

Тестова програма «Циклічний алгоритм»

**---------------------------------------------------------------------------**

**| TOKEN TABLE |**

**---------------------------------------------------------------------------**

**| line number | token | value | token code | type of token |**

**---------------------------------------------------------------------------**

**| 1 | StartProgram | 0 | 0 | StartProgram |**

**---------------------------------------------------------------------------**

**| 1 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 2 | Variable | 0 | 2 | Variable |**

**---------------------------------------------------------------------------**

**| 2 | Int32 | 0 | 3 | Integer |**

**---------------------------------------------------------------------------**

**| 2 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_i | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | , | 0 | 36 | Comma |**

**---------------------------------------------------------------------------**

**| 2 | \_j | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 2 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 4 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 4 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 4 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 5 | Scan | 0 | 5 | Input |**

**---------------------------------------------------------------------------**

**| 5 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 5 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 7 | For | 0 | 10 | For |**

**---------------------------------------------------------------------------**

**| 7 | \_i | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 7 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 7 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 7 | To | 0 | 11 | To |**

**---------------------------------------------------------------------------**

**| 7 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 7 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 7 | 1 | 1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 7 | Do | 0 | 13 | Number |**

**---------------------------------------------------------------------------**

**| 8 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 8 | \_i | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | \*\* | 0 | 23 | Mul |**

**---------------------------------------------------------------------------**

**| 8 | \_i | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 8 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 10 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 10 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 10 | 0 | 0 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 10 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 11 | For | 0 | 10 | For |**

**---------------------------------------------------------------------------**

**| 11 | \_i | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 11 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 11 | 1 | 1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 11 | To | 0 | 11 | To |**

**---------------------------------------------------------------------------**

**| 11 | \_a | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 11 | Do | 0 | 13 | Identifier |**

**---------------------------------------------------------------------------**

**| 12 | For | 0 | 10 | For |**

**---------------------------------------------------------------------------**

**| 12 | \_j | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 12 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 12 | 1 | 1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 12 | To | 0 | 11 | To |**

**---------------------------------------------------------------------------**

**| 12 | \_b | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 12 | Do | 0 | 13 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ==> | 0 | 20 | Assign |**

**---------------------------------------------------------------------------**

**| 13 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 13 | ++ | 0 | 21 | Add |**

**---------------------------------------------------------------------------**

**| 13 | 1 | 1 | 19 | Number |**

**---------------------------------------------------------------------------**

**| 13 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 15 | Print | 0 | 6 | Output |**

**---------------------------------------------------------------------------**

**| 15 | \_x | 0 | 18 | Identifier |**

**---------------------------------------------------------------------------**

**| 15 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

**| 17 | EndBlok | 0 | 4 | End |**

**---------------------------------------------------------------------------**

**| 17 | ; | 0 | 35 | Semicolon |**

**---------------------------------------------------------------------------**

Додаток Б (Код на мові C)

Тестова програма «Лінійний алгоритм»

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_x = 0, \_y = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

printf("%d\n", \_a + \_b);

printf("%d\n", \_a - \_b);

printf("%d\n", \_a \* \_b);

printf("%d\n", \_a / (\_b == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : \_b));

printf("%d\n", \_a % (\_b == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : \_b));

\_x = (\_a - \_b) \* 10 + (\_a + \_b) / (10 == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : 10);

\_y = \_x + \_x % (10 == 0 ? (printf("Error: Division by zero\n"), exit(1), 0) : 10);

printf("%d\n", \_x);

printf("%d\n", \_y);

return 0;

}

Тестова програма «Алгоритм з розгалуженням»

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_c = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

printf("Enter \_c:");

scanf\_s("%d", &\_c);

if (\_a > \_b) {

if (\_a > \_b) {

printf("%d\n", \_a);

}

else {

printf("%d\n", \_c);

}

}

else {

if (\_b > \_c) {

printf("%d\n", \_b);

}

else {

printf("%d\n", \_c);

}

}

if ((\_a == \_b) && (\_a == \_c) && (\_b == \_c)) {

printf("%d\n", 1);

}

else {

printf("%d\n", 0);

}

if ((\_a < 0) || (\_b < 0) || (\_c < 0)) {

printf("%d\n", -1);

}

else {

printf("%d\n", 0);

}

if (!(\_a < (\_b + \_c))) {

printf("%d\n", 10);

}

else {

printf("%d\n", 0);

}

return 0;

}

Тестова програма «Циклічний алгоритм»

#include <stdio.h>

#include <stdlib.h>

int main() {

int \_a = 0, \_b = 0, \_x = 0, \_i = 0, \_j = 0;

printf("Enter \_a:");

scanf\_s("%d", &\_a);

printf("Enter \_b:");

scanf\_s("%d", &\_b);

for (int \_i = \_a; \_i < \_b + 1; \_i++) {

printf("%d\n", \_i \* \_i);

}

\_x = 0;

for (int \_i = 1; \_i < \_a; \_i++) {

for (int \_j = 1; \_j < \_b; \_j++) {

\_x = \_x + 1;

}

}

printf("%d\n", \_x);

return 0;

}

Додаток В (Абстрактне синтаксичне дерево для тестових прикладів)

**Тестова програма «Лінійний алгоритм»**

|-- program

| |-- var

| | |-- YYYYYY

| | |-- var

| | | |-- XXXXXX

| | | |-- var

| | | | |-- BBBBBB

| | | | |-- var

| | | | | |-- AAAAAA

| |-- statement

| | |-- statement

| | | |-- statement

| | | | |-- statement

| | | | | |-- statement

| | | | | | |-- statement

| | | | | | | |-- statement

| | | | | | | | |-- statement

| | | | | | | | | |-- statement

| | | | | | | | | | |-- input

| | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | |-- input

| | | | | | | | | | | |-- BBBBBB

| | | | | | | | | | |-- output

| | | | | | | | | | | |-- +

| | | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | | | |-- BBBBBB

| | | | | | | | | | |-- output

| | | | | | | | | | | |-- -

| | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | | |-- BBBBBB

| | | | | | | | | | |-- output

| | | | | | | | | | | |-- \*

| | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | |-- output

| | | | | | | | | | | |-- /

| | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | |-- output

| | | | | | | | | | | |-- %

| | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | |-- <==

| | | | | | | | | | | |-- XXXXXX

| | | | | | | | | | | |-- + \*

| | | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | | | |-- BBBBBB

| | | | | | | | | | | |-- / + AAAAAA

| | | | | | | | | | | |-- output

| | | | | | | | | | | |-- YYYYYY

**Тестова програма «Алгоритм з розгалуженням»**

|-- program

| |-- var

| | |-- CCCCCC

| | |-- var

| | | |-- BBBBBB

| | | |-- var

| | | | |-- AAAAAA

| |-- statement

| | |-- statement

| | | |-- statement

| | | | |-- statement

| | | | | |-- statement

| | | | | | |-- statement

| | | | | | | |-- statement

| | | | | | | | |-- statement

| | | | | | | | | |-- input

| | | | | | | | | | |-- AAAAAA

| | | | | | | | | |-- input

| | | | | | | | | | |-- BBBBBB

| | | | | | | | |-- input

| | | | | | | | | |-- CCCCCC

| | | | | | | |-- if

| | | | | | | | |-- > >

| | | | | | | | | |-- AAAAAA

| | | | | | | | | |-- BBBBBB

| | | | | | | | |-- branches

| | | | | | | | | |-- compound

| | | | | | | | | | |-- if

| | | | | | | | | | | |-- > >

| | | | | | | | | | | | |-- AAAAAA

| | | | | | | | | | | | |-- CCCCCC

| | | | | | | | | | | |-- branches

| | | | | | | | | | | | |-- compound

| | | | | | | | | | | | | |-- goto

| | | | | | | | | | | | | | |-- Abigger

| | | | | | | | | | | | |-- compound

| | | | | | | | | | | | | |-- statement

**Тестова програма «Циклічний алгоритм»**

|-- program

| |-- var

| | |-- CCCCC2

| | |-- var

| | | |-- CCCCC1

| | | |-- var

| | | | |-- XXXXXX

| | | | |-- var

| | | | | |-- BBBBBB

| | | | | |-- var

| | | | | | |-- AAAAA2

| | | | | | |-- var

| | | | | | | |-- AAAAAA

| |-- statement

| | |-- statement

| | | |-- statement

| | | | |-- statement

| | | | | |-- statement

| | | | | | |-- statement

| | | | | | | |-- statement

| | | | | | | | |-- statement

| | | | | | | | | |-- statement

| | | | | | | | | | |-- statement

| | | | | | | | | | | |-- statement

| | | | | | | | | | | | |-- statement

| | | | | | | | | | | | | |-- for

| | | | | | | | | | | | | | |-- >= AAAAAA

| | | | | | | | | | | | | |-- >

| | | | | | | | | | | | |-- addstmt

| | | | | | | | | | | | | |-- output

| | | | | | | | | | | | | | |-- CCCCCC

Додаток Г (Документований текст програмних модулів (лістинги))

**CodeGenerator.cpp  
Модуль генерації коду, який перетворює оброблені лексеми на вихідний код мовою C.**

#include "CodeGenerator.h"

static int pos = 3;

static int offset = 0;

// набір функцій для рекурсивного спуску

// на кожне правило - окрема функція

void writeOffset(FILE\* outFile);

void variable\_declaration(FILE\* outFile);

void variable\_list(FILE\* outFile);

void program\_body(FILE\* outFile);

void statement(FILE\* outFile);

void assignment(FILE\* outFile);

void label(FILE\* outFile);

void arithmetic\_expression(FILE\* outFile);

void term(FILE\* outFile);

void factor(FILE\* outFile);

void input(FILE\* outFile);

void output(FILE\* outFile);

void conditional(FILE\* outFile);

void logical\_expression(FILE\* outFile);

void and\_expression(FILE\* outFile);

void comparison(FILE\* outFile);

void for\_statement(FILE\* outFile);

void forTo\_statement(FILE\* outFile);

void forDownTo\_statement(FILE\* outFile);

void while\_statement(FILE\* outFile);

void repeat\_statement(FILE\* outFile);

void goto\_statement(FILE\* outFile);

void compound\_statement(FILE\* outFile);

//<програма> = ‘StartProgram’ <ідентифікатор> ‘;’ ‘Variable’ <оголошення змінних> ‘;’ <тіло програми> ‘EndBlok’ ';'

void generateCCode(FILE\* outFile)

{

fprintf(outFile, "#include <stdio.h>\n#include <stdlib.h>\n\n");

fprintf(outFile, "int main() {\n");

offset++;

variable\_declaration(outFile);

program\_body(outFile);

writeOffset(outFile);

fprintf(outFile, "return 0;\n");

fprintf(outFile, "}\n");

}

void writeOffset(FILE\* outFile) {

for (int i = 0; i < offset; i++)

fprintf(outFile, "\t");

}

// <оголошення змінних> = [<тип даних> <список змінних>]

void variable\_declaration(FILE\* outFile)

{

if (TokenTable[pos].type == Type)

{

writeOffset(outFile);

fprintf(outFile, "int ");

pos++;

variable\_list(outFile);

fprintf(outFile, ";\n\n");

pos++;

}

}

// <список змінних> = <ідентифікатор> ['==>' <арифметичний вираз>] { ‘,’ <ідентифікатор> ['==>' <арифметичний вираз>]}

void variable\_list(FILE\* outFile)

{

fprintf(outFile, TokenTable[pos++].name);

if (TokenTable[pos].type == Assign) {

fprintf(outFile, " = ");

pos++;

arithmetic\_expression(outFile);

}

else

fprintf(outFile, " = 0");

while (TokenTable[pos].type == Comma)

{

fprintf(outFile, ", ");

pos++;

fprintf(outFile, TokenTable[pos++].name);

if (TokenTable[pos].type == Assign) {

fprintf(outFile, " = ");

pos++;

arithmetic\_expression(outFile);

}

else

fprintf(outFile, " = 0");

}

}

// <тіло програми> = { <оператор> }

void program\_body(FILE\* outFile)

{

do

{

statement(outFile);

} while (TokenTable[pos].type != End);

}

// <оператор> = <присвоєння> | <ввід> | <вивід> | <умовний оператор> | <умовний перехід> | <мітка> | <оператор циклу> | <складений оператор>

void statement(FILE\* outFile)

{

writeOffset(outFile);

switch (TokenTable[pos].type)

{

case Input: input(outFile); break;

case Output: output(outFile); break;

case If: conditional(outFile); break;

case Goto: goto\_statement(outFile); break;

case For: for\_statement(outFile); break;

case While: while\_statement(outFile); break;

case Repeat: repeat\_statement(outFile); break;

case Identifier: assignment(outFile); break;

case Label: label(outFile); break;

case StartBody: compound\_statement(outFile); break;

default: assignment(outFile);

}

}

// <присвоєння> = <ідентифікатор> '==>' <арифметичний вираз> ';'

void assignment(FILE\* outFile)

{

fprintf(outFile, TokenTable[pos++].name);

fprintf(outFile, " = ");

pos++;

arithmetic\_expression(outFile);

fprintf(outFile, ";\n");

pos++;

}

// <мітка> = <ідентифікатор> ':'

void label(FILE\* outFile) {

fprintf(outFile, TokenTable[pos++].name);

fprintf(outFile, "\n");

}

// <арифметичний вираз> = <доданок> { ‘++’ <доданок> | ‘--’ <доданок> }

void arithmetic\_expression(FILE\* outFile)

{

term(outFile);

while (TokenTable[pos].type == Add || TokenTable[pos].type == Sub)

{

if (TokenTable[pos].type == Add)

fprintf(outFile, " + ");

else

fprintf(outFile, " - ");

pos++;

term(outFile);

}

}

// <доданок> = <множник> { ‘\*\*’ <доданок> | ‘Div’ <доданок> | ‘Mod’ <доданок> }

void term(FILE\* outFile)

{

factor(outFile);

while (TokenTable[pos].type == Mul || TokenTable[pos].type == Div || TokenTable[pos].type == Mod)

{

if (TokenTable[pos].type == Mul) {

fprintf(outFile, " \* ");

pos++;

factor(outFile);

}

else if (TokenTable[pos].type == Div) {

fprintf(outFile, " / ");

fprintf(outFile, "(");

pos++;

int denom = pos;

factor(outFile);

fprintf(outFile, " == 0 ? (printf(\"Error: Division by zero\\n\"), exit(1), 0) : ");

pos = denom;

factor(outFile);

fprintf(outFile, ")");

}

else {

fprintf(outFile, " %% ");

fprintf(outFile, "(");

pos++;

int denom = pos;

factor(outFile);

fprintf(outFile, " == 0 ? (printf(\"Error: Division by zero\\n\"), exit(1), 0) : ");

pos = denom;

factor(outFile);

fprintf(outFile, ")");

}

//pos++;

}

}

// <множник> = <ідентифікатор> | <число> | ‘(’ <арифметичний вираз> ‘)’

void factor(FILE\* outFile)

{

if (TokenTable[pos].type == Identifier)

fprintf(outFile, TokenTable[pos++].name);

else if (TokenTable[pos].type == Number)

fprintf(outFile, "%d", TokenTable[pos++].value);

else

if (TokenTable[pos].type == LBraket)

{

fprintf(outFile, "(");

pos++;

arithmetic\_expression(outFile);

fprintf(outFile, ")");

pos++;

}

}

// <ввід> = 'Scan' <ідентифікатор> ';'

void input(FILE\* outFile)

{

pos++;

fprintf(outFile, "printf(\"Enter ");

fprintf(outFile, TokenTable[pos].name);

fprintf(outFile, ":\");\n");

writeOffset(outFile);

fprintf(outFile, "scanf\_s(\"%%d\", &");

fprintf(outFile, TokenTable[pos++].name);

fprintf(outFile, ");\n");

pos++;

}

// <вивід> = ‘Print’ <арифметичний вираз> ';'

void output(FILE\* outFile)

{

fprintf(outFile, "printf(\"%%d\\n\", ");

pos++;

arithmetic\_expression(outFile);

//fprintf(outFile, TokenTable[pos++].name);

fprintf(outFile, ");\n");

pos++;

}

// <умовний оператор> ==> 'If' '(' <логічний вираз> ')' <оператор> ['Else' <оператор>]

void conditional(FILE\* outFile)

{

fprintf(outFile, "if ");

pos++;

logical\_expression(outFile);

fprintf(outFile, " {\n");

offset++;

statement(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "}\n");

if (TokenTable[pos].type == Else)

{

pos++;

writeOffset(outFile);

fprintf(outFile, "else {\n");

offset++;

statement(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "}\n");

}

}

// <логічний вираз> = <вираз І> { ‘Or’ <вираз І> }

void logical\_expression(FILE\* outFile)

{

and\_expression(outFile);

while (TokenTable[pos].type == Or)

{

fprintf(outFile, " || ");

pos++;

and\_expression(outFile);

}

}

// <вираз І> = <порівняння> { ‘And’ <вираз І> }

void and\_expression(FILE\* outFile)

{

comparison(outFile);

while (TokenTable[pos].type == And)

{

fprintf(outFile, " && ");

pos++;

comparison(outFile);

}

}

// <порівняння> = <операція порівняння> | ‘Not’ ‘(’ <логічний вираз> ‘)’ | ‘(’ <логічний вираз> ‘)’

// <операція порівняння> = <арифметичний вираз> <менше-більше> <арифметичний вираз>

// <менше-більше> = ‘Eq’ | ‘Neq’ | ‘Ls’ | ‘Gr’

void comparison(FILE\* outFile)

{

if (TokenTable[pos].type == Not)

{

// Варіант: ! (<логічний вираз>)

fprintf(outFile, "!(");

pos++;

pos++;

logical\_expression(outFile);

fprintf(outFile, ")");

pos++;

}

else if (TokenTable[pos].type == LBraket)

{

// Варіант: ( <логічний вираз> )

fprintf(outFile, "(");

pos++;

logical\_expression(outFile);

fprintf(outFile, ")");

pos++;

}

else

{

// Варіант: <арифметичний вираз> <менше-більше> <арифметичний вираз>

arithmetic\_expression(outFile);

if (TokenTable[pos].type == Greate || TokenTable[pos].type == Less ||

TokenTable[pos].type == Equality || TokenTable[pos].type == NotEquality)

{

switch (TokenTable[pos].type)

{

case Greate: fprintf(outFile, " > "); break;

case Less: fprintf(outFile, " < "); break;

case Equality: fprintf(outFile, " == "); break;

case NotEquality: fprintf(outFile, " != "); break;

}

pos++;

arithmetic\_expression(outFile);

}

}

}

// <оператор циклу> ==> <оператор циклу forTo> | <оператор циклу forDownTo>

void for\_statement(FILE\* outFile) {

fprintf(outFile, "for (int ");

pos++;

int iter = pos;

fprintf(outFile, TokenTable[iter].name);

fprintf(outFile, " = ");

pos += 2;

arithmetic\_expression(outFile);

fprintf(outFile, "; ");

fprintf(outFile, TokenTable[iter].name);

if (TokenTable[pos].type == To) {

fprintf(outFile, " < ");

pos++;

arithmetic\_expression(outFile);

fprintf(outFile, "; ");

fprintf(outFile, TokenTable[iter].name);

forTo\_statement(outFile);

}

else {

fprintf(outFile, " > ");

pos++;

arithmetic\_expression(outFile);

fprintf(outFile, "; ");

fprintf(outFile, TokenTable[iter].name);

forDownTo\_statement(outFile);

}

}

// <оператор циклу forTo> ==> 'For' <ідентифікатор> '==>' <арифметичний вираз> 'To' <арифметичний вираз> 'Do' <оператор>

void forTo\_statement(FILE\* outFile) {

fprintf(outFile, "++) {\n");

pos++;

offset++;

statement(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "}\n");

}

// <оператор циклу forDownTo> ==> 'For' <ідентифікатор> '==>' <арифметичний вираз> 'DownTo' <арифметичний вираз> 'Do' <оператор>

void forDownTo\_statement(FILE\* outFile) {

fprintf(outFile, "--) {\n");

pos++;

offset++;

statement(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "}\n");

}

// <оператор циклу while> ==> 'While' '(' <логічний вираз> ')' 'Do' <оператор>

void while\_statement(FILE\* outFile) {

fprintf(outFile, "while ");

pos++;

logical\_expression(outFile);

pos++;

fprintf(outFile, "\n");

statement(outFile);

}

// <оператор циклу repeat> ==> 'Repeat' <оператор> 'Until' '(' <логічний вираз> ')' ';'

void repeat\_statement(FILE\* outFile) {

pos++;

fprintf(outFile, "do {\n");

offset++;

statement(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "} while ");

pos++;

logical\_expression(outFile);

pos++;

fprintf(outFile, ";\n");

}

// <умовний перехід> ==> 'Goto' <мітка> ';'

void goto\_statement(FILE\* outFile) {

fprintf(outFile, "goto ");

pos++;

fprintf(outFile, TokenTable[pos].name);

fprintf(outFile, ";\n");

pos+=2;

}

// <складений оператор> = ‘StartBlok’ <тіло програми> ‘EndBlok’ ';'

void compound\_statement(FILE\* outFile)

{

fprintf(outFile, "{\n");

pos++;

offset++;

program\_body(outFile);

offset--;

writeOffset(outFile);

fprintf(outFile, "}\n");

pos+=2;

}

**CodeGenerator.h  
Заголовковий файл, що містить оголошення функцій і змінних, необхідних для генерації вихідного коду мовою C. У ньому визначені зовнішні змінні для роботи з таблицею лексем, ідентифікаторів і їх кількістю, а також функція generateCCode, яка виконує генерацію коду на основі синтаксично оброблених даних.**

#pragma once

#include <stdio.h>

#include <string.h>

#include <stdlib.h>

#include "header.h"

// таблиця лексем

extern Token\* TokenTable;

// кількість лексем

extern unsigned int TokensNum;

// таблиця ідентифікаторів

extern Id\* IdTable;

// кількість ідентифікаторів

extern unsigned int IdNum;

void generateCCode(FILE\* outFile);

**SyntaxAnalyzer.h  
Заголовковий файл, що містить оголошення змінних і функції для синтаксичного аналізу. Він визначає зовнішні змінні для роботи з таблицею лексем і зберігання помилок. Основна функція parser відповідає за перевірку структури вхідної програми відповідно до граматики.**

#pragma once

#include "stdio.h"

#include "stdlib.h"

#include <string.h>

#include "header.h"

extern unsigned int errors;

extern char\* ErrorsFileName;

extern FILE\* ErrorsFile;

// Вхідна таблиця лексем

extern Token\* TokenTable;

// кількість лексем

extern unsigned int TokensNum;

void parser();

**SyntaxAnalyzer.cpp  
Синтаксичний аналізатор, який перевіряє правильність структури програми відповідно до заданої граматики**

#include "SyntaxAnalyzer.h"

unsigned int pos = 0;

void match(TypeOfTokens expectedType);

void program();

void variable\_declaration();

void variable\_list();

void program\_body();

void statement();

void assignment();

void label();

void arithmetic\_expression();

void term();

void factor();

void input();

void output();

void conditional();

void logical\_expression();

void and\_expression();

void comparison();

void for\_statement();

void forTo\_statement();

void forDownTo\_statement();

void while\_statement();

void repeat\_statement();

void goto\_statement();

void compound\_statement();

// Переводить тип токена в текстовий варіант

const char\* getType(TypeOfTokens tokenType) {

switch (tokenType)

{

case StartProgram:

return "StartProgram";

break;

case StartBody:

return "StartBlok";

break;

case End:

return "EndBlok";

break;

case Variable:

return "Variable";

break;

case Type:

return "Int32";

break;

case Input:

return "Scan";

break;

case Output:

return "Print";

break;

case If:

return "If";

break;

case Else:

return "Else";

break;

case Goto:

return "Goto";

break;

case For:

return "For";

break;

case To:

return "To";

break;

case DownTo:

return "DownTo";

break;

case Do:

return "Do";

break;

case While:

return "While";

break;

case Repeat:

return "Repeat";

break;

case Until:

return "Until";

break;

case Label:

return "Label";

break;

case Identifier:

return "Identifier";

break;

case Number:

return "Number";

break;

case Assign:

return "==>";

break;

case Add:

return "++";

break;

case Sub:

return "--";

break;

case Mul:

return "\*\*";

break;

case Div:

return "Div";

break;

case Mod:

return "Mod";

break;

case Equality:

return "Eq";

break;

case NotEquality:

return "Neq";

break;

case Greate:

return "Ls";

break;

case Less:

return "Gr";

break;

case Not:

return "Not";

break;

case And:

return "And";

break;

case Or:

return "Or";

break;

case LBraket:

return "(";

break;

case RBraket:

return ")";

break;

case Semicolon:

return ";";

break;

case Comma:

return ",";

break;

case Unknown:

return "Unknown";

break;

default:

return "Undefined";

break;

}

}

void parser()

{

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "at")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return;

}

program();

fclose(ErrorsFile);

}

//Розглянемо рекурсивні процедури, що реалізують синтаксичний аналізатор :

void match(TypeOfTokens expectedType)

{

if (TokenTable[pos].type != expectedType) {

const char\* type = getType(expectedType);

if (pos < TokensNum) {

printf("Syntax error in line %d. Expected '%s'\n", TokenTable[pos].line, type);

fprintf(ErrorsFile, "Syntax error in line %d. Expected '%s'\n", TokenTable[pos].line, type);

}

else {

printf("Syntax error in line %d. Expected '%s'\n", TokenTable[TokensNum - 1].line + 1, type);

fprintf(ErrorsFile, "Syntax error in line %d. Expected '%s'\n", TokenTable[TokensNum - 1].line + 1, type);

}

errors++;

}

pos++;

}

// <програма> = ‘StartProgram’ <ідентифікатор> ‘;’ ‘Variable’ <оголошення змінних> ‘;’ <тіло програми> ‘EndBlok’ ';'

void program()

{

match(StartProgram);

match(Semicolon);

match(Variable);

variable\_declaration();

match(Semicolon);

program\_body();

match(End);

match(Semicolon);

while (pos < TokensNum) {

printf("Syntax error in line %d. Statement after end operator.\n", TokenTable[pos].line);

fprintf(ErrorsFile, "Syntax error in line %d. Statement after end operator.\n", TokenTable[pos].line);

errors++;

pos++;

}

}

// <оголошення змінних> = [<тип даних> <список змінних>]

void variable\_declaration()

{

if (TokenTable[pos].type == Type)

{

pos++;

variable\_list();

}

}

// <список змінних> = <ідентифікатор> ['==>' <арифметичний вираз>] { ‘,’ <ідентифікатор> ['==>' <арифметичний вираз>] }

void variable\_list()

{

match(Identifier);

if (TokenTable[pos].type == Assign) {

pos++;

match(Number);

}

while (TokenTable[pos].type == Comma)

{

pos++;

match(Identifier);

if (TokenTable[pos].type == Assign) {

pos++;

match(Number);

}

}

}

// <тіло програми> = { <оператор> }

void program\_body()

{

do

{

if (pos > TokensNum) break;

statement();

} while (TokenTable[pos].type != End);

}

// <оператор> = <присвоєння> | <ввід> | <вивід> | <умовний оператор> | <безумовний перехід> | <мітка> | <оператор циклу forTo>

// | <оператор циклу forDownTo> | <оператор циклу while> | <оператор циклу repeat> | <складений оператор>

void statement()

{

switch (TokenTable[pos].type)

{

case Input: input(); break;

case Output: output(); break;

case If: conditional(); break;

case Goto: goto\_statement(); break;

case For: for\_statement(); break;

case While: while\_statement(); break;

case Repeat: repeat\_statement(); break;

case Identifier: assignment(); break;

case Label: label(); break;

case StartBody: compound\_statement(); break;

default: {

if (pos < TokensNum) {

printf("Syntax error in line %d. Expected statement.\n", TokenTable[pos].line);

fprintf(ErrorsFile, "Syntax error in line %d. Expected statement.\n", TokenTable[pos].line);

}

else {

printf("Syntax error in line %d. Expected statement.\n", TokenTable[TokensNum - 1].line + 1);

fprintf(ErrorsFile, "Syntax error in line %d. Expected statement.\n", TokenTable[TokensNum - 1].line + 1);

}

errors++;

pos++;

}

}

}

// <присвоєння> = <ідентифікатор> '==>' <арифметичний вираз> ';'

void assignment()

{

match(Identifier);

match(Assign);

arithmetic\_expression();

match(Semicolon);

}

// <мітка> ==> <ідентифікатор> ':'

void label()

{

match(Label);

}

// <арифметичний вираз> = <доданок> { ‘++’ <доданок> | ‘--’ <доданок> }

void arithmetic\_expression()

{

term();

while (TokenTable[pos].type == Add || TokenTable[pos].type == Sub)

{

pos++;

term();

}

}

// <доданок> = <множник> { ‘\*\*’ <доданок> | ‘Div’ <доданок> | ‘Mod’ <доданок> }

void term()

{

factor();

while (TokenTable[pos].type == Mul || TokenTable[pos].type == Div || TokenTable[pos].type == Mod)

{

pos++;

factor();

}

}

// <множник> = <ідентифікатор> | <число> | ‘(’ <арифметичний вираз> ‘)’

void factor()

{

if (TokenTable[pos].type == Identifier)

{

match(Identifier);

}

else if (TokenTable[pos].type == Number)

{

match(Number);

}

else if (TokenTable[pos].type == LBraket)

{

match(LBraket);

arithmetic\_expression();

match(RBraket);

}

else

{

if (pos < TokensNum) {

printf("Syntax error in line %d. Expected multiplier\n", TokenTable[pos].line);

fprintf(ErrorsFile, "Syntax error in line %d. Expected multiplier\n", TokenTable[pos].line);

}

else {

printf("Syntax error in line %d. Expected multiplier\n", TokenTable[TokensNum - 1].line + 1);

fprintf(ErrorsFile, "Syntax error in line %d. Expected multiplier\n", TokenTable[TokensNum - 1].line + 1);

}

errors++;

}

}

// <ввід> = 'Scan' <ідентифікатор> ';'

void input()

{

match(Input);

arithmetic\_expression();

match(Semicolon);

}

// <вивід> = ‘Print’ <арифметичний вираз> ';'

void output()

{

match(Output);

arithmetic\_expression();

match(Semicolon);

}

// <умовний оператор> ==> 'If' '(' <логічний вираз> ')' <оператор> ['Else' <оператор>]

void conditional()

{

match(If);

match(LBraket);

logical\_expression();

match(RBraket);

statement();

if (TokenTable[pos].type == Else)

{

pos++;

statement();

}

}

// <логічний вираз> = <вираз І> { ‘Or’ <вираз І> }

void logical\_expression()

{

and\_expression();

while (TokenTable[pos].type == Or)

{

pos++;

and\_expression();

}

}

// <вираз І> = <порівняння> { ‘And’ <вираз І> }

void and\_expression()

{

comparison();

while (TokenTable[pos].type == And)

{

pos++;

and\_expression();

}

}

// <порівняння> = <операція порівняння> | ‘Not’ ‘(’ <логічний вираз> ‘)’ | ‘(’ <логічний вираз> ‘)’

// <операція порівняння> = <арифметичний вираз> <менше-більше> <арифметичний вираз>

// <менше-більше> = ‘Eq’ | ‘Neq’ | ‘Ls’ | ‘Gr’

void comparison()

{

if (TokenTable[pos].type == Not)

{

// Варіант: Not(<логічний вираз>)

pos++;

match(LBraket);

logical\_expression();

match(RBraket);

}

else if (TokenTable[pos].type == LBraket)

{

// Варіант: ( <логічний вираз> )

pos++;

logical\_expression();

match(RBraket);

}

else

{

// Варіант: <арифметичний вираз> <менше-більше> <арифметичний вираз>

arithmetic\_expression();

if (TokenTable[pos].type == Greate || TokenTable[pos].type == Less || TokenTable[pos].type == Equality || TokenTable[pos].type == NotEquality)

{

pos++;

arithmetic\_expression();

}

else

{

if (pos < TokensNum) {

printf("Syntax error in line %d. Expected comparison operation\n", TokenTable[pos].line);

fprintf(ErrorsFile, "Syntax error in line %d. Expected comparison operation\n", TokenTable[pos].line);

}

else {

printf("Syntax error in line %d. Expected comparison operation\n", TokenTable[TokensNum - 1].line + 1);

fprintf(ErrorsFile, "Syntax error in line %d. Expected comparison operation\n", TokenTable[TokensNum - 1].line + 1);

}

errors++;

}

}

}

// <оператор циклу> ==> <оператор циклу forTo> | <оператор циклу forDownTo>

void for\_statement() {

match(For);

match(Identifier);

match(Assign);

arithmetic\_expression();

if (TokenTable[pos].type == To)

forTo\_statement();

else

forDownTo\_statement();

}

// <оператор циклу forTo> ==> 'For' <ідентифікатор> '==>' <арифметичний вираз> 'To' <арифметичний вираз> 'Do' <оператор>

void forTo\_statement() {

match(To);

arithmetic\_expression();

match(Do);

statement();

}

// <оператор циклу forDownTo> ==> 'For' <ідентифікатор> '==>' <арифметичний вираз> 'DownTo' <арифметичний вираз> 'Do' <оператор>

void forDownTo\_statement() {

match(DownTo);

arithmetic\_expression();

match(Do);

statement();

}

// <оператор циклу while> ==> 'While' '(' <логічний вираз> ')' 'Do' <оператор>

void while\_statement() {

match(While);

match(LBraket);

logical\_expression();

match(RBraket);

match(Do);

statement();

}

// <оператор циклу repeat> ==> 'Repeat' <оператор> 'Until' '(' <логічний вираз> ')' ';'

void repeat\_statement() {

match(Repeat);

statement();

match(Until);

match(LBraket);

logical\_expression();

match(RBraket);

match(Semicolon);

}

// <безумовний перехід> ==> 'Goto' <ідентифікатор> ';'

void goto\_statement() {

match(Goto);

match(Identifier);

match(Semicolon);

}

// <складений оператор> = ‘StartBlok’ <тіло програми> ‘EndBlok’ ';'

void compound\_statement()

{

match(StartBody);

program\_body();

match(End);

match(Semicolon);

}

**LexicalAnalyzer.h  
Заголовковий файл, що оголошує функції для лексичного аналізу. Визначає зовнішні змінні для обробки помилок та файлів. Містить функції:**

#pragma once

#include <stdio.h>

#include <string.h>

#include <stdlib.h>

#include "header.h"

extern unsigned int errors;

extern char\* ErrorsFileName;

extern FILE\* ErrorsFile;

unsigned int GetTokens(FILE\* F, Token TokenTable[]);

void PrintTokens(Token TokenTable[], unsigned int TokensNum);

void PrintTokensToFile(char\* FileName, Token TokenTable[], unsigned int TokensNum);

**LexicalAnalyzer.cpp  
Лексичний аналізатор, що розбиває вхідний текст на лексеми, виявляє синтаксичні одиниці та фіксує помилки.**

#include "LexicalAnalyzer.h"

// функція отримує лексеми з вхідного файлу F і записує їх у таблицю лексем TokenTable

// результат функції - кількість лексем

unsigned int GetTokens(FILE\* F, Token TokenTable[])

{

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "at")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return 0;

}

States state = Start;

Token TempToken;

// кількість лексем

unsigned int NumberOfTokens = 0;

char ch, buf[16];

int line = 1;

// читання першого символу з файлу

ch = getc(F);

// пошук лексем

while (1)

{

switch (state)

{

// стан Start - початок виділення чергової лексеми

// якщо поточний символ маленька літера, то переходимо до стану Letter

// якщо поточний символ цифра, то переходимо до стану Digit

// якщо поточний символ пробіл, символ табуляції або переходу на новий рядок, то переходимо до стану Separators

// якщо поточний символ { то є ймовірність, що це коментар, переходимо до стану SComment

// якщо поточний символ EOF (ознака кінця файлу), то переходимо до стану EndOfFile

// якщо поточний символ відмінний від попередніх, то переходимо до стану Another

case Start:

{

if (ch == EOF)

state = EndOfFile;

else

if (ch <= 'z' && ch >= 'a' || ch <= 'Z' && ch >= 'A')

state = KeyWords;

else

if (ch == '\_')

state = Identifiers;

else

if (ch <= '9' && ch >= '0')

state = Digit;

else

if (ch == ' ' || ch == '\t' || ch == '\n')

state = Separators;

else

if (ch == '{')

state = Comment;

else

state = Another;

break;

}

// стан Finish - кінець виділення чергової лексеми і запис лексеми у таблицю лексем

case Finish:

{

if (NumberOfTokens < MAX\_TOKENS)

{

if (TempToken.type == Unknown) {

//Логування лексичних помилок

printf("Lexical error in line %d. '%s' is unknown.\n", line, TempToken.name);

fprintf(ErrorsFile, "Lexical error in line %d. '%s' is unknown.\n", line, TempToken.name);

errors++;

}

TokenTable[NumberOfTokens++] = TempToken;

if (ch != EOF)

state = Start;

else

state = EndOfFile;

}

else

{

printf("\n\t\t\ttoo many tokens !!!\n");

fclose(ErrorsFile);

return NumberOfTokens - 1;

}

break;

}

// стан EndOfFile - кінець файлу, можна завершувати пошук лексем

case EndOfFile:

{

fclose(ErrorsFile);

return NumberOfTokens;

}

// стан KeyWords - поточний символ - велика літера, поточна лексема - ключове слово

case KeyWords:

{

buf[0] = ch;

int j = 1;

ch = getc(F);

while (((ch <= 'Z' && ch >= 'A') || (ch <= 'z' && ch >= 'a') || (ch <= '9' && ch >= '0')) && j < MAX\_LEXEM\_NAME - 1)

{

buf[j++] = ch;

ch = getc(F);

}

buf[j] = '\0';

TypeOfTokens temp\_type = Unknown;

if (!strcmp(buf, "StartProgram"))

temp\_type = StartProgram;

else if (!strcmp(buf, "StartBlok"))

temp\_type = StartBody;

else if (!strcmp(buf, "Variable"))

temp\_type = Variable;

else if (!strcmp(buf, "Int32"))

temp\_type = Type;

else if (!strcmp(buf, "EndBlok"))

temp\_type = End;

else if (!strcmp(buf, "Scan"))

temp\_type = Input;

else if (!strcmp(buf, "Print"))

temp\_type = Output;

else if (!strcmp(buf, "If"))

temp\_type = If;

else if (!strcmp(buf, "Else"))

temp\_type = Else;

else if (!strcmp(buf, "Goto"))

temp\_type = Goto;

else if (!strcmp(buf, "For"))

temp\_type = For;

else if (!strcmp(buf, "To"))

temp\_type = To;

else if (!strcmp(buf, "DownTo"))

temp\_type = DownTo;

else if (!strcmp(buf, "Do"))

temp\_type = Do;

else if (!strcmp(buf, "While"))

temp\_type = While;

else if (!strcmp(buf, "Repeat"))

temp\_type = Repeat;

else if (!strcmp(buf, "Until"))

temp\_type = Until;

else if (!strcmp(buf, "Div"))

temp\_type = Div;

else if (!strcmp(buf, "Mod"))

temp\_type = Mod;

else if (!strcmp(buf, "Eq"))

temp\_type = Equality;

else if (!strcmp(buf, "Neq"))

temp\_type = NotEquality;

else if (!strcmp(buf, "Ls"))

temp\_type = Less;

else if (!strcmp(buf, "Gr"))

temp\_type = Greate;

else if (!strcmp(buf, "Not"))

temp\_type = Not;

else if (!strcmp(buf, "And"))

temp\_type = And;

else if (!strcmp(buf, "Or"))

temp\_type = Or;

strcpy\_s(TempToken.name, buf);

TempToken.type = temp\_type;

TempToken.value = 0;

TempToken.line = line;

state = Finish;

break;

}

// стан Identifiers - поточний символ - символ '\_', поточна лексема - ідентифікатор

case Identifiers:

{

ch = getc(F);

if (ch <= 'Z' && ch >= 'A' || ch <= 'z' && ch >= 'a')

{

buf[0] = '\_';

buf[1] = ch;

int j = 2;

ch = getc(F);

TypeOfTokens temp\_type = Identifier;

while (((ch <= 'Z' && ch >= 'A') || (ch <= 'z' && ch >= 'a') || (ch <= '9' && ch >= '0')) && j < MAX\_LEXEM\_NAME - 1)

{

buf[j++] = ch;

ch = getc(F);

}

buf[j] = '\0';

if (strlen(buf) > MAX\_IDENTIFIER\_LENGTH)

temp\_type = Unknown;

if (temp\_type == Identifier && ch == ':') {

temp\_type = Label;

buf[j++] = ch;

buf[j] = '\0';

ch = getc(F);

}

strcpy\_s(TempToken.name, buf);

TempToken.type = temp\_type;

TempToken.value = 0;

TempToken.line = line;

state = Finish;

}

else

{

strcpy\_s(TempToken.name, "\_");

TempToken.type = Unknown;

TempToken.value = 0;

TempToken.line = line;

state = Finish;

}

break;

}

// стан Digit - поточний символ - цифра, поточна лексема - число

case Digit:

{

buf[0] = ch;

int j = 1;

ch = getc(F);

TypeOfTokens temp\_type = Number;

while (((ch <= 'z' && ch >= 'a') || (ch <= 'Z' && ch >= 'A') || (ch <= '9' && ch >= '0') || ch == '.') && j < MAX\_LEXEM\_NAME - 1)

{

if (ch == '.') {

ch = getc(F);

if (ch <= '9' && ch >= '0') {

fseek(F, ftell(F) - 2, SEEK\_SET);

ch = getc(F);

}

else {

fseek(F, ftell(F) - 2, SEEK\_SET);

ch = getc(F);

break;

}

}

if (!(ch <= '9' && ch >= '0'))

temp\_type = Unknown;

buf[j++] = ch;

ch = getc(F);

}

buf[j] = '\0';

if (strlen(buf) == 1 && !(buf[0] <= '9' && buf[0] >= '0')) {

if (buf[0] == '-') {

strcpy\_s(TempToken.name, "-");

}

else if (buf[0] == '+') {

strcpy\_s(TempToken.name, "+");

}

TempToken.type = Unknown;

TempToken.value = 0;

TempToken.line = line;

state = Finish;

}

else {

strcpy\_s(TempToken.name, buf);

TempToken.type = temp\_type;

TempToken.value = temp\_type == Unknown ? 0 : atoi(buf);

TempToken.line = line;

state = Finish;

}

break;

}

// стан Separators - поточний символ пробіл, символ табуляції або переходу на новий рядок - видаляємо їх

case Separators:

{

if (ch == '\n')

line++;

ch = getc(F);

state = Start;

break;

}

// стан Comment - поточний символ '{', отже це коментар, видаляємо усі символи до '}'

case Comment:

{

long commentStart = ftell(F);

int tLine = line;

while (1)

{

ch = getc(F);

if (ch == EOF)

{

fseek(F, commentStart, SEEK\_SET);

line = tLine;

strcpy\_s(TempToken.name, "{");

TempToken.type = Unknown;

TempToken.value = 0;

TempToken.line = line;

ch = getc(F);

state = Start;

break;

}

else if (ch == '\n')

{

line++;

}

else if (ch == '}')

{

ch = getc(F);

state = Start;

break;

}

}

break;

}

// стан Another - поточний символ - символ, відмінний від попередніх

case Another:

{

TempToken.value = 0;

TempToken.line = line;

state = Finish;

if (ch == '+') {

ch = getc(F);

if (ch <= '9' && ch >= '0')

{

fseek(F, ftell(F) - 2, SEEK\_SET);

ch = getc(F);

state = Digit;

}

else if (ch == '+')

{

strcpy\_s(TempToken.name, "++");

TempToken.type = Add;

ch = getc(F);

}

else

{

strcpy\_s(TempToken.name, "+");

TempToken.type = Unknown;

}

}

else if (ch == '-') {

ch = getc(F);

if (ch <= '9' && ch >= '0')

{

fseek(F, ftell(F) - 2, SEEK\_SET);

ch = getc(F);

state = Digit;

}

else if (ch == '-')

{

strcpy\_s(TempToken.name, "--");

TempToken.type = Sub;

ch = getc(F);

}

else

{

strcpy\_s(TempToken.name, "-");

TempToken.type = Unknown;

}

}

else if (ch == '\*') {

ch = getc(F);

if (ch == '\*')

{

strcpy\_s(TempToken.name, "\*\*");

TempToken.type = Mul;

ch = getc(F);

}

else

{

strcpy\_s(TempToken.name, "\*");

TempToken.type = Unknown;

}

}

else if (ch == '=') {

ch = getc(F);

if (ch == '=')

{

ch = getc(F);

if (ch == '>')

{

strcpy\_s(TempToken.name, "==>");

TempToken.type = Assign;

ch = getc(F);

}

else

{

strcpy\_s(TempToken.name, "==");

TempToken.type = Unknown;

}

}

else

{

strcpy\_s(TempToken.name, "=");

TempToken.type = Unknown;

}

}

else if (ch == '(') {

strcpy\_s(TempToken.name, "(");

TempToken.type = LBraket;

ch = getc(F);

}

else if (ch == ')') {

strcpy\_s(TempToken.name, ")");

TempToken.type = RBraket;

ch = getc(F);

}

else if (ch == ';') {

strcpy\_s(TempToken.name, ";");

TempToken.type = Semicolon;

ch = getc(F);

}

else if (ch == ',') {

strcpy\_s(TempToken.name, ",");

TempToken.type = Comma;

ch = getc(F);

}

else {

TempToken.name[0] = ch;

TempToken.name[1] = '\0';

TempToken.type = Unknown;

ch = getc(F);

}

}

}

}

fclose(ErrorsFile);

}

// функція друкує таблицю лексем на екран

void PrintTokens(Token TokenTable[], unsigned int TokensNum)

{

char type\_tokens[16];

printf("\n\n---------------------------------------------------------------------------\n");

printf("| TOKEN TABLE |\n");

printf("---------------------------------------------------------------------------\n");

printf("| line number | token | value | token code | type of token |\n");

printf("---------------------------------------------------------------------------");

for (unsigned int i = 0; i < TokensNum; i++)

{

switch (TokenTable[i].type)

{

case StartProgram:

strcpy\_s(type\_tokens, "StartProgram");

break;

case Variable:

strcpy\_s(type\_tokens, "Variable");

break;

case Type:

strcpy\_s(type\_tokens, "Integer");

break;

case Identifier:

strcpy\_s(type\_tokens, "Identifier");

break;

case Label:

strcpy\_s(type\_tokens, "Label");

break;

case StartBody:

strcpy\_s(type\_tokens, "Start");

break;

case End:

strcpy\_s(type\_tokens, "End");

break;

case Input:

strcpy\_s(type\_tokens, "Input");

break;

case Output:

strcpy\_s(type\_tokens, "Output");

break;

case If:

strcpy\_s(type\_tokens, "If");

break;

case Else:

strcpy\_s(type\_tokens, "Else");

break;

case Goto:

strcpy\_s(type\_tokens, "Goto");

break;

case For:

strcpy\_s(type\_tokens, "For");

break;

case To:

strcpy\_s(type\_tokens, "To");

break;

case DownTo:

strcpy\_s(type\_tokens, "DownTo");

break;

case While:

strcpy\_s(type\_tokens, "While");

break;

case Repeat:

strcpy\_s(type\_tokens, "Repeat");

break;

case Until:

strcpy\_s(type\_tokens, "Until");

break;

case Assign:

strcpy\_s(type\_tokens, "Assign");

break;

case Add:

strcpy\_s(type\_tokens, "Add");

break;

case Sub:

strcpy\_s(type\_tokens, "Sub");

break;

case Mul:

strcpy\_s(type\_tokens, "Mul");

break;

case Div:

strcpy\_s(type\_tokens, "Div");

break;

case Mod:

strcpy\_s(type\_tokens, "Mod");

break;

case Equality:

strcpy\_s(type\_tokens, "Equality");

break;

case NotEquality:

strcpy\_s(type\_tokens, "NotEquality");

break;

case Greate:

strcpy\_s(type\_tokens, "Greate");

break;

case Less:

strcpy\_s(type\_tokens, "Less");

break;

case Not:

strcpy\_s(type\_tokens, "Not");

break;

case And:

strcpy\_s(type\_tokens, "And");

break;

case Or:

strcpy\_s(type\_tokens, "Or");

break;

case LBraket:

strcpy\_s(type\_tokens, "LBraket");

break;

case RBraket:

strcpy\_s(type\_tokens, "RBraket");

break;

case Number:

strcpy\_s(type\_tokens, "Number");

break;

case Semicolon:

strcpy\_s(type\_tokens, "Semicolon");

break;

case Comma:

strcpy\_s(type\_tokens, "Comma");

break;

case Unknown:

strcpy\_s(type\_tokens, "Unknown");

break;

}

printf("\n|%12d |%16s |%11d |%11d | %-13s |\n",

TokenTable[i].line,

TokenTable[i].name,

TokenTable[i].value,

TokenTable[i].type,

type\_tokens);

printf("---------------------------------------------------------------------------");

}

printf("\n");

}

// функція друкує таблицю лексем у файл

void PrintTokensToFile(char\* FileName, Token TokenTable[], unsigned int TokensNum)

{

FILE\* F;

if ((fopen\_s(&F, FileName, "wt")) != 0)

{

printf("Error: Can not create file: %s\n", FileName);

return;

}

char type\_tokens[16];

fprintf(F, "---------------------------------------------------------------------------\n");

fprintf(F, "| TOKEN TABLE |\n");

fprintf(F, "---------------------------------------------------------------------------\n");

fprintf(F, "| line number | token | value | token code | type of token |\n");

fprintf(F, "---------------------------------------------------------------------------");

for (unsigned int i = 0; i < TokensNum; i++)

{

switch (TokenTable[i].type)

{

case StartProgram:

strcpy\_s(type\_tokens, "StartProgram");

break;

case Variable:

strcpy\_s(type\_tokens, "Variable");

break;

case Type:

strcpy\_s(type\_tokens, "Integer");

break;

case Identifier:

strcpy\_s(type\_tokens, "Identifier");

break;

case Label:

strcpy\_s(type\_tokens, "Label");

break;

case StartBody:

strcpy\_s(type\_tokens, "Start");

break;

case End:

strcpy\_s(type\_tokens, "End");

break;

case Input:

strcpy\_s(type\_tokens, "Input");

break;

case Output:

strcpy\_s(type\_tokens, "Output");

break;

case If:

strcpy\_s(type\_tokens, "If");

break;

case Else:

strcpy\_s(type\_tokens, "Else");

break;

case Goto:

strcpy\_s(type\_tokens, "Goto");

break;

case For:

strcpy\_s(type\_tokens, "For");

break;

case To:

strcpy\_s(type\_tokens, "To");

break;

case DownTo:

strcpy\_s(type\_tokens, "DownTo");

break;

case While:

strcpy\_s(type\_tokens, "While");

break;

case Repeat:

strcpy\_s(type\_tokens, "Repeat");

break;

case Until:

strcpy\_s(type\_tokens, "Until");

break;

case Assign:

strcpy\_s(type\_tokens, "Assign");

break;

case Add:

strcpy\_s(type\_tokens, "Add");

break;

case Sub:

strcpy\_s(type\_tokens, "Sub");

break;

case Mul:

strcpy\_s(type\_tokens, "Mul");

break;

case Div:

strcpy\_s(type\_tokens, "Div");

break;

case Mod:

strcpy\_s(type\_tokens, "Mod");

break;

case Equality:

strcpy\_s(type\_tokens, "Equality");

break;

case NotEquality:

strcpy\_s(type\_tokens, "NotEquality");

break;

case Greate:

strcpy\_s(type\_tokens, "Greate");

break;

case Less:

strcpy\_s(type\_tokens, "Less");

break;

case Not:

strcpy\_s(type\_tokens, "Not");

break;

case And:

strcpy\_s(type\_tokens, "And");

break;

case Or:

strcpy\_s(type\_tokens, "Or");

break;

case LBraket:

strcpy\_s(type\_tokens, "LBraket");

break;

case RBraket:

strcpy\_s(type\_tokens, "RBraket");

break;

case Number:

strcpy\_s(type\_tokens, "Number");

break;

case Semicolon:

strcpy\_s(type\_tokens, "Semicolon");

break;

case Comma:

strcpy\_s(type\_tokens, "Comma");

break;

case Unknown:

strcpy\_s(type\_tokens, "Unknown");

break;

}

fprintf(F, "\n|%12d |%16s |%11d |%11d | %-13s |\n",

TokenTable[i].line,

TokenTable[i].name,

TokenTable[i].value,

TokenTable[i].type,

type\_tokens);

fprintf(F, "---------------------------------------------------------------------------");

}

fclose(F);

}

**main.cpp  
Програма є точкою входу для транслятора. Виконує ініціалізацію, обробку аргументів, запуск лексичного, синтаксичного аналізаторів і генератора коду.**

#include <stdio.h>

#include <string.h>

#include <stdlib.h>

#include "header.h"

#include "LexicalAnalyzer.h"

#include "SyntaxAnalyzer.h"

#include "CodeGenerator.h"

// таблиця лексем

Token \* TokenTable;

// кількість лексем

unsigned int TokensNum;

// таблиця ідентифікаторів

Id\* IdTable;

// кількість ідентифікаторів

unsigned int IdNum;

unsigned int errors = 0;

char\* ErrorsFileName;

FILE\* ErrorsFile;

unsigned int IdIdentification(Id IdTable[], Token TokenTable[], unsigned int tokenCount);

int main(int argc, char\* argv[])

{

char InputFile[32] = "";

FILE\* InFile;

if (argc != 2) {

printf("p18 to C translator\n~~~~~~~~~~~~~~~~~~~~~~~~\nInput file name: ");

gets\_s(InputFile);

}

else {

strcpy\_s(InputFile, argv[1]);

}

int i = 0;

const char format[4] = "p18";

while (InputFile[i++] != '.');

for (int j = 0; j < strlen(format);) {

if (InputFile[i++] != format[j++]) {

printf("Unknown file format.\n");

return 1;

}

}

if ((fopen\_s(&InFile, InputFile, "rt")) != 0)

{

printf("Error: Can not open file: %s\n", InputFile);

return 1;

}

char ErrorFile[32] = "Errors.txt";

ErrorsFileName = ErrorFile;

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "wt")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return 0;

}

else fclose(ErrorsFile);

// таблиця лесем

TokenTable = new Token[MAX\_TOKENS];

// таблиця ідентифікаторів

IdTable = new Id[MAX\_IDENTIFIER\_COUNT];

TokensNum = GetTokens(InFile, TokenTable);

fclose(InFile);

if (errors == 0) {

char TokenFile[32];

i = 0;

while (InputFile[i] != '.')

{

TokenFile[i] = InputFile[i];

i++;

}

TokenFile[i] = '\0';

strcat\_s(TokenFile, ".token");

PrintTokensToFile(TokenFile, TokenTable, TokensNum);

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "at")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return 0;

}

fprintf(ErrorsFile, "Lexical analysis completed successfully\n");

printf("\nThe program is lexical correct. List of tokens in the file %s\n", TokenFile);

fclose(ErrorsFile);

parser();

if (errors == 0) {

printf("\nThe program is syntax correct.\n");

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "at")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return 0;

}

fprintf(ErrorsFile, "\nSyntax analysis completed successfully\n");

int ids = IdIdentification(IdTable, TokenTable, TokensNum);

fclose(ErrorsFile);

if (errors == 0) {

if ((fopen\_s(&ErrorsFile, ErrorsFileName, "at")) != 0)

{

printf("Error: Can not create file: %s\n", ErrorsFileName);

return 0;

}

fprintf(ErrorsFile, "\nSemantic analysis completed successfully, %d identifier found\n", ids);

printf("\nThe program is semantic correct. %d identifier found\n", ids);

fclose(ErrorsFile);

char OutputFile[32];

i = 0;

while (InputFile[i] != '.')

{

OutputFile[i] = InputFile[i];

i++;

}

OutputFile[i] = '\0';

strcat\_s(OutputFile, ".c");

FILE\* outFile;

fopen\_s(&outFile, OutputFile, "w");

if (!outFile)

{

printf("Failed to open output file.\n");

exit(1);

}

// генерація вихідного С коду

generateCCode(outFile);

printf("\nC code generated to %s\n", OutputFile);

}

else {

printf("The program has semantic errors.\n");

}

}

else {

printf("The program has syntax errors.\n");

}

}

else {

printf("The program has lexical errors.\n");

}

delete[]TokenTable;

delete[]IdTable;

if (argc != 2) system("pause");

return 0;

}

// функція записує оголошені ідентифікатори в таблицю ідентифікаторів IdTable

// повертає кількість ідентифікаторів

// перевіряє чи усі використані ідентифікатори оголошені

unsigned int IdIdentification(Id IdTable[], Token TokenTable[], unsigned int tokenCount)

{

unsigned int idCount = 0;

unsigned int i = 0;

while (TokenTable[i++].type != Variable);

if (TokenTable[i++].type == Type)

{

while (TokenTable[i].type != Semicolon)

{

if (TokenTable[i].type == Identifier)

{

int yes = 0;

for (unsigned int j = 0; j < idCount; j++)

{

if (!strcmp(TokenTable[i].name, IdTable[j].name))

{

yes = 1;

break;

}

}

if (yes == 1)

{

yes = 0;

printf("Identifier '%s' is already declared!\n", TokenTable[i].name);

fprintf(ErrorsFile, "Identifier '%s' is already declared!\n", TokenTable[i].name);

errors++;

idCount--;

}

if (idCount < MAX\_IDENTIFIER\_COUNT)

{

strcpy\_s(IdTable[idCount++].name, TokenTable[i++].name);

}

else

{

printf("Too many identifiers!\n");

fprintf(ErrorsFile, "Too many identifiers!\n");

}

}

else

i++;

}

}

for (; i < tokenCount; i++)

{

if (TokenTable[i].type == Identifier && TokenTable[i - 1].type != Goto)

{

int yes = 0;

for (unsigned int j = 0; j < idCount; j++)

{

if (!strcmp(TokenTable[i].name, IdTable[j].name))

{

yes = 1;

break;

}

}

if (yes == 0) {

printf("In line %d, an undeclared identifier '%s'!\n", TokenTable[i].line, TokenTable[i].name);

fprintf(ErrorsFile, "In line %d, an undeclared identifier '%s'!\n", TokenTable[i].line, TokenTable[i].name);

errors++;

}

}

}

return idCount; // Повертає кількість ідентифікаторів

}

**header.h  
Заголовковий файл, що містить оголошення структур, перерахувань, прототипів функцій і констант, які використовуються в трансляторі.**

#pragma once

#define MAX\_TOKENS 1000

#define MAX\_LEXEM\_NAME 16

#define MAX\_IDENTIFIER\_LENGTH 6

#define MAX\_IDENTIFIER\_COUNT 10

// перерахування, яке описує всі можливі типи лексем

enum TypeOfTokens

{

StartProgram, // StartProgram

StartBody, // StartBlok

Variable, // Variable

Type, // Int32

End, // EndBlok

Input, // Scan

Output, // Print

If, // If

Else, // Else

Goto, // Goto

For, // For

To, // To

DownTo, // DownTo

Do, // Do

While, // While

Repeat, // Repeat

Until, // Until

Label, // Label

Identifier, // Identifier

Number, // Number

Assign, // ==>

Add, // ++

Sub, // --

Mul, // \*\*

Div, // Div

Mod, // Mod

Equality, // Eq

NotEquality, // Neq

Greate, // Ls

Less, // Gr

Not, // Not

And, // And

Or, // Or

LBraket, // (

RBraket, // )

Semicolon, // ;

Comma, // ,

Unknown

};

// структура для зберігання інформації про лексему

struct Token

{

char name[16]; // ім'я лексеми

int value; // значення лексеми (для цілих констант)

int line; // номер рядка

TypeOfTokens type; // тип лексеми

};

// структура для зберігання інформації про ідентифікатор

struct Id

{

char name[16];

};

// перерахування, яке описує стани лексичного аналізатора

enum States

{

Start, // початок виділення чергової лексеми

Finish, // кінець виділення чергової лексеми

KeyWords, // опрацювання слів (ключові слова)

Identifiers,// опрацювання слів (ідентифікатори)

Digit, // опрацювання цифри

Separators, // видалення пробілів, символів табуляції і переходу на новий рядок

Another, // опрацювання інших символів

EndOfFile, // кінець файлу

Comment // видалення коментаря

};

Додаток Д (тестові програми)

**Тестова програма «*Лінійний алгоритм*»**

StartProgram;

Variable Int32 \_a, \_b, \_x, \_y;

Scan \_a;

Scan \_b;

Print \_a ++ \_b;

Print \_a -- \_b;

Print \_a \*\* \_b;

Print \_a Div \_b;

Print \_a Mod \_b;

\_x ==> (\_a -- \_b) \*\* 10 ++ (\_a ++ \_b) Div 10;

\_y ==> \_x ++ \_x Mod 10;

Print \_x;

Print \_y;

EndBlok;

**Тестова програма «Алгоритм з розгалуженням»**

StartProgram;

Variable Int32 \_a, \_b, \_c;

Scan \_a;

Scan \_b;

Scan \_c;

If(\_a Gr \_b)

If(\_a Gr \_b)

Print \_a;

Else

Print \_c;

Else

If(\_b Gr \_c)

Print \_b;

Else

Print \_c;

If ((\_a Eq \_b) And (\_a Eq \_c) And (\_b Eq \_c))

Print 1;

Else

Print 0;

If ((\_a Ls 0) Or (\_b Ls 0) Or (\_c Ls 0))

Print -1;

Else

Print 0;

If (Not (\_a Gr (\_b ++ \_c)))

Print 10;

Else

Print 0;

EndBlok;

**Тестова програма «Циклічний алгоритм»**

StartProgram;

Variable Int32 \_a, \_b, \_x, \_i, \_j;

Scan \_a;

Scan \_b;

For \_i ==> \_a To \_b Do

Print \_i \*\* \_i;

\_x ==> 0;

For \_i ==> 1 To \_a Do

For \_j ==> 1 To \_b Do

\_x ==> \_x ++ 1;

Print \_x;

EndBlok;

Додаток Е (Алгоритм транслятора P18)