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 №22**

**Виконав:** ст. гр. КІ-309

Цюсьмак В.А.

**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](#_Toc153318762)

[Завдання до курсового проекту 4](#_Toc153318763)

[Вступ 6](#_Toc153318764)

[1. Огляд методів та способів проектування трансляторів 7](#_Toc153318765)

[2. Формальний опис вхідної мови програмування 10](#_Toc153318766)

[2.1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура 10](#_Toc153318767)

[2.2. Опис термінальних символів та ключових слів 12](#_Toc153318768)

[3. Розробка транслятора вхідної мови програмування 14](#_Toc153318769)

[3.1. Вибір технології програмування 14](#_Toc153318770)

[3.2. Проектування таблиць транслятора 15](#_Toc153318771)

[3.3. Розробка лексичного аналізатора 17](#_Toc153318772)

[3.3.1. Розробка блок-схеми алгоритму 18](#_Toc153318773)

[3.3.2. Опис програми реалізації лексичного аналізатора 18](#_Toc153318774)

[3.4. Розробка синтаксичного та семантичного аналізатора 20](#_Toc153318775)

[3.4.1. Опис програми реалізації синтаксичного та семантичного аналізатора 21](#_Toc153318776)

[3.4.2. Розробка граф-схеми алгоритму 21](#_Toc153318777)

[3.5. Розробка генератора коду 22](#_Toc153318778)

[3.5.1. Розробка граф-схеми алгоритму 23](#_Toc153318779)

[3.5.2. Опис програми реалізації генератора коду 24](#_Toc153318780)

[4. Опис програми 25](#_Toc153318781)

[4.1. Опис інтерфейсу та інструкція користувачеві 28](#_Toc153318782)

[5. Відлагодження та тестування програми 29](#_Toc153318783)

[5.1. Виявлення лексичних та синтаксичних помилок 29](#_Toc153318784)

[5.2. Виявлення семантичних помилок 30](#_Toc153318785)

[5.3. Загальна перевірка коректності роботи транслятора 30](#_Toc153318786)

[5.4. Тестова програма №1 32](#_Toc153318787)

[5.5. Тестова програма №2 33](#_Toc153318788)

[5.6. Тестова програма №3 34](#_Toc153318789)

[Висновки 36](#_Toc153318790)

[Список використаної літератури 37](#_Toc153318791)

[Додатки 38](#_Toc153318792)

Завдання до курсового проекту

**Варіант 22**

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 .t22

Опис вхідної мови програмування:

* Тип даних: INTEGER\_2
* Блок тіла програми: NAME <name>; DATA…; BODY END
* Оператор вводу: INPUT ()
* Оператор виводу: OUTPUT ()
* Оператори: IF ELSE (C)

GOTO (C)

FOR-TO-DO (Паскаль)

FOR-DOWNTO-DO (Паскаль)

WHILE (Бейсік)

REPEAT-UNTIL (Паскаль)

* Регістр ключових слів: Up
* Регістр ідентифікаторів: Up-Low16 перший символ \_
* Операції арифметичні: +, -, \*, DIV, MOD
* Операції порівняння: =, <>, >, <
* 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).

topRule = "**NAME**", identifier , ";", varsBlok, ";", "**BODY**", operators, "**END**";

varsBlok = "**DATA**", "**INTEGER\_2**", identifier, [{ commaAndIdentifier }];

identifier = “\_”, up\_letter, { low\_letter | number } {16};

commaAndIdentifier = ",", identifier;

codeBlok = "**BODY**", write | read | assignment | ifStatement | goto\_statement | labelRule | forToOrDownToDoRule | while | repeatUntil, "**END**";

operators = write | read | assignment | ifStatement | goto\_statement | labelRule | forToOrDownToDoRule | while | repeatUntil;

read = "**INPUT**", "(", identifier, ")";

write = "**OUTPUT**", "(", equation | stringRule, ")";

assignment = identifier, "**<==**", equation;

cycle\_counter = identifier;

cycle\_counter\_last\_value = equation;

ifStatement = "**IF**", "(", equation, ")", codeBlok, ["**ELSE**", codeBlok];

goto\_statement = "**GOTO**", ident ;

labelRule = identifier, ":";

forToOrDownToDoRule = "**FOR**", cycle\_counter, "**<==**", equation , "**TO**" | "**DOWNTO**", cycle\_counter\_last\_value, "**DO",** codeBlok;

while = "**WHILE**", "(", equation, ")", "**BODY**", operators | whileContinue | whileExit, "**END**", "**WHILE**";

whileContinue = "**CONTINUE**", "**WHILE**";

whileExit = "**EXIT**", "**WHILE**";

repeatUntil = "**REPEAT**", operators, "**UNTIL**", "(", equation, ")";

equation = signedNumber | identifier | notRule [{ operationAndIdentOrNumber | equation }];

notRule = notOperation, signedNumber | identifier | equation;

operationAndIdentOrNumber = mult | arithmetic | logic | compare signedNumber | identifier | equation;

arithmetic = "**+**" | "**-**";

mult = "**\***" | "**DIV**" | "**MOD**";

logic = "**AND**" | "**OR**";

notOperation = "**NOT**";

compare = "**=**" | "**<>**" | "**<**" | "**>**";

comment = "**LComment**", text , "**RComment**" ;

LComment = "**##**";

RComment = "**##**";

text = { low\_letter | up\_letter | number };

signedNumber = [ sign ] digit [{digit}];

sign = "+" | "-";

low\_letter = "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" | "j" | "k" | "l" | "m" | "n" | "o" | "p" | "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x" | "y" | "z";

up\_letter = "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" | "J" | "K" | "L" | "M" | "N" | "O" | "P" | "Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X" | "Y" | "Z";

digit = "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9";

* 1. Опис термінальних символів та ключових слів

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

|  |  |
| --- | --- |
| Термінальний символ або ключове слово | Значення |
| NAME | Початок програми |
| BODY | Початок тексту програми |
| DATA | Початок блоку опису змінних |
| END | Кінець розділу операторів |
| INPUT | Оператор вводу змінних |
| OUTPUT | 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 |
| CONTINUE | Оператор циклу |
| EXIT | Оператор циклу |
| REPEAT | Початок тіла циклу |
| UNTIL | Оператор циклу |
| + | Оператор додавання |
| - | Оператор віднімання |
| \* | Оператор множення |
| DIV | Оператор ділення |
| MOD | Оператор знаходження залишку від ділення |
| = | Оператор перевірки на рівність |
| <> | Оператор перевірки на нерівність |
| < | Оператор перевірки чи менше |
| > | Оператор перевірки чи більше |
| NOT | Оператор логічного заперечення |
| AND | Оператор кон’юнкції |
| OR | Оператор диз’юнкції |
| INTEGER\_2 | 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,

Exit,

Continue,

End,

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** |
| Program | NAME |
| Start | BODY |
| Vars | DATA |
| End | END |
| VarType | INTEGER\_2 |
| Read | INPUT |
| Write | OUTPUT |
| Assignment | <== |
| If | IF |
| Else | ELSE |
| Goto | GOTO |
| Colon | : |
| Label |  |
| For | FOR |
| To | TO |
| DownTo | DOWNTO |
| Do | DO |
| While | WHILE |
| Continue | CONTINUE |
| Exit | EXIT |
| Repeat | REPEAT |
| Until | UNTIL |
| Addition | + |
| Subtraction | - |
| Multiplication | \* |
| Division | DIV |
| Mod | MOD |
| Equal | = |
| NotEqual | <> |
| Less | < |
| Greate | > |
| Not | NOT |
| And | AND |
| Or | OR |
| Identifier |  |
| Number |  |
| Unknown |  |
| Comma | , |
| Semicolon | ; |
| LBraket | ( |
| RBraket | ) |
| LComment | ## |
| RComment | ## |
| Comment |  |

* 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;
  + 1. Розробка блок-схеми алгоритму

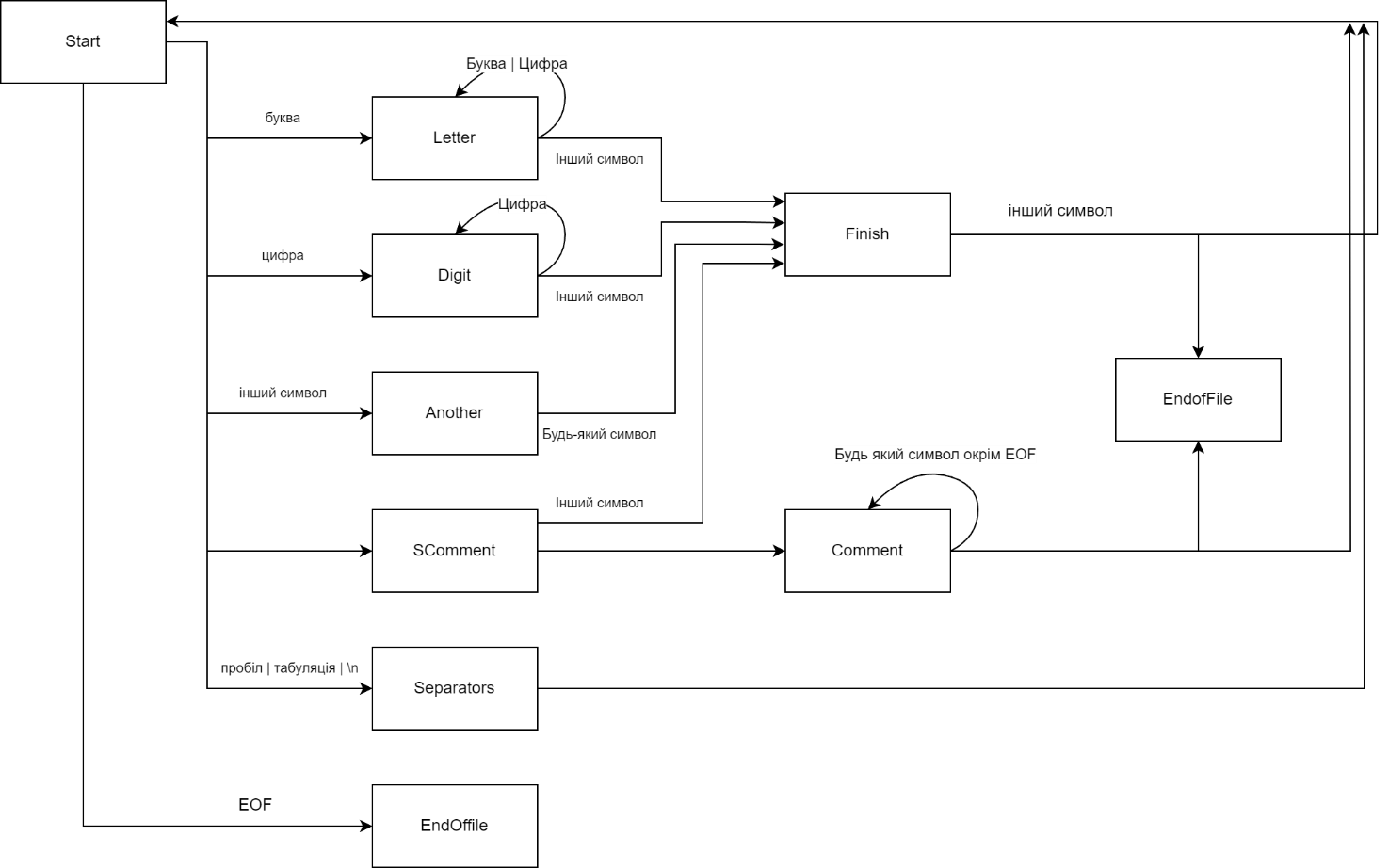


Рис. 3.1 Блок-схема роботи лексичного аналізатора

* + 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.

* 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. 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.

* + 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:

enum TypeOfTokens

{

Mainprogram,

StartProgram,

Variable,

Type,

EndProgram,

Input,

Output,

If,

Else,

Goto,

Label,

For,

To,

DownTo,

Do,

While,

Exit,

Continue,

End,

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(Mainprogram);

match(StartProgram);

match(Variable);

variable\_declaration();

match(Semicolon);

program\_body();

match(EndProgram);

}

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 t22. 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: "CWork\_t22.exe <ім’я програми>.t22"

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 lexems.txt, 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 error.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

Виявлення лексичних помилок відбувається на стадії лексичного аналізу. Під час розбиття вхідної програми на окремі лексеми відбувається перевірка чи відповідає вхідна лексема граматиці. Якщо ця лексема є в граматиці то вона ідентифікується і в таблиці лексем визначається. У випадку неспівпадіння лексемі присвоюється тип "невпізнаної лексеми". Повідомлення про такі помилки можна побачити лише після виконання процедури перевірки таблиці лексем, яка знаходиться в файлі.

Виявлення синтаксичних помилок відбувається на стадії перевірки програми на коректність окремо від синтаксичного аналізу. При цьому перевіряється окремо кожне твердження яке може бути або виразом, або оператором (циклу, вводу/виводу), або оголошенням, та перевіряється структура програми в цілому.

Приклад виявлення:

***Текст програми з помилками***

##Prog1##

NAME prog1;

DATA INTEGER32\_t \_Aaaaaaaaaaaa aaaaa,\_Bbbbbbbbbbbbbbbbb,\_Xxxxxxxxxxxxxxxxx,\_Yyyyyyyyyyyyyyyyy;

BODY

INPUT \_Aaaaaaaaaaaaaaaaa;

INPUT \_Bbbbbbbbbbbbbbbbb

OUTPUT \_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa \* \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa DIV \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa MOD \_Bbbbbbbbbbbbbbbbb;

\_Xxxxxxxxxxxxxxxxx<==(\_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb) \* 10 + (\_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb) DIV 10;

\_Yyyyyyyyyyyyyyyyy<==\_Xxxxxxxxxxxxxxxxx + (\_Xxxxxxxxxxxxxxxxx MOD 10);

OUTPUT \_Xxxxxxxxxxxxxxxxx;

OUTPUT \_Yyyyyyyyyyyyyyyyy;

END

***Текст файлу з повідомленнями про помилки***

Lexical Error: line 3, lexem \_Aaaaaaaaaaaa is Unknown

Lexical Error: line 3, lexem aaaaa is Unknown

Syntax error in line 3 : another type of lexeme was expected.

Syntax error: type Unknown

Expected Type: Identifier

* 1. Виявлення семантичних помилок

Суттю виявлення семантичних помилок є перевірка числових констант на відповідність типу INTEGER\_2, тобто знаковому цілому числу з відповідним діапазоном значень і перевірку на коректність використання змінних INTEGER\_2 у цілочисельних і логічних виразах.

* 1. Загальна перевірка коректності роботи транслятора

Для того щоб здійснити перевірку коректності роботи транслятора необхідно завантажити коректну до заданої вхідної мови програму.

***Текст коректної програми***

##Prog1##

NAME prog1;

DATA INTEGER32\_t \_Aaaaaaaaaaaaaaaaa,\_Bbbbbbbbbbbbbbbbb,\_Xxxxxxxxxxxxxxxxx,\_Yyyyyyyyyyyyyyyyy;

BODY

INPUT \_Aaaaaaaaaaaaaaaaa;

INPUT \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa \* \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa DIV \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa MOD \_Bbbbbbbbbbbbbbbbb;

\_Xxxxxxxxxxxxxxxxx<==(\_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb) \* 10 + (\_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb) DIV 10;

\_Yyyyyyyyyyyyyyyyy<==\_Xxxxxxxxxxxxxxxxx + (\_Xxxxxxxxxxxxxxxxx MOD 10);

OUTPUT \_Xxxxxxxxxxxxxxxxx;

OUTPUT \_Yyyyyyyyyyyyyyyyy;

END

Оскільки дана програма відповідає граматиці то результати виконання лексичного, синтаксичного аналізів, а також генератора коду будуть позитивними.

В результаті буде отримано с файл, який є результатом виконання трансляції з заданої вхідної мови на мову С даної програми (його вміст наведений в Додатку А).

Після виконання компіляції даного файлу на виході отримаєм наступний результат роботи програми:

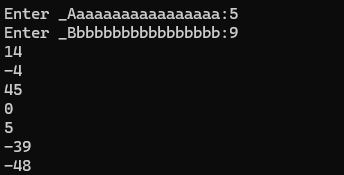


Рис. 5.1 Результат виконання коректної програми

При перевірці отриманого результату, можна зробити висновок про правильність роботи програми, а отже і про правильність роботи транслятора.

* 1. Тестова програма №1

***Текст програми***

##Prog1##

NAME prog1;

DATA INTEGER32\_t \_Aaaaaaaaaaaaaaaaa,\_Bbbbbbbbbbbbbbbbb,\_Xxxxxxxxxxxxxxxxx,\_Yyyyyyyyyyyyyyyyy;

BODY

INPUT \_Aaaaaaaaaaaaaaaaa;

INPUT \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa \* \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa DIV \_Bbbbbbbbbbbbbbbbb;

OUTPUT \_Aaaaaaaaaaaaaaaaa MOD \_Bbbbbbbbbbbbbbbbb;

\_Xxxxxxxxxxxxxxxxx<==(\_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb) \* 10 + (\_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb) DIV 10;

\_Yyyyyyyyyyyyyyyyy<==\_Xxxxxxxxxxxxxxxxx + (\_Xxxxxxxxxxxxxxxxx MOD 10);

OUTPUT \_Xxxxxxxxxxxxxxxxx;

OUTPUT \_Yyyyyyyyyyyyyyyyy;

END

***Результат виконання***

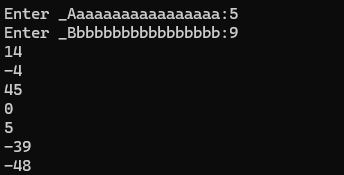


Рис. 5.2 Результат виконання тестової програми №1

* 1. Тестова програма №2

***Текст програми***

##Prog2##

NAME prog2;

DATA INTEGER32\_t \_Aaaaaaaaaaaaaaaaa,\_Bbbbbbbbbbbbbbbbb,\_Ccccccccccccccccc;

BODY

INPUT \_Aaaaaaaaaaaaaaaaa;

INPUT \_Bbbbbbbbbbbbbbbbb;

INPUT \_Ccccccccccccccccc;

IF(\_Aaaaaaaaaaaaaaaaa > \_Bbbbbbbbbbbbbbbbb)

BODY

IF(\_Aaaaaaaaaaaaaaaaa > \_Ccccccccccccccccc)

BODY

GOTO Abigger;

END

ELSE

BODY

OUTPUT \_Ccccccccccccccccc;

GOTO OutofIF;

Abigger:

OUTPUT \_Aaaaaaaaaaaaaaaaa;

GOTO OutofIF;

END

END

IF(\_Bbbbbbbbbbbbbbbbb < \_Ccccccccccccccccc)

BODY

OUTPUT \_Ccccccccccccccccc;

END

ELSE

BODY

OUTPUT \_Bbbbbbbbbbbbbbbbb;

END

OutofIF:

IF((\_Aaaaaaaaaaaaaaaaa = \_Bbbbbbbbbbbbbbbbb) AND (\_Aaaaaaaaaaaaaaaaa = \_Ccccccccccccccccc) AND (\_Bbbbbbbbbbbbbbbbb = \_Ccccccccccccccccc))

BODY

OUTPUT 1;

END

ELSE

BODY

OUTPUT 0;

END

IF((\_Aaaaaaaaaaaaaaaaa < 0) OR (\_Bbbbbbbbbbbbbbbbb < 0) OR (\_Ccccccccccccccccc < 0))

BODY

OUTPUT -1;

END

ELSE

BODY

OUTPUT 0;

END

IF(NOT(\_Aaaaaaaaaaaaaaaaa < (\_Bbbbbbbbbbbbbbbbb + \_Ccccccccccccccccc)))

BODY

OUTPUT(10);

END

ELSE

BODY

OUTPUT(0);

END

END

***Результат виконання***

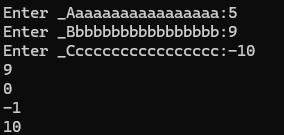


Рис. 5.3 Результат виконання тестової програми №2

* 1. Тестова програма №3

***Текст програми***

##Prog3##

NAME prog3;

DATA INTEGER32\_t \_Aaaaaaaaaaaaaaaaa,\_Aaaaaaaaaaaaaaaa2,\_Bbbbbbbbbbbbbbbbb,\_Xxxxxxxxxxxxxxxxx,\_Cccccccccccccccc1,\_Cccccccccccccccc2;

BODY

INPUT \_Aaaaaaaaaaaaaaaaa;

INPUT \_Bbbbbbbbbbbbbbbbb;

FOR \_Aaaaaaaaaaaaaaaa2<==\_Aaaaaaaaaaaaaaaaa TO \_Bbbbbbbbbbbbbbbbb DO

OUTPUT \_Aaaaaaaaaaaaaaaa2 \* \_Aaaaaaaaaaaaaaaa2;

FOR \_Aaaaaaaaaaaaaaaa2<==\_Bbbbbbbbbbbbbbbbb TO \_Aaaaaaaaaaaaaaaaa DO

OUTPUT \_Aaaaaaaaaaaaaaaa2 \* \_Aaaaaaaaaaaaaaaa2;

\_Xxxxxxxxxxxxxxxxx<==0;

\_Cccccccccccccccc1<==0;

WHILE \_Cccccccccccccccc1 < \_Aaaaaaaaaaaaaaaaa

BODY

\_Cccccccccccccccc2<==0;

WHILE \_Cccccccccccccccc2 < \_Bbbbbbbbbbbbbbbbb

BODY

\_Xxxxxxxxxxxxxxxxx<==\_Xxxxxxxxxxxxxxxxx + 1;

\_Cccccccccccccccc2<==\_Cccccccccccccccc2 + 1;

END

END WHILE

\_Cccccccccccccccc1<==\_Cccccccccccccccc1 + 1;

END

END WHILE

OUTPUT \_Xxxxxxxxxxxxxxxxx;

\_Xxxxxxxxxxxxxxxxx<==0;

\_Cccccccccccccccc1<==1;

REPEAT

BODY

\_Cccccccccccccccc2<==1;

REPEAT

BODY

\_Xxxxxxxxxxxxxxxxx<==\_Xxxxxxxxxxxxxxxxx + 1;

\_Cccccccccccccccc2<==\_Cccccccccccccccc2 + 1;

END

UNTIL NOT(\_Cccccccccccccccc2 > \_Bbbbbbbbbbbbbbbbb)

\_Cccccccccccccccc1<==\_Cccccccccccccccc1 + 1;

END

UNTIL NOT(\_Cccccccccccccccc1 > \_Aaaaaaaaaaaaaaaaa)

OUTPUT \_Xxxxxxxxxxxxxxxxx;

END

***Результат виконання***

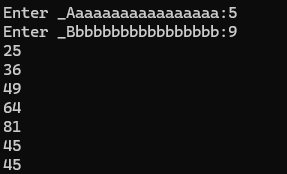


Рис. 5.4 Результат виконання тестової програми №3

Висновки

В процесі виконання курсового проекту було виконано наступне:

1. Складено формальний опис мови програмування t22, в термінах розширеної нотації Бекуса-Наура, виділено усі термінальні символи та ключові слова.

2. Створено компілятор мови програмування t22, а саме:

2.1. Розроблено прямий лексичний аналізатор, орієнтований на розпізнавання лексем, що є заявлені в формальному описі мови програмування.

2.2. Розроблено синтаксичний аналізатор на основі низхідного методу. Складено деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура

2.3. Розроблено генератор коду, відповідні процедури якого викликаються після перевірки синтаксичним аналізатором коректності запису чергового оператора, мови програмування t22. Вихідним кодом генератора є програма на мові C.

3. Проведене тестування компілятора на тестових програмах за наступними пунктами:

3.1. На виявлення лексичних помилок.

3.2. На виявлення синтаксичних помилок.

3.3. Загальна перевірка роботи компілятора.

Тестування не виявило помилок в роботі компілятор, і всі помилки в тестових програмах на мові t22 були успішно виявлені і відповідно оброблені.

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

Список використаної літератури

1. C Programming Language Tutorial - GeeksforGeeks

URL: [C Programming Language Tutorial - GeeksforGeeks](https://www.geeksforgeeks.org/c-programming-language/?ref=header_outind)

1. Error Handling in Compiler Design

URL: [Error Handling in Compiler Design - GeeksforGeeks](https://www.geeksforgeeks.org/error-handling-compiler-design/?ref=lbp)

1. Symbol Table in Compiler

URL: [Symbol Table in Compiler - GeeksforGeeks](https://www.geeksforgeeks.org/symbol-table-compiler/?ref=lbp)

1. Вікіпедія

URL: [Wikipedia](https://www.wikipedia.org/)

1. Stack Overflow

URL: [Stack Overflow - Where Developers Learn, Share, & Build Careers](https://stackoverflow.com/)

Додатки

**А.Таблиці лексем для тестових прикладів**

Тестова програма “Лінійний алгоритм”

---------------------------------------------------------------------------

| TOKEN TABLE |

---------------------------------------------------------------------------

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

| 2 | NAME | 0 | 0 | MainProgram |

| 2 | prog1 | 0 | 1 | Unknown |

| 3 | DATA | 0 | 3 | Variable |

| 3 | INTEGER32\_t | 0 | 4 | Integer |

| 3 | \_Aaaaaaaaaaaa | 0 | 43 | Unknown |

| 3 | aaaaa | 0 | 43 | Unknown |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Yyyyyyyyyyyyyyyyy | 0 | 23 | Identifier |

| 3 | ; | 0 | 40 | Semicolon |

| 4 | BODY | 0 | 2 | StartProgram |

| 5 | INPUT | 0 | 6 | Input |

| 5 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 5 | ; | 0 | 40 | Semicolon |

| 6 | INPUT | 0 | 6 | Input |

| 6 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 7 | OUTPUT | 0 | 7 | Output |

| 7 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 7 | + | 0 | 26 | Add |

| 7 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 7 | ; | 0 | 40 | Semicolon |

| 8 | OUTPUT | 0 | 7 | Output |

| 8 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 8 | - | 0 | 27 | Sub |

| 8 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 8 | ; | 0 | 40 | Semicolon |

| 9 | OUTPUT | 0 | 7 | Output |

| 9 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 9 | \* | 0 | 28 | Mul |

| 9 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 9 | ; | 0 | 40 | Semicolon |

| 10 | OUTPUT | 0 | 7 | Output |

| 10 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 10 | DIV | 0 | 29 | Div |

| 10 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 10 | ; | 0 | 40 | Semicolon |

| 11 | OUTPUT | 0 | 7 | Output |

| 11 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 11 | MOD | 0 | 30 | Mod |

| 11 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 11 | ; | 0 | 40 | Semicolon |

| 13 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 13 | <== | 0 | 25 | Assign |

| 13 | ( | 0 | 38 | LBraket |

| 13 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 13 | - | 0 | 27 | Sub |

| 13 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 13 | ) | 0 | 39 | RBraket |

| 13 | \* | 0 | 28 | Mul |

| 13 | 10 | 10 | 24 | Number |

| 13 | + | 0 | 26 | Add |

| 13 | ( | 0 | 38 | LBraket |

| 13 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 13 | + | 0 | 26 | Add |

| 13 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 13 | ) | 0 | 39 | RBraket |

| 13 | DIV | 0 | 29 | Div |

| 13 | 10 | 10 | 24 | Number |

| 13 | ; | 0 | 40 | Semicolon |

| 14 |\_Yyyyyyyyyyyyyyyyy | 0 | 23 | Identifier |

| 14 | <== | 0 | 25 | Assign |

| 14 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 14 | + | 0 | 26 | Add |

| 14 | ( | 0 | 38 | LBraket |

| 14 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 14 | MOD | 0 | 30 | Mod |

| 14 | 10 | 10 | 24 | Number |

| 14 | ) | 0 | 39 | RBraket |

| 14 | ; | 0 | 40 | Semicolon |

| 15 | OUTPUT | 0 | 7 | Output |

| 15 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 15 | ; | 0 | 40 | Semicolon |

| 16 | OUTPUT | 0 | 7 | Output |

| 16 |\_Yyyyyyyyyyyyyyyyy | 0 | 23 | Identifier |

| 16 | ; | 0 | 40 | Semicolon |

| 17 | END | 0 | 5 | EndProgram |

Тестова програма “Алгоритм з розгалуженням”

---------------------------------------------------------------------------

| TOKEN TABLE |

---------------------------------------------------------------------------

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

| 2 | NAME | 0 | 0 | MainProgram |

| 2 | prog2 | 0 | 1 | Unknown |

| 3 | DATA | 0 | 3 | Variable |

| 3 | INTEGER32\_t | 0 | 4 | Integer |

| 3 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 3 | ; | 0 | 40 | Semicolon |

| 4 | BODY | 0 | 2 | StartProgram |

| 5 | INPUT | 0 | 6 | Input |

| 5 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 5 | ; | 0 | 40 | Semicolon |

| 6 | INPUT | 0 | 6 | Input |

| 6 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 6 | ; | 0 | 40 | Semicolon |

| 7 | INPUT | 0 | 6 | Input |

| 7 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 7 | ; | 0 | 40 | Semicolon |

| 8 | IF | 0 | 8 | If |

| 8 | ( | 0 | 38 | LBraket |

| 8 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 8 | > | 0 | 33 | Greate |

| 8 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 8 | ) | 0 | 39 | RBraket |

| 9 | BODY | 0 | 2 | StartProgram |

| 10 | IF | 0 | 8 | If |

| 10 | ( | 0 | 38 | LBraket |

| 10 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 10 | > | 0 | 33 | Greate |

| 10 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 10 | ) | 0 | 39 | RBraket |

| 11 | BODY | 0 | 2 | StartProgram |

| 12 | GOTO | 0 | 11 | Goto |

| 12 | Abigger | 0 | 23 | Identifier |

| 12 | ; | 0 | 40 | Semicolon |

| 13 | END | 0 | 5 | EndProgram |

| 14 | ELSE | 0 | 10 | Else |

| 15 | BODY | 0 | 2 | StartProgram |

| 16 | OUTPUT | 0 | 7 | Output |

| 16 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 16 | ; | 0 | 40 | Semicolon |

| 17 | GOTO | 0 | 11 | Goto |

| 17 | OutofIF | 0 | 23 | Identifier |

| 17 | ; | 0 | 40 | Semicolon |

| 18 | Abigger | 0 | 12 | Label |

| 19 | OUTPUT | 0 | 7 | Output |

| 19 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 19 | ; | 0 | 40 | Semicolon |

| 20 | GOTO | 0 | 11 | Goto |

| 20 | OutofIF | 0 | 23 | Identifier |

| 20 | ; | 0 | 40 | Semicolon |

| 21 | END | 0 | 5 | EndProgram |

| 22 | END | 0 | 5 | EndProgram |

| 23 | IF | 0 | 8 | If |

| 23 | ( | 0 | 38 | LBraket |

| 23 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 23 | < | 0 | 34 | Less |

| 23 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 23 | ) | 0 | 39 | RBraket |

| 24 | BODY | 0 | 2 | StartProgram |

| 25 | OUTPUT | 0 | 7 | Output |

| 25 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 25 | ; | 0 | 40 | Semicolon |

| 26 | END | 0 | 5 | EndProgram |

| 27 | ELSE | 0 | 10 | Else |

| 28 | BODY | 0 | 2 | StartProgram |

| 29 | OUTPUT | 0 | 7 | Output |

| 29 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 29 | ; | 0 | 40 | Semicolon |

| 30 | END | 0 | 5 | EndProgram |

| 31 | OutofIF | 0 | 12 | Label |

| 33 | IF | 0 | 8 | If |

| 33 | ( | 0 | 38 | LBraket |

| 33 | ( | 0 | 38 | LBraket |

| 33 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 33 | = | 0 | 31 | Equality |

| 33 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 33 | ) | 0 | 39 | RBraket |

| 33 | AND | 0 | 36 | And |

| 33 | ( | 0 | 38 | LBraket |

| 33 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 33 | = | 0 | 31 | Equality |

| 33 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 33 | ) | 0 | 39 | RBraket |

| 33 | AND | 0 | 36 | And |

| 33 | ( | 0 | 38 | LBraket |

| 33 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 33 | = | 0 | 31 | Equality |

| 33 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 33 | ) | 0 | 39 | RBraket |

| 33 | ) | 0 | 39 | RBraket |

| 34 | BODY | 0 | 2 | StartProgram |

| 35 | OUTPUT | 0 | 7 | Output |

| 35 | 1 | 1 | 24 | Number |

| 35 | ; | 0 | 40 | Semicolon |

| 36 | END | 0 | 5 | EndProgram |

| 37 | ELSE | 0 | 10 | Else |

| 38 | BODY | 0 | 2 | StartProgram |

| 39 | OUTPUT | 0 | 7 | Output |

| 39 | 0 | 0 | 24 | Number |

| 39 | ; | 0 | 40 | Semicolon |

| 40 | END | 0 | 5 | EndProgram |

| 41 | IF | 0 | 8 | If |

| 41 | ( | 0 | 38 | LBraket |

| 41 | ( | 0 | 38 | LBraket |

| 41 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 41 | < | 0 | 34 | Less |

| 41 | 0 | 0 | 24 | Number |

| 41 | ) | 0 | 39 | RBraket |

| 41 | OR | 0 | 37 | Or |

| 41 | ( | 0 | 38 | LBraket |

| 41 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 41 | < | 0 | 34 | Less |

| 41 | 0 | 0 | 24 | Number |

| 41 | ) | 0 | 39 | RBraket |

| 41 | OR | 0 | 37 | Or |

| 41 | ( | 0 | 38 | LBraket |

| 41 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 41 | < | 0 | 34 | Less |

| 41 | 0 | 0 | 24 | Number |

| 41 | ) | 0 | 39 | RBraket |

| 41 | ) | 0 | 39 | RBraket |

| 42 | BODY | 0 | 2 | StartProgram |

| 43 | OUTPUT | 0 | 7 | Output |

| 43 | - | 0 | 27 | Sub |

| 43 | 1 | 1 | 24 | Number |

| 43 | ; | 0 | 40 | Semicolon |

| 44 | END | 0 | 5 | EndProgram |

| 45 | ELSE | 0 | 10 | Else |

| 46 | BODY | 0 | 2 | StartProgram |

| 47 | OUTPUT | 0 | 7 | Output |

| 47 | 0 | 0 | 24 | Number |

| 47 | ; | 0 | 40 | Semicolon |

| 48 | END | 0 | 5 | EndProgram |

| 49 | IF | 0 | 8 | If |

| 49 | ( | 0 | 38 | LBraket |

| 49 | NOT | 0 | 35 | Not |

| 49 | ( | 0 | 38 | LBraket |

| 49 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 49 | < | 0 | 34 | Less |

| 49 | ( | 0 | 38 | LBraket |

| 49 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 49 | + | 0 | 26 | Add |

| 49 |\_Ccccccccccccccccc | 0 | 23 | Identifier |

| 49 | ) | 0 | 39 | RBraket |

| 49 | ) | 0 | 39 | RBraket |

| 49 | ) | 0 | 39 | RBraket |

| 50 | BODY | 0 | 2 | StartProgram |

| 51 | OUTPUT | 0 | 7 | Output |

| 51 | ( | 0 | 38 | LBraket |

| 51 | 10 | 10 | 24 | Number |

| 51 | ) | 0 | 39 | RBraket |

| 51 | ; | 0 | 40 | Semicolon |

| 52 | END | 0 | 5 | EndProgram |

| 53 | ELSE | 0 | 10 | Else |

| 54 | BODY | 0 | 2 | StartProgram |

| 55 | OUTPUT | 0 | 7 | Output |

| 55 | ( | 0 | 38 | LBraket |

| 55 | 0 | 0 | 24 | Number |

| 55 | ) | 0 | 39 | RBraket |

| 55 | ; | 0 | 40 | Semicolon |

| 56 | END | 0 | 5 | EndProgram |

| 57 | END | 0 | 5 | EndProgram |

Тестова програма “Циклічний алгоритм”

---------------------------------------------------------------------------

| TOKEN TABLE |

---------------------------------------------------------------------------

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

| 2 | NAME | 0 | 0 | MainProgram |

| 2 | prog3 | 0 | 1 | Unknown |

| 3 | DATA | 0 | 3 | Variable |

| 3 | INTEGER32\_t | 0 | 4 | Integer |

| 3 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 3 | , | 0 | 42 | Comma |

| 3 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 3 | ; | 0 | 40 | Semicolon |

| 4 | BODY | 0 | 2 | StartProgram |

| 5 | INPUT | 0 | 6 | Input |

| 5 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 5 | ; | 0 | 40 | Semicolon |

| 6 | INPUT | 0 | 6 | Input |

| 6 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 6 | ; | 0 | 40 | Semicolon |

| 7 | FOR | 0 | 13 | For |

| 7 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 7 | <== | 0 | 25 | Assign |

| 7 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 7 | TO | 0 | 14 | To |

| 7 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 7 | DO | 0 | 16 | Do |

| 8 | OUTPUT | 0 | 7 | Output |

| 8 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 8 | \* | 0 | 28 | Mul |

| 8 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 8 | ; | 0 | 40 | Semicolon |

| 10 | FOR | 0 | 13 | For |

| 10 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 10 | <== | 0 | 25 | Assign |

| 10 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 10 | TO | 0 | 14 | To |

| 10 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 10 | DO | 0 | 16 | Do |

| 11 | OUTPUT | 0 | 7 | Output |

| 11 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 11 | \* | 0 | 28 | Mul |

| 11 |\_Aaaaaaaaaaaaaaaa2 | 0 | 23 | Identifier |

| 11 | ; | 0 | 40 | Semicolon |

| 13 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 13 | <== | 0 | 25 | Assign |

| 13 | 0 | 0 | 24 | Number |

| 13 | ; | 0 | 40 | Semicolon |

| 14 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 14 | <== | 0 | 25 | Assign |

| 14 | 0 | 0 | 24 | Number |

| 14 | ; | 0 | 40 | Semicolon |

| 15 | WHILE | 0 | 17 | While |

| 15 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 15 | < | 0 | 34 | Less |

| 15 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 16 | BODY | 0 | 2 | StartProgram |

| 17 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 17 | <== | 0 | 25 | Assign |

| 17 | 0 | 0 | 24 | Number |

| 17 | ; | 0 | 40 | Semicolon |

| 18 | WHILE | 0 | 17 | While |

| 18 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 18 | < | 0 | 34 | Less |

| 18 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 19 | BODY | 0 | 2 | StartProgram |

| 20 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 20 | <== | 0 | 25 | Assign |

| 20 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 20 | + | 0 | 26 | Add |

| 20 | 1 | 1 | 24 | Number |

| 20 | ; | 0 | 40 | Semicolon |

| 21 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 21 | <== | 0 | 25 | Assign |

| 21 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 21 | + | 0 | 26 | Add |

| 21 | 1 | 1 | 24 | Number |

| 21 | ; | 0 | 40 | Semicolon |

| 22 | END | 0 | 5 | EndProgram |

| 23 | END | 0 | 20 | End |

| 23 | WHILE | 0 | 17 | While |

| 24 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 24 | <== | 0 | 25 | Assign |

| 24 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 24 | + | 0 | 26 | Add |

| 24 | 1 | 1 | 24 | Number |

| 24 | ; | 0 | 40 | Semicolon |

| 25 | END | 0 | 5 | EndProgram |

| 26 | END | 0 | 20 | End |

| 26 | WHILE | 0 | 17 | While |

| 27 | OUTPUT | 0 | 7 | Output |

| 27 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 27 | ; | 0 | 40 | Semicolon |

| 29 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 29 | <== | 0 | 25 | Assign |

| 29 | 0 | 0 | 24 | Number |

| 29 | ; | 0 | 40 | Semicolon |

| 30 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 30 | <== | 0 | 25 | Assign |

| 30 | 1 | 1 | 24 | Number |

| 30 | ; | 0 | 40 | Semicolon |

| 31 | REPEAT | 0 | 21 | Repeat |

| 32 | BODY | 0 | 2 | StartProgram |

| 33 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 33 | <== | 0 | 25 | Assign |

| 33 | 1 | 1 | 24 | Number |

| 33 | ; | 0 | 40 | Semicolon |

| 34 | REPEAT | 0 | 21 | Repeat |

| 35 | BODY | 0 | 2 | StartProgram |

| 36 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 36 | <== | 0 | 25 | Assign |

| 36 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 36 | + | 0 | 26 | Add |

| 36 | 1 | 1 | 24 | Number |

| 36 | ; | 0 | 40 | Semicolon |

| 37 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 37 | <== | 0 | 25 | Assign |

| 37 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 37 | + | 0 | 26 | Add |

| 37 | 1 | 1 | 24 | Number |

| 37 | ; | 0 | 40 | Semicolon |

| 38 | END | 0 | 5 | EndProgram |

| 39 | UNTIL | 0 | 22 | Until |

| 39 | NOT | 0 | 35 | Not |

| 39 | ( | 0 | 38 | LBraket |

| 39 |\_Cccccccccccccccc2 | 0 | 23 | Identifier |

| 39 | > | 0 | 33 | Greate |

| 39 |\_Bbbbbbbbbbbbbbbbb | 0 | 23 | Identifier |

| 39 | ) | 0 | 39 | RBraket |

| 40 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 40 | <== | 0 | 25 | Assign |

| 40 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 40 | + | 0 | 26 | Add |

| 40 | 1 | 1 | 24 | Number |

| 40 | ; | 0 | 40 | Semicolon |

| 41 | END | 0 | 5 | EndProgram |

| 42 | UNTIL | 0 | 22 | Until |

| 42 | NOT | 0 | 35 | Not |

| 42 | ( | 0 | 38 | LBraket |

| 42 |\_Cccccccccccccccc1 | 0 | 23 | Identifier |

| 42 | > | 0 | 33 | Greate |

| 42 |\_Aaaaaaaaaaaaaaaaa | 0 | 23 | Identifier |

| 42 | ) | 0 | 39 | RBraket |

| 43 | OUTPUT | 0 | 7 | Output |

| 43 |\_Xxxxxxxxxxxxxxxxx | 0 | 23 | Identifier |

| 43 | ; | 0 | 40 | Semicolon |

| 45 | END | 0 | 5 | EndProgram |

**Б . С код отриманий на виході транслятора для тестових прикладів**

**Prog1.c**

#include <stdio.h>

#include <stdlib.h>

int main()

{

int \_Aaaaaaaaaaaaaaaaa, \_Bbbbbbbbbbbbbbbbb, \_Xxxxxxxxxxxxxxxxx, \_Yyyyyyyyyyyyyyyyy;

printf("Enter \_Aaaaaaaaaaaaaaaaa:");

scanf("%d", &\_Aaaaaaaaaaaaaaaaa);

printf("Enter \_Bbbbbbbbbbbbbbbbb:");

scanf("%d", &\_Bbbbbbbbbbbbbbbbb);

printf("%d\n", \_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb);

printf("%d\n", \_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb);

printf("%d\n", \_Aaaaaaaaaaaaaaaaa \* \_Bbbbbbbbbbbbbbbbb);

printf("%d\n", \_Aaaaaaaaaaaaaaaaa / \_Bbbbbbbbbbbbbbbbb);

printf("%d\n", \_Aaaaaaaaaaaaaaaaa % \_Bbbbbbbbbbbbbbbbb);

\_Xxxxxxxxxxxxxxxxx = (\_Aaaaaaaaaaaaaaaaa - \_Bbbbbbbbbbbbbbbbb) \* 10 + (\_Aaaaaaaaaaaaaaaaa + \_Bbbbbbbbbbbbbbbbb) / 10;

\_Yyyyyyyyyyyyyyyyy = \_Xxxxxxxxxxxxxxxxx + (\_Xxxxxxxxxxxxxxxxx % 10);

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

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

system("pause");

return 0;

}

**Prog2.c**

#include <stdio.h>

#include <stdlib.h>

int main()

{

int \_Aaaaaaaaaaaaaaaaa, \_Bbbbbbbbbbbbbbbbb, \_Ccccccccccccccccc;

printf("Enter \_Aaaaaaaaaaaaaaaaa:");

scanf("%d", &\_Aaaaaaaaaaaaaaaaa);

printf("Enter \_Bbbbbbbbbbbbbbbbb:");

scanf("%d", &\_Bbbbbbbbbbbbbbbbb);

printf("Enter \_Ccccccccccccccccc:");

scanf("%d", &\_Ccccccccccccccccc);

if ((\_Aaaaaaaaaaaaaaaaa > \_Bbbbbbbbbbbbbbbbb))

{

if ((\_Aaaaaaaaaaaaaaaaa > \_Ccccccccccccccccc))

{

goto Abigger;

}

else

{

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

goto OutofIF;

Abigger:

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

goto OutofIF;

}

}

if ((\_Bbbbbbbbbbbbbbbbb < \_Ccccccccccccccccc))

{

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

}

else

{

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

}

OutofIF:

if (((\_Aaaaaaaaaaaaaaaaa == \_Bbbbbbbbbbbbbbbbb) && (\_Aaaaaaaaaaaaaaaaa == \_Ccccccccccccccccc) && (\_Bbbbbbbbbbbbbbbbb == \_Ccccccccccccccccc)))

{

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

}

else

{

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

}

if (((\_Aaaaaaaaaaaaaaaaa < 0) || (\_Bbbbbbbbbbbbbbbbb < 0) || (\_Ccccccccccccccccc < 0)))

{

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

}

else

{

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

}

if ((!(\_Aaaaaaaaaaaaaaaaa < (\_Bbbbbbbbbbbbbbbbb + \_Ccccccccccccccccc))))

{

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

}

else

{

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

}

system("pause");

return 0;

}

**Prog3.c**

#include <stdio.h>

#include <stdlib.h>

int main()

{

int \_Aaaaaaaaaaaaaaaaa, \_Aaaaaaaaaaaaaaaa2, \_Bbbbbbbbbbbbbbbbb, \_Xxxxxxxxxxxxxxxxx, \_Cccccccccccccccc1, \_Cccccccccccccccc2;

printf("Enter \_Aaaaaaaaaaaaaaaaa:");

scanf("%d", &\_Aaaaaaaaaaaaaaaaa);

printf("Enter \_Bbbbbbbbbbbbbbbbb:");

scanf("%d", &\_Bbbbbbbbbbbbbbbbb);

for (int \_Aaaaaaaaaaaaaaaa2 = \_Aaaaaaaaaaaaaaaaa; \_Aaaaaaaaaaaaaaaa2 <= \_Bbbbbbbbbbbbbbbbb; \_Aaaaaaaaaaaaaaaa2++)

printf("%d\n", \_Aaaaaaaaaaaaaaaa2 \* \_Aaaaaaaaaaaaaaaa2);

for (int \_Aaaaaaaaaaaaaaaa2 = \_Bbbbbbbbbbbbbbbbb; \_Aaaaaaaaaaaaaaaa2 <= \_Aaaaaaaaaaaaaaaaa; \_Aaaaaaaaaaaaaaaa2++)

printf("%d\n", \_Aaaaaaaaaaaaaaaa2 \* \_Aaaaaaaaaaaaaaaa2);

\_Xxxxxxxxxxxxxxxxx = 0;

\_Cccccccccccccccc1 = 0;

while (\_Cccccccccccccccc1 < \_Aaaaaaaaaaaaaaaaa)

{

{

\_Cccccccccccccccc2 = 0;

while (\_Cccccccccccccccc2 < \_Bbbbbbbbbbbbbbbbb)

{

{

\_Xxxxxxxxxxxxxxxxx = \_Xxxxxxxxxxxxxxxxx + 1;

\_Cccccccccccccccc2 = \_Cccccccccccccccc2 + 1;

}

}

\_Cccccccccccccccc1 = \_Cccccccccccccccc1 + 1;

}

}

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

\_Xxxxxxxxxxxxxxxxx = 0;

\_Cccccccccccccccc1 = 1;

do

{

\_Cccccccccccccccc2 = 1;

do

{

\_Xxxxxxxxxxxxxxxxx = \_Xxxxxxxxxxxxxxxxx + 1;

\_Cccccccccccccccc2 = \_Cccccccccccccccc2 + 1;

}

while (!(\_Cccccccccccccccc2 > \_Bbbbbbbbbbbbbbbbb));

\_Cccccccccccccccc1 = \_Cccccccccccccccc1 + 1;

}

while (!(\_Cccccccccccccccc1 > \_Aaaaaaaaaaaaaaaaa));

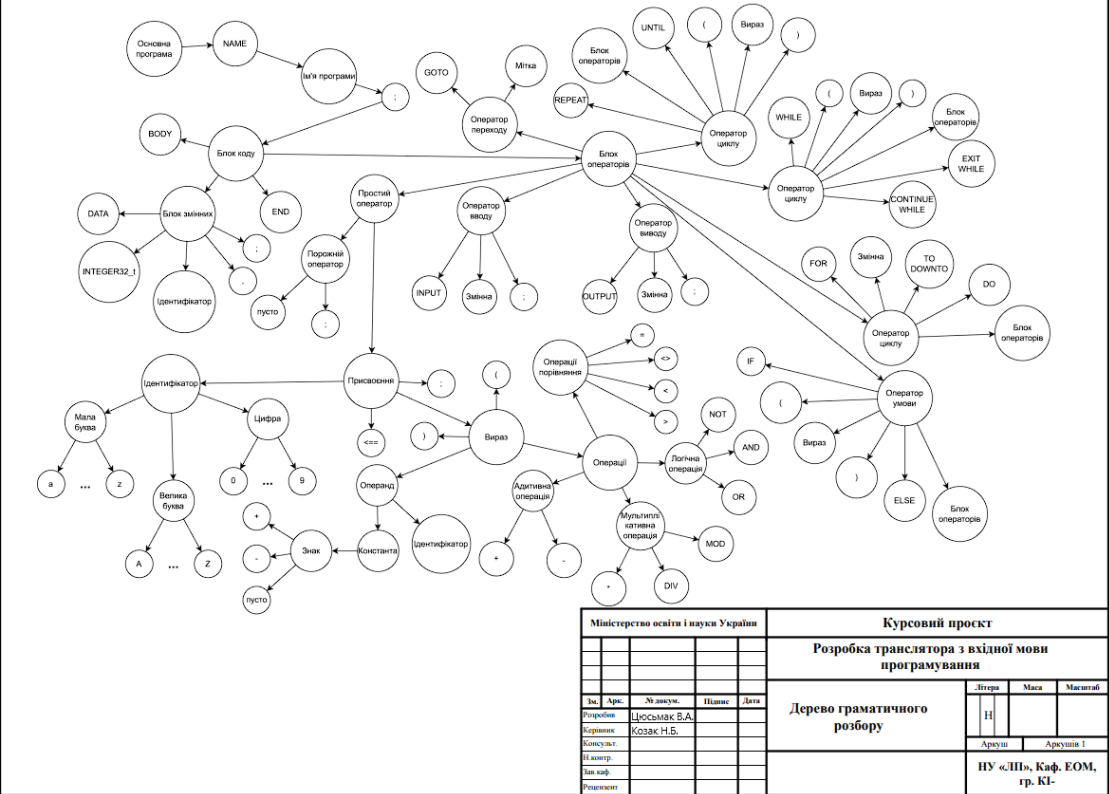
printf("%d\n", \_Xxxxxxxxxxxxxxxxx);

system("pause");

return 0;

}

**В.** **Абстрактне синтаксичне дерево для тестових прикладів (за наявності).**



**Г.** **Документований текст програмних модулів (лістинги)**

**translator.h**

#pragma once

#define MAX\_TOKENS 1000

#define MAX\_IDENTIFIER 10

// перерахування, яке описує всі можливі типи лексем enum TypeOfTokens { Mainprogram, ProgramName, StartProgram,

Variable,

Type,

EndProgram,

Input,

Output,

If,   
Then,   
Else,   
  
Goto,   
Label,  
  
For,   
To,   
DownTo,   
Do,   
  
While,   
Exit,   
Continue,   
End,  
  
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[32]; // 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[32]; };

// 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 };

// перерахування, яке описує всі можливі вузли абстрактного синтаксичного дерева enum TypeOfNodes { program\_node, var\_node, input\_node, output\_node,

if\_node,  
then\_node,  
  
goto\_node,  
label\_node,  
  
for\_to\_node,  
for\_downto\_node,  
  
while\_node,  
exit\_while\_node,  
continue\_while\_node,  
  
repeat\_until\_node,  
  
id\_node,  
num\_node,  
assign\_node,  
add\_node,  
sub\_node,  
mul\_node,  
div\_node,  
mod\_node,  
or\_node,  
and\_node,  
not\_node,  
cmp\_node,  
statement\_node,  
compount\_node

};

// структура, яка описує вузол абстрактного синтаксичного дерева (AST) struct ASTNode { TypeOfNodes nodetype; // Тип вузла char name[32]; // Ім'я вузла struct ASTNode\* left; // Лівий нащадок struct ASTNode\* right; // Правий нащадок };

// 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);

// 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);

// синтаксичний аналіз методом рекурсивного спуску // вхідні дані - глобальна таблиця лексем TokenTable void Parser(FILE\* errFile);

// функція синтаксичного аналізу і створення абстрактного синтаксичного дерева ASTNode\* ParserAST();

// функція знищення дерева void destroyTree(ASTNode\* root);

// функція для друку AST у вигляді дерева на екран void PrintAST(ASTNode\* node, int level);

// функція для друку AST у вигляді дерева у файл void PrintASTToFile(ASTNode\* node, int level, FILE\* outFile);

// Рекурсивна функція для генерації коду з AST void generateCodefromAST(ASTNode\* node, FILE\* output);

// функція для генерації коду void generateCCode(FILE\* outFile);

void compile\_to\_exe(const char\* source\_file, const char\* output\_file);

**ast.cpp**

#define \_CRT\_SECURE\_NO\_WARNINGS #include <stdio.h> #include <stdlib.h> #include <string.h> #include "translator.h" #include

// таблиця лексем extern Token\* TokenTable; // кількість лексем extern unsigned int TokensNum;

static int pos = 0;

// функція створення вузла AST ASTNode\* createNode(TypeOfNodes type, const char\* name, ASTNode\* left, ASTNode\* right) { ASTNode\* node = (ASTNode\*)malloc(sizeof(ASTNode)); node->nodetype = type; strcpy\_s(node->name, name); node->left = left; node->right = right; return node; }

// функція знищення дерева void destroyTree(ASTNode\* root) { if (root == NULL) return;

// Рекурсивно знищуємо ліве і праве піддерево  
destroyTree(root->left);  
destroyTree(root->right);  
  
// Звільняємо пам'ять для поточного вузла  
free(root);

}

// набір функцій для рекурсивного спуску // на кожне правило - окрема функція ASTNode\* program(); ASTNode\* variable\_declaration(); ASTNode\* variable\_list(); ASTNode\* program\_body(); ASTNode\* statement(); ASTNode\* assignment(); ASTNode\* arithmetic\_expression(); ASTNode\* term(); ASTNode\* factor(); ASTNode\* input(); ASTNode\* output(); ASTNode\* conditional();

ASTNode\* goto\_statement(); ASTNode\* label\_statement(); ASTNode\* for\_to\_do(); ASTNode\* for\_downto\_do(); ASTNode\* while\_statement(); ASTNode\* repeat\_until();

ASTNode\* logical\_expression(); ASTNode\* and\_expression(); ASTNode\* comparison(); ASTNode\* compound\_statement();

// функція синтаксичного аналізу і створення абстрактного синтаксичного дерева ASTNode\* ParserAST() { ASTNode\* tree = program();

printf("\nParsing completed. AST created.\n");  
  
return tree;

}

static void match(TypeOfTokens expectedType) { if (TokenTable[pos].type == expectedType) pos++; else { printf("\nSyntax error in line %d: Expected another type of lexeme.\n", TokenTable[pos].line); std::cout << "AST Type: " << TokenTable[pos].type << std::endl; std::cout << "AST Expected type:" << expectedType << std::endl; exit(10); } }

// <програма> = 'start' 'var' <оголошення змінних> ';' <тіло програми> 'stop' ASTNode\* program() { match(Mainprogram); match(ProgramName); match(Variable); ASTNode\* declarations = variable\_declaration(); match(Semicolon); match(StartProgram); ASTNode\* body = program\_body(); match(EndProgram); return createNode(program\_node, "program", declarations, body); }

// <оголошення змінних> = [<тип даних> <список змінних>] ASTNode\* variable\_declaration() { if (TokenTable[pos].type == Type) { pos++; return variable\_list(); } return NULL; }

// <список змінних> = <ідентифікатор> { ',' <ідентифікатор> } ASTNode\* variable\_list() { match(Identifier); ASTNode\* id = createNode(id\_node, TokenTable[pos - 1].name, NULL, NULL); ASTNode\* list = list = createNode(var\_node, "var", id, NULL); while (TokenTable[pos].type == Comma) { match(Comma); match(Identifier); id = createNode(id\_node, TokenTable[pos - 1].name, NULL, NULL); list = createNode(var\_node, "var", id, list); } return list; }

// <тіло програми> = <оператор> ';' { <оператор> ';' } ASTNode\* program\_body() { ASTNode\* stmt = statement(); //match(Semicolon); ASTNode\* body = stmt; while (TokenTable[pos].type != EndProgram) { ASTNode\* nextStmt = statement(); body = createNode(statement\_node, "statement", body, nextStmt); } return body; }

// <оператор> = <присвоєння> | <ввід> | <вивід> | <умовний оператор> | <складений оператор> ASTNode\* statement() { switch (TokenTable[pos].type) { case Input: return input(); case Output: return output(); case If: return conditional(); case StartProgram: return compound\_statement(); case Goto: return goto\_statement(); case Label: return label\_statement(); case For: { int temp\_pos = pos + 1; while (TokenTable[temp\_pos].type != To && TokenTable[temp\_pos].type != DownTo && temp\_pos < TokensNum) { temp\_pos++; } if (TokenTable[temp\_pos].type == To) { return for\_to\_do(); } else if (TokenTable[temp\_pos].type == DownTo) { return for\_downto\_do(); } else { printf("Error: Expected 'To' or 'DownTo' after 'For'\n"); exit(1); } } case While: return while\_statement(); case Exit: match(Exit); match(While); return createNode(exit\_while\_node, "exit-while", NULL, NULL); case Continue: match(Continue); match(While); return createNode(continue\_while\_node, "continue-while", NULL, NULL); case Repeat: return repeat\_until(); default: return assignment(); } }

// <присвоєння> = <ідентифікатор> ':=' <арифметичний вираз> ASTNode\* assignment() { ASTNode\* id = createNode(id\_node, TokenTable[pos].name, NULL, NULL); match(Identifier); match(Assign); ASTNode\* expr = arithmetic\_expression(); match(Semicolon); return createNode(assign\_node, "<==", id, expr); }

// <арифметичний вираз> = <доданок> { ('+' | '-') <доданок> } ASTNode\* arithmetic\_expression() { ASTNode\* left = term(); while (TokenTable[pos].type == Add || TokenTable[pos].type == Sub) { TypeOfTokens op = TokenTable[pos].type; match(op); ASTNode\* right = term(); if (op == Add) left = createNode(add\_node, "+", left, right); else left = createNode(sub\_node, "-", left, right); } return left; }

// <доданок> = <множник> { ('*' | '/') <множник> } ASTNode* term() { ASTNode\* left = factor(); while (TokenTable[pos].type == Mul || TokenTable[pos].type == Div || TokenTable[pos].type == Mod) { TypeOfTokens op = TokenTable[pos].type; match(op); ASTNode\* right = factor(); if (op == Mul) left = createNode(mul\_node, "\*", left, right); if (op == Div) left = createNode(div\_node, "/", left, right); if (op == Mod) left = createNode(mod\_node, "%", left, right); } return left; }

// <множник> = <ідентифікатор> | <число> | '(' <арифметичний вираз> ')' ASTNode\* factor() { if (TokenTable[pos].type == Identifier) { ASTNode\* id = createNode(id\_node, TokenTable[pos].name, NULL, NULL); match(Identifier); return id; } else if (TokenTable[pos].type == Number) { ASTNode\* num = createNode(num\_node, TokenTable[pos].name, NULL, NULL); match(Number); return num; } else if (TokenTable[pos].type == LBraket) { match(LBraket); ASTNode\* expr = arithmetic\_expression(); match(RBraket); return expr; } else { printf("\nSyntax error in line %d: A multiplier was expected.\n", TokenTable[pos].line); exit(11); } }

// <ввід> = 'input' <ідентифікатор> ASTNode\* input() { match(Input); ASTNode\* id = createNode(id\_node, TokenTable[pos].name, NULL, NULL); match(Identifier); match(Semicolon); return createNode(input\_node, "input", id, NULL); }

// <вивід> = 'output' <ідентифікатор> ASTNode\* output() { match(Output); // Match the "Output" token

ASTNode\* expr = NULL;  
// Check for a negative number  
if (TokenTable[pos].type == Sub && TokenTable[pos + 1].type == Number)  
{  
 pos++; // Skip the 'Sub' token  
 expr = createNode(sub\_node, "-", createNode(num\_node, "0", NULL, NULL),  
 createNode(num\_node, TokenTable[pos].name, NULL, NULL));  
 match(Number); // Match the number token  
}  
else  
{  
 // Parse the arithmetic expression  
 expr = arithmetic\_expression();  
}  
match(Semicolon); // Ensure the statement ends with a semicolon  
  
// Create the output node with the parsed expression as its left child  
return createNode(output\_node, "output", expr, NULL);

}

// <умовний оператор> = 'if' <логічний вираз> <оператор> [ 'else' <оператор> ] ASTNode\* conditional() { match(If); ASTNode\* condition = logical\_expression(); ASTNode\* ifBranch = statement(); ASTNode\* elseBranch = NULL; if (TokenTable[pos].type == Else) { match(Else); elseBranch = statement(); } return createNode(if\_node, "if", condition, createNode(statement\_node, "branches", ifBranch, elseBranch)); }

ASTNode\* goto\_statement() { match(Goto); if (TokenTable[pos].type == Identifier) { ASTNode\* label = createNode(label\_node, TokenTable[pos].name, NULL, NULL); match(Identifier); match(Semicolon); return createNode(goto\_node, "goto", label, NULL); } else { printf("Syntax error: Expected a label after 'goto' at line %d.\n", TokenTable[pos].line); exit(1); } }

ASTNode\* label\_statement() { match(Label); ASTNode\* label = createNode(label\_node, TokenTable[pos - 1].name, NULL, NULL); return label; }

ASTNode\* for\_to\_do() { match(For);

if (TokenTable[pos].type != Identifier)  
{  
 printf("Syntax error: Expected variable name after 'for' at line %d.\n", TokenTable[pos].line);  
 exit(1);  
}  
ASTNode\* var = createNode(id\_node, TokenTable[pos].name, NULL, NULL);  
match(Identifier);  
match(Assign);  
ASTNode\* start = arithmetic\_expression();  
match(To);  
ASTNode\* end = arithmetic\_expression();  
match(Do);  
ASTNode\* body = statement();  
// Повертаємо вузол циклу for-to  
return createNode(for\_to\_node, "for-to",  
 createNode(assign\_node, "<==", var, start),   
 createNode(statement\_node, "body", end, body));

}

ASTNode\* for\_downto\_do() { // Очікуємо "for" match(For);

// Очікуємо ідентифікатор змінної циклу  
if (TokenTable[pos].type != Identifier)  
{  
 printf("Syntax error: Expected variable name after 'for' at line %d.\n", TokenTable[pos].line);  
 exit(1);  
}  
ASTNode\* var = createNode(id\_node, TokenTable[pos].name, NULL, NULL);  
match(Identifier);  
match(Assign);  
ASTNode\* start = arithmetic\_expression();  
match(DownTo);  
ASTNode\* end = arithmetic\_expression();  
match(Do);  
ASTNode\* body = statement();  
// Повертаємо вузол циклу for-to  
return createNode(for\_downto\_node, "for-downto",  
 createNode(assign\_node, "<==", var, start),  
 createNode(statement\_node, "body", end, body));

}

ASTNode\* while\_statement() { match(While); ASTNode\* condition = logical\_expression();

// Parse the body of the While loop  
ASTNode\* body = NULL;  
while (1) // Process until "End While"  
{  
 if (TokenTable[pos].type == End)  
 {  
 match(End);  
 match(While);  
 break; // End of the While loop  
 }  
 else  
 {  
 // Delegate to the `statement` function  
 ASTNode\* stmt = statement();  
 body = createNode(statement\_node, "statement", body, stmt);  
 }  
}  
  
return createNode(while\_node, "while", condition, body);

}

// Updated variable validation logic ASTNode\* validate\_identifier() { const char\* identifierName = TokenTable[pos].name;

// Check if the identifier was declared  
bool declared = false;  
for (unsigned int i = 0; i < TokensNum; i++)  
{  
 if (TokenTable[i].type == Variable && !strcmp(TokenTable[i].name, identifierName))  
 {  
 declared = true;  
 break;  
 }  
}  
  
if (!declared && (pos == 0 || TokenTable[pos - 1].type != Goto))  
{  
 printf("Syntax error: Undeclared identifier '%s' at line %d.\n", identifierName, TokenTable[pos].line);  
 exit(1);  
}  
  
match(Identifier);  
return createNode(id\_node, identifierName, NULL, NULL);

}

ASTNode\* repeat\_until() { match(Repeat); ASTNode\* body = NULL; ASTNode\* stmt = statement(); body = createNode(statement\_node, "body", body, stmt); //pos++; match(Until); ASTNode\* condition = logical\_expression(); return createNode(repeat\_until\_node, "repeat-until", body, condition); }

// <логічний вираз> = <вираз І> { '|' <вираз І> } ASTNode\* logical\_expression() { ASTNode\* left = and\_expression(); while (TokenTable[pos].type == Or) { match(Or); ASTNode\* right = and\_expression(); left = createNode(or\_node, "|", left, right); } return left; }

// <вираз І> = <порівняння> { '&' <порівняння> } ASTNode\* and\_expression() { ASTNode\* left = comparison(); while (TokenTable[pos].type == And) { match(And); ASTNode\* right = comparison(); left = createNode(and\_node, "&", left, right); } return left; }

// <порівняння> = <операція порівняння> | ‘!‘ ‘(‘ <логічний вираз> ‘)‘ | ‘(‘ <логічний вираз> ‘)‘ // <операція порівняння> = <арифметичний вираз> <менше-більше> <арифметичний вираз> // <менше-більше> = ‘>‘ | ‘<‘ | ‘=‘ | ‘<>‘ ASTNode\* comparison() { if (TokenTable[pos].type == Not) { // Варіант: ! (<логічний вираз>) match(Not); match(LBraket); ASTNode\* expr = logical\_expression(); match(RBraket); return createNode(not\_node, "!", expr, NULL); } else if (TokenTable[pos].type == LBraket) { // Варіант: ( <логічний вираз> ) match(LBraket); ASTNode\* expr = logical\_expression(); match(RBraket); return expr; // Повертаємо вираз у дужках як піддерево } else { // Варіант: <арифметичний вираз> <менше-більше> <арифметичний вираз> ASTNode\* left = arithmetic\_expression(); if (TokenTable[pos].type == Greate || TokenTable[pos].type == Less || TokenTable[pos].type == Equality || TokenTable[pos].type == NotEquality) { TypeOfTokens op = TokenTable[pos].type; char operatorName[16]; strcpy\_s(operatorName, TokenTable[pos].name); match(op); ASTNode\* right = arithmetic\_expression(); return createNode(cmp\_node, operatorName, left, right); } else { printf("\nSyntax error: A comparison operation is expected.\n"); exit(12); } } }

// <складений оператор> = 'start' <тіло програми> 'stop' ASTNode\* compound\_statement() { match(StartProgram); ASTNode\* body = program\_body(); match(EndProgram); return createNode(compount\_node, "compound", body, NULL); }

// функція для друку AST у вигляді дерева на екран void PrintAST(ASTNode\* node, int level) { if (node == NULL) return;

// Відступи для позначення рівня вузла  
for (int i = 0; i < level; i++)  
 printf("| ");  
  
// Виводимо інформацію про вузол  
printf("|-- %s", node->name);  
printf("\n");  
  
// Рекурсивний друк лівого та правого піддерева  
if (node->left || node->right)  
{  
 PrintAST(node->left, level + 1);  
 PrintAST(node->right, level + 1);  
}

}

// функція для друку AST у вигляді дерева у файл void PrintASTToFile(ASTNode\* node, int level, FILE\* outFile) { if (node == NULL) return;

// Відступи для позначення рівня вузла  
for (int i = 0; i < level; i++)  
 fprintf(outFile, "| ");  
  
// Виводимо інформацію про вузол  
fprintf(outFile, "|-- %s", node->name);  
fprintf(outFile, "\n");  
  
// Рекурсивний друк лівого та правого піддерева  
if (node->left || node->right)  
{  
 PrintASTToFile(node->left, level + 1, outFile);  
 PrintASTToFile(node->right, level + 1, outFile);  
}

}

**lexer.cpp**

#include <stdlib.h> #include <stdio.h> #include <string.h> #include "translator.h" #include

// 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) { States state = Start; Token TempToken; // кількість лексем unsigned int NumberOfTokens = 0; char ch, buf[32]; 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 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') || ch == '\_')  
 state = Letter;  
 else  
 if (ch <= '9' && ch >= '0')  
 state = Digit;  
 else  
 if (ch == ' ' || ch == '\t' || ch == '\n')  
 state = Separators;  
 else  
 if (ch == '#')  
 state = SComment;  
 else  
 state = Another;  
 break;  
 }  
  
 // стан Finish - кінець виділення чергової лексеми і запис лексеми у таблицю лексем  
 case Finish:  
 {  
 if (NumberOfTokens < MAX\_TOKENS)  
 {TokenTable[NumberOfTokens++] = TempToken;  
 if (ch != EOF)  
 state = Start;  
 else  
 state = EndOfFile;  
 }  
 else  
 {  
 printf("\n\t\t\ttoo many tokens !!!\n");  
 return NumberOfTokens - 1;  
 }  
 break;  
 }  
  
 // стан EndOfFile - кінець файлу, можна завершувати пошук лексем  
 case EndOfFile:  
 {  
 return NumberOfTokens;  
 }  
  
 // стан Letter - поточний символ - маленька літера, поточна лексема - ключове слово або ідентифікатор  
 case Letter:  
 {  
 buf[0] = ch;  
 int j = 1;  
  
 ch = getc(F);  
  
 while (((ch >= 'a' && ch <= 'z') || (ch >= 'A' && ch <= 'Z') ||  
 (ch >= '0' && ch <= '9') || ch == '\_' || ch == ':' || ch == '-') && j < 31)  
 {  
 buf[j++] = ch;  
 ch = getc(F);  
 }  
 buf[j] = '\0';   
  
 TypeOfTokens temp\_type = Unknown;  
  
 if (!strcmp(buf, "END"))  
 {  
 char next\_buf[16];  
 int next\_j = 0;  
  
 while (ch == ' ' || ch == '\t')  
 {  
 ch = getc(F);  
 }  
  
 while (((ch >= 'a' && ch <= 'z') || (ch >= 'A' && ch <= 'Z')) && next\_j < 31)  
 {  
 next\_buf[next\_j++] = ch;  
 ch = getc(F);  
 }  
 next\_buf[next\_j] = '\0';  
  
 if (!strcmp(next\_buf, "WHILE"))  
 {  
 temp\_type = End;  
 strcpy\_s(TempToken.name, buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 TokenTable[NumberOfTokens++] = TempToken;  
  
 temp\_type = While;  
 strcpy\_s(TempToken.name, next\_buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 TokenTable[NumberOfTokens++] = TempToken;  
  
 state = Start;  
 break;  
 }  
 else  
 {  
 temp\_type = EndProgram;  
 strcpy\_s(TempToken.name, buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 state = Finish;  
  
 for (int k = next\_j - 1; k >= 0; k--)  
 {  
 ungetc(next\_buf[k], F);  
 }  
 break;  
 }  
 }  
  
 if (!strcmp(buf, "NAME"))  
 {  
 char next\_buf[32];  
 int next\_j = 0;  
  
 while (ch == ' ' || ch == '\t')  
 {  
 ch = getc(F);  
 }  
  
 while (((ch >= 'a' && ch <= 'z') || (ch >= 'A' && ch <= 'Z') || (ch >= '0' && ch <= '9' || ch == ';')) && next\_j < 31)  
 {  
 next\_buf[next\_j++] = ch;  
 ch = getc(F);  
 }  
 next\_buf[next\_j] = '\0';  
  
 if (next\_buf[strlen(next\_buf) - 1] == ';')  
 {  
 temp\_type = Mainprogram;  
 strcpy\_s(TempToken.name, buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 TokenTable[NumberOfTokens++] = TempToken;  
  
 next\_buf[strlen(next\_buf) - 1] = '\0';  
 temp\_type = ProgramName;  
 strcpy\_s(TempToken.name, next\_buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 TokenTable[NumberOfTokens++] = TempToken;  
  
 state = Start;  
 break;  
 }  
 }  
  
 else if (!strcmp(buf, "BODY")) temp\_type = StartProgram;  
 else if (!strcmp(buf, "DATA")) temp\_type = Variable;  
 else if (!strcmp(buf, "INTEGER32\_t")) temp\_type = Type;  
 else if (!strcmp(buf, "INPUT")) temp\_type = Input;  
 else if (!strcmp(buf, "OUTPUT")) temp\_type = Output;  
  
 else if (!strcmp(buf, "DIV")) temp\_type = Div;  
 else if (!strcmp(buf, "MOD")) temp\_type = Mod;  
  
 else if (!strcmp(buf, "NOT")) temp\_type = Not;  
 else if (!strcmp(buf, "AND")) temp\_type = And;  
 else if (!strcmp(buf, "OR")) temp\_type = Or;  
  
 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, "EXIT")) temp\_type = Exit;  
 else if (!strcmp(buf, "WHILE")) temp\_type = While;  
 else if (!strcmp(buf, "CONTINUE")) temp\_type = Continue;  
 else if (!strcmp(buf, "REPEAT")) temp\_type = Repeat;  
 else if (!strcmp(buf, "UNTIL")) temp\_type = Until;  
 if (temp\_type == Unknown && TokenTable[NumberOfTokens - 1].type == Goto)  
 {  
 temp\_type = Identifier;  
 }  
 else if (buf[strlen(buf) - 1] == ':')  
 {  
 buf[strlen(buf) - 1] = '\0';   
 temp\_type = Label;  
 }  
 else if (buf[0] == '\_' && (strlen(buf) == 18))  
 {   
 bool valid = true;  
  
 if (!(buf[1] >= 'A' && buf[1] <= 'Z')) valid = false;   
 for (int i = 2; i < 18; i++)  
 {  
 if (!(buf[i] >= 'a' && buf[i] <= 'z') && !(buf[i] >= '0' && buf[i] <= '9'))  
 {  
 valid = false;   
 break;  
 }  
 }  
 if (valid)  
 {  
 temp\_type = Identifier;   
 }  
 }  
 strcpy\_s(TempToken.name, buf);  
 TempToken.type = temp\_type;  
 TempToken.value = 0;  
 TempToken.line = line;  
 if (temp\_type == Unknown)  
 {  
 fprintf(errFile, "Lexical Error: line %d, lexem %s is Unknown\n", line, TempToken.name);  
 }  
 state = Finish;  
 break;  
 }  
  
 case Digit:  
 {  
 buf[0] = ch;  
 int j = 1;  
  
 ch = getc(F);  
  
 while ((ch <= '9' && ch >= '0') && j < 15)  
 {  
 buf[j++] = ch;  
 ch = getc(F);  
 }  
 buf[j] = '\0';  
  
 strcpy\_s(TempToken.name, buf);  
 TempToken.type = Number;  
 TempToken.value = atoi(buf);  
 TempToken.line = line;  
 state = Finish;  
 break;  
 }  
  
 case Separators:  
 {  
 if (ch == '\n')  
 line++;  
  
 ch = getc(F);  
  
 state = Start;  
 break;  
 }  
  
 case SComment:  
 {  
 ch = getc(F);  
 if (ch == '#')  
 state = Comment;  
 break;  
 }  
  
 case Comment:  
 {  
 while (1)  
 {  
 ch = getc(F);  
  
 if (ch == '#')  
 {  
 ch = getc(F);  
 if (ch == '#')  
 {  
 state = Start;   
 ch = getc(F);   
 break;  
 }  
 }  
 if (ch == EOF)  
 {  
 printf("Error: Comment not closed!\n");  
 state = EndOfFile;  
 break;  
 }  
 }  
 break;  
 }  
  
 case Another:  
 {  
 switch (ch)  
 {  
  
 case '(':  
 {  
 strcpy\_s(TempToken.name, "(");  
 TempToken.type = LBraket;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 case ')':  
 {  
 strcpy\_s(TempToken.name, ")");  
 TempToken.type = RBraket;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 case ';':  
 {  
 strcpy\_s(TempToken.name, ";");  
 TempToken.type = Semicolon;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
 case ',':  
 {  
 strcpy\_s(TempToken.name, ",");  
 TempToken.type = Comma;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 case ':':  
 {  
 char next = getc(F);   
 strcpy\_s(TempToken.name, ":");  
 TempToken.type = Colon;   
 ungetc(next, F);   
   
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
 case '+':  
 {  
  
 strcpy\_s(TempToken.name, "+");  
 TempToken.type = Add;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
  
 break;  
 }  
  
 case '-':  
 {  
 strcpy\_s(TempToken.name, "-");  
 TempToken.type = Sub;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 case '\*':  
 {  
 strcpy\_s(TempToken.name, "\*");  
 TempToken.type = Mul;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
   
 case '!':  
 {  
 ch = getc(F);  
 if (ch == '!')  
 {  
 strcpy\_s(TempToken.name, "!!");  
 TempToken.type = Not;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 }  
 else  
 {  
 strcpy\_s(TempToken.name, "!");  
 TempToken.type = Unknown;  
 TempToken.value = 0;  
 TempToken.line = line;  
 fprintf(errFile, "Lexical Error: line %d, lexem %s is Unknown\n", line, TempToken.name);  
 state = Finish;  
 }  
 break;  
 }  
  
 case '<':  
 {  
 ch = getc(F);  
 if (ch == '=')  
 {  
 ch = getc(F);  
 if (ch == '=')  
 {  
 strcpy\_s(TempToken.name, "<==");  
 TempToken.type = Assign;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 }  
 }  
 else if (ch == '>')  
 {  
 strcpy\_s(TempToken.name, "<>");  
 TempToken.type = NotEquality;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 }  
 else  
 {  
 strcpy\_s(TempToken.name, "<");  
 TempToken.type = Less;  
 TempToken.value = 0;  
 TempToken.line = line;  
 state = Finish;   
 }  
 break;  
 }  
  
 case '>':  
 {  
 strcpy\_s(TempToken.name, ">");  
 TempToken.type = Greate;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 case '=':  
 {  
 strcpy\_s(TempToken.name, "=");  
 TempToken.type = Equality;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
  
 default:  
 {  
 TempToken.name[0] = ch;  
 TempToken.name[1] = '\0';  
 TempToken.type = Unknown;  
 TempToken.value = 0;  
 TempToken.line = line;  
 ch = getc(F);  
 state = Finish;  
 break;  
 }  
 }  
 }  
 }  
}

}

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 Mainprogram: strcpy\_s(type\_tokens, "MainProgram"); break; 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 EndProgram: strcpy\_s(type\_tokens, "EndProgram"); 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 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 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 Do: strcpy\_s(type\_tokens, "Do"); break; case While: strcpy\_s(type\_tokens, "While"); break; case Exit: strcpy\_s(type\_tokens, "Exit"); break; case Continue: strcpy\_s(type\_tokens, "Continue"); break; case End: strcpy\_s(type\_tokens, "End"); break; case Repeat: strcpy\_s(type\_tokens, "Repeat"); break; case Until: strcpy\_s(type\_tokens, "Until"); break; case Label: strcpy\_s(type\_tokens, "Label"); break; case Unknown: default: 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");

}

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 Mainprogram: strcpy\_s(type\_tokens, "MainProgram"); break; 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 EndProgram: strcpy\_s(type\_tokens, "EndProgram"); 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 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 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 Do: strcpy\_s(type\_tokens, "Do"); break; case While: strcpy\_s(type\_tokens, "While"); break; case Exit: strcpy\_s(type\_tokens, "Exit"); break; case Continue: strcpy\_s(type\_tokens, "Continue"); break; case End: strcpy\_s(type\_tokens, "End"); break; case Repeat: strcpy\_s(type\_tokens, "Repeat"); break; case Until: strcpy\_s(type\_tokens, "Until"); break; case Label: strcpy\_s(type\_tokens, "Label"); break; case Unknown: default: 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**

#include <stdio.h> #include <string.h> #include <stdlib.h> #include "translator.h"

// таблиця лексем Token\* TokenTable; // кількість лексем unsigned int TokensNum;

// таблиця ідентифікаторів Id\* IdTable; // кількість ідентифікаторів unsigned int IdNum;

// Function to validate file extension int hasValidExtension(const char\* fileName, const char\* extension) { const char\* dot = strrchr(fileName, '.'); if (!dot || dot == fileName) return 0; // No extension found return strcmp(dot, extension) == 0; }

int main(int argc, char\* argv[]) { // виділення пам'яті під таблицю лексем TokenTable = new Token[MAX\_TOKENS];

// виділення пам'яті під таблицю ідентифікаторів  
IdTable = new Id[MAX\_IDENTIFIER];  
  
char InputFile[32] = "";  
  
FILE\* InFile;  
  
if (argc != 2)  
{  
 printf("Input file name: ");  
 gets\_s(InputFile);  
}  
else  
{  
 strcpy\_s(InputFile, argv[1]);  
}  
  
// Check if the input file has the correct extension  
if (!hasValidExtension(InputFile, ".t22"))  
{  
 printf("Error: Input file has invalid extension.\n");  
 return 1;  
}  
  
if ((fopen\_s(&InFile, InputFile, "rt")) != 0)  
{  
 printf("Error: Cannot open file: %s\n", InputFile);  
 return 1;  
}  
  
char NameFile[32] = "";  
int i = 0;  
while (InputFile[i] != '.' && InputFile[i] != '\0')  
{  
 NameFile[i] = InputFile[i];  
 i++;  
}  
NameFile[i] = '\0';  
  
char TokenFile[32];  
strcpy\_s(TokenFile, NameFile);  
strcat\_s(TokenFile, ".token");  
  
char ErrFile[32];  
strcpy\_s(ErrFile, NameFile);  
strcat\_s(ErrFile, "\_errors.txt");  
  
FILE\* errFile;  
if (fopen\_s(&errFile, ErrFile, "w") != 0)  
{  
 printf("Error: Cannot open file for writing: %s\n", ErrFile);  
 return 1;  
}  
  
TokensNum = GetTokens(InFile, TokenTable, errFile);  
  
PrintTokensToFile(TokenFile, TokenTable, TokensNum);  
fclose(InFile);  
  
printf("\nLexical analysis completed: %d tokens. List of tokens in the file %s\n", TokensNum, TokenFile);  
printf("\nList of errors in the file %s\n", ErrFile);  
  
Parser(errFile);  
fclose(errFile);  
ASTNode\* ASTree = ParserAST();  
  
char AST[32];  
strcpy\_s(AST, NameFile);  
strcat\_s(AST, ".ast");  
// Open output file  
FILE\* ASTFile;  
fopen\_s(&ASTFile, AST, "w");  
if (!ASTFile)  
{  
 printf("Failed to open output file.\n");  
 exit(1);  
}  
PrintASTToFile(ASTree, 0, ASTFile);  
printf("\nAST has been created and written to %s.\n", AST);  
  
char OutputFile[32];  
strcpy\_s(OutputFile, NameFile);  
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 has been generated and written to %s.\n", OutputFile);  
  
fclose(outFile);  
  
fopen\_s(&outFile, OutputFile, "r");  
char ExecutableFile[32];  
strcpy\_s(ExecutableFile, NameFile);  
strcat\_s(ExecutableFile, ".exe");  
compile\_to\_exe(OutputFile, ExecutableFile);  
  
char OutputFileFromAST[32];  
strcpy\_s(OutputFileFromAST, NameFile);  
strcat\_s(OutputFileFromAST, "\_fromAST.c");  
  
FILE\* outFileFromAST;  
fopen\_s(&outFileFromAST, OutputFileFromAST, "w");  
if (!outFileFromAST)  
{  
 printf("Failed to open output file.\n");  
 exit(1);  
}  
generateCodefromAST(ASTree, outFileFromAST);  
printf("\nC code has been generated and written to %s.\n", OutputFileFromAST);  
  
fclose(outFileFromAST);  
  
fopen\_s(&outFileFromAST, OutputFileFromAST, "r");  
char ExecutableFileFromAST[32];  
strcpy\_s(ExecutableFileFromAST, NameFile);  
strcat\_s(ExecutableFileFromAST, "\_fromAST.exe");  
compile\_to\_exe(OutputFileFromAST, ExecutableFileFromAST);  
  
// Close the file  
\_fcloseall();  
  
destroyTree(ASTree);  
  
delete[] TokenTable;  
delete[] IdTable;  
  
return 0;

}

**Parser.cpp**

#include <stdio.h> #include <string.h> #include <stdlib.h> #include "translator.h" #include #include

// таблиц¤ лексем extern Token\* TokenTable; // к≥льк≥сть лексем extern unsigned int TokensNum;

// таблиц¤ ≥дентиф≥катор≥в extern Id\* IdTable; // к≥льк≥сть ≥дентиф≥катор≥в extern unsigned int IdNum;

static int pos = 0;

// наб≥р функц≥й дл¤ рекурсивного спуску // на кожне правило - окрема функц≥¤ void program(FILE\* errFile); void variable\_declaration(FILE\* errFile); void variable\_list(FILE\* errFile); void program\_body(FILE\* errFile); void statement(FILE\* errFile); void assignment(FILE\* errFile); void arithmetic\_expression(FILE\* errFile); void term(FILE\* errFile); void factor(FILE\* errFile); void input(FILE\* errFile); void output(FILE\* errFile); void conditional(FILE\* errFile);

void goto\_statement(FILE\* errFile); void label\_statement(FILE\* errFile); void for\_to\_do(FILE\* errFile); void for\_downto\_do(FILE\* errFile); void while\_statement(FILE\* errFile); void repeat\_until(FILE\* errFile);

void logical\_expression(FILE\* errFile); void and\_expression(FILE\* errFile); void comparison(FILE\* errFile); void compound\_statement(FILE\* errFile); std::string TokenTypeToString(TypeOfTokens type);

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

void Parser(FILE\* errFile) { program(errFile); fprintf(errFile, "\nNo errors found.\n"); }

void match(TypeOfTokens expectedType, FILE\* errFile) { if (TokenTable[pos].type == expectedType) pos++; else { fprintf(errFile, "\nSyntax error in line %d : another type of lexeme was expected.\n", TokenTable[pos].line); fprintf(errFile, "\nSyntax error: type %s\n", TokenTypeToString(TokenTable[pos].type).c\_str()); fprintf(errFile, "Expected Type: %s ", TokenTypeToString(expectedType).c\_str()); exit(10); } }

void program(FILE\* errFile) { match(Mainprogram, errFile); match(ProgramName, errFile); match(Variable, errFile); variable\_declaration(errFile); match(Semicolon, errFile); match(StartProgram, errFile); program\_body(errFile); match(EndProgram, errFile); }

void variable\_declaration(FILE\* errFile) { if (TokenTable[pos].type == Type) { pos++; variable\_list(errFile); } }

void variable\_list(FILE\* errFile) { match(Identifier, errFile); while (TokenTable[pos].type == Comma) { pos++; match(Identifier, errFile); } }

void program\_body(FILE\* errFile) { do { statement(errFile); } while (TokenTable[pos].type != EndProgram); }

void statement(FILE\* errFile) { switch (TokenTable[pos].type) { case Input: input(errFile); break; case Output: output(errFile); break; case If: conditional(errFile); break; case Label: label\_statement(errFile); break; case StartProgram: compound\_statement(errFile); break; case Goto: goto\_statement(errFile); break; case For: { int temp\_pos = pos + 1; while (TokenTable[temp\_pos].type != To && TokenTable[temp\_pos].type != DownTo && temp\_pos < TokensNum) { temp\_pos++; } if (TokenTable[temp\_pos].type == To) { for\_to\_do(errFile); } else if (TokenTable[temp\_pos].type == DownTo) { for\_downto\_do(errFile); } else { printf("Error: Expected 'To' or 'DownTo' after 'For'\n"); } break; } case While: while\_statement(errFile); break; case Exit: pos += 2; break; case Continue: pos += 2; break; case Repeat: repeat\_until(errFile); break; default: assignment(errFile); break; } }

void assignment(FILE\* errFile) { match(Identifier, errFile); match(Assign, errFile); arithmetic\_expression(errFile); match(Semicolon, errFile); }

void arithmetic\_expression(FILE\* errFile) { term(errFile); while (TokenTable[pos].type == Add || TokenTable[pos].type == Sub) { pos++; term(errFile); } }

void term(FILE\* errFile) { factor(errFile); while (TokenTable[pos].type == Mul || TokenTable[pos].type == Div || TokenTable[pos].type == Mod) { pos++; factor(errFile); } }

void factor(FILE\* errFile) { if (TokenTable[pos].type == Identifier) { match(Identifier, errFile); } else if (TokenTable[pos].type == Number) { match(Number, errFile); } else if (TokenTable[pos].type == LBraket) { match(LBraket, errFile); arithmetic\_expression(errFile); match(RBraket, errFile); } else { printf("\nSyntax error in line %d : A multiplier was expected.\n", TokenTable[pos].line); exit(11); } }

void input(FILE\* errFile) { match(Input, errFile); match(Identifier, errFile); match(Semicolon, errFile); }

void output(FILE\* errFile) { match(Output, errFile); if (TokenTable[pos].type == Sub) { pos++; if (TokenTable[pos].type == Number) { match(Number, errFile); } } else { arithmetic\_expression(errFile); } match(Semicolon, errFile); }

void conditional(FILE\* errFile) { match(If, errFile); logical\_expression(errFile); statement(errFile); if (TokenTable[pos].type == Else) { pos++; statement(errFile); } }

void goto\_statement(FILE\* errFile) { match(Goto, errFile); if (TokenTable[pos].type == Identifier) { pos++; match(Semicolon, errFile); } else { printf("Error: Expected a label after 'goto' at line %d.\n", TokenTable[pos].line); exit(1); } }

void label\_statement(FILE\* errFile) { match(Label, errFile); }

void for\_to\_do(FILE\* errFile) { match(For, errFile); match(Identifier, errFile); match(Assign, errFile); arithmetic\_expression(errFile); match(To, errFile); arithmetic\_expression(errFile); match(Do, errFile); statement(errFile); }

void for\_downto\_do(FILE\* errFile) { match(For, errFile); match(Identifier, errFile); match(Assign, errFile); arithmetic\_expression(errFile); match(DownTo, errFile); arithmetic\_expression(errFile); match(Do, errFile); statement(errFile); }

void while\_statement(FILE\* errFile) { match(While, errFile); logical\_expression(errFile);

while (1)   
{  
 if (TokenTable[pos].type == End)  
 {  
 pos++;   
 match(While, errFile);   
 break;   
 }  
 else  
 {  
 statement(errFile);  
 if (TokenTable[pos].type == Semicolon)  
 {  
 pos++;  
 }  
 }  
}

}

void repeat\_until(FILE\* errFile) { match(Repeat, errFile); statement(errFile); match(Until, errFile); logical\_expression(errFile); }

void logical\_expression(FILE\* errFile) { and\_expression(errFile); while (TokenTable[pos].type == Or) { pos++; and\_expression(errFile); } }

void and\_expression(FILE\* errFile) { comparison(errFile); while (TokenTable[pos].type == And) { pos++; comparison(errFile); } }

void comparison(FILE\* errFile) { if (TokenTable[pos].type == Not) { pos++; match(LBraket, errFile); logical\_expression(errFile); match(RBraket, errFile); } else if (TokenTable[pos].type == LBraket) { pos++; logical\_expression(errFile); match(RBraket, errFile); } else { arithmetic\_expression(errFile); if (TokenTable[pos].type == Greate || TokenTable[pos].type == Less || TokenTable[pos].type == Equality || TokenTable[pos].type == NotEquality) { pos++; arithmetic\_expression(errFile); } else { fprintf(errFile, "\nSyntax error in line %d : A comparison operation is expected.\n", TokenTable[pos].line); exit(12); } } }

void compound\_statement(FILE\* errFile) { match(StartProgram, errFile); program\_body(errFile); match(EndProgram, errFile); }

unsigned int IdIdentification(Id IdTable[], Token TokenTable[], unsigned int tokenCount, FILE\* errFile) { 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)  
 {  
 printf("\nidentifier \"%s\" is already declared !\n", TokenTable[i].name);  
 return idCount;  
 }  
  
  
 if (idCount < MAX\_IDENTIFIER)  
 {  
 strcpy\_s(IdTable[idCount++].name, TokenTable[i++].name);  
 }  
 else  
 {  
 printf("\nToo many identifiers !\n");  
 return idCount;  
 }  
 }  
 else  
 i++;  
 }  
}  
  
  
for (; i < tokenCount; i++)  
{  
 if (TokenTable[i].type == Identifier && TokenTable[i + 1].type != Colon)  
 {  
 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)  
 {  
 if (idCount < MAX\_IDENTIFIER)  
 {  
 strcpy\_s(IdTable[idCount++].name, TokenTable[i].name);  
 }  
 else  
 {  
 printf("\nToo many identifiers!\n");  
 return idCount;  
 }  
 }  
 }  
  
  
}  
  
return idCount;

}

std::string TokenTypeToString(TypeOfTokens type)

{

switch (type)

{

case Mainprogram: return "Mainprogram";

case StartProgram: return "StartProgram";

case Variable: return "Variable";

case Type: return "Type";

case EndProgram: return "EndProgram";

case Input: return "Input";

case Output: return "Output";

case If: return "If";

case Else: return "Else";

case Goto: return "Goto";

case Label: return "Label";

case For: return "For";

case To: return "To";

case DownTo: return "DownTo";

case Do: return "Do";

case While: return "While";

case Exit: return "Exit";

case Continue: return "Continue";

case End: return "End";

case Repeat: return "Repeat";

case Until: return "Until";

case Identifier: return "Identifier";

case Number: return "Number";

case Assign: return "Assign";

case Add: return "Add";

case Sub: return "Sub";

case Mul: return "Mul";

case Div: return "Div";

case Mod: return "Mod";

case Equality: return "Equality";

case NotEquality: return "NotEquality";

case Greate: return "Greate";

case Less: return "Less";

case Not: return "Not";

case And: return "And";

case Or: return "Or";

case LBraket: return "LBraket";

case RBraket: return "RBraket";

case Semicolon: return "Semicolon";

case Colon: return "Colon";

case Comma: return "Comma";

case Unknown: return "Unknown";

default: return "InvalidType";

}

}