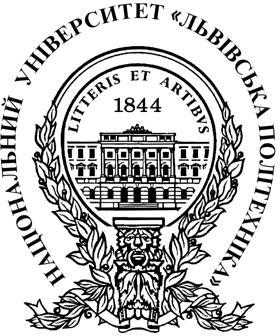
Міністерство освіти і науки України

Національний університет "Львівська Політехніка"

Кафедра ЕОМ



## Пояснювальна записка

до курсового проєкту "СИСТЕМНЕ ПРОГРАМУВАННЯ"

на тему: “РОЗРОБКА СИСТЕМНИХ ПРОГРАМНИХ МОДУЛІВ ТА КОМПОНЕНТ СИСТЕМ ПРОГРАМУВАННЯ”

Індивідуальне завдання 7

“РОЗРОБКА ТРАНСЛЯТОРА З ВХІДНОЇ МОВИ ПРОГРАМУВАННЯ”

Виконав студент групи КІ-307:

Гагін. С. Я

Перевірив:

Козак Н.Б.

Львів-2024

ЗАВДАННЯ НА КУРСОВИЙ ПРОЄКТ

1. Цільова мова транслятора - мова програмування С або асемблер для 32/64 розрядного процесора.
2. Для отримання виконуваного файлу на виході розробленого транслятора скористатися середовищем Microsoft Visual Studio або будь-яким іншим.
3. Мова розробки транслятора: C/C++.
4. Реалізувати графічну 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 С або асемблера;*

*об’єктний файл; виконуваний файл.*

1. 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.

Деталізація завдання на проектування:

1. В кожному завданні передбачається блок оголошення змінних; змінні зберігають значення цілих чисел і, в залежності від варіанту, можуть бути 16/32 розрядними. За потребою можна реалізувати логічний тип даних.
2. Необхідно реалізувати арифметичні операції - додавання, віднімання, множення, ділення, залишок від ділення; операції порівняння - перевірка на рівність і нерівність, більше і менше; логічні операції - заперечення, “логічне І” і “логічне АБО”.

Пріоритет операцій наступний - круглі дужки (), логічне заперечення, мультиплікативні (множення, ділення, залишок від ділення), адитивні (додавання, віднімання), відношення (більше, менше), перевірка на рівність і нерівність, логічне І, логічне АБО.

1. За допомогою оператора вводу можна зчитати з клавіатури значення змінної; за допомогою оператора виводу можна вивести на екран значення змінної, виразу чи цілої константи.
2. В кожному завданні обов’язковим є оператор присвоєння за допомогою якого можна реалізувати обчислення виразів з використанням заданих операцій і операції круглі дужки (); у якості операндів можуть бути цілі константи, змінні, а також інші вирази.
3. В кожному завданні обов’язковим є оператор типу “блок” (складений оператор), його вигляд має бути таким, як і блок тіла програми.
4. Необхідно реалізувати задані варіантом оператори, синтаксис операторів наведено у таблиці 1.1. Синтаксис вхідної мови має забезпечити реалізацію обчислень лінійних алгоритмів, алгоритмів з розгалуженням і циклічних алгоритмів. Опис формальної мови студент погоджує з викладачем.
5. Оператори можуть бути довільної вкладеності і в будь-якій послідовності.
6. Для перевірки роботи розробленого транслятора, необхідно написати три тестові програми на вхідній мові програмування.

Деталізований опис власної мови програмування:

Блок тіла програми:

**Program**

**var … ;**

**start**

## finish

Оператори вводу-виводу: **get, put**

Оператор присвоєння:<-

Оператори:

**If- else**

**goto**

**for-to**

**for-downto**

**while**

**repeat-untli**

Регістр ідентифікаторів: **Up4**

Операції:

арифметичні: **+ , -, \*, /, %**

порівняння: **eg, ne, >= ,<=**

логічні: **! , and , or**

Тип даних: **int16\_t**

Коментар: **%%…%%**

Програма на вхідній мові програмування має починатись з ключового слова **program**, далі йде <name>, далі має йти розділ опису змінних **var** після має відкриватись блок коду програми з ключового слова **start.** Між розділом **start** і ключовим словом **finish** розміщуються оператори програми. Операторів є 9: оператор вводу даних **get**, оператор виводу даних **put**, оператор присвоєння **<-,** умовний оператор **If- else**, оператор безумовного переходу **goto** та оператори циклів: **for-to**, **for-downto**, **while**, **repeat-untli**. Кожен оператор має завершуватись символом крапка з комою **;**. Оператор присвоєння дозволяє присвоїти деякій змінній значення арифметичного виразу. Допустимі арифметичні операції - **+ , \*, /, %**. Операндами можуть бути змінні, цілі додатні константи і інші вирази, взяті в дужки. В умовному операторі використовуються логічні вирази, допустимі такі операції порівняння **eg, ne, >= ,<=** і такі логічні операції **! , and , or**. Ідентифікатори (імена змінних) можуть бути довжиною 4 символи, складаються лише з великих латинських літер або цифр. Тип даних лише один – **int16\_t,** при оголошення декількох змінних вони записуються через кому, вкінці опису змінних ставиться символ крапка з комою **;**. Коментарі починаються з двох символів відсоток %% та закінчуються двома символами відсотка.

Приклади оголошення змінних:

var int16\_t NUMA;

var int16\_t NUMA, NUMB;

var int16\_t NUMA, NUMB, RESU;

Приклади ідентифікаторів:

NUMA, MAXX, RESU.

Приклади арифметичних виразів:

0 - 75

NUMY + 68

NUMA + NUMB \*NUMC + 76

NUMC\* (NUMA + NUMB) - 50 / NUMB

Приклади логічних виразів:

NUMA >= NUMB

NUMA>=NUMB and NUMA >=NUMA

! (NUMA <= NUMC )

NUMA ne NUMB

MINN <= 0 - 100

АНОТАЦІЯ

У даному курсовому проекті розроблено програмне забезпечення - транслятор з вхідної мови програмування.

Для реалізації транслятора визначено граматику вхідної мови програмування у термінах розширеної нотації Бекуса-Наура.

Реалізовано лексичний, синтаксичний, семантичний аналізатор. На етапі синтаксичного і семантичного аналізу відбувається перевірка програми на вхідній мові програмування на наявність помилок.

Перед генеруванням вихідного коду програма на вхідній мові програмування перетворюється у двійкове абстрактне синтаксичне дерево, обходячи яке генератор коду будує вихідний код на мові програмування С.

Розроблене програмне забезпечення налаштоване і протестоване на тестових прикладах.

ВСТУП 8

1. ОГЛЯД МЕТОДІВ ТА СПОСОБІВ ПРОЄКТУВАННЯ ТРАНСЛЯТОРІВ 9
2. [ФОРМАЛЬНИЙ ОПИС ВХІДНОЇ МОВИ ПРОГРАМУВАННЯ 1](#bookmark19" \o "Current Document)2
   1. [Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура. 1](#bookmark21" \o "Current Document)3
   2. [Опис термінальних символів та ключових слів 1](#bookmark26" \o "Current Document)4
3. [РОЗРОБКА ТРАНСЛЯТОРА З ВХІДНОЇ МОВИ ПРОГРАМУВАННЯ 1](#bookmark29" \o "Current Document)6
   1. [Вибір технології програмування 1](#bookmark32" \o "Current Document)6
   2. [Проектування таблиць транслятора та вибір структур даних 1](#bookmark35" \o "Current Document)6
   3. [Розробка лексичного аналізатора 1](#bookmark38" \o "Current Document)9
      1. Розробка алгоритму роботи лексичного аналізатора 20
      2. Опис програми реалізації лексичного аналізатора 22
   4. [Розробка синтаксичного та семантичного аналізатора](#bookmark43" \o "Current Document) 23
      1. Розробка дерева граматичного розбору 24
      2. [Розробка алгоритму роботи синтаксичного і семантичного аналізатора................................................................................................................](#bookmark46" \o "Current Document)25
      3. [Опис програми реалізації синтаксичного та семантичного аналізатора.](#bookmark47" \o "Current Document)  33
   5. [Розробка генератора коду](#bookmark49" \o "Current Document) 34
      1. Розробка алгоритму роботи генератора коду 34
      2. Опис програми реалізації генератора коду 49
4. [НАЛАГОДЖЕННЯ ТА ТЕСТУВАННЯ РОЗРОБЛЕНОГО ТРАНСЛЯТОРА ... 36](#bookmark53" \o "Current Document)
   1. [Опис інтерфейсу та інструкції користувачу](#bookmark58" \o "Current Document) 49
   2. [Виявлення лексичних і синтаксичних помилок](#bookmark61" \o "Current Document) 50
   3. [Перевірка роботи транслятора за допомогою тестових задач](#bookmark64" \o "Current Document) 53

ВИСНОВОК 62

[СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ](#bookmark69" \o "Current Document) 63

[ДОДАТКИ](#bookmark72" \o "Current Document) 64

### ВСТУП

У сучасному світі технології невпинно розвиваються, проникаючи в усі сфери людської діяльності — від науки та техніки до мистецтва та дозвілля. Програмне забезпечення стало основою багатьох процесів, а створення інструментів для його розробки сприяє ефективнішій реалізації ідей і проєктів. Однією з найважливіших складових таких інструментів є транслятори, які дозволяють перетворювати вихідний код однієї мови програмування в іншу, забезпечуючи його виконання на різних апаратних платформах.  
Цей курсовий проєкт спрямований на створення компілятора для спеціально розробленої мови програмування. Основним завданням є забезпечення можливості трансляції вхідного коду в мову низького рівня для подальшого виконання або інтеграції у більші системи. У рамках проєкту будуть реалізовані ключові етапи компіляції: лексичний, синтаксичний і семантичний аналіз, а також генерація коду.

### Актуальність

Зростання складності програмного забезпечення і кількості платформ, на яких воно виконується, підкреслює необхідність у гнучких інструментах розробки. Компілятори забезпечують не лише спрощення розробки, але й підвищення її продуктивності, дозволяючи швидко адаптувати програмний код до різних середовищ. У сучасних умовах, коли гнучкість і портативність стали ключовими вимогами, створення нового транслятора під авторську мову програмування є актуальним завданням, що сприяє розвитку індустрії.

### Мета проєкту

Метою роботи є розробка компілятора для авторської мови програмування, який перетворює її код у формат, придатний для виконання. Для цього буде реалізовано:

* лексичний аналізатор для розпізнавання структурних елементів мови;
* синтаксичний аналізатор для перевірки правильності побудови виразів;
* семантичний аналізатор для перевірки коректності значень і операцій;
* генератор коду для створення вихідного файлу.

1. **ОГЛЯД МЕТОДІВ ТА СПОСОБІВ ПРОЄКТУВАННЯ**

**ТРАНСЛЯТОРІВ**

**1.1 Лексичний аналіз**

На цьому етапі текст програми розбивається на лексеми: ключові слова, ідентифікатори, літерали, роздільники та оператори. Основні методи реалізації:

* Регулярні вирази — для опису шаблонів лексем.
* Скінченні автомати — для автоматичного розпізнавання лексем.

**1.2 Синтаксичний аналіз**

Синтаксичний аналіз перевіряє структуру програми відповідно до граматики мови та формує дерево розбору. Основні методи:

* Зверху вниз (Top-down parsing):
* Рекурсивний спуск (Recursive Descent Parsing).
* LL-аналізатори (табличні методи).
* Знизу вгору (Bottom-up parsing):
* LR-аналізатори (LR(0), SLR, LALR, Canonical LR).

**1.3 Семантичний аналіз**

Семантичний аналіз перевіряє логічну коректність програми, враховуючи:

* Відповідність типів даних.
* Область видимості змінних.
* Коректність викликів функцій.

Методи реалізації:

* Атрибутні граматики.
* Таблиці символів — для збереження інформації про ідентифікатори.

**1.4 Генерація коду**

Цей етап перетворює вихідний код у проміжний або машинний код. Методи:

- Однопрохідна генерація — для простих мов.

- Двопрохідна генерація — створення проміжного коду з подальшою оптимізацією.

- Багатопрохідна генерація — використовується для складних мов, включає додаткові етапи оптимізації.

**1.5 Оптимізація коду**

Мета оптимізації — підвищити ефективність виконання програми та зменшити її розмір. Основні підходи:

* Локальна оптимізація — у межах одного блоку коду.
* Глобальна оптимізація — враховує взаємодію між блоками.
* Машинно-незалежна оптимізація — спрощення обчислень.
* Машинно-залежна оптимізація — призначення регістрів, налаштування інструкцій.

**Класифікація трансляторів**

1. За способом роботи:

- Компілятори — перетворюють код у машинний.

- Інтерпретатори — виконують код безпосередньо.

2. За кількістю проходів:

- Однопрохідні.

- Багатопрохідні.

3. За цільовим кодом:

- Машинний код.

- Проміжний код (LLVM IR, байт-код).

**Основні підходи до проєктування**

**1. Модульний підхід**

Розділення транслятора на незалежні компоненти, такі як лексичний і синтаксичний аналізатори, генератор коду тощо.

**2. Фазовий підхід**

Чітке розділення етапів із передачею проміжних результатів між ними.

**3. Однопрохідний дизайн**

Ефективний для простих мов, коли всі етапи виконуються в одному проході.

**4. Багатопрохідний дизайн**

Використовується для складних мов і передбачає глибоку оптимізацію коду через кілька етапів аналізу й генерації.

1. ФОРМАЛЬНИЙ ОПИС ВХІДНОЇ МОВИ ПРОГРАМУВАННЯ
   1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура.

Для задання синтаксису мов програмування використовують форму Бекуса- Наура або розширену форму Бекуса-Наура — це спосіб запису правил контекстно- вільної граматики, тобто форма опису формальної мови. Саме її типово використовують для запису правил мов програмування та протоколів комунікації.

БНФ визначає скінченну кількість символів (нетерміналів). Крім того, вона визначає правила заміни символу на якусь послідовність букв (терміналів) і символів. Процес отримання ланцюжка букв можна визначити поетапно: спочатку є один символ (символи зазвичай знаходяться у кутових дужках, а їх назва не несе жодної інформації). Потім цей символ замінюється на деяку послідовність букв і символів, відповідно до одного з правил. Потім процес повторюється (на кожному кроці один із символів замінюється на послідовність, згідно з правилом). Зрештою , виходить ланцюжок , що складається з букв і не містить символів. Це означає , що отриманий ланцюжок може бути виведений з початкового символу .

Нотація БНФ є набором «продукцій», кожна з яких відповідає зразку:

<символ> ::= <вираз, що містить символи>

=

де вираз, що містить символи це послідовність символів або послідовності символів, розділених вертикальною рискою |, що повністю перелічують можливий вибір символ з лівої частини формули.

У розширеній формі нотації Бекуса — Наура вирази, що можна пропускати або які можуть повторятись слід записувати у фігурних дужках { ... }:, а можлива поява може відображатися застосуванням квадратних дужок [ ... ]:.

Опис вхідної мови програмування у термінах розширеної форми Бекуса- Наура:

program ::= "**program**", {"**var**", variable\_declaration, "**;**"}, "**start**", {statement, "**;**"}, "**finish**";

variable\_declaration ::= "**int16\_t**", variable\_list;

variable\_list ::= identifier, {"**,**", **identifier**};

identifier ::= up{4};

up\_low ::= **up** | **low** | **digit**;

up ::= "**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**";

low ::= **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**" ;

statement ::= input\_statement | output\_statement | assign\_statement | if\_else\_statement | goto\_statement | label\_point | for\_statement | while\_statement | repeat\_until\_statement | compound\_statement;

input\_statement ::= "**get**", identifier;

output\_statement ::= "**put**", arithmetic\_expression;

arithmetic\_expression ::= low\_priority\_expression {low\_priority\_operator, low\_priority\_expression};

low\_priority\_operator ::= "**+**" | "**-**";

low\_priority\_expression ::= middle\_priority\_expression {middle\_priority\_operator, middle\_priority\_expression};

middle\_priority\_operator ::= "**\***" | "**/**" | "**%**";

middle\_priority\_expression ::= identifier | number | "**(**", arithmetic\_expression, "**)**";

number ::= ["**-**"], digit {5};

assign\_statement ::= identifier, "**<-**", arithmetic\_expression;

if\_else\_statement ::= "**if**", "**(**", logical\_expression, "**)**", statement, ["**;**", "**else**", statement];

logical\_expression ::= and\_expression {or\_operator, and\_expression};

or\_operator ::= "**or**";

and\_expression ::= comparison {and\_operator, and\_expression};

and\_operator ::= "**and**";

comparison ::= comparison\_expression | [not\_operator] "**(**", logical\_expression, "**)**";

not\_operator ::= "**!**";

comparison\_expression ::= arithmetic\_expression comparison\_operator arithmetic\_expression;

comparison\_operator ::= "eg" | "ne" | "**<=**" | "**>=**";

goto\_statement ::= "**goto**", identifier;

label\_point ::= identifier, "**:**";

for\_to\_statement ::= "**for**", assign\_statement, "**to**" | "**downto**", arithmetic\_expression, "**do**", statement;

while\_statement ::= "**while**", logical\_expression, {statement, "**;**"}, "**wend**";

repeat\_until\_statement ::= "**repeat**", {statement, "**;**"}, "**until**", "**(**", logical\_expression, "**)**";

compoundStatement ::= "**start**", {statement, "**;**"}, "**finish**";

* 1. Опис термінальних символів та ключових слів.

Визначаємо термінальні символи і ключові слова:

* **program** - початок програми
* **start** - початок блоку
* **var** - оголошення змінних
* **finish** - кінець програми
* **int16\_t** - тип даних
* **get** - оператор вводу
* **put** - оператор виводу
* **if, else** - умовний оператор
* **goto** – оператор безумного переходу
* **for, to, do** – оператор циклу for
* **for, downto, do** – оператор циклу for
* **while, wend** – оператор циклу while
* **repeat, until** – оператор циклу repeat
* **<-** - оператор присвоєння
* **+** - додавання
* **-** - віднімання
* **\*** - множення
* **/** - ділення
* **%** - остача від ділення
* **>=** - більше
* **<=** - менше
* **eg** - рівність
* **ne** - нерівність
* **!** - заперечення
* **and** - логічне І
* **or** - логічне АБО
* **;** - кінець оператора
* **,** - розділювач змінних
* **(**- відкрита дужка
* **)** - закрита дужка
* **%%** - початок коментаря
* **a...z** - маленькі латинські букви
* **0...9** - цифри
* символи табуляції, переходу на новий рядок, пробіл

1. РОЗРОБКА ТРАНСЛЯТОРА З ВХІДНОЇ МОВИ  
   ПРОГРАМУВАННЯ
   1. Вибір технології програмування.

Перед тим як розпочинати створювати програму, для більш швидкого і ефективного її написання, необхідно розробити алгоритм її функціонування, та вибрати технологію програмування, середовище програмування.

Для виконання поставленого завдання найбільш доцільно буде використати середовище програмування Microsoft Visual Studio 2022, та мову програмування C/C++.

Для якісного і зручного використання розробленої програми користувачем, було прийнято рішення створення консольного інтерфейсу.

* 1. Проектування таблиць транслятора та вибір структур даних.

Використання таблиць значно полегшує створення трансляторів, а тому створимо необхідні структури даних для зберігання інформації про лексеми:

enum TypeOfToken {

StartProgram, // program

StartBlock, // start

Variable, // var

Type, // int16\_t

EndBlock, // finish

Input, // get

Output, // put

If, // if

Then, // then

Else, // else

Goto, // goto

For, // for

To, // to

Downto, // downto

Do, // do

While, // while

Continue, // continue

Exit, // exit

WEnd, // wend

Repeat, // repest

Until, // until

Identifier, // Identifier

Number, // number

Float, // float (incorrect)

Assign, // <-

Add, // +

Sub, // -

Mul, // \*

Div, // /

Mod, // %

Equality, // eg

NotEquality, // en

Greate, // >=

Less, // <=

Not, // !

And, // and

Or, // or

LBracket, // (

RBracket, // )

Semicolon, // ;

Colon, // :

Comma, // ,

Unknown\_

};

struct Token {

char name[16]; // ім'я лексеми

int value; // значення (для констант)

int line; // номер рядка

enum TypeOfToken type; // тип лексеми

};

struct id {

char name[16];

//unsigned int pos; //for labels

};

У програмі будемо зберігати таблицю лексем і таблицю та кількість лексем, ідентифікаторів та міток:

struct Token\* TokenTable; // Таблиця лексем

unsigned int TokensNum; // Кількість лексем

struct id\* idTable; // Таблиця ідентифікаторів

unsigned int idNum; // кількість ідентифікаторів

struct id\* labelTable; // Таблиця міток

unsigned int labelNum; // кількість міток

* 1. Розробка лексичного аналізатора.

Основна задача лексичного аналізу - розбити вихідний текст, що складається з послідовності 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. Вхідна програма проглядається послідовно з початку до кінця. Базові елементи, або лексичнї одиниці, розділяються пробілами, знаками операцїй i спеціальними символами (новий рядок, знак табуляції), і таким чином виділяються та розпізнаються ідентифікатори, літерали і термінальні символи (операції, ключові слова).

При виділенні лексеми вона розпізнається та записується у таблицю лексем за допомогою відповідного номера лексеми, що є унікальним для кожної лексеми із усього можливого їх набору. Це дає можливість наступним фазам компіляції звертатись лексеми не як до послідовності символів, а як до унікального номера 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 текучої позиції аналізу. Також в таблиці лексем ведуться записи, щодо рядка відповідної лексеми - для місця помилки - та додаткова інформація.

Лексична фаза відкидає коментарі, оскільки вони не мають ніякого впливу на виконання програми, отже ж й на синтаксичний розбір та генерацію коду.

Розділимо лексеми на типи або лексичні класи:

• Ключові слова (program, start, var, finish, int16\_t, get, put, if, else, goto, for, to, downto, do, while, wend, repest, until, eg, ne, and, or)

* Ідентифікатори (4 великі літери або цифри)
* Числові константи (ціле число)
* Оператор присвоєння (<-)
* Знаки операції (+, -, \*, eg, ne, >=, <=, !, and, or)
* Розділювачі (, ;)
* Дужки (**(,** **)**)
* Невідома лексема (символи і ланцюжки символів, які не підпадають під вищеописані правила).
  + 1. **Розробка алгоритму роботи лексичного аналізатора.**

Розробимо алгоритм роботи лексичного аналізатора на основі скінченного автомату. Лексичний аналізатор працює за принципом скінченного автомату з такими станами:

* **Start** - початок виділення чергової лексеми;
* **Finish** - кінець виділення чергової лексеми;
* **EndOfFile** - кінець файлу, завершення розпізнавання лексем;
* **Letter** - перший символ буква, розпізнавання слів (ключові слова і ідентифікатори);
* **Digit** - перший символ цифра, розпізнавання числових констант;
* **Separators** - видалення пробілів, символів табуляції і переходу на новий рядок;
* **Comment** - видалення тексту коментаря;
* **Another** - опрацювання інших символів.

У стані **Letter** читаємо по одному символи з файлу і виділяємо ланцюжок символів, який починається з букви чи символу \_, а далі можуть слідувати букви або цифри. Кінець ланцюжка - якщо прочитаний символ відмінний від букви чи цифри. Виділений ланцюжок порівнюємо з ключовими словами, якщо співпадінь немає, вважаємо його ідентифікатором при умові, що довжина ланцюжка не більше 2-х символів і ланцюжок складається з великих літер чи цифр, інакше це невизначена лексема. Переходимо до стану **Finish**.

У стані **Digit** читаємо по одному символи з файлу і виділяємо ланцюжок символів, який починається з крапки, мінуса або ж цифри, далі ж йдуть лише цифри або крапки, вважаємо цей ланцюжок числовою константою. Кінець ланцюжка - якщо прочитаний символ відмінний від цифри. Переходимо до стану **Finish**.

У стан **Comment** ми потрапляємо, якщо натикаємось на відсоток перевіряємо, щоб були 2 символи, далі перебуваючи у стані **Comment,** читаємо символи, поки не зустрінеться символ %% після цього переходимо до виділення нової лексеми - до стану **Start**.

У стані **Separators** читаємо наступний символ і переходимо до виділення нової лексеми - до стану **Start**. Тобто пропускаємо усі пробіли, символи табуляції і переходу на новий рядок.

У стані **Another** порівнюємо поточний прочитаний символ з символами, що позначають знаки операцій, розділювачі і круглі дужки і визначаємо одну з лексем. Є кілька лексем, які вимагають ще читання наступного символу з файлу - це оператор присвоєння “<-” і операцій “<=”, “>=”. Якщо співпадіння не виявлено, то поточний символ - невідома лексема, читаємо наступний символ і переходимо до стану **Finish**.

У стані **Finish** записуємо поточну лексему у таблицю лексем і переходимо до виділення нової лексеми, до стану **Start**.

У стані **EndOfFile** завершуємо обробку вхідного файлу, усі символи з файлу прочитані, усі лексеми записані у таблицю лексем.

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

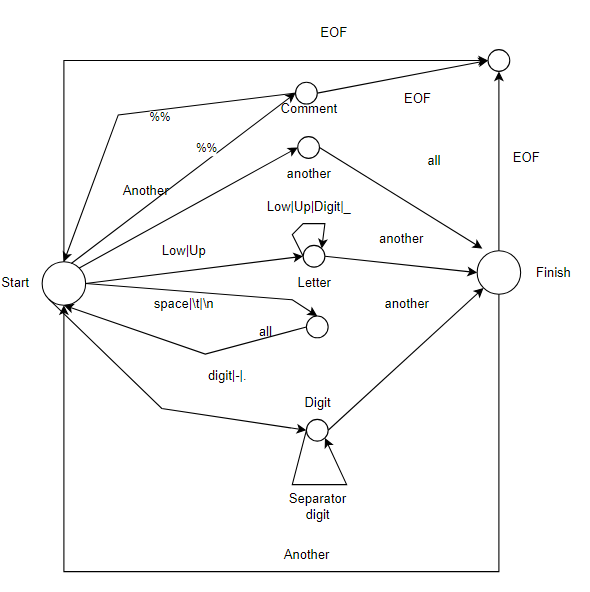


Рис. 3.1. Граф-схема алгоритму роботи лексичного аналізатора.

* + 1. **Опис програми реалізації лексичного аналізатора.**

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

Створимо структуру даних для зберігання стану аналізатора:

enum States {

Start, // початковий стан

Finish, // фінальний стан

Letter, // опрацювання слів (ключові слова та ідентифікаторів)

Digit, // опрацювання цифр

Separator, // опрацювання роздільників

Another, // опрацювання інших символів

EndOFile, // кінець файлу

Comment // ігнорування коментаря

};

Напишемо функцію, яка реалізує лексичний аналіз:

unsigned int getTokens(FILE\* F);

І функції, які друкують список лексем:

void printTokens(void);

void fprintTokens(FILE\* F);

* 1. Розробка синтаксичного та семантичного аналізатора.

Синтаксичний аналіз - це процес, що визначає, чи належить деяка послідовність лексем граматиці мови програмування. В принципі, для будь-якої граматики можна побудувати синтаксичний аналізатор, але граматики, які використовуються на практиці, мають спеціальну форму. Наприклад, відомо, що для будь-якої контекстно-вільної граматики може бути побудований аналізатор, складність якого не перевищує O(n3) для вхідного рядка довжиною n, але в більшості випадків для заданої мови програмування ми можемо побудувати таку граматику, що дозволить сконструювати і більш швидкий аналізатор.

Аналізатори реальних мов зазвичай мають лінійну складність; це досягається за рахунок перегляду вхідної програми зліва направо із загляданням уперед на один термінальний символ (лексичний клас).

Вхід синтаксичного аналізатора - це послідовність лексем і таблиці представлень, які є виходом лексичного аналізатора.

На виході синтаксичного аналізатора отримуємо дерево граматичного розбору і таблиці ідентифікаторів та типів, які є входом для наступного перегляду компілятора (наприклад, це може бути перегляд, який здійснює контроль типів - семантичний аналіз).

* + 1. **Розробка дерева граматичного розбору.**

Схема дерева розбору виглядає наступним чином:

Program

├── "program"

├── {VariableDeclaration ";"}

│ ├── VariableDeclaration

│ │ ├── "int16\_t"

│ │ └── VariableList

│ │ ├── Identifier

│ │ │ └── Up {4}

│ │ │ ├── Up

│ │ │ │ └── "A" | "B" | ... | "Z"

│ │ │ └── Digit

│ │ │ └── "0" | "1" | ... | "9"

│ │ └── {"," Identifier}

├── "start"

├── {Statement ";"}

│ ├── Statement

│ │ ├── Input Statement

│ │ │ ├── "get"

│ │ │ └── Identifier

│ │ ├── OutputStatement

│ │ │ ├── "put"

│ │ │ └── ArithmeticExpression

│ │ │ ├── LowPriorityExpression

│ │ │ │ ├── MiddlePriorityExpression

│ │ │ │ │ ├── Identifier

│ │ │ │ │ ├── Number

│ │ │ │ │ │ └── Digit {4}

│ │ │ │ │ └── "(" ArithmeticExpression ")"

│ │ │ │ └── {MiddlePriorityOperator MiddlePriorityExpression}

│ │ │ └── {LowPriorityOperator LowPriorityExpression}

│ │ ├── AssignStatement

│ │ │ ├── Identifier

│ │ │ ├── "<-"

│ │ │ └── ArithmeticExpression

│ │ ├── IfElseStatement

│ │ │ ├── "if"

│ │ │ ├── "(" LogicalExpression ")"

│ │ │ │ ├── AndExpression

│ │ │ │ │ ├── Comparison

│ │ │ │ │ │ ├── ComparisonExpression

│ │ │ │ │ │ │ ├── ArithmeticExpression

│ │ │ │ │ │ │ ├── ComparisonOperator

│ │ │ │ │ │ │ └── ArithmeticExpression

│ │ │ │ │ │ └── [NotOperator] "(" LogicalExpression ")"

│ │ │ │ │ └── {AndOperator AndExpression}

│ │ │ │ └── {OrOperator AndExpression}

│ │ │ ├── Statement

│ │ │ └── [";" "else" Statement]

│ │ ├── GotoStatement

│ │ │ ├── "goto"

│ │ │ └── Identifier

│ │ ├── LabelPoint

│ │ │ ├── Identifier

│ │ │ └── ":"

│ │ ├── ForToStatement

│ │ │ ├── "for"

│ │ │ ├── AssignStatement

│ │ │ ├── "to" | "DOWNTO"

│ │ │ ├── ArithmeticExpression

│ │ │ ├── "do"

│ │ │ └── Statement

│ │ ├── WhileStatement

│ │ │ ├── "while"

│ │ │ ├── LogicalExpression

│ │ │ ├── {Statement ";"}

│ │ │ └── "wend"

│ │ ├── RepeatUntilStatement

│ │ │ ├── "repeat"

│ │ │ ├── {Statement ";"}

│ │ │ └── "until" "(" LogicalExpression ")"

│ │ └── CompoundStatement

│ │ ├── "start"

│ │ ├── {Statement ";"}

│ │ └── "finish"

└── "finish"

*Рис. 3.2. Дерево граматичного розбору.*

**3.4.2. Розробка алгоритму роботи синтаксичного і семантичного аналізатора.**

Одним з найбільш простих і найбільш популярних методів низхідного синтаксичного аналізу є метод рекурсивного спуску (recursive descent method).

Метод заснований на тому, що в склад синтаксичного аналізатора входить множина рекурсивних процедур граматичного розбору, по одній для кожного правила граматики.

Визначимо назви процедур, що відповідають нетерміналам граматики таким чином:

// program ::= "program", {"var", variable\_declaration, ";"}, "start", {statement, ";"}, "finish";

void program();

// programBody ::= { statement, ";" };

void programBody();

// variable\_declaration ::= "int16\_t", variable\_list;

void variableDeclaration();

// variable\_list ::= identifier, {",", identifier};

void variableList();

// statement ::= input\_statement | output\_statement | assign\_statement | if\_else\_statement | goto\_statement | label\_point | for\_statement | while\_statement | repeat\_until\_statement | compound\_statement;

void statement();

// input\_statement ::= "get", identifier;

void inputStatement();

// output\_statement ::= "put", arithmetic\_expression;

void outputStatement();

// arithmetic\_expression :: = low\_priority\_expression{ low\_priority\_operator, low\_priority\_expression };

void arithmeticExpression();

// low\_priority\_expression :: = middle\_priority\_expression{ middle\_priority\_operator, middle\_priority\_expression };

void lowPriorityExpression();

// low\_priority\_operator :: = "+" | "-";

//void lowPriorityOperator();

// middle\_priority\_expression :: = identifier | number | "(", arithmetic\_expression, ")";

void middlePriorityExpression();

// middle\_priority\_operator :: = "\*" | "/" | "%";

//void middlePriorityOperator();

// assign\_statement :: = arithmetic\_expression, "<-", identifier;

void assignStatement();

// if\_else\_statement :: = "if", "(", logical\_expression, ")", statement, [";", "ELSE", statement];

void ifStatement();

// logical\_expression :: = and\_expression{ or\_operator, and\_expression };

void logicalExpression();

// or\_operator :: = "or";

//void orOperator();

// and\_expression :: = comparison{ and\_operator, and\_expression };

void andExpression();

// and\_operator :: = "and";

// void andOperator();

// comparison :: = comparison\_expression | [not\_operator] "(", logical\_expression, ")";

void comparison();

// not\_operator :: = "!";

// void notOperator();

// comparison\_expression :: = arithmetic\_expression comparison\_operator arithmetic\_expression;

void comparisonExpression();

// comparison\_operator :: = "eg" | "ne" | "<=" | ">=";

// void comparisonOperator();

// goto\_statement :: = "goto", identifier;

void gotoStatement();

// label\_point :: = identifier, ":";

void labelPoint();

// for\_to\_statement ::= "for", assign\_statement, "to" | "downto", arithmetic\_expression, "do", statement;

void forStatement();

// while\_statement :: = "while", logical\_expression, { statement, ";" }, "wend";

void whileStatement();

// repeat\_until\_statement :: = "repeat", { statement, ";" }, "until", "(", logical\_expression, ")";

void repeatStatement();

// compoundStatement :: = "start", { statement, ";" }, "finish";

void compoundStatement();

Блок-схема алгоритму роботи синтаксичного аналізатора виглядатиме наступним чином:

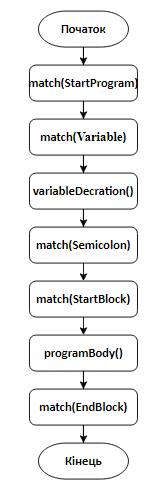


Рис. 3.3. Блок-сема алгоритму роботи синтаксичного аналізатора.

Синтаксичний аналізатор буде читати лексеми з таблиці лексем і аналізувати їх. Нам знадобиться допоміжна функція

void match(enum TypeOfToken expectedType);

яка перевіряє, чи поточна лексема збігається з очікуваною.

Блок-схема алгоритму функції variable\_declaration(), яка перевіряє чи правильно описані змінні виглядає наступним чином:

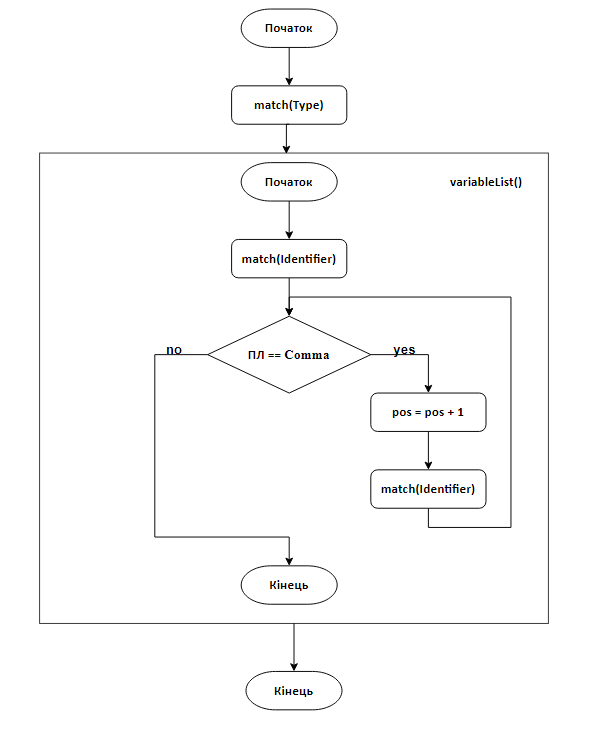


Рис. 3.4. Блок-сема алгоритму роботи функції variableDeclaration().

Блок-схема алгоритму функції programBody(), яка перевіряє чи правильно написані оператори вхідної мови програмування зображена на рисунку 3.5.

У наведених блок-схемах використано позначення ПЛ, яке позначає поточну лексему та НЛ, що позначає наступну лексему. У процесі перегляду таблиці лексем, після аналізу поточної лексеми, необхідно переміщатися на наступну лексему.

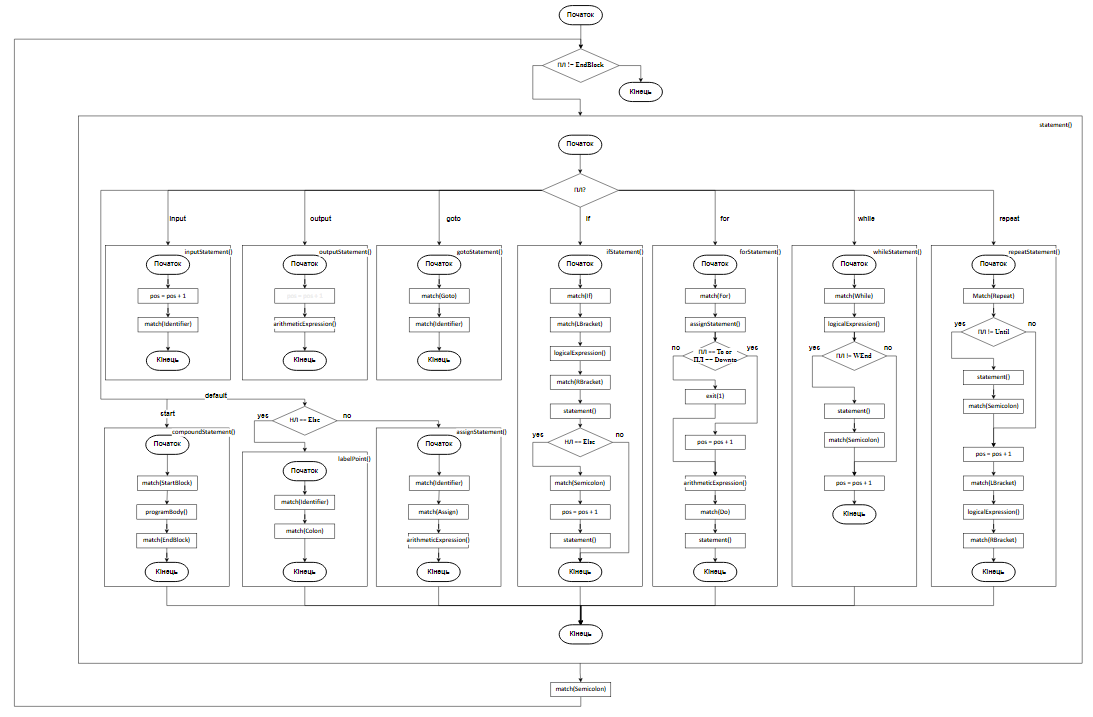


Рис. 3.5. Блок-сема алгоритму роботи функції programBody().

На етапі семантичного аналізу нам необхідно вирішити задачу ідентифікації ідентифікаторів. Алгоритм ідентифікації складається з двох частин:

* перша частина алгоритму опрацьовує оголошення ідентифікаторів;
* друга частина алгоритму опрацьовує використання ідентифікаторів.

Опрацювання оголошення ідентифікатора. Нехай лексичний аналізатор видав чергову лексему, що є ідентифікатором. Лексичний аналізатор сформував структуру, що містить атрибути виділеної лексеми, такі як ім’я ідентифікатора, його тип і лексичний клас. Далі вся ця інформація передається семантичному аналізатору. Припустимо, що в даний момент опрацьовується оголошення ідентифікатора. Основна семантична дія в цьому випадку полягає в занесенні інформації про ідентифікатор у таблицю ідентифікаторів.

Опрацювання використання ідентифікатора. Припустимо, що уже побудовано (цілком чи частково) таблицю ідентифікаторів. Далі вся ця інформація передається фазі використання ідентифікаторів. Таким чином, відомо, що опрацьовується використання ідентифікатора. Для того, щоб одержати інформацію про тип ідентифікатора нам достатньо прочитати певне поле таблиці ідентифікаторів.

**3.4.3. Опис програми реалізації синтаксичного та семантичного аналізатора.**

Структура синтаксичного аналізатора буде такою:

*//* Вхідна таблиця лексем

extern Token\* TokenTable;

int pos = 0;

void Parser() {

program();

printf("\nThe program is syntax correct.\n");

fprintf(errorFile, "\nThe program is syntax correct.\n");

}

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

Семантичний аналіз у нашому випадку буде реалізований у функції, яка викликає функції, що отримують списки ідентифікаторів та міток:

void Semantic() {

idNum = IdIdentification(idTable, TokenTable, TokensNum);

labelNum = LabelIdentification(labelTable, TokenTable, TokensNum);

printf("\nThe program is semantic correct.\n");

printf("\n%d labels found\n", labelNum);

fprintf(errorFile, "\nThe program is semantic correct.\n");

fprintf(errorFile, "\n%d labels found\n", labelNum);

printIdentifiers(labelNum, labelTable);

fprintIdentifiers(errorFile, labelNum, labelTable);

printf("\n%d identifiers found\n", idNum);

fprintf(errorFile, "\n%d identifiers found\n", idNum);

printIdentifiers(idNum, idTable);

fprintIdentifiers(errorFile, idNum, idTable);

}  
 Семантичний аналізатор у цьому проекті виконує перевірку логічної коректності програми, зокрема перевіряє правильність використання ідентифікаторів і міток. Основна функція Semantic() відповідає за ідентифікацію всіх змінних і міток у програмі та перевірку їхнього використання.  
 Функція IdIdentification() обробляє всі оголошення змінних і додає їх до таблиці ідентифікаторів, генеруючи помилку у разі використання змінної до її оголошення. Аналогічно, функція LabelIdentification() перевіряє правильність використання міток, забезпечуючи їх наявність і правильне застосування, інакше виникають помилки. Перевірка типів даних гарантує коректність типів у виразах і операціях, забезпечуючи їхню відповідність під час присвоєння та інших операцій. Також аналізується область видимості змінних, щоб уникнути помилок доступу до неіснуючих або невидимих змінних.

3.5. Розробка генератора коду.

Генерація вихідного коду передбачає спочатку перетворення програми у якесь проміжне представлення, а тоді вже генерацію з проміжного представлення у вихідний код. У якості проміжного представлення виберемо абстрактне синтаксичне дерево.

Абстрактне синтаксичне дерево (AST) — це структура даних, яка представляє синтаксичну структуру вихідного коду програми у вигляді дерева. AST використовується в компіляторах, інтерпретаторах та інструментах статичного аналізу для обробки коду.

AST представляє тільки важливу для аналізу і виконання інформацію, ігноруючи зайві деталі (наприклад, круглі дужки чи крапки з комою). Це спрощений, але точний опис логіки програми.

Вузли дерева представляють конструкції мови програмування (оператори, вирази, змінні, функції тощо). Гілки відповідають підконструкціям або елементам цих конструкцій.

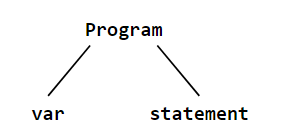
Кожен вузол відповідає певному типу конструкції коду (наприклад, оператору додавання, виклику функції, оголошенню змінної).

AST є спрощеною версією синтаксичного дерева. Воно не включає зайві вузли, що відповідають елементам, які не впливають на логіку програми (наприклад, дужки чи крапки з комою).

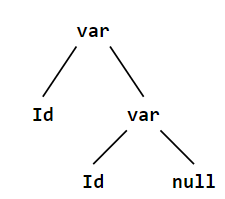
* + 1. **Розробка алгоритму роботи генератора коду.**

Будемо використовувати бінарні дерева, а отже вузол у нас має два нащадки, відповідно нарисуємо типові варіанти побудови дерева.

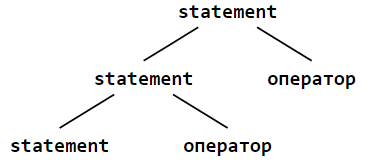
Програма має вигляд:



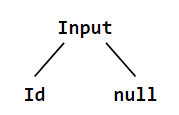
Оголошення змінних:



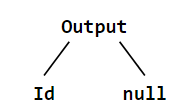
Тіло програми:



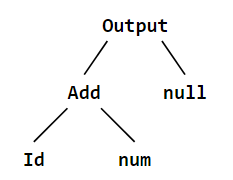
Оператор вводу:



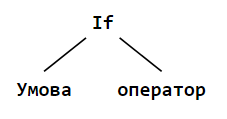
Оператор виводу:



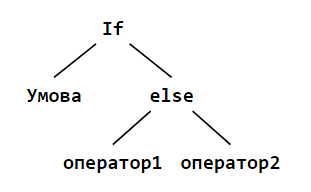
Також оператор виводу може мати за лівого нащадка різні арифметичні вирази, наприклад:



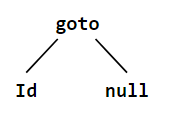
Умовний оператор (IF() оператор;):



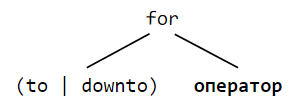
Умовний оператор (IF() оператор1; else оператор2;):



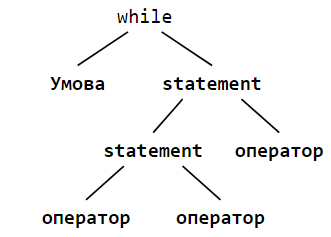
Оператор безумовного переходу:



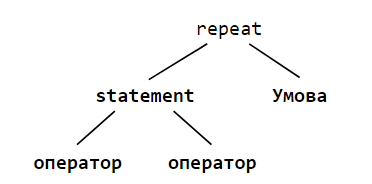
Оператор циклу for:



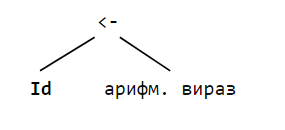
Оператор циклу while:



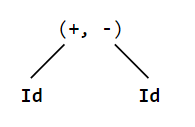
Оператор циклу repeat:



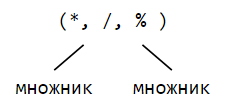
Оператор присвоєння:



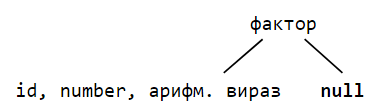
Арифметичний вираз:



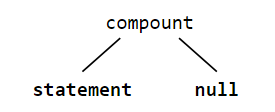
Доданок:



Множник:



Складений оператор:



Генератор коду буде обходити створене дерево і, маючи усію необхідну інформацію, генерувати вихідний код на мові програмування С у текстовий файл. Кожен вузол у дереві буде позначати якусь конструкцію, для якої генерується певний код на мові програмування С. Опрацювання кожного з вузлів дерева передбачає рекурсивний виклик функції генерування коду для лівого і правого нащадків.

Блок-схема алгоритму роботи генератора коду зображена на рисунку 3.6.

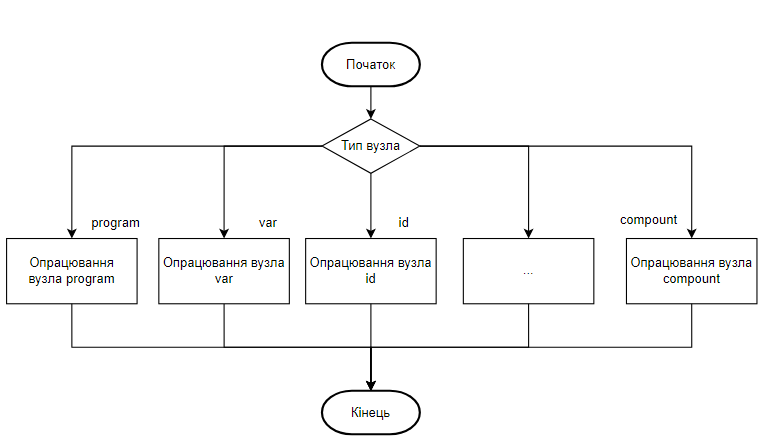


Рис. 3.6. Блок-сема алгоритму роботи генератора коду.

Розглянемо на прикладі вузла program детальніше алгоритм обходу дерева, який зображено на рисунку 3.7. Вузол позначає програму, зліва будемо зберігати інформацію про оголошені змінні, справа про оператори програми. Опрацювання вузла полягає у друці у файл необхідних шаблонів на мові програмування С, а також рекурсивного виклику для опрацювання лівого і правого нащадків. Лівий нащадок - оголошення змінних (вузол var), правий - тіло програми (вузол statement).

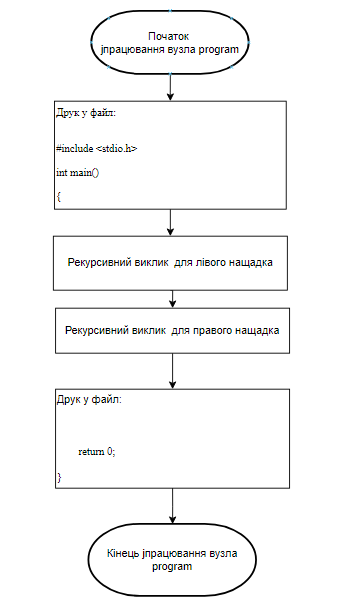


Рис. 3.7. Блок-сема алгоритму опрацювання вузла program.

* + 1. **Опис програми реалізації генератора коду.**

Генерувати вихідний код будемо з абстрактного синтаксичного дерева.

Створимо таку структуру даних для зберігання вузлів дерева:

enum TypeOfNode {

program\_node,

id\_node,

var\_node,

statement\_node,

input\_node,

output\_node,

add\_node,

sub\_node,

mul\_node,

div\_node,

mod\_node,

number\_node,

assign\_node,

if\_node,

else\_node,

or\_node,

and\_node,

not\_node,

eq\_node,

neq\_node,

gr\_node,

ls\_node,

goto\_node,

label\_node,

for\_node,

to\_node,

downto\_node,

while\_node,

repeat\_node,

compound\_node

};

struct astNode {

enum TypeOfNode type;

char name[16];

struct astNode\* left;

struct astNode\* right;

};

Функція створення вузла дерева:

// функція створення вузла AST

struct astNode\* createNode(enum TypeOfNode type, const char\* name, struct astNode\* left, struct astNode\* right);

Функція створення абстрактного синтаксичного дерева реалізована методом рекурсивного спуску:

// функція синтаксичного аналізу і створення абстрактного синтаксичного дерева

struct astNode\* astParser() {

pos = 0;

struct astNode\* tree = program();

printf("AST created.\n");

return tree;

}

Також напишемо функції для друку абстрактного синтаксичного дерева:

*//* функція для друку AST у вигляді дерева на екран

void printAST(struct astNode\* node, int level)

// функція для друку AST у вигляді дерева у файл

void fPrintAST(FILE\* outFile, struct astNode\* node, int level);

Напишемо рекурсивну функцію, яка буде генерувати код на мові програмування С з абстрактного синтаксичного дерева:

// Рекурсивна функція для генерації коду з AST

void codegen(FILE\* outFile, struct astNode\* node){

if (node == NULL) return;

switch (node->type) {

case ...

case ...

default: {

exit(1);

printf("Undescribed node type: %d\n", node->type);

break;

}

}

Тепер розглянемо варіанти генерації коду для можливих вузлів дерева.

Отже опрацювання вузла program\_node буде виглядати таким чином:

{

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

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

codegen(outFile, node->left); // for declaration

fprintf(outFile, "\n");

codegen(outFile, node->right); // for statements

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

break;

}

При генерації вихідного коду для блоку оголошення змінних будемо опрацьовувати вузли var\_node:

{

fprintf(outFile, "int ");

codegen(outFile, node->left);

fprintf(outFile, ";\n");

codegen(outFile, node->right);

break;

}

Опрацювання вузлів id\_node і number\_node буде полягати у друці імені вузла (ім’я ідентифікатора):

{

fprintf(outFile, "%s", node->name);

break;

}

При генерації вихідного коду для блоку тіло програми будемо опрацьовувати вузли statement\_node:

{

codegen(outFile, node->left);

codegen(outFile, node->right);

break;

}

Опрацювання вузла input\_node буде виглядати таким чином:

{

fprintf(outFile, "printf(\"Enter ");

codegen(outFile, node->left);

fprintf(outFile, ": \");\n");

fprintf(outFile, "scanf(\"%%d\", &");

codegen(outFile, node->left);

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

break;

}

А вузла output\_node буде виглядати таким чином:

{

fprintf(outFile, "printf(\"%%d\\n\", ");

codegen(outFile, node->left);

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

break;

}

Опрацювання вузла if\_node буде виглядати таким чином:

{

fprintf(outFile, "if (");

codegen(outFile, node->left);

fprintf(outFile, ") ");

codegen (outFile, node->right);

break;

}

Опрацювання вузла else\_node буде виглядати таким чином:

{

codegen(outFile, node->left);

fprintf(outFile, "else ");

codegen(outFile, node->right);

break;

}

Отже опрацювання вузла assign\_node буде виглядати таким чином:

{

codegen(outFile, node->left);

fprintf(outFile, " = ");

codegen(outFile, node->right);

fprintf(outFile, ";\n");

break;

}

Отже опрацювання вузла not\_node буде виглядати таким чином:

{

fprintf(outFile, "!(");

codegen(outFile, node->left);

fprintf(outFile, ")");

break;

}

Опрацювання вузлів or\_node, and\_node, eq\_node, neq\_node, gr\_node, ls\_node, а також вузлів add\_node, sub\_node, mul\_node, div\_node, mod\_node буде полягати у друці знаку операції і круглих дужок справа і зліва від знаку операції.

{

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, <необхідний знак операції>);

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

Опрацювання вузла for\_node буде виглядати таким чином:

{

fprintf(outFile, "for(\n");

codegen(outFile, node->left);

fprintf(outFile, "\n) ");

codegen(outFile, node->right);

break;

}

Опрацювання вузла to\_node буде виглядати таким чином:

{

codegen(outFile, node->left);

codegen(outFile, node->left->left);

fprintf(outFile, " <= ");

codegen(outFile, node->right);

fprintf(outFile, ";\n++");

codegen(outFile, node->left->left);

break;

}

Опрацювання вузла downto\_node буде виглядати таким чином:

{

codegen(outFile, node->left);

codegen(outFile, node->left->left);

fprintf(outFile, " >= ");

codegen(outFile, node->right);

fprintf(outFile, ";\n--");

codegen(outFile, node->left->left);

break;

}

Опрацювання вузла while\_node буде виглядати таким чином:

{

fprintf(outFile, "while(");

codegen(outFile, node->left);

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

codegen(outFile, node->right);

fprintf(outFile, "}\n");

break;

}

Опрацювання вузла repeat\_node буде виглядати таким чином:

{

fprintf(outFile, "do {\n");

codegen(outFile, node->left);

fprintf(outFile, "} while(");

codegen(outFile, node->right);

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

break;

}

Опрацювання вузла compound\_node буде виглядати таким чином:

{

fprintf(outFile, "{\n");

codegen(outFile, node->left);

codegen(outFile, node->right);

fprintf(outFile, "}\n");

break;

}

4. НАЛАГОДЖЕННЯ ТА ТЕСТУВАННЯ РОЗРОБЛЕНОГО

ТРАНСЛЯТОРА

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

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

* 1. Опис інтерфейсу та інструкції користувачу.

Розроблений транслятор має простий консольний інтерфейс.

При запуску програми необхідно ввести ім’я файлу з текстом програми на вхідній мові програмування:

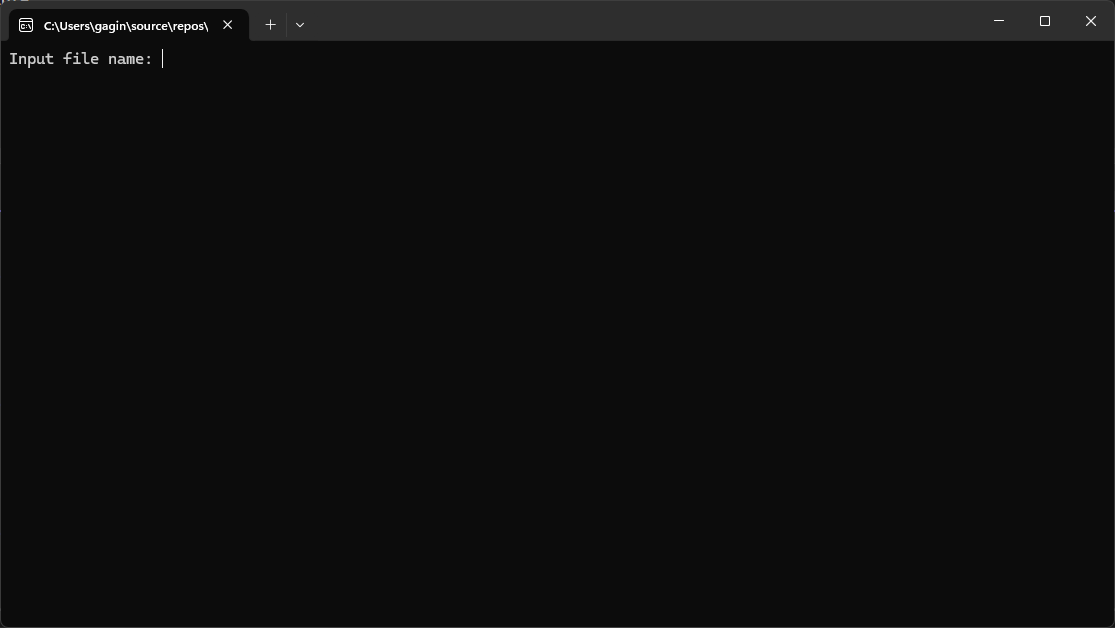
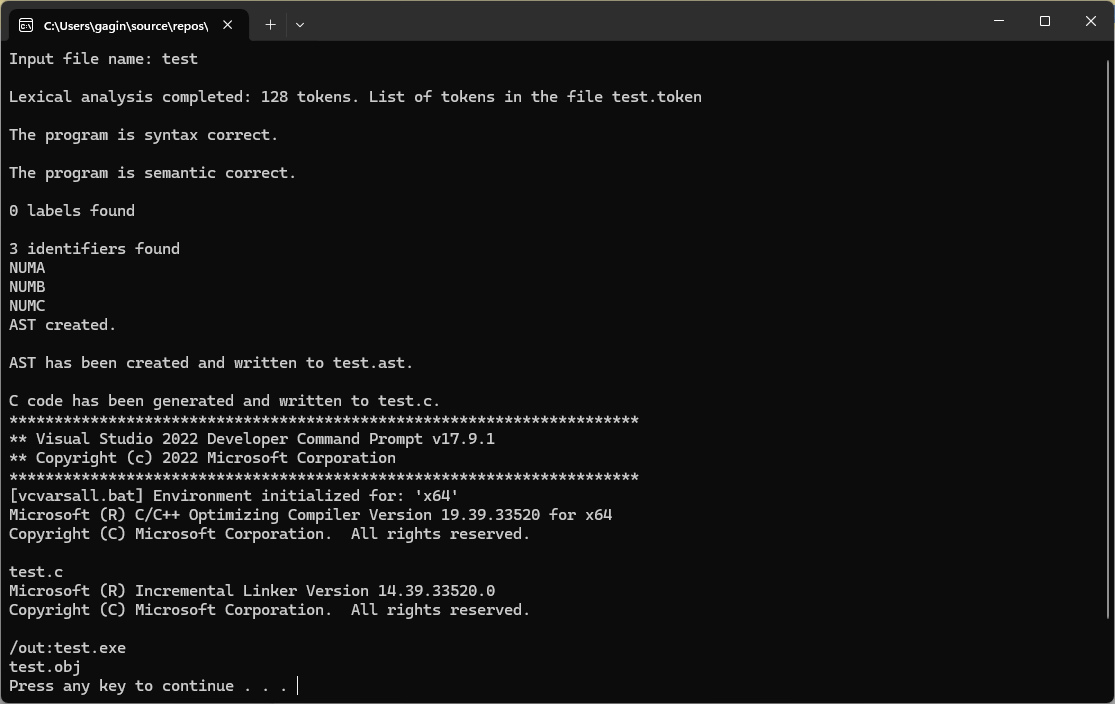


Рис. 4.1. Інтерфейс розробленого транслятора.

Після вводу імені програми отримуємо результати роботи транслятора:

Рис. 4.2.Результати роботи розробленого транслятора.

* 1. Виявлення лексичних і синтаксичних помилок.

Помилки у вхідній програмі виявляються на етапі синтаксичного і семантичного аналізу.

**Нехай, у програмі змінна X матиме некоректну назву:**

Лінійний алгоритм

program CALC

var

int16\_t NUMA, NUMB, X, RESY;

start

get NUMA;

get NUMB;

put NUMA + NUMB;

put NUMA - NUMB;

put NUMA \* NUMB;

put NUMA / NUMB;

put NUMA % NUMB;

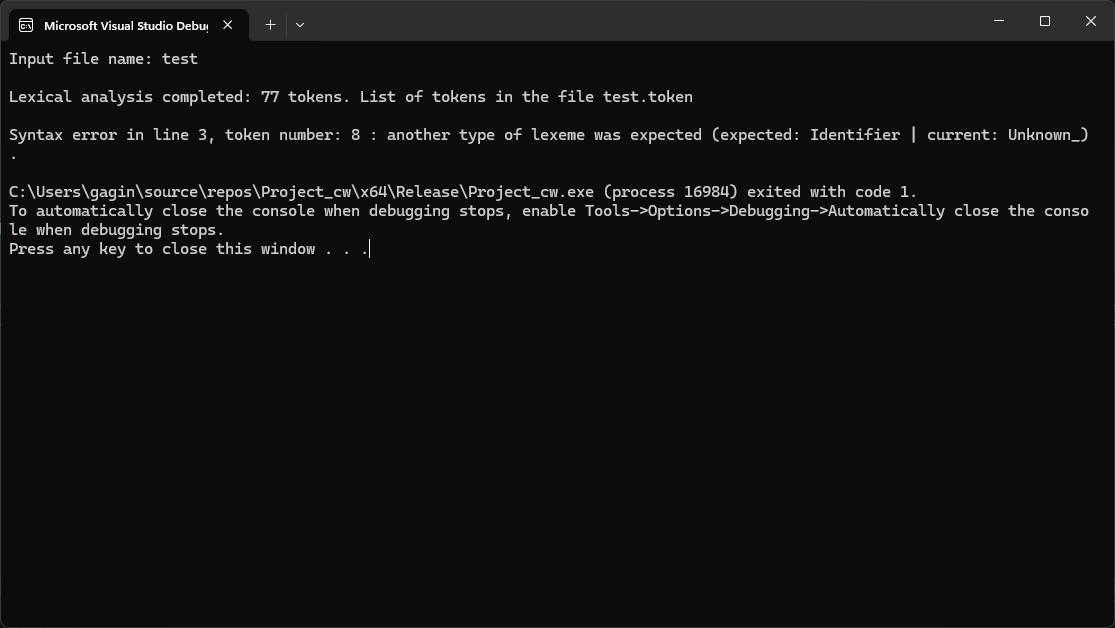
X <- (NUMA - NUMB) \* 10 + (NUMA + NUMB) / 10;

RESY <- RESX + RESX % 10;

put X;

put RESY;

finish



*Рис. 4.3. Вивід інформації про синтаксичну помилку.*

**Зробимо семантичну помилку, а саме не оголосимо змінну X:**

Лінійний алгоритм

program CALC

var

int16\_t NUMA, NUMB, RESY;

start

get NUMA;

get NUMB;

put NUMA + NUMB;

put NUMA - NUMB;

put NUMA \* NUMB;

put NUMA / NUMB;

put NUMA % NUMB;

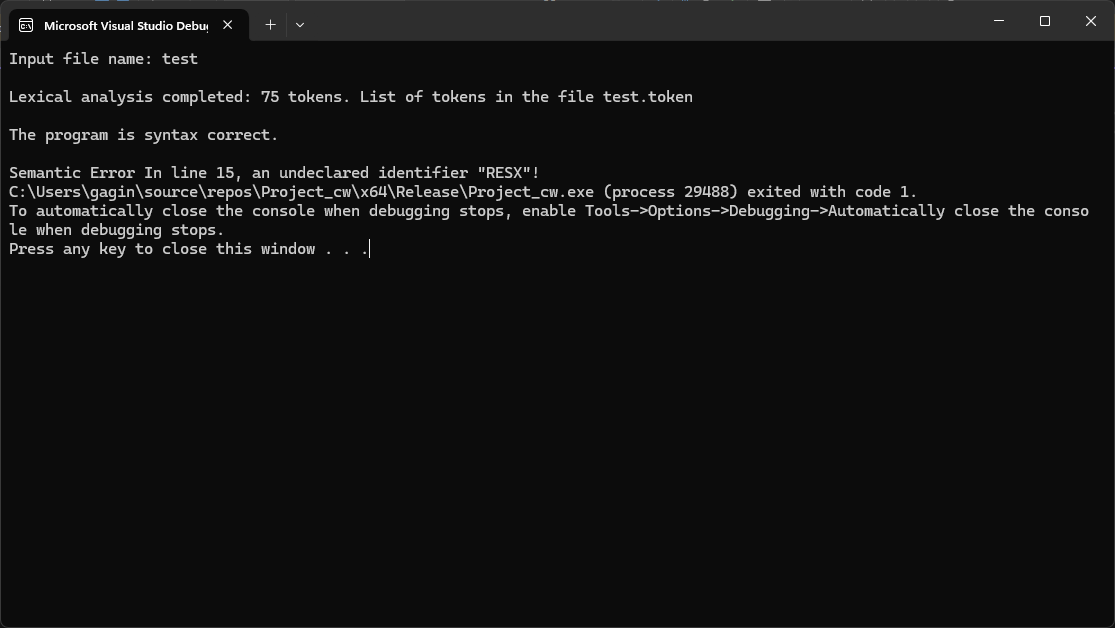
RESX <- (NUMA - NUMB) \* 10 + (NUMA + NUMB) / 10;

RESY <- RESX + RESX % 10;

put RESX;

put RESY;

finish



*Рис. 4.4. Вивід інформації про семантичну помилку.*

* 1. Перевірка роботи транслятора за допомогою тестових задач.

Тестова програма «*Лінійний алгоритм*»

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

2. Вивести на екран:

А + В (результат операції додавання);

А - В (результат операції віднімання);

А \* В (результат операції множення);

А / В (результат операції ділення);

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

3. Обрахувати значення виразів

Х = (А - В) \* 10 + (А + В) / 10

Y = Х + Х % 10

1. Вивести значення Х і У на екран.

Напишемо програму на вхідній мові програмування:

Лінійний алгоритм

program CALC

var

int16\_t NUMA, NUMB, RESX, RESY;

start

get NUMA;

get NUMB;

put NUMA + NUMB;

put NUMA - NUMB;

put NUMA \* NUMB;

put NUMA / NUMB;

put NUMA % NUMB;

RESX <- (NUMA - NUMB) \* 10 + (NUMA + NUMB) / 10;

RESY <- RESX + RESX % 10;

put RESX;

put RESY;

finish

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

int main() {

int RESY;

int RESX;

int NUMB;

int NUMA;

printf("Enter NUMA: ");

scanf("%d", &NUMA);

printf("Enter NUMB: ");

scanf("%d", &NUMB);

printf("%d\n", (NUMA + NUMB));

printf("%d\n", (NUMA - NUMB));

printf("%d\n", (NUMA \* NUMB));

printf("%d\n", (NUMA / NUMB));

printf("%d\n", (NUMA % NUMB));

RESX = (((NUMA - NUMB) \* 10) + ((NUMA + NUMB) / 10));

RESY = (RESX + (RESX % 10));

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

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

return 0;

}

Також на виході маємо виконуваний файл test1.exe, запустивши який маємо:

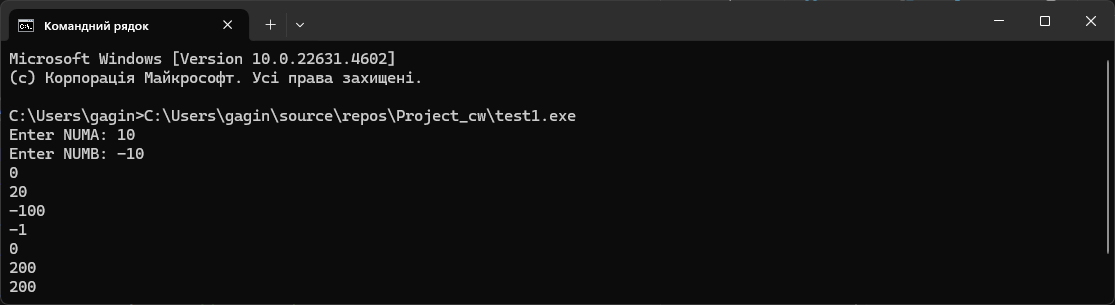


Рис. 4.5. Результати виконання test1.exe.

**Тестова програма «Алгоритм з розгалуженням»**

1. Ввести три числа А, В, С (імена змінних можуть бути іншими і мають відповідати правилам запису ідентифікаторів згідно індивідуального завдання). Використання вкладеного умовного оператора:
2. Знайти найбільше з них і вивести його на екран. Використання простого умовного оператора:
3. Вивести на екран число 1, якщо усі числа однакові (логічний вираз в умовному операторі має виглядати так: «(А=В) і (А=С) і (В=С)»), інакше вивести 0.
4. Вивести на екран число -1, якщо хоча б одне з чисел від’ємне (логічний вираз в умовному операторі має виглядати так: «(А<0) або (В<0) або (С<0)»), інакше вивести 0.
5. Вивести на екран число 10, якщо число А більше за суму чисел В і С (логічний вираз в умовному операторі має виглядати так: «!(А<(В+С))»), інакше вивести 0

Напишемо програму на вхідній мові програмування:

Алгоритм з розгалуженням

program MAXC

var

int16\_t VALA, VALB, VALC;

start

get VALA;

get VALB;

get VALC;

if (VALA >= VALB)

if (VALA >= VALC)

put VALA;

else

put VALC;

else

if (VALB >= VALC)

put VALB;

else

put VALC;

if ((VALA eg VALB) and (VALA eg VALC) and (VALB eg VALC))

put 1;

else

put 0;

if ((VALA <= 0) or (VALB <= 0) or (VALC <= 0))

put -1;

else

put 0;

if (!(VALA <= (VALB + VALC)))

put 10;

else

put 0;

finish

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

int main() {

int VALC;

int VALB;

int VALA;

printf("Enter VALA: ");

scanf("%d", &VALA);

printf("Enter VALB: ");

scanf("%d", &VALB);

printf("Enter VALC: ");

scanf("%d", &VALC);

if ((VALA > VALB)) if ((VALA > VALC)) printf("%d\n", VALA);

else printf("%d\n", VALC);

else if ((VALB > VALC)) printf("%d\n", VALB);

else printf("%d\n", VALC);

if (((VALA == VALB) && ((VALA == VALC) && (VALB == VALC)))) printf("%d\n", 1);

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

if (((VALA < 0) || ((VALB < 0) || (VALC < 0)))) printf("%d\n", -1);

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

if (!((VALA < (VALB + VALC)))) printf("%d\n", 10);

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

return 0;

}

Також на виході маємо виконуваний файл test2.exe, запустивши який маємо:

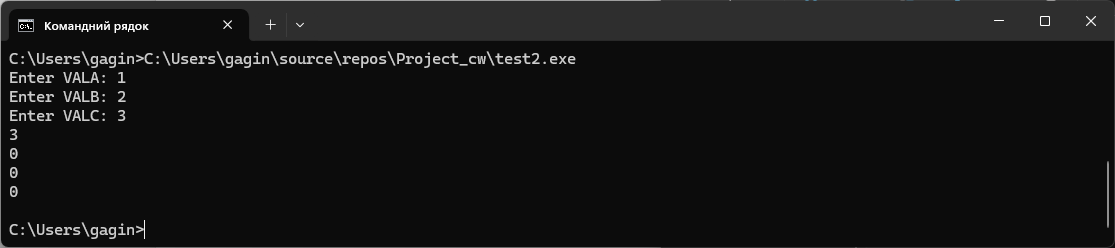


Рис. 4.6. Результати виконання test2.exe.

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

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

Використання простого оператора циклу:

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

Використання вкладеного оператора циклу:

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

Х = 0

Цикл від 1 до А з кроком 1

Цикл від 1 до В з кроком 1

Х = Х + 1

4. Вивести значення Х на екран.

Напишемо програму на вхідній мові програмування:

Тестова програма «Циклічний алгоритм»

program SQUA

var

int16\_t AVAL, BVAL, XVAL, IVAL, JVAL;

start

get AVAL;

get BVAL;

for XVAL <- AVAL to BVAL do

put XVAL \* XVAL;

XVAL <- 0;

for IVAL <- 1 to AVAL do

for JVAL <- 1 to BVAL do

XVAL <- XVAL + 1;

put XVAL;

finish

В результаті отримаємо файл з програмою на мові програмування С:

#include <stdio.h>

int main() {

int JVAL;

int IVAL;

int XVAL;

int BVAL;

int AVAL;

printf("Enter AVAL: ");

scanf("%d", &AVAL);

printf("Enter BVAL: ");

scanf("%d", &BVAL);

for(

XVAL = AVAL;

XVAL <= BVAL;

++XVAL

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

XVAL = 0;

for(

IVAL = 1;

IVAL <= AVAL;

++IVAL

) for(

JVAL = 1;

JVAL <= BVAL;

++JVAL

) XVAL = (XVAL + 1);

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

return 0;

}

Також на виході маємо виконуваний файл test3.exe, запустивши який маємо:

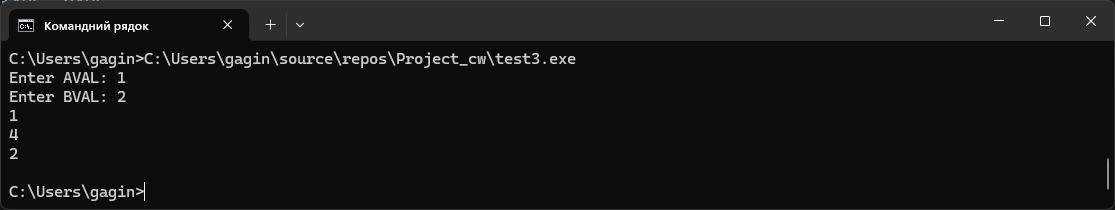


Рис. 4.7. Результати виконання test3.exe.

**ВИСНОВОК**

У рамках курсового проєкту було розроблено транслятор для вхідної мови програмування, який реалізує основні етапи трансляції. Першим етапом є **лексичний аналіз**, який виконується за допомогою спеціального алгоритму, що розділяє текст програми на окремі лексеми. Цей аналіз включає формування таблиці, що містить інформацію про типи лексем, їх значення та номери рядків. Лексичний аналізатор побудований на основі скінченого автомату, здатного розпізнавати ключові слова, ідентифікатори, константи, оператори та роздільники.

1. **Лексичний аналіз:**

Алгоритм лексичного аналізу здійснює розподіл тексту програми на лексеми, створюючи таблицю, що містить інформацію про їх типи, значення та номер рядка.

Лексичний аналізатор базується на скінченому автоматі, який здатен розпізнавати ключові слова, ідентифікатори, константи, оператори та роздільники.

1. **Синтаксичний та семантичний аналіз:**

Синтаксичний аналізатор перевіряє правильність структури програми за граматикою, будує дерево розбору та формує таблиці ідентифікаторів і типів.

Семантичний аналізатор гарантує логічну коректність програми, перевіряючи типи даних, області видимості змінних і правильність викликів функцій.

1. **Генерація коду:**

Генератор коду перетворює абстрактне синтаксичне дерево у вихідний код на мові С, обробляючи дерево розбору і генеруючи відповідний код для кожного елемента.

1. **Тестування:**

Транслятор був протестований на програмах різної структури (лінійних, з розгалуженнями, циклічних), що дозволило виявити й усунути лексичні, синтаксичні та семантичні помилки.

СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ

* Основи проектування трансляторів: Конспект лекцій : [Електронний ресурс] : навч. посіб. для студ. спеціальності 123 - «Комп’ютерна інженерія» / О. І. Марченко ; КПІ ім. Ігоря Сікорського. - Київ: КПІ ім. Ігоря Сікорського, 2021. - 108 с.
* Формальні мови, граматики та автомати: Навчальний посібник / Гавриленко С.Ю. - Харків: НТУ «ХПІ», 2021. - 133 с.
* Сопронюк Т.М. Системне програмування. Частина І. Елементи теорії формальних мов: Навчальний посібник у двох частинах. - Чернівці: ЧНУ, 2008. - 84 c.
* Сопронюк Т.М. Системне програмування. Частина ІІ. Елементи теорії компіляції: Навчальний посібник у двох частинах. - Чернівці: ЧНУ, 2008. - 84 c.
* Alfred V. Aho, Monica S. Lam, Ravi Seth, Jeffrey D. Ullma. Compilers, principles, techniques, and tools, Second Edition, New York, 2007. - 1038 c.
* Системне програмування (курсовий проект) [Електронний ресурс] - Режим доступу до ресурсу: <https://vns.lpnu.ua/course/view.php?id=11685>.
* MIT OpenCourseWare. Computer Language Engineering [Електронний ресурс] - Режим доступу до ресурсу: [https://ocw.mit.edu/courses/6-035-computer- language-engineering-spring-2010](https://ocw.mit.edu/courses/6-035-computer-language-engineering-spring-2010)

**ДОДАТКИ**

А. Таблиці лексем для тестових прикладів

Тестова програма «Лінійний алгоритм»

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

| TOKEN TABLE |

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

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

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

| 1 | program | 0 | 0 | StartProgram |

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

| 1 | CALC | 0 | 21 | Identifier |

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

| 2 | var | 0 | 2 | Variable |

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

| 3 | int16\_t | 0 | 3 | Type |

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

| 3 | NUMA | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | NUMB | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | RESX | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | RESY | 0 | 21 | Identifier |

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

| 3 | ; | 0 | 39 | Semicolon |

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

| 4 | start | 0 | 1 | StartBlock |

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

| 6 | get | 0 | 5 | Input |

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

| 6 | NUMA | 0 | 21 | Identifier |

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

| 6 | ; | 0 | 39 | Semicolon |

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

| 7 | get | 0 | 5 | Input |

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

| 7 | NUMB | 0 | 21 | Identifier |

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

| 7 | ; | 0 | 39 | Semicolon |

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

| 9 | put | 0 | 6 | Output |

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

| 9 | NUMA | 0 | 21 | Identifier |

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

| 9 | + | 0 | 25 | Add |

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

| 9 | NUMB | 0 | 21 | Identifier |

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

| 9 | ; | 0 | 39 | Semicolon |

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

| 10 | put | 0 | 6 | Output |

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

| 10 | NUMA | 0 | 21 | Identifier |

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

| 10 | - | 0 | 26 | Sub |

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

| 10 | NUMB | 0 | 21 | Identifier |

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

| 10 | ; | 0 | 39 | Semicolon |

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

| 11 | put | 0 | 6 | Output |

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

| 11 | NUMA | 0 | 21 | Identifier |

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

| 11 | \* | 0 | 27 | Mul |

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

| 11 | NUMB | 0 | 21 | Identifier |

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

| 11 | ; | 0 | 39 | Semicolon |

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

| 12 | put | 0 | 6 | Output |

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

| 12 | NUMA | 0 | 21 | Identifier |

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

| 12 | / | 0 | 28 | Div |

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

| 12 | NUMB | 0 | 21 | Identifier |

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

| 12 | ; | 0 | 39 | Semicolon |

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

| 13 | put | 0 | 6 | Output |

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

| 13 | NUMA | 0 | 21 | Identifier |

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

| 13 | % | 0 | 29 | Mod |

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

| 13 | NUMB | 0 | 21 | Identifier |

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

| 13 | ; | 0 | 39 | Semicolon |

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

| 15 | RESX | 0 | 21 | Identifier |

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

| 15 | <- | 0 | 24 | Assign |

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

| 15 | ( | 0 | 37 | LBracket |

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

| 15 | NUMA | 0 | 21 | Identifier |

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

| 15 | - | 0 | 26 | Sub |

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

| 15 | NUMB | 0 | 21 | Identifier |

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

| 15 | ) | 0 | 38 | RBracket |

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

| 15 | \* | 0 | 27 | Mul |

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

| 15 | 10 | 10 | 22 | Number |

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

| 15 | + | 0 | 25 | Add |

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

| 15 | ( | 0 | 37 | LBracket |

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

| 15 | NUMA | 0 | 21 | Identifier |

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

| 15 | + | 0 | 25 | Add |

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

| 15 | NUMB | 0 | 21 | Identifier |

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

| 15 | ) | 0 | 38 | RBracket |

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

| 15 | / | 0 | 28 | Div |

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

| 15 | 10 | 10 | 22 | Number |

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

| 15 | ; | 0 | 39 | Semicolon |

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

| 16 | RESY | 0 | 21 | Identifier |

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

| 16 | <- | 0 | 24 | Assign |

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

| 16 | RESX | 0 | 21 | Identifier |

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

| 16 | + | 0 | 25 | Add |

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

| 16 | RESX | 0 | 21 | Identifier |

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

| 16 | % | 0 | 29 | Mod |

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

| 16 | 10 | 10 | 22 | Number |

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

| 16 | ; | 0 | 39 | Semicolon |

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

| 18 | put | 0 | 6 | Output |

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

| 18 | RESX | 0 | 21 | Identifier |

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

| 18 | ; | 0 | 39 | Semicolon |

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

| 19 | put | 0 | 6 | Output |

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

| 19 | RESY | 0 | 21 | Identifier |

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

| 19 | ; | 0 | 39 | Semicolon |

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

| 20 | finish | 0 | 4 | EndBlock |

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

Тестова програма «Алгоритм з розгалуженням»

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

| TOKEN TABLE |

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

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

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

| 1 | program | 0 | 0 | StartProgram |

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

| 1 | MAXC | 0 | 21 | Identifier |

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

| 2 | var | 0 | 2 | Variable |

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

| 3 | int16\_t | 0 | 3 | Type |

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

| 3 | VALA | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | VALB | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | VALC | 0 | 21 | Identifier |

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

| 3 | ; | 0 | 39 | Semicolon |

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

| 4 | start | 0 | 1 | StartBlock |

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

| 5 | get | 0 | 5 | Input |

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

| 5 | VALA | 0 | 21 | Identifier |

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

| 5 | ; | 0 | 39 | Semicolon |

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

| 6 | get | 0 | 5 | Input |

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

| 6 | VALB | 0 | 21 | Identifier |

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

| 6 | ; | 0 | 39 | Semicolon |

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

| 7 | get | 0 | 5 | Input |

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

| 7 | VALC | 0 | 21 | Identifier |

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

| 7 | ; | 0 | 39 | Semicolon |

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

| 9 | if | 0 | 7 | If |

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

| 9 | ( | 0 | 37 | LBracket |

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

| 9 | VALA | 0 | 21 | Identifier |

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

| 9 | >= | 0 | 32 | Greate |

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

| 9 | VALB | 0 | 21 | Identifier |

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

| 9 | ) | 0 | 38 | RBracket |

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

| 10 | if | 0 | 7 | If |

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

| 10 | ( | 0 | 37 | LBracket |

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

| 10 | VALA | 0 | 21 | Identifier |

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

| 10 | >= | 0 | 32 | Greate |

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

| 10 | VALC | 0 | 21 | Identifier |

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

| 10 | ) | 0 | 38 | RBracket |

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

| 11 | put | 0 | 6 | Output |

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

| 11 | VALA | 0 | 21 | Identifier |

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

| 11 | ; | 0 | 39 | Semicolon |

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

| 12 | else | 0 | 9 | Else |

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

| 13 | put | 0 | 6 | Output |

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

| 13 | VALC | 0 | 21 | Identifier |

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

| 13 | ; | 0 | 39 | Semicolon |

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

| 14 | else | 0 | 9 | Else |

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

| 15 | if | 0 | 7 | If |

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

| 15 | ( | 0 | 37 | LBracket |

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

| 15 | VALB | 0 | 21 | Identifier |

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

| 15 | >= | 0 | 32 | Greate |

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

| 15 | VALC | 0 | 21 | Identifier |

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

| 15 | ) | 0 | 38 | RBracket |

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

| 16 | put | 0 | 6 | Output |

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

| 16 | VALB | 0 | 21 | Identifier |

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

| 16 | ; | 0 | 39 | Semicolon |

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

| 17 | else | 0 | 9 | Else |

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

| 18 | put | 0 | 6 | Output |

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

| 18 | VALC | 0 | 21 | Identifier |

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

| 18 | ; | 0 | 39 | Semicolon |

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

| 20 | if | 0 | 7 | If |

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

| 20 | ( | 0 | 37 | LBracket |

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

| 20 | ( | 0 | 37 | LBracket |

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

| 20 | VALA | 0 | 21 | Identifier |

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

| 20 | eg | 0 | 30 | Equality |

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

| 20 | VALB | 0 | 21 | Identifier |

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

| 20 | ) | 0 | 38 | RBracket |

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

| 20 | and | 0 | 35 | And |

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

| 20 | ( | 0 | 37 | LBracket |

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

| 20 | VALA | 0 | 21 | Identifier |

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

| 20 | eg | 0 | 30 | Equality |

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

| 20 | VALC | 0 | 21 | Identifier |

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

| 20 | ) | 0 | 38 | RBracket |

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

| 20 | and | 0 | 35 | And |

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

| 20 | ( | 0 | 37 | LBracket |

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

| 20 | VALB | 0 | 21 | Identifier |

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

| 20 | eg | 0 | 30 | Equality |

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

| 20 | VALC | 0 | 21 | Identifier |

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

| 20 | ) | 0 | 38 | RBracket |

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

| 20 | ) | 0 | 38 | RBracket |

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

| 21 | put | 0 | 6 | Output |

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

| 21 | 1 | 1 | 22 | Number |

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

| 21 | ; | 0 | 39 | Semicolon |

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

| 22 | else | 0 | 9 | Else |

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

| 23 | put | 0 | 6 | Output |

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

| 23 | 0 | 0 | 22 | Number |

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

| 23 | ; | 0 | 39 | Semicolon |

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

| 25 | if | 0 | 7 | If |

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

| 25 | ( | 0 | 37 | LBracket |

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

| 25 | ( | 0 | 37 | LBracket |

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

| 25 | VALA | 0 | 21 | Identifier |

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

| 25 | <= | 0 | 33 | Less |

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

| 25 | 0 | 0 | 22 | Number |

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

| 25 | ) | 0 | 38 | RBracket |

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

| 25 | or | 0 | 36 | Or |

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

| 25 | ( | 0 | 37 | LBracket |

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

| 25 | VALB | 0 | 21 | Identifier |

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

| 25 | <= | 0 | 33 | Less |

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

| 25 | 0 | 0 | 22 | Number |

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

| 25 | ) | 0 | 38 | RBracket |

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

| 25 | or | 0 | 36 | Or |

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

| 25 | ( | 0 | 37 | LBracket |

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

| 25 | VALC | 0 | 21 | Identifier |

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

| 25 | <= | 0 | 33 | Less |

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

| 25 | 0 | 0 | 22 | Number |

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

| 25 | ) | 0 | 38 | RBracket |

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

| 25 | ) | 0 | 38 | RBracket |

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

| 26 | put | 0 | 6 | Output |

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

| 26 | -1 | -1 | 22 | Number |

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

| 26 | ; | 0 | 39 | Semicolon |

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

| 27 | else | 0 | 9 | Else |

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

| 28 | put | 0 | 6 | Output |

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

| 28 | 0 | 0 | 22 | Number |

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

| 28 | ; | 0 | 39 | Semicolon |

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

| 30 | if | 0 | 7 | If |

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

| 30 | ( | 0 | 37 | LBracket |

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

| 30 | ! | 0 | 34 | Not |

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

| 30 | ( | 0 | 37 | LBracket |

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

| 30 | VALA | 0 | 21 | Identifier |

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

| 30 | <= | 0 | 33 | Less |

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

| 30 | ( | 0 | 37 | LBracket |

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

| 30 | VALB | 0 | 21 | Identifier |

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

| 30 | + | 0 | 25 | Add |

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

| 30 | VALC | 0 | 21 | Identifier |

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

| 30 | ) | 0 | 38 | RBracket |

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

| 30 | ) | 0 | 38 | RBracket |

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

| 30 | ) | 0 | 38 | RBracket |

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

| 31 | put | 0 | 6 | Output |

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

| 31 | 10 | 10 | 22 | Number |

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

| 31 | ; | 0 | 39 | Semicolon |

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

| 32 | else | 0 | 9 | Else |

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

| 33 | put | 0 | 6 | Output |

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

| 33 | 0 | 0 | 22 | Number |

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

| 33 | ; | 0 | 39 | Semicolon |

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

| 34 | finish | 0 | 4 | EndBlock |

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

Тестова програма «Циклічний алгоритм»

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

| TOKEN TABLE |

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

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

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

| 1 | program | 0 | 0 | StartProgram |

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

| 1 | SQUA | 0 | 21 | Identifier |

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

| 2 | var | 0 | 2 | Variable |

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

| 3 | int16\_t | 0 | 3 | Type |

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

| 3 | AVAL | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | BVAL | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | XVAL | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | IVAL | 0 | 21 | Identifier |

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

| 3 | , | 0 | 41 | Comma |

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

| 3 | JVAL | 0 | 21 | Identifier |

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

| 3 | ; | 0 | 39 | Semicolon |

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

| 4 | start | 0 | 1 | StartBlock |

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

| 5 | get | 0 | 5 | Input |

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

| 5 | AVAL | 0 | 21 | Identifier |

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

| 5 | ; | 0 | 39 | Semicolon |

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

| 6 | get | 0 | 5 | Input |

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

| 6 | BVAL | 0 | 21 | Identifier |

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

| 6 | ; | 0 | 39 | Semicolon |

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

| 8 | for | 0 | 11 | For |

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

| 8 | XVAL | 0 | 21 | Identifier |

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

| 8 | <- | 0 | 24 | Assign |

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

| 8 | AVAL | 0 | 21 | Identifier |

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

| 8 | to | 0 | 12 | To |

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

| 8 | BVAL | 0 | 21 | Identifier |

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

| 8 | do | 0 | 14 | Do |

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

| 9 | put | 0 | 6 | Output |

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

| 9 | XVAL | 0 | 21 | Identifier |

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

| 9 | \* | 0 | 27 | Mul |

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

| 9 | XVAL | 0 | 21 | Identifier |

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

| 9 | ; | 0 | 39 | Semicolon |

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

| 11 | XVAL | 0 | 21 | Identifier |

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

| 11 | <- | 0 | 24 | Assign |

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

| 11 | 0 | 0 | 22 | Number |

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

| 11 | ; | 0 | 39 | Semicolon |

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

| 12 | for | 0 | 11 | For |

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

| 12 | IVAL | 0 | 21 | Identifier |

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

| 12 | <- | 0 | 24 | Assign |

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

| 12 | 1 | 1 | 22 | Number |

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

| 12 | to | 0 | 12 | To |

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

| 12 | AVAL | 0 | 21 | Identifier |

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

| 12 | do | 0 | 14 | Do |

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

| 13 | for | 0 | 11 | For |

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

| 13 | JVAL | 0 | 21 | Identifier |

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

| 13 | <- | 0 | 24 | Assign |

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

| 13 | 1 | 1 | 22 | Number |

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

| 13 | to | 0 | 12 | To |

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

| 13 | BVAL | 0 | 21 | Identifier |

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

| 13 | do | 0 | 14 | Do |

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

| 14 | XVAL | 0 | 21 | Identifier |

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

| 14 | <- | 0 | 24 | Assign |

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

| 14 | XVAL | 0 | 21 | Identifier |

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

| 14 | + | 0 | 25 | Add |

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

| 14 | 1 | 1 | 22 | Number |

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

| 14 | ; | 0 | 39 | Semicolon |

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

| 16 | put | 0 | 6 | Output |

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

| 16 | XVAL | 0 | 21 | Identifier |

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

| 16 | ; | 0 | 39 | Semicolon |

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

| 17 | finish | 0 | 4 | EndBlock |

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

Б. Абстрактне синтаксичне дерево для тестових прикладів

Тестова програма «Лінійний алгоритм»

|-- Program(0)

| |-- var(2)

| | |-- RESY(1)

| | |-- var(2)

| | | |-- RESX(1)

| | | |-- var(2)

| | | | |-- NUMB(1)

| | | | |-- var(2)

| | | | | |-- NUMA(1)

| |-- statement(3)

| | |-- statement(3)

| | | |-- statement(3)

| | | | |-- statement(3)

| | | | | |-- statement(3)

| | | | | | |-- statement(3)

| | | | | | | |-- statement(3)

| | | | | | | | |-- statement(3)

| | | | | | | | | |-- statement(3)

| | | | | | | | | | |-- statement(3)

| | | | | | | | | | | |-- input(4)

| | | | | | | | | | | | |-- NUMA(1)

| | | | | | | | | | | |-- input(4)

| | | | | | | | | | | | |-- NUMB(1)

| | | | | | | | | | |-- output(5)

| | | | | | | | | | | |-- Add(6)

| | | | | | | | | | | | |-- NUMA(1)

| | | | | | | | | | | | |-- NUMB(1)

| | | | | | | | | |-- output(5)

| | | | | | | | | | |-- Sub(7)

| | | | | | | | | | | |-- NUMA(1)

| | | | | | | | | | | |-- NUMB(1)

| | | | | | | | |-- output(5)

| | | | | | | | | |-- Mul(8)

| | | | | | | | | | |-- NUMA(1)

| | | | | | | | | | |-- NUMB(1)

| | | | | | | |-- output(5)

| | | | | | | | |-- Div(9)

| | | | | | | | | |-- NUMA(1)

| | | | | | | | | |-- NUMB(1)

| | | | | | |-- output(5)

| | | | | | | |-- Mod(10)

| | | | | | | | |-- NUMA(1)

| | | | | | | | |-- NUMB(1)

| | | | | |-- assign(12)

| | | | | | |-- RESX(1)

| | | | | | |-- Add(6)

| | | | | | | |-- Mul(8)

| | | | | | | | |-- Sub(7)

| | | | | | | | | |-- NUMA(1)

| | | | | | | | | |-- NUMB(1)

| | | | | | | | |-- 10(11)

| | | | | | | |-- Div(9)

| | | | | | | | |-- Add(6)

| | | | | | | | | |-- NUMA(1)

| | | | | | | | | |-- NUMB(1)

| | | | | | | | |-- 10(11)

| | | | |-- assign(12)

| | | | | |-- RESY(1)

| | | | | |-- Add(6)

| | | | | | |-- RESX(1)

| | | | | | |-- Mod(10)

| | | | | | | |-- RESX(1)

| | | | | | | |-- 10(11)

| | | |-- output(5)

| | | | |-- RESX(1)

| | |-- output(5)

| | | |-- RESY(1)

Тестова програма «Алгоритм з розгалуженням»

|-- Program(0)

| |-- var(2)

| | |-- VALC(1)

| | |-- var(2)

| | | |-- VALB(1)

| | | |-- var(2)

| | | | |-- VALA(1)

| |-- statement(3)

| | |-- statement(3)

| | | |-- statement(3)

| | | | |-- statement(3)

| | | | | |-- statement(3)

| | | | | | |-- statement(3)

| | | | | | | |-- input(4)

| | | | | | | | |-- VALA(1)

| | | | | | | |-- input(4)

| | | | | | | | |-- VALB(1)

| | | | | | |-- input(4)

| | | | | | | |-- VALC(1)

| | | | | |-- if(13)

| | | | | | |-- Greate(20)

| | | | | | | |-- VALA(1)

| | | | | | | |-- VALB(1)

| | | | | | |-- else(14)

| | | | | | | |-- if(13)

| | | | | | | | |-- Greate(20)

| | | | | | | | | |-- VALA(1)

| | | | | | | | | |-- VALC(1)

| | | | | | | | |-- else(14)

| | | | | | | | | |-- output(5)

| | | | | | | | | | |-- VALA(1)

| | | | | | | | | |-- output(5)

| | | | | | | | | | |-- VALC(1)

| | | | | | | |-- if(13)

| | | | | | | | |-- Greate(20)

| | | | | | | | | |-- VALB(1)

| | | | | | | | | |-- VALC(1)

| | | | | | | | |-- else(14)

| | | | | | | | | |-- output(5)

| | | | | | | | | | |-- VALB(1)

| | | | | | | | | |-- output(5)

| | | | | | | | | | |-- VALC(1)

| | | | |-- if(13)

| | | | | |-- and(16)

| | | | | | |-- Equality(18)

| | | | | | | |-- VALA(1)

| | | | | | | |-- VALB(1)

| | | | | | |-- and(16)

| | | | | | | |-- Equality(18)

| | | | | | | | |-- VALA(1)

| | | | | | | | |-- VALC(1)

| | | | | | | |-- Equality(18)

| | | | | | | | |-- VALB(1)

| | | | | | | | |-- VALC(1)

| | | | | |-- else(14)

| | | | | | |-- output(5)

| | | | | | | |-- 1(11)

| | | | | | |-- output(5)

| | | | | | | |-- 0(11)

| | | |-- if(13)

| | | | |-- or(15)

| | | | | |-- Less(21)

| | | | | | |-- VALA(1)

| | | | | | |-- 0(11)

| | | | | |-- or(15)

| | | | | | |-- Less(21)

| | | | | | | |-- VALB(1)

| | | | | | | |-- 0(11)

| | | | | | |-- Less(21)

| | | | | | | |-- VALC(1)

| | | | | | | |-- 0(11)

| | | | |-- else(14)

| | | | | |-- output(5)

| | | | | | |-- -1(11)

| | | | | |-- output(5)

| | | | | | |-- 0(11)

| | |-- if(13)

| | | |-- not(17)

| | | | |-- Less(21)

| | | | | |-- VALA(1)

| | | | | |-- Add(6)

| | | | | | |-- VALB(1)

| | | | | | |-- VALC(1)

| | | |-- else(14)

| | | | |-- output(5)

| | | | | |-- 10(11)

| | | | |-- output(5)

| | | | | |-- 0(11)

Тестова програма «Циклічний алгоритм»

|-- Program(0)

| |-- var(2)

| | |-- JVAL(1)

| | |-- var(2)

| | | |-- IVAL(1)

| | | |-- var(2)

| | | | |-- XVAL(1)

| | | | |-- var(2)

| | | | | |-- BVAL(1)

| | | | | |-- var(2)

| | | | | | |-- AVAL(1)

| |-- statement(3)

| | |-- statement(3)

| | | |-- statement(3)

| | | | |-- statement(3)

| | | | | |-- statement(3)

| | | | | | |-- input(4)

| | | | | | | |-- AVAL(1)

| | | | | | |-- input(4)

| | | | | | | |-- BVAL(1)

| | | | | |-- for(24)

| | | | | | |-- to(25)

| | | | | | | |-- assign(12)

| | | | | | | | |-- XVAL(1)

| | | | | | | | |-- AVAL(1)

| | | | | | | |-- BVAL(1)

| | | | | | |-- output(5)

| | | | | | | |-- Mul(8)

| | | | | | | | |-- XVAL(1)

| | | | | | | | |-- XVAL(1)

| | | | |-- assign(12)

| | | | | |-- XVAL(1)

| | | | | |-- 0(11)

| | | |-- for(24)

| | | | |-- to(25)

| | | | | |-- assign(12)

| | | | | | |-- IVAL(1)

| | | | | | |-- 1(11)

| | | | | |-- AVAL(1)

| | | | |-- for(24)

| | | | | |-- to(25)

| | | | | | |-- assign(12)

| | | | | | | |-- JVAL(1)

| | | | | | | |-- 1(11)

| | | | | | |-- BVAL(1)

| | | | | |-- assign(12)

| | | | | | |-- XVAL(1)

| | | | | | |-- Add(6)

| | | | | | | |-- XVAL(1)

| | | | | | | |-- 1(11)

| | |-- output(5)

| | | |-- XVAL(1)

В. С код, отриманий на виході транслятора для тестових прикладів

Тестова програма «Лінійний алгоритм»

#include <stdio.h>

int main() {

int RESY;

int RESX;

int NUMB;

int NUMA;

printf("Enter NUMA: ");

scanf("%d", &NUMA);

printf("Enter NUMB: ");

scanf("%d", &NUMB);

printf("%d\n", (NUMA + NUMB));

printf("%d\n", (NUMA - NUMB));

printf("%d\n", (NUMA \* NUMB));

printf("%d\n", (NUMA / NUMB));

printf("%d\n", (NUMA % NUMB));

RESX = (((NUMA - NUMB) \* 10) + ((NUMA + NUMB) / 10));

RESY = (RESX + (RESX % 10));

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

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

return 0;

}

Тестова програма «Алгоритм з розгалуженням»

#include <stdio.h>

int main() {

int VALC;

int VALB;

int VALA;

printf("Enter VALA: ");

scanf("%d", &VALA);

printf("Enter VALB: ");

scanf("%d", &VALB);

printf("Enter VALC: ");

scanf("%d", &VALC);

if ((VALA > VALB)) if ((VALA > VALC)) printf("%d\n", VALA);

else printf("%d\n", VALC);

else if ((VALB > VALC)) printf("%d\n", VALB);

else printf("%d\n", VALC);

if (((VALA == VALB) && ((VALA == VALC) && (VALB == VALC)))) printf("%d\n", 1);

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

if (((VALA < 0) || ((VALB < 0) || (VALC < 0)))) printf("%d\n", -1);

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

if (!((VALA < (VALB + VALC)))) printf("%d\n", 10);

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

return 0;

}

Тестова програма «Циклічний алгоритм»

#include <stdio.h>

int main() {

int JVAL;

int IVAL;

int XVAL;

int BVAL;

int AVAL;

printf("Enter AVAL: ");

scanf("%d", &AVAL);

printf("Enter BVAL: ");

scanf("%d", &BVAL);

for(

XVAL = AVAL;

XVAL <= BVAL;

++XVAL

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

XVAL = 0;

for(

IVAL = 1;

IVAL <= AVAL;

++IVAL

) for(

JVAL = 1;

JVAL <= BVAL;

++JVAL

) XVAL = (XVAL + 1);

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

return 0;

}

Г. Документований текст програмних модулів (лістинги)

Файл Ast.cpp

#include "Ast.hpp"

extern struct Token\* TokenTable;

extern int pos;

namespace AST {

void deleteNode(struct astNode\* node) {

if (node == nullptr) return;

deleteNode(node->left);

deleteNode(node->right);

free(node);

}

struct astNode\* createNode(enum TypeOfNode type, const char\* name, struct astNode\* left, struct astNode\* right) {

struct astNode\* node = (struct astNode\*)malloc(sizeof(struct astNode));

node->type = type;

strcpy\_s(node->name, name);

node->left = left;

node->right = right;

return node;

}

void printAST(struct astNode\* node, int level) {

if (node == nullptr)

return;

for (int i = 0; i < level; i++)

printf("| ");

printf("|-- %s(%d)", node->name, node->type);

printf("\n");

if (node->left || node->right)

{

printAST(node->left, level + 1);

printAST(node->right, level + 1);

}

}

void fPrintAST(FILE\* outFile, struct astNode\* node, int level) {

if (node == nullptr)

return;

for (int i = 0; i < level; i++)

fprintf(outFile, "| ");

fprintf(outFile, "|-- %s(%d)", node->name, node->type);

fprintf(outFile, "\n");

if (node->left || node->right)

{

fPrintAST(outFile, node->left, level + 1);

fPrintAST(outFile, node->right, level + 1);

}

}

void match(enum TypeOfToken expectedType) {

if (TokenTable[pos].type == expectedType)

pos++;

else {

printf("\nSyntax error in line %d : another type of lexeme was expected (expected: %s | current: %s).\n", TokenTable[pos].line, lexemeTypeName(expectedType), lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

struct astNode\* astParser() {

pos = 0;

struct astNode\* tree = program();

printf("AST created.\n");

return tree;

}

struct astNode\* program() {

match(StartProgram);

match(Identifier);

struct astNode\* declaration = nullptr;

if (TokenTable[pos].type == Variable) {

++pos; // for Variable

declaration = variableDeclaration();

match(Semicolon);

}

match(StartBlock);

struct astNode\* body = programBody();

match(EndBlock);

return createNode(program\_node, "Program", declaration, body);

}

struct astNode\* variableDeclaration() {

match(Type);

return variableList();

}

struct astNode\* variableList() {

match(Identifier);

struct astNode\* id = createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr);

struct astNode\* list = list = createNode(var\_node, "var", id, nullptr);

while (TokenTable[pos].type == Comma)

{

match(Comma);

match(Identifier);

id = createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr);

list = createNode(var\_node, "var", id, list);

}

return list;

}

struct astNode\* programBody() {

if (TokenTable[pos].type != EndBlock) {

struct astNode\* stmt = statement();

match(Semicolon);

struct astNode\* body = stmt;

while (TokenTable[pos].type != EndBlock)

{

struct astNode\* nextStmt = statement();

match(Semicolon);

body = createNode(statement\_node, "statement", body, nextStmt);

}

return body;

}

return nullptr;

}

struct astNode\* statement() {

switch (TokenTable[pos].type) {

case Input: return inputStatement();

case Output: return outputStatement();

case If: return ifStatement();

case Goto: return gotoStatement();

case For: return forStatement();

case While: return whileStatement();

case Repeat: return repeatStatement();

case StartBlock:return compoundStatement();

default: {

if (TokenTable[pos + 1].type == Colon)

return labelPoint();

else

return assignStatement();

}

}

}

struct astNode\* inputStatement() {

match(Input);

match(Identifier);

return createNode(input\_node, "input", createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr), nullptr);

}

struct astNode\* outputStatement() {

match(Output);

return createNode(output\_node, "output", arithmeticExpression(), nullptr);

}

struct astNode\* arithmeticExpression() {

struct astNode\* left = lowPriorityExpression();

if (TokenTable[pos].type == Add || TokenTable[pos].type == Sub) {

enum TypeOfToken op = TokenTable[pos].type;

++pos; // for Add or Sub

struct astNode\* right = arithmeticExpression();

return createNode(op == Add ? add\_node : sub\_node, lexemeTypeName(op), left, right);

}

return left;

}

struct astNode\* lowPriorityExpression() {

struct astNode\* left = middlePriorityExpression();

if (TokenTable[pos].type == Mul || TokenTable[pos].type == Mod || TokenTable[pos].type == Div) {

enum TypeOfToken op = TokenTable[pos].type;

++pos; // for Mul or Mod or Div

struct astNode\* right = lowPriorityExpression();

return createNode(op == Mul ? mul\_node : op == Div ? div\_node : mod\_node, lexemeTypeName(op), left, right);

}

return left;

}

struct astNode\* middlePriorityExpression() {

switch (TokenTable[pos].type) {

case Identifier: match(Identifier); return createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr);

case Number: match(Number); return createNode(number\_node, TokenTable[pos - 1].name, nullptr, nullptr);

case LBracket: {

++pos; // for (

struct astNode\* expr = arithmeticExpression();

match(RBracket);

return expr;

}

default: {

printf("\nSyntax error in line %d, token number: %d : middle priority operation was expected (current: %s).\n", TokenTable[pos].line, pos, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

}

struct astNode\* assignStatement() {

match(Identifier);

struct astNode\* left = createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr);

match(Assign);

struct astNode\* right = arithmeticExpression();

return createNode(assign\_node, "assign", left, right);

}

struct astNode\* ifStatement() {

match(If);

match(LBracket);

struct astNode\* expr = logicalExpression();

match(RBracket);

struct astNode\* stmt = statement();

if (TokenTable[pos + 1].type == Else) {

match(Semicolon);

++pos; // for ELSE

struct astNode\* elseStmt = statement();

stmt = createNode(else\_node, "else", stmt, elseStmt);

}

return createNode(if\_node, "if", expr, stmt);

}

struct astNode\* logicalExpression() {

struct astNode\* left = andExpression();

if (TokenTable[pos].type == Or) {

++pos; // for Or

struct astNode\* right = logicalExpression();

return createNode(or\_node, "or", left, right);

}

return left;

}

struct astNode\* andExpression() {

struct astNode\* left = comparison();

if (TokenTable[pos].type == And) {

++pos; // for And

struct astNode\* right = andExpression();

return createNode(and\_node, "and", left, right);

}

return left;

}

struct astNode\* comparison() {

struct astNode\* comp;

switch (TokenTable[pos].type) {

case Not: {

++pos; // for Not

match(LBracket);

comp = createNode(not\_node, "not", logicalExpression(), nullptr);

match(RBracket);

break;

}

case LBracket: {

++pos; // for (

comp = logicalExpression();

match(RBracket);

break;

}

default: {

comp = comparisonExpression();

}

}

return comp;

}

struct astNode\* comparisonExpression() {

struct astNode\* left = arithmeticExpression();

if (TokenTable[pos].type == Equality ||

TokenTable[pos].type == NotEquality ||

TokenTable[pos].type == Greate ||

TokenTable[pos].type == Less

) {

enum TypeOfToken op = TokenTable[pos].type;

++pos; // for Equality or NotEquality or Greate or Less

struct astNode\* right = arithmeticExpression();

return createNode(op == Equality ? eq\_node : op == NotEquality ? neq\_node : op == Greate ? gr\_node : ls\_node, lexemeTypeName(op), left, right);

}

else {

printf("\nSyntax error in line %d : Comparison operator was Expected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

struct astNode\* gotoStatement() {

match(Goto);

match(Identifier);

return createNode(goto\_node, "goto", createNode(id\_node, TokenTable[pos - 1].name, nullptr, nullptr), nullptr);

}

struct astNode\* labelPoint() {

match(Identifier);

match(Colon);

return createNode(label\_node, "label", createNode(id\_node, TokenTable[pos - 2].name, nullptr, nullptr), nullptr);

}

struct astNode\* forStatement() {

match(For);

struct astNode\* assign = assignStatement();

switch (TokenTable[pos].type) {

case To: {

++pos; // for To

struct astNode\* expr = arithmeticExpression();

match(Do);

return createNode(for\_node, "for", createNode(to\_node, "to", assign, expr), statement());

}

case Downto: {

++pos; // for Downto

struct astNode\* expr = arithmeticExpression();

match(Do);

return createNode(for\_node, "for", createNode(to\_node, "downto", assign, expr), statement());

}

default: {

printf("\nSyntax error in line %d : TO | DOWNTO operator was xpected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

}

struct astNode\* whileStatement() {

match(While);

struct astNode\* expr = logicalExpression();

struct astNode\* body = whileBody();

match(WEnd);

return createNode(while\_node, "while", expr, body);

}

struct astNode\* whileBody() {

if (TokenTable[pos].type != WEnd) {

struct astNode\* stmt = statement();

match(Semicolon);

struct astNode\* body = stmt;

while (TokenTable[pos].type != WEnd)

{

struct astNode\* nextStmt = statement();

match(Semicolon);

body = createNode(statement\_node, "statement", body, nextStmt);

}

return body;

}

return nullptr;

}

struct astNode\* repeatStatement() {

match(Repeat);

struct astNode\* body = repeatBody();

match(Until);

match(LBracket);

struct astNode\* expr = logicalExpression();

match(RBracket);

return createNode(repeat\_node, "repeat", body, expr);

}

struct astNode\* repeatBody() {

if (TokenTable[pos].type != Until) {

struct astNode\* stmt = statement();

match(Semicolon);

struct astNode\* body = stmt;

while (TokenTable[pos].type != Until)

{

struct astNode\* nextStmt = statement();

match(Semicolon);

body = createNode(statement\_node, "statement", body, nextStmt);

}

return body;

}

return nullptr;

}

struct astNode\* compoundStatement() {

match(StartBlock);

struct astNode\* body = programBody();

match(EndBlock);

return createNode(compound\_node, "compound", body, nullptr);

}

}

Файл Ast.hpp

#pragma once

#include "header.hpp"

#include <string.h>

#include <stdio.h>

#include <stdlib.h>

#include <malloc.h>

extern struct Token\* TokenTable;

extern int pos;

namespace AST {

struct astNode\* program();

struct astNode\* programBody();

struct astNode\* variableDeclaration();

struct astNode\* variableList();

struct astNode\* statement();

struct astNode\* inputStatement();

struct astNode\* outputStatement();

struct astNode\* arithmeticExpression();

struct astNode\* lowPriorityExpression();

struct astNode\* middlePriorityExpression();

struct astNode\* assignStatement();

struct astNode\* ifStatement();

struct astNode\* logicalExpression();

struct astNode\* andExpression();

struct astNode\* comparison();

struct astNode\* comparisonExpression();

struct astNode\* gotoStatement();

struct astNode\* labelPoint();

struct astNode\* forStatement();

struct astNode\* whileStatement();

struct astNode\* whileBody();

struct astNode\* repeatStatement();

struct astNode\* repeatBody();

struct astNode\* compoundStatement();

void deleteNode(struct astNode\* node);

struct astNode\* createNode(enum TypeOfNode type, const char\* name, struct astNode\* left, struct astNode\* right);

void printAST(struct astNode\* node, int level);

void fPrintAST(FILE\* outFile, struct astNode\* node, int level);

void match(enum TypeOfToken expectedType);

struct astNode\* astParser();

}

Файл Codegen.cpp

#include "Codegen.hpp"

namespace Codegen {

void codegen(FILE\* outFile, struct astNode\* node) {

if (node == 0) {

return;

}

switch (node->type) {

case program\_node: {

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

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

codegen(outFile, node->left); // for declaration

fprintf(outFile, "\n");

codegen(outFile, node->right); // for statements

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

break;

}

case var\_node: {

fprintf(outFile, "int ");

codegen(outFile, node->left);

fprintf(outFile, ";\n");

codegen(outFile, node->right);

break;

}

case number\_node:

case id\_node: {

fprintf(outFile, "%s", node->name);

break;

}

case statement\_node: {

codegen(outFile, node->left);

codegen(outFile, node->right);

break;

}

case input\_node: {

fprintf(outFile, "printf(\"Enter ");

codegen(outFile, node->left);

fprintf(outFile, ": \");\n");

fprintf(outFile, "scanf(\"%%d\", &");

codegen(outFile, node->left);

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

break;

}

case output\_node: {

fprintf(outFile, "printf(\"%%d\\n\", ");

codegen(outFile, node->left);

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

break;

}

case add\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " + ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case sub\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " - ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case mul\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " \* ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case div\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " / ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case mod\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " %% ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case assign\_node: {

codegen(outFile, node->left);

fprintf(outFile, " = ");

codegen(outFile, node->right);

fprintf(outFile, ";\n");

break;

}

case if\_node: {

fprintf(outFile, "if (");

codegen(outFile, node->left);

fprintf(outFile, ") ");

codegen(outFile, node->right);

break;

}

case else\_node: {

codegen(outFile, node->left);

fprintf(outFile, "else ");

codegen(outFile, node->right);

break;

}

case or\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " || ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case and\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " && ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case not\_node: {

fprintf(outFile, "!(");

codegen(outFile, node->left);

fprintf(outFile, ")");

break;

}

case eq\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " == ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case neq\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " != ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case gr\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " > ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case ls\_node: {

fprintf(outFile, "(");

codegen(outFile, node->left);

fprintf(outFile, " < ");

codegen(outFile, node->right);

fprintf(outFile, ")");

break;

}

case goto\_node: {

fprintf(outFile, "goto ");

codegen(outFile, node->left);

fprintf(outFile, ";\n");

break;

}

case label\_node: {

codegen(outFile, node->left);

fprintf(outFile, ":\n");

break;

}

case for\_node: {

fprintf(outFile, "for(\n");

codegen(outFile, node->left);

fprintf(outFile, "\n) ");

codegen(outFile, node->right);

break;

}

case to\_node: {

codegen(outFile, node->left);

codegen(outFile, node->left->left);

fprintf(outFile, " <= ");

codegen(outFile, node->right);

fprintf(outFile, ";\n++");

codegen(outFile, node->left->left);

break;

}

case downto\_node: {

codegen(outFile, node->left);

codegen(outFile, node->left->left);

fprintf(outFile, " >= ");

codegen(outFile, node->right);

fprintf(outFile, ";\n--");

codegen(outFile, node->left->left);

break;

}

case while\_node: {

fprintf(outFile, "while(");

codegen(outFile, node->left);

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

codegen(outFile, node->right);

fprintf(outFile, "}\n");

break;

}

case repeat\_node: {

fprintf(outFile, "do {\n");

codegen(outFile, node->left);

fprintf(outFile, "} while(");

codegen(outFile, node->right);

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

break;

}

case compound\_node: {

fprintf(outFile, "{\n");

codegen(outFile, node->left);

codegen(outFile, node->right);

fprintf(outFile, "}\n");

break;

}

default: {

exit(1);

printf("Undescribed node type: %d\n", node->type);

break;

}

}

}

}

Файл Codegen.hpp

#pragma once

#include "header.hpp"

#include <stdio.h>

#include <stdlib.h>

namespace Codegen {

void codegen(FILE\* outFile, struct astNode\* node);

}

Файл Header.cpp

#include "header.hpp"

const char\* lexemeTypeName(enum TypeOfToken type) {

switch (type) {

case StartProgram: return "StartProgram";

case StartBlock: return "StartBlock";

case Variable: return "Variable";

case Type: return "Type";

case EndBlock: return "EndBlock";

case Input: return "Input";

case Output: return "Output";

case If: return "If";

case Else: return "Else";

case Goto: return "Goto";

case For: return "For";

case To: return "To";

case Downto: return "Downto";

case Do: return "Do";

case While: return "While";

case WEnd: return "WEnd";

case Repeat: return "Repeat";

case Until: return "Until";

case Identifier: return "Identifier";

case Number: return "Number";

case Float: return "Float";

case Assign: return "Assign";

case Add: return "Add";

case Sub: return "Sub";

case Mod: return "Mod";

case Mul: return "Mul";

case Div: return "Div";

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 LBracket: return "LBracket";

case RBracket: return "RBracket";

case Semicolon: return "Semicolon";

case Colon: return "Colon";

case Comma: return "Comma";

case Unknown\_: return "Unknown\_";

}

return "Forgotten";

}

Файл header.hpp

#pragma once

#define MAX\_TOKENS 1000

#define MAX\_IDENTIFIER 10

enum TypeOfToken {

StartProgram, // program

StartBlock, // start

Variable, // var

Type, // int16\_t

EndBlock, // finish

Input, // get

Output, // put

If, // if

Then, // then

Else, // else

Goto, // goto

For, // for

To, // to

Downto, // downto

Do, // do

While, // while

Continue, // continue

Exit, // exit

WEnd, // wend

Repeat, // repest

Until, // until

Identifier, // Identifier

Number, // number

Float, // float (incorrect)

Assign, // <-

Add, // +

Sub, // -

Mul, // \*

Div, // /

Mod, // %

Equality, // eg

NotEquality, // en

Greate, // >=

Less, // <=

Not, // !

And, // and

Or, // or

LBracket, // (

RBracket, // )

Semicolon, // ;

Colon, // :

Comma, // ,

Unknown\_

};

struct Token {

char name[16]; // ім'я лексеми

int value; // значення (для констант)

int line; // номер рядка

enum TypeOfToken type; // тип лексеми

};

struct id {

char name[16];

//unsigned int pos; //for labels

};

enum States {

Start, // початковий стан

Finish, // фінальний стан

Letter, // опрацювання слів (ключові слова та ідентифікаторів)

Digit, // опрацювання цифр

Separator, // опрацювання роздільників

Another, // опрацювання інших символів

EndOFile, // кінець файлу

Comment // ігнорування коментаря

};

enum TypeOfNode {

program\_node,

id\_node,

var\_node,

statement\_node,

input\_node,

output\_node,

add\_node,

sub\_node,

mul\_node,

div\_node,

mod\_node,

number\_node,

assign\_node,

if\_node,

else\_node,

or\_node,

and\_node,

not\_node,

eq\_node,

neq\_node,

gr\_node,

ls\_node,

goto\_node,

label\_node,

for\_node,

to\_node,

downto\_node,

while\_node,

repeat\_node,

compound\_node

};

struct astNode {

enum TypeOfNode type;

char name[16];

struct astNode\* left;

struct astNode\* right;

};

const char\* lexemeTypeName(enum TypeOfToken type);

Файл LexicAnalyzer.hpp

#pragma once

#include "header.hpp"

#include <stdio.h>

namespace LexicAnalyzer {

unsigned int getTokens(FILE\* F);

void printTokens(void);

void fprintTokens(FILE\* F);

}

Файл LexicAnalyzercpp.cpp

#include "LexicAnalyzer.hpp"

#include <stdbool.h>

#include <string.h>

#include <stdlib.h>

extern struct Token\* TokenTable; // Таблиця лексем

extern unsigned int TokensNum; // Кількість лексем

unsigned int LexicAnalyzer::getTokens(FILE\* F) {

enum States state = Start;

struct Token tempToken;

char ch, buf[16];

unsigned int tokenCount = 0;

int line = 1;

int tokenLength = 0;

ch = getc(F);

while (true) {

switch (state) {

case Start: {

if (ch == EOF) {

state = EndOFile;

}

else if (('0' <= ch && ch <= '9') || ch == '-' || ch == '.') {

state = Digit;

}

else if (ch == '%') {

ch = getc(F);

if (ch == '%') {

state = Comment; // Це коментар

}

else {

ungetc(ch, F); // Повертаємо символ назад, бо це оператор модуль

strcpy\_s(tempToken.name, "%");

tempToken.type = Mod; // Оператор модуль

tempToken.value = 0;

tempToken.line = line;

state = Finish; // Завершити обробку токена

}

}

else if (('a' <= ch && ch <= 'z') || ('A' <= ch && ch <= 'Z')) {

state = Letter;

}

else if (ch == ' ' || ch == '\t' || ch == '\n') {

state = Separator;

}

else {

state = Another;

}

break;

}

case Digit: {

buf[0] = ch;

int j = 1;

ch = getc(F);

while (((ch <= '9' && ch >= '0') || ch == '.') && j < 10) {

buf[j++] = ch;

ch = getc(F);

}

buf[j] = '\0';

if (!strcmp(buf, "-")) {

strcpy\_s(tempToken.name, "-");

tempToken.type = Sub;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

break;

}

if (!strcmp(buf, ".")) {

strcpy\_s(tempToken.name, ".");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

break;

}

short dotCounter = 0, currCharIndex = 0;

while (buf[currCharIndex] != '\0') {

if (buf[currCharIndex] == '.') {

++dotCounter;

}

++currCharIndex;

}

if (dotCounter > 1) {

strcpy\_s(tempToken.name, buf);

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

break;

}

if (dotCounter == 1) {

strcpy\_s(tempToken.name, buf);

tempToken.type = Float;

tempToken.value = atof(buf);

tempToken.line = line;

state = Finish;

break;

}

strcpy\_s(tempToken.name, buf);

tempToken.type = Number;

tempToken.value = atoi(buf);

tempToken.line = line;

state = Finish;

break;

}

case Separator: {

if (ch == '\n') {

line++;

}

ch = getc(F);

state = Start;

break;

}

case Comment:

while (true) {

if (ch == EOF) {

state = EndOFile;

break;

}

if (ch == '%') {

ch = getc(F); // Читаємо наступний символ

if (ch == '%') { // Закриваємо коментар

state = Start; // Повертаємося до початку

ch = getc(F); // Продовжуємо обробку

break;

}

}

ch = getc(F); // Продовжуємо читання символів

}

break;

case Finish: {

if (tokenCount < MAX\_TOKENS) {

TokenTable[tokenCount++] = tempToken;

if (ch != EOF) {

state = Start;

}

else {

state = EndOFile;

}

}

else {

printf("\n\t\t\ttoo many tokens !!!\n");

return TokensNum = tokenCount - 1;

}

break;

}

case EndOFile: {

return TokensNum = tokenCount;

}

case Letter: {

buf[0] = ch;

int j = 1;

ch = getc(F);

while ((('a' <= ch && ch <= 'z') || (ch == '\_') || ('A' <= ch && ch <= 'Z') || ('0' <= ch && ch <= '9')) && j < 15) {

buf[j++] = ch;

ch = getc(F);

}

buf[j] = '\0';

enum TypeOfToken tempType = Unknown\_;

if (!strcmp(buf, "program")) {

tempType = StartProgram;

}

else if (!strcmp(buf, "start")) {

tempType = StartBlock;

}

else if (!strcmp(buf, "var")) {

tempType = Variable;

}

else if (!strcmp(buf, "int16\_t")) {

tempType = Type;

}

else if (!strcmp(buf, "finish")) {

tempType = EndBlock;

}

else if (!strcmp(buf, "get")) {

tempType = Input;

}

else if (!strcmp(buf, "put")) {

tempType = Output;

}

else if (!strcmp(buf, "if")) {

tempType = If;

}

else if (!strcmp(buf, "else")) {

tempType = Else;

}

else if (!strcmp(buf, "goto")) {

tempType = Goto;

}

else if (!strcmp(buf, "for")) {

tempType = For;

}

else if (!strcmp(buf, "to")) {

tempType = To;

}

else if (!strcmp(buf, "downto")) {

tempType = Downto;

}

else if (!strcmp(buf, "do")) {

tempType = Do;

}

else if (!strcmp(buf, "while")) {

tempType = While;

}

else if (!strcmp(buf, "wend")) {

tempType = WEnd;

}

else if (!strcmp(buf, "repeat")) {

tempType = Repeat;

}

else if (!strcmp(buf, "until")) {

tempType = Until;

}

else if (!strcmp(buf, "eg")) {

tempType = Equality;

}

else if (!strcmp(buf, "ne")) {

tempType = NotEquality;

}

else if (!strcmp(buf, "and")) {

tempType = And;

}

else if (!strcmp(buf, "or")) {

tempType = Or;

}

else if (strlen(buf) == 4) {

tempType = Identifier;

}

strcpy\_s(tempToken.name, buf);

tempToken.type = tempType;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

break;

}

case Another: {

switch (ch) {

case '(': {

strcpy\_s(tempToken.name, "(");

tempToken.type = LBracket;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case ')': {

strcpy\_s(tempToken.name, ")");

tempToken.type = RBracket;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case ',': {

strcpy\_s(tempToken.name, ",");

tempToken.type = Comma;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case ';': {

strcpy\_s(tempToken.name, ";");

tempToken.type = Semicolon;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case ':': {

strcpy\_s(tempToken.name, ":");

tempToken.type = Colon;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '|': {

strcpy\_s(tempToken.name, "|");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '&': {

strcpy\_s(tempToken.name, "&");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '^': {

strcpy\_s(tempToken.name, "^");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '!': {

strcpy\_s(tempToken.name, "!");

tempToken.type = Not;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '+': {

strcpy\_s(tempToken.name, "+");

tempToken.type = Add;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '\*': {

strcpy\_s(tempToken.name, "\*");

tempToken.type = Mul;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

case '/': {

strcpy\_s(tempToken.name, "/");

tempToken.type = Div;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}

/\*case '%': {

strcpy\_s(tempToken.name, "%");

tempToken.type = Mod;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

break;

}\*/

case '<': {

ch = getc(F);

if (ch == '=') {

strcpy\_s(tempToken.name, "<=");

tempToken.type = Less;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

}

else if (ch == '-') {

strcpy\_s(tempToken.name, "<-");

tempToken.type = Assign;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

}

else {

strcpy\_s(tempToken.name, "<");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

}

break;

}

case '>': {

ch = getc(F);

if (ch == '=') {

strcpy\_s(tempToken.name, ">=");

tempToken.type = Greate;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

}

else {

strcpy\_s(tempToken.name, ">");

tempToken.type = Unknown\_;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

}

break;

}

case '=': {

strcpy\_s(tempToken.name, "=");

tempToken.type = Equality;

tempToken.value = 0;

tempToken.line = line;

state = Finish;

ch = getc(F);

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;

}

}

}

}

}

return TokensNum = tokenCount;

}

// Функція друкує таблицю лексем на екран

void LexicAnalyzer::printTokens(void) {

char type\_tokens[16];

printf("---------------------------------------------------------------------------\n");

printf("| TOKEN TABLE |\n");

printf("---------------------------------------------------------------------------\n");

printf("| line number | token | value | token code | type of token |\n");

printf("---------------------------------------------------------------------------\n");

for (unsigned int i = 0; i < TokensNum; i++) {

strcpy\_s(type\_tokens, lexemeTypeName(TokenTable[i].type));

printf("|%12d |%16s |%11d |%11d | %-13s |\n",

TokenTable[i].line,

TokenTable[i].name,

TokenTable[i].value,

TokenTable[i].type,

type\_tokens);

printf("---------------------------------------------------------------------------\n");

}

}

void LexicAnalyzer::fprintTokens(FILE\* F) {

char type\_tokens[16];

fprintf(F, "\n\n---------------------------------------------------------------------------\n");

fprintf(F, "| TOKEN TABLE |\n");

fprintf(F, "---------------------------------------------------------------------------\n");

fprintf(F, "| line number | token | value | token code | type of token |\n");

fprintf(F, "---------------------------------------------------------------------------\n");

for (unsigned int i = 0; i < TokensNum; i++) {

strcpy\_s(type\_tokens, lexemeTypeName(TokenTable[i].type));

fprintf(F, "|%12d |%16s |%11d |%11d | %-13s |\n",

TokenTable[i].line,

TokenTable[i].name,

TokenTable[i].value,

TokenTable[i].type,

type\_tokens);

fprintf(F, "---------------------------------------------------------------------------\n");

}

}

Файл main.cpp

#include <iostream>

#include <windows.h>

#include "LexicAnalyzer.hpp"

#include "Parser.hpp"

#include "Ast.hpp"

#include "Codegen.hpp"

#define LANGUAGE "test3"

struct Token\* TokenTable; // Таблиця лексем

unsigned int TokensNum; // Кількість лексем

struct id\* idTable; // Таблиця ідентифікаторів

unsigned int idNum; // кількість ідентифікаторів

struct id\* labelTable; // Таблиця міток

unsigned int labelNum; // кількість міток

FILE\* errorFile;

int main(int argc, char\* argv[])

{

// виділення пам'яті під таблицю лесем

TokenTable = (struct Token\*)malloc(MAX\_TOKENS \* sizeof(struct Token));

// виділення пам'яті під таблицю ідентифікаторів

idTable = (struct id\*)malloc(MAX\_IDENTIFIER \* sizeof(struct id));

labelTable = (struct id\*)malloc(MAX\_IDENTIFIER \* sizeof(struct id));

char InputFile[32] = "";

FILE\* InFile, \* TokenFileP;

if (argc != 2)

{

printf("Input file name: ");

gets\_s(InputFile);

}

else

{

strcpy\_s(InputFile, argv[1]);

}

if (!strstr(InputFile, LANGUAGE)) {

free(TokenTable);

free(idTable);

free(labelTable);

\_fcloseall();

printf("Program file name should contain \"%s\"", LANGUAGE);

exit(1);

}

if ((fopen\_s(&InFile, InputFile, "rt")) != 0)

{

printf("Error: Can not 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], errorFileName[32];

strcpy\_s(TokenFile, NameFile);

strcat\_s(TokenFile, ".token");

strcpy\_s(errorFileName, NameFile);

strcat\_s(errorFileName, ".errorlist");

// лексичний аналіз

if ((fopen\_s(&TokenFileP, TokenFile, "wt")) != 0)

{

printf("Error: Can not create file: %s\n", TokenFile);

free(TokenTable);

free(idTable);

free(labelTable);

\_fcloseall();

return 1;

}

if ((fopen\_s(&errorFile, errorFileName, "wt")) != 0)

{

printf("Error: Can not create file: %s\n", errorFileName);

free(TokenTable);

free(idTable);

free(labelTable);

\_fcloseall();

return 1;

}

TokensNum = LexicAnalyzer::getTokens(InFile);

LexicAnalyzer::fprintTokens(TokenFileP);

fclose(InFile);

fclose(TokenFileP);

printf("\nLexical analysis completed: %d tokens. List of tokens in the file %s\n", TokensNum, TokenFile);

//PrintTokens(TokenTable, TokensNum);

// синтаксичний аналіз

Parser::Parser();

Parser::Semantic();

// створення абстрактного синтаксичного дерева

struct astNode\* ast = AST::astParser();

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");

free(TokenTable);

free(idTable);

free(labelTable);

AST::deleteNode(ast);

exit(1);

}

AST::fPrintAST(ASTFile, ast, 0);

printf("\nAST has been created and written to %s.\n", AST);

char OutputFile[32];

strcpy\_s(OutputFile, NameFile);

strcat\_s(OutputFile, ".c");

// Open output file

FILE\* outFile;

fopen\_s(&outFile, OutputFile, "w");

if (!outFile)

{

printf("Failed to open output file.\n");

free(TokenTable);

free(idTable);

free(labelTable);

AST::deleteNode(ast);

exit(1);

}

// Generate C code from AST

Codegen::codegen(outFile, ast);

// генерація вихідного С коду

printf("\nC code has been generated and written to %s.\n", OutputFile);

// Close the file

\_fcloseall();

free(TokenTable);

free(idTable);

free(labelTable);

char setVar[256] = "\"C:\\Program Files\\Microsoft Visual Studio\\2022\\Community\\VC\\Auxiliary\\Build\\vcvars64.bat\"";

char createExe[128];

sprintf\_s(createExe, "cl %s", OutputFile);

strcat\_s(setVar, " && ");

strcat\_s(setVar, createExe);

system(setVar);

system("pause");

return 0;

}

Файл Parser.cpp

#include "Parser.hpp"

#include <stdbool.h>

#include <stdlib.h>

#include <string.h>

#include <stdio.h>

extern struct Token\* TokenTable;

extern unsigned int TokensNum;

extern struct id\* idTable;

extern unsigned int idNum;

extern struct id\* labelTable;

extern unsigned int labelNum;

extern FILE\* errorFile;

int pos = 0;

namespace Parser {

void Parser() {

program();

printf("\nThe program is syntax correct.\n");

fprintf(errorFile, "\nThe program is syntax correct.\n");

}

void Semantic() {

idNum = IdIdentification(idTable, TokenTable, TokensNum);

labelNum = LabelIdentification(labelTable, TokenTable, TokensNum);

printf("\nThe program is semantic correct.\n");

printf("\n%d labels found\n", labelNum);

fprintf(errorFile, "\nThe program is semantic correct.\n");

fprintf(errorFile, "\n%d labels found\n", labelNum);

printIdentifiers(labelNum, labelTable);

fprintIdentifiers(errorFile, labelNum, labelTable);

printf("\n%d identifiers found\n", idNum);

fprintf(errorFile, "\n%d identifiers found\n", idNum);

printIdentifiers(idNum, idTable);

fprintIdentifiers(errorFile, idNum, idTable);

}

void match(enum TypeOfToken expectedType) {

if (TokenTable[pos].type == expectedType)

pos++;

else {

printf("\nSyntax error in line %d, token number: %d : another type of lexeme was expected (expected: %s | current: %s).\n", TokenTable[pos].line, pos, lexemeTypeName(expectedType), lexemeTypeName(TokenTable[pos].type));

fprintf(errorFile, "\nSyntax error in line %d, token number: %d : another type of lexeme was expected (expected: %s | current: %s).\n", TokenTable[pos].line, pos, lexemeTypeName(expectedType), lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

// функція записує оголошені ідентифікатори в таблицю ідентифікаторів idTable

// повертає кількість ідентифікаторів

// перевіряє чи усі використані ідентифікатори оголошені

unsigned int IdIdentification(struct id idTable[], struct Token TokenTable[], unsigned int tokenCount) {

unsigned int idCount = 0;

unsigned int i = 0;

while (TokenTable[i++].type != Variable && TokenTable[i].type != EndBlock);

while (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;

}

++i;

}

for (; i < tokenCount; ++i) {

if (

TokenTable[i].type == Identifier

&& TokenTable[i - 1].type != Goto

&& TokenTable[i + 1].type != Colon

) {

bool yes = 0;

for (unsigned int j = 0; j < idCount; ++j) {

if (!strcmp(TokenTable[i].name, idTable[j].name)) {

yes = 1;

break;

}

}

if (!yes) {

printf("\nSemantic Error In line %d, an undeclared identifier \"%s\"!", TokenTable[i].line, TokenTable[i].name);

fprintf(errorFile, "\nSemantic Error In line %d, an undeclared identifier \"%s\"!", TokenTable[i].line, TokenTable[i].name);

exit(1);

}

}

}

return idCount; // Повертає кількість ідентифікаторів

}

void program() {

match(StartProgram);

match(Identifier); // Перевірка, що після program йде ідентифікатор

// Обробка змінних

if (TokenTable[pos].type == Variable) {

++pos; // для variable

variableDeclaration();

match(Semicolon);

}

match(StartBlock);

programBody();

match(EndBlock);

}

void variableDeclaration() {

match(Type);

variableList();

}

void variableList() {

match(Identifier);

while (TokenTable[pos].type == Comma) {

++pos; // for Comma

match(Identifier);

}

}

void statement() {

switch (TokenTable[pos].type) {

case Input: inputStatement(); break;

case Output: outputStatement(); break;

case If: ifStatement(); break;

case Goto: gotoStatement(); break;

case For: forStatement(); break;

case While: whileStatement(); break;

case Repeat: repeatStatement(); break;

case StartBlock:compoundStatement(); break;

default: {

if (TokenTable[pos + 1].type == Colon)

labelPoint();

else

assignStatement();

}

}

}

void inputStatement() {

++pos; // for get

match(Identifier);

}

void outputStatement() {

++pos; // for put

arithmeticExpression();

}

void arithmeticExpression() {

lowPriorityExpression();

//lowPriorityOperator

while (TokenTable[pos].type == Add || TokenTable[pos].type == Sub) {

++pos; // for + or -

lowPriorityExpression();

}

}

void lowPriorityExpression() {

middlePriorityExpression();

// middlePriorityOperator

while (TokenTable[pos].type == Mul || TokenTable[pos].type == Mod || TokenTable[pos].type == Div) {

++pos; // for \* or % or /

middlePriorityExpression();

}

}

void middlePriorityExpression() {

switch (TokenTable[pos].type) {

case Identifier: ++pos; break;

case Number: ++pos; break;

case LBracket: {

++pos; // for (

arithmeticExpression();

match(RBracket);

break;

}

default: {

printf("\nSyntax error in line %d, token number: %d : middle priority operation was expected (current: %s).\n", TokenTable[pos].line, pos, lexemeTypeName(TokenTable[pos].type));

fprintf(errorFile, "\nSyntax error in line %d, token number: %d : middle priority operation was expected (current: %s).\n", TokenTable[pos].line, pos, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

}

}

void assignStatement() {

match(Identifier);

match(Assign);

arithmeticExpression();

}

void ifStatement() {

match(If);

match(LBracket);

logicalExpression();

match(RBracket);

statement();

if (TokenTable[pos + 1].type == Else) {

match(Semicolon);

++pos; // for else

statement();

}

}

void logicalExpression() {

andExpression();

while (TokenTable[pos].type == Or) {

++pos; // for or

andExpression();

}

}

void andExpression() {

comparison();

while (TokenTable[pos].type == And) {

++pos; // for and

andExpression();

}

}

void comparison() {

if (TokenTable[pos].type == Not) {

++pos; // for !

match(LBracket);

logicalExpression();

match(RBracket);

}

else if (TokenTable[pos].type == LBracket) {

++pos; // for (

logicalExpression();

match(RBracket);

}

else {

comparisonExpression();

}

}

void comparisonExpression() {

arithmeticExpression();

if (TokenTable[pos].type == Equality ||

TokenTable[pos].type == NotEquality ||

TokenTable[pos].type == Greate ||

TokenTable[pos].type == Less

) {

++pos; // for "eg" | "ne" | "<=" | ">="

}

else {

printf("\nSyntax error in line %d : Comparison operator was Expected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

fprintf(errorFile, "\nSyntax error in line %d : Comparison operator was Expected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

arithmeticExpression();

}

void gotoStatement() {

match(Goto);

match(Identifier);

}

void labelPoint() {

match(Identifier);

match(Colon);

}

void forStatement() {

match(For);

assignStatement();

if (TokenTable[pos].type == To || TokenTable[pos].type == Downto) {

++pos; // for to or downto

}

else {

printf("\nSyntax error in line %d : TO or DOWNTO was expected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

fprintf(errorFile, "\nSyntax error in line %d : TO or DOWNTO was expected, %s token gained).\n", TokenTable[pos].line, lexemeTypeName(TokenTable[pos].type));

exit(1);

}

arithmeticExpression();

match(Do);

statement();

}

void whileStatement() {

match(While);

logicalExpression();

while (TokenTable[pos].type != WEnd) {

statement();

match(Semicolon);

}

++pos; // for wend

}

void repeatStatement() {

match(Repeat);

while (TokenTable[pos].type != Until) {

statement();

match(Semicolon);

}

++pos; // for until

match(LBracket);

logicalExpression();

match(RBracket);

}

void compoundStatement() {

match(StartBlock);

programBody();

match(EndBlock);

}

void programBody() {

while (TokenTable[pos].type != EndBlock) {

statement();

match(Semicolon);

}

}

void printIdentifiers(int num, struct id\* table) {

for (int i = 0; i < num; i++) {

printf("%s\n", table[i].name);

}

}

void fprintIdentifiers(FILE\* F, int num, struct id\* table) {

for (int i = 0; i < num; i++) {

fprintf(F, "%s\n", table[i].name);

}

}

unsigned int LabelIdentification(struct id labelTable[], struct Token TokenTable[], unsigned int tokenCount) {

unsigned int labelNum = 0;

unsigned int i = 0;

unsigned int start = 0;

while (TokenTable[start++].type != StartProgram);

i = start;

while (TokenTable[i].type != EndBlock) {

if (TokenTable[i].type == Identifier && TokenTable[i + 1].type == Colon) {

strcpy\_s(labelTable[labelNum++].name, TokenTable[i].name);

}

++i;

}

i = start;

while (TokenTable[i].type != EndBlock) {

if (TokenTable[i].type == Identifier && TokenTable[i - 1].type == Goto) {

bool found = false;

for (unsigned int j = 0; j < labelNum; j++) {

if (strcmp(labelTable[j].name, TokenTable[i].name) == 0) {

found = true;

break;

}

}

if (!found) {

printf("\n Semantic error: In line %d label %s is not defined\n", TokenTable[i].line, TokenTable[i].name);

fprintf(errorFile, "\n Semantic error: In line %d label %s is not defined\n", TokenTable[i].line, TokenTable[i].name);

exit(1);

}

}

++i;

}

return labelNum;

}

}

Файл Parser.hpp  
#pragma once

#include "header.hpp"

#include <stdbool.h>

#include <stdlib.h>

#include <string.h>

#include <stdio.h>

extern struct Token\* TokenTable;

extern unsigned int TokensNum;

extern struct id\* idTable;

extern unsigned int idNum;

extern struct id\* labelTable;

extern unsigned int labelNum;

namespace Parser {

void Parser();

void Semantic();

void match(enum TypeOfToken expectedType);

unsigned int IdIdentification(struct id idTable[], struct Token TokenTable[], unsigned int tokenCount);

unsigned int LabelIdentification(struct id labelTable[], struct Token TokenTable[], unsigned int tokenCount);

void printIdentifiers(int num, struct id\* table);

void fprintIdentifiers(FILE\* F, int num, struct id\* table);

// program ::= "program", {"var", variable\_declaration, ";"}, "start", {statement, ";"}, "finish";

void program();

// programBody ::= { statement, ";" };

void programBody();

// variable\_declaration ::= "int16\_t", variable\_list;

void variableDeclaration();

// variable\_list ::= identifier, {",", identifier};

void variableList();

// statement ::= input\_statement | output\_statement | assign\_statement | if\_else\_statement | goto\_statement | label\_point | for\_statement | while\_statement | repeat\_until\_statement | compound\_statement;

void statement();

// input\_statement ::= "get", identifier;

void inputStatement();

// output\_statement ::= "put", arithmetic\_expression;

void outputStatement();

// arithmetic\_expression :: = low\_priority\_expression{ low\_priority\_operator, low\_priority\_expression };

void arithmeticExpression();

// low\_priority\_expression :: = middle\_priority\_expression{ middle\_priority\_operator, middle\_priority\_expression };

void lowPriorityExpression();

// low\_priority\_operator :: = "+" | "-";

//void lowPriorityOperator();

// middle\_priority\_expression :: = identifier | number | "(", arithmetic\_expression, ")";

void middlePriorityExpression();

// middle\_priority\_operator :: = "\*" | "/" | "%";

//void middlePriorityOperator();

// assign\_statement :: = arithmetic\_expression, "<-", identifier;

void assignStatement();

// if\_else\_statement :: = "if", "(", logical\_expression, ")", statement, [";", "ELSE", statement];

void ifStatement();

// logical\_expression :: = and\_expression{ or\_operator, and\_expression };

void logicalExpression();

// or\_operator :: = "or";

//void orOperator();

// and\_expression :: = comparison{ and\_operator, and\_expression };

void andExpression();

// and\_operator :: = "and";

// void andOperator();

// comparison :: = comparison\_expression | [not\_operator] "(", logical\_expression, ")";

void comparison();

// not\_operator :: = "!";

// void notOperator();

// comparison\_expression :: = arithmetic\_expression comparison\_operator arithmetic\_expression;

void comparisonExpression();

// comparison\_operator :: = "eg" | "ne" | "<=" | ">=";

// void comparisonOperator();

// goto\_statement :: = "goto", identifier;

void gotoStatement();

// label\_point :: = identifier, ":";

void labelPoint();

// for\_to\_statement ::= "for", assign\_statement, "to" | "downto", arithmetic\_expression, "do", statement;

void forStatement();

// while\_statement :: = "while", logical\_expression, { statement, ";" }, "wend";

void whileStatement();

// repeat\_until\_statement :: = "repeat", { statement, ";" }, "until", "(", logical\_expression, ")";

void repeatStatement();

// compoundStatement :: = "start", { statement, ";" }, "finish";

void compoundStatement();

}