ЗВІТ З ЛАБОРАТОРНОЇ РОБОТИ № 2

за курсом “Мовні технології”

студента групи Пз-22-3

Казанікова Іллі Івановича

2023/2024 н.р.

1. **Постановка задачі**

Реалізувати лексичний аналізатор php

1. **Опис розв’язку**

Код виконує лексичний аналіз введеного коду, розпізнаючи різні типи лексем, такі як ключові слова, ідентифікатори, числа, рядкові та символьні константи, оператори та коментарі. Розглянемо кожну частину коду текстом.

Опис розв’язку

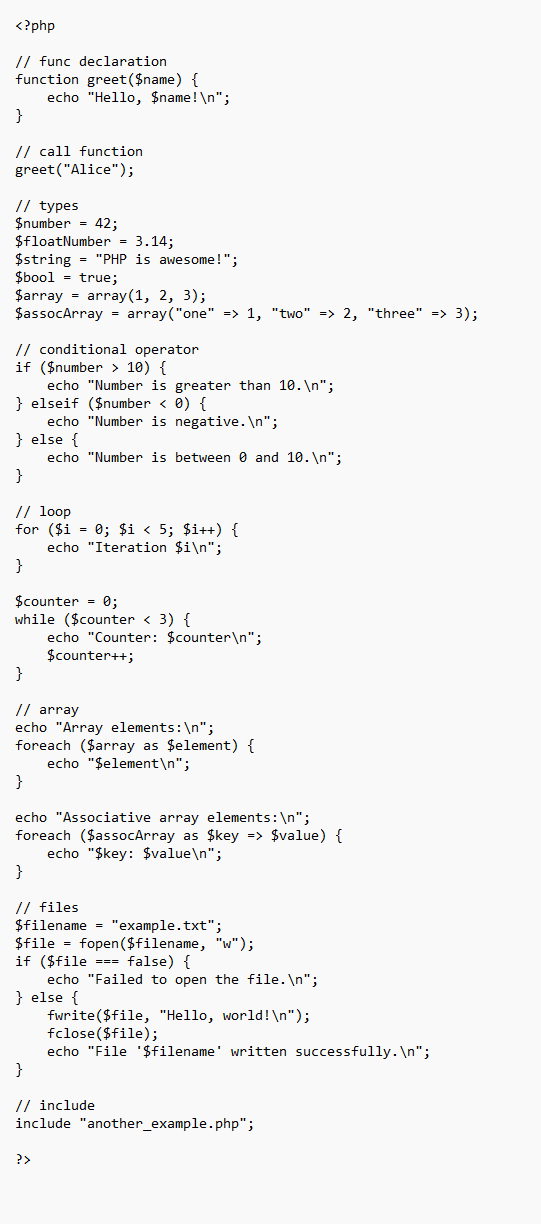
1. Ініціалізація:
   * Функція isKeyword використовує статичний контейнер unordered\_set для зберігання ключових слів. Це дозволяє швидко перевіряти, чи є дане слово ключовим, оскільки доступ до елементів у множині здійснюється за константний час.
   * У функції recognizeLexemes визначається перелік станів автомату (State), які використовуються для розпізнавання різних типів лексем: початковий стан (START), ідентифікатор (IN\_IDENTIFIER), число (IN\_NUMBER), рядкова константа (IN\_STRING), символьна константа (IN\_CHAR), оператор (IN\_OPERATOR) та коментар (IN\_COMMENT).
2. Обробка символів:
   * У початковому стані (START) кожен символ перевіряється, щоб визначити, який тип лексеми починається. Якщо це літера або підкреслення, то починається ідентифікатор, якщо цифра — число, якщо лапки — рядкова або символьна константа, якщо спеціальний символ — оператор або розділювач.
   * Коли автомат переходить у відповідний стан (наприклад, IN\_IDENTIFIER для ідентифікаторів), він продовжує додавати символи до поточної лексеми доти, доки не буде знайдено символ, який не належить цьому типу лексеми. Потім лексема завершується і обробляється відповідним чином (наприклад, перевіряється, чи є ідентифікатор ключовим словом).
   * Якщо автомат перебуває в стані коментаря (IN\_COMMENT), він додає всі символи до коментаря доти, доки не зустріне символ нового рядка, після чого коментар виводиться.
3. Розпізнавання та вивід:
   * Після завершення зчитування кожної лексеми автомат повертається до початкового стану (START), готовий до обробки наступної лексеми.
   * Лексеми різних типів виводяться на екран з відповідним позначенням: ключові слова, ідентифікатори, числа, рядкові константи, символьні константи, оператори, коментарі та розділювачі.
4. У функції main файл example.txt відкривається для зчитування. Вміст файлу зчитується пострічково та додається до рядка code.
5. Після завершення зчитування файл закривається, а функція recognizeLexemes викликається для аналізу коду.

**3. Опис інтерфейсу**

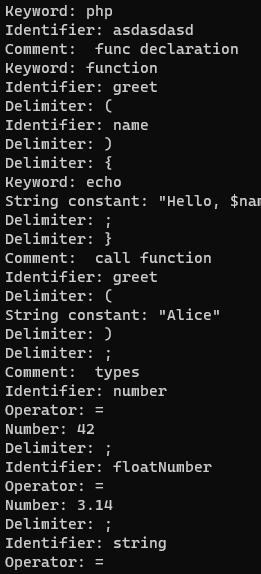
Створити файл з кодом php

**4. Тестування програми**

**Вхідний файл:**

****

**Фрагмент результату:**

****

**5. Аналіз помилок**

**немає**

**6. додаток:**

#include <iostream>

#include <fstream>

#include <string>

#include <cctype>

#include <unordered\_set>

bool checkKeyword(const std::string& word) {

static const std::unordered\_set<std::string> keywords = {

"if", "else", "for", "while", "function", "class", "include", "require", "return", "echo",

"false", "true", "break", "continue", "switch", "case", "default", "try", "catch", "throw",

"namespace", "use", "public", "private", "protected", "static", "final", "abstract", "interface",

"extends", "implements", "new", "global", "const", "var", "define", "die", "exit", "include\_once",

"require\_once", "yield", "yield from", "instanceof", "php", "as"

};

return keywords.count(word) > 0;

}

void analyze(const std::string& code) {

enum class State { START, IN\_IDENTIFIER, IN\_NUMBER, IN\_STRING, IN\_CHAR, IN\_OPERATOR, IN\_COMMENT, IN\_MULTILINE\_COMMENT };

State state = State::START;

std::string lexeme;

std::string comment;

char stringDelimiter = '\0';

char charDelimiter = '\0';

for (size\_t i = 0; i < code.size(); ++i) {

char c = code[i];

switch (state) {

case State::START:

if (std::isalpha(c) || c == '\_') {

state = State::IN\_IDENTIFIER;

lexeme += c;

}

else if (std::isdigit(c)) {

state = State::IN\_NUMBER;

lexeme += c;

}

else if (c == '\"' || c == '\'') {

state = (c == '\"') ? State::IN\_STRING : State::IN\_CHAR;

stringDelimiter = c;

lexeme += c;

}

else if (c == '/') {

if (i + 1 < code.size() && code[i + 1] == '/') {

state = State::IN\_COMMENT;

++i;

}

else if (i + 1 < code.size() && code[i + 1] == '\*') {

state = State::IN\_MULTILINE\_COMMENT;

++i;

}

else {

state = State::IN\_OPERATOR;

lexeme += c;

}

}

else if (std::isspace(c)) {

}

else if (c == '+' || c == '-' || c == '\*' || c == '/' || c == '%' || c == '=' || c == '!' || c == '&' || c == '|') {

state = State::IN\_OPERATOR;

lexeme += c;

}

else if (c == ';' || c == ',' || c == '.' || c == ')' || c == '(' || c == '{' || c == '}') {

std::cout << "Delimiter: " << c << std::endl;

}

break;

case State::IN\_IDENTIFIER:

if (std::isalnum(c) || c == '\_') {

lexeme += c;

}

else {

if (checkKeyword(lexeme)) {

std::cout << "Keyword: " << lexeme << std::endl;

}

else {

std::cout << "Identifier: " << lexeme << std::endl;

}

lexeme.clear();

state = State::START;

--i;

}

break;

case State::IN\_NUMBER:

if (std::isdigit(c) || c == '.') {

lexeme += c;

}

else {

std::cout << "Number: " << lexeme << std::endl;

lexeme.clear();

state = State::START;

--i;

}

break;

case State::IN\_STRING:

lexeme += c;

if (c == stringDelimiter && lexeme.back() != '\\') {

std::cout << "String constant: " << lexeme << std::endl;

lexeme.clear();

state = State::START;

stringDelimiter = '\0';

}

break;

case State::IN\_CHAR:

lexeme += c;

if (c == charDelimiter && lexeme.back() != '\\') {

std::cout << "Character constant: " << lexeme << std::endl;

lexeme.clear();

state = State::START;

charDelimiter = '\0';

}

break;

case State::IN\_OPERATOR:

if (c == '+' || c == '-' || c == '\*' || c == '/' || c == '%' || c == '=' || c == '!' || c == '&' || c == '|') {

lexeme += c;

}

std::cout << "Operator: " << lexeme << std::endl;

lexeme.clear();

state = State::START;

break;

case State::IN\_COMMENT:

if (c == '\n') {

std::cout << "Comment: " << comment << std::endl;

comment.clear();

state = State::START;

}

else {

comment += c;

}

break;

case State::IN\_MULTILINE\_COMMENT:

comment += c;

if (c == '\*' && i + 1 < code.size() && code[i + 1] == '/') {

comment += '/';

++i;

std::cout << "Multiline Comment: " << comment << std::endl;

comment.clear();

state = State::START;

}

break;

}

}

}

int main() {

std::ifstream file("example.txt");

if (!file.is\_open()) {

std::cerr << "Failed to open the file" << std::endl;

return 1;

}

std::string code;

std::string line;

while (std::getline(file, line)) {

code += line + '\n';

}

file.close();

analyze(code);

return 0;

}