Національний університет “Львівська політехніка”

Інститут комп’ютерної техніки, автоматики та метрології

Кафедра СКС

****

**Курсова робота**

з предмету “Системне програмування”

На тему: “Розробка системних програмних модулів та компонент систем програмування”

Індивідуальне завдання: “Розробка транслятора з вхідної мови програмування”

Варіант № 13

Виконав: ст. групи КІ-308

Мірошниченко К. В.

Прийняв:   
викладач каф. СКС  
Козак Н.Б.

Львів – 2024

**Анотація**

Курсова робота з дисципліни "Системне програмування" вміщає в собі весь матеріал, який ми вивчали протягом даного курсу. Вона узагальнює матеріал і закріпляє навички які ми набули при вивченні матеріалу.

В ній ми маємо продемонструвати розробку транслятора з вхідної мови програмування, яка нам була задана, на мову асемблер, компілювати код і створити виконавчий файл. Транслятор повинен виконувати: лексичний аналіз, синтаксичний аналіз, семантичний аналіз, виводити список помилок при наявності та попереджень.

В цій курсовій роботі буде використовуватися лексичний аналізатор на базі скінченного автомата та на основі магазинного автомата.

Зміст

[**Анотація** 2](#_Toc501563928)

[**Завдання на курсову роботу** 4](#_Toc501563929)

[**1. Огляд методів та способів проектування трансляторів** 6](#_Toc501563930)

[**2. Формальний опис вхідної мови програмування** 8](#_Toc501563931)

[**2.1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура** 8](#_Toc501563932)

[**2.2. Термінальні символи та ключові слова.** 9](#_Toc501563933)

**Завдання на курсову роботу**

Тема: Розробка транслятора з вхідної мови програмування

* Початок програми: program
* Початок блоку даних: variable
* Типи даних: integer\_4
* Оператор введення: input
* Оператор виведення: output
* Блок тіла програми: program; start variable …; stop
* Оператори:
  + Умовні оператори: if-else (С)
  + Переходи: goto (С)
  + Цикли: for-to (Паскаль), for-downto (Паскаль), while (Бейсік), repeat-until (Паскаль)
* Регістр ключових слів: Low
* Регістр ідентифікаторів: Up-Low12 (перший символ Up)
* Арифметичні операції: +, -, \*, /, %
* Операції порівняння: ==, !=, >=, <=
* Логічні операції: !!, &&, ||
* Коментарі: // ... //
* Ідентифікатори змінних, числові та рядкові константи
* Оператор присвоєння: ==>

Для отримання виконавчого файлу з вихідного асемблерного коду потрібно використовувати masm32.exe.

**Вступ**

Компілятор (англ. Compiler від англ. to compile збирати в ціле) - комп'ютерна програма, що перетворює (компілює) програмний код, написаний певною мовою програмування, на семантично еквівалентний код в іншій мові програмування, який, як правило, необхідний для виконання програми машиною.

Транслятор – це той самий компілятор, з тею різницею, що генерує він не об’єктний код, а код на іншій мові програмування.

Процес компіляції як правило складається з декількох етапів: лексичного, синтаксичного та семантичного (типозалежного) аналізів, генерації проміжного коду, оптимізації та генерації результуючого машинного коду. Крім цього, програма як правило залежить від сервісів, наданих операційною системою і сторонніми бібліотеками (наприклад, файловий ввід-вивід або графічний інтерфейс), і машинний код програми необхідно пов'язати з цими сервісами. Для зв'язування зі статичними бібліотеками виконується редактор зв'язків або компонувальник, а з операційною системою і динамічними бібліотеками зв'язування виконується на початку виконання програми завантажувача.

Основні задачі, які виконуються різними компіляторами та трансляторами, по суті, одні і ті ж. Розуміючи ці задачі, існує можливість створювати транслятори для різних початкових мов і цільових машин з використанням одних і тих же базових технологій.

В даній курсовій роботі створюватимется транслятор мови програмування заданої варіантом, який включає:

- лексичний аналізатор, здатний розпізнавати лексеми, що є описані в формальному описі мови програмування.

- синтаксичний аналізатор на основі методу зсув-згортка.

- генератор коду, що генерує код який відповідає кожній конструкції вхідної мови.

Також повинне бути забеспечене виявлення лексичних та синтаксичних помилок, виконана загальна перевірка роботи компілятора.

1. ОГЛЯД МЕТОДІВ ТА СПОСОБІВ ПРОЄКТУВАННЯ ТРАНСЛЯТОРІВ

Створення компіляторів відбувається в певних конкретних умовах: для різних мов, для різних цільових платформ, з різними вимогами для створення компіляторів. Є такі методи створення компіляторів:

1. Прямий метод - цільовою мовою і мовою реалізації є асемблер.

2. Метод розкрути - саме цей метод і використовується у даній курсовій роботі, тобто вибирається інструмент (в даній курсовій це мова асемблер), для якого вже існує компілятор.

3.Використання крос-трансляторів.

4.З використанням віртуальних машин – дає спосіб отримати переносимо програму.

5. Компіляція на ходу.

В даній курсовій роботі згідно із завданням для непарних варіантів необхідно реалізувати висхідний метод граматичного розбору.

При такій стратегії дерево синтаксичного аналізу будується, рухаючись від листя (вхідної програми, яка розглядається як рядок символів) до кореня дерева (аксіоми граматики). Аналізатор (розпізнавач) шукає частину рядка, яку можна звести до нетермінального символу. Таку частину рядка називають фразою. У більшості висхідних розпізнавачів відшукується найлівіша фраза, що безпосередньо зводиться до нетермінального символу (така фраза називається основою). Основа заміняється нетермінальним символом. У отриманому рядку знову відшукується основа, заміняється нетермінальним символом і т.д.

У загальному випадку процедуру побудови висхiдного розпiзнавача за заданою граматикою можна описати в такий спосiб:

1. Визначити для даної граматики функцiї ВПЕРВ i ВПIСЛЯ.
2. Побудувати детермiновану таблицю переходiв, що має по одному стовпцю для кожного граматичного символу i по одному рядку для кожного граматичного входження i маркера дна.
3. Якщо таблиця, побудована на кроцi 2, виходить недетермiнованою (має бiльше одного стану), то потрiбно перетворити цю таблицю в детермiновану, розглядаючи її як недетермiновану таблицю переходiв кiнцевого автомата з початковим станом *h*0.
4. Стани, отриманi на кроцi 3 (крiм стану, що вiдповiдає порожнiй множинi), варто використовувати як магазиннi символи. Отримана таблиця переходiв може мiстити переходи в порожню множину. Такi елементи варто розумiти як забороненi i розглядати переходи в них як помилки.
5. Керуючу таблицю заповнюють рядок за рядком вiдповiдно до множини граматичних входжень, що позначають рядки.

Фази лексичного (ЛА) та синтаксичного (СА) аналізів розкладають початкову програму на частини. Генерація коду проміжною мовою (ГПК), оптимізація (ОК) та генерація коду (ГК) асемблера синтезують програму на вихідній мові. Керування таблицями (КТ) та обробка помилками (ОП) використовуються на всіх фазах трансляції.

Лексичний аналіз об`єднує літери в лексеми - службові слова, ідентифікатори, знаки операцій та пунктуації. Лексеми можна кодувати цілими числами, наприклад, do-одиницею, "+" - двійкою, ідентифікатор - трійкою, константу - четвіркою тощо. До коду лексем ідентифікаторів і констант додається ще одна величина - вид чи значення лексеми.

Синтаксичний аналіз групує лексеми в синтаксичні структури, які можуть бути складовими інших синтаксичних структур; наприклад, А+В може входити в оператор чи вираз. Як рекурсивні структури даних, вони зображуються (явно чи неявно) у формі дерева з лексемами в вузлах і з позначенням синтаксичних конструкцій у внутрішніх вузлах. Крони піддерев відтворюють частини програми відповідної синтаксичної конструкції.

1. ФОРМАЛЬНИЙ ОПИС ВХІДНОЇ МОВИ ПРОГРАМУВАННЯ

2.1. Деталізований опис вхідної мови в термінах розширеної нотації Бекуса-Наура

Розширена форма Бекуса-Наура (РБНФ) — це формальна система визначення синтаксису, в якій синтаксичні категорії послідовно визначаються через інші. Вона використовується для опису контекстно-вільних формальних граматик і була запропонована Никлаусом Віртом. РБНФ є розширеною версією форм Бекуса-Наура (БНФ), що відрізняється більш компактними конструкціями, що спрощують опис при тій самій виразності.

php

Копировать код

<program> ::= program <name> start variable <variable\_block> <code\_block> stop.

<variable\_block> ::= <declarations> {;<declarations>} .

<declarations> ::= <type\_i> <declaration\_i> {,<declaration\_i>} | <type\_b> <declaration\_b> {,<declaration\_b>} .

<type\_i> ::= integer\_4 .

<type\_b> ::= Boolean .

<declaration\_i> ::= <ident> [ ==> <const\_i>] .

<declaration\_b> ::= <ident> [==> <const\_b>] .

<ident> ::= <letter> {<l\_or\_n>} .

<name> ::= <letter> {<l\_or\_n>} .

<l\_or\_n> ::= <letter> | <number> .

<letter> ::= A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|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 .

<number> ::= 0|1|2|3|4|5|6|7|8|9 .

<const\_i> ::= [-]<number>{<number>} .

<const\_b> ::= True | False .

<code\_block> ::= <statement> {<statement>} .

<statement> ::= <equation> | <cycle> | <read> | <write> .

<equation> ::= <ident> ==> <expression\_i> | <expression\_b> ; .

<expression\_i> ::= <term> { + <term> | - <term> } .

<term> ::= <ident> | <const\_i> | <factor> .

<factor> ::= { \* <term> | / <term> | % <term> | <brackets> } .

<brackets> ::= ( <expression\_i> ) .

<expression\_b> ::= <term\_b> { == <term\_b> | != <term\_b> | >= <term\_b> | <= <term\_b> } .

<term\_b> ::= <ident> | <const\_b> | <factor\_b> .

<factor\_b> ::= <term\_b> { || <term\_b> | <&&> | <!!> | <brackets\_b> } .

<and> ::= { &&<term\_b> } .

<not> ::= !! <factor\_b> .

<brackets\_b> ::= ( <expression\_b> ) .

<cycle> ::= for <equation> to <const\_i> do [start] <code\_block> [stop] .

<read> ::= input ( <ident> ) ; .

<write> ::= output ( <expression\_i> | <string> ) ; .

<string> ::= “ {<l\_or\_n>} ” .

2.2. Термінальні символи та ключові слова

* Термінальні символи — це всі символи, що використовуються у конструкціях мови та є «кінцевими» в граматичних правилах. Ось перелік ключових слів та термінальних символів у вхідній мові програмування:
* **program** — початок програми, визначає назву програми.
* **start** — початок блоку програми.
* **variable** — блок змінних.
* **stop** — кінець програми або циклу.
* **input** — оператор введення даних.
* **output** — оператор виведення даних (змінні або рядкові константи).
* **==>** — оператор присвоєння.
* **for ... to ... do** — оператор циклу з умовами.
* **+** — операція додавання.
* **-** — операція віднімання.
* **\*** — операція множення.
* **/** — операція ділення.
* % — операція знаходження залишку від ділення.
* **==** — операція перевірки на рівність.
* **!=** — перевірка на нерівність.
* **>=** — операція «менше або рівно».
* **<=** — операція «більше або рівно».
* !! — операція логічного заперечення.
* **&&** — логічна кон'юнкція (AND).
* **||** — логічна диз'юнкція (OR).
* **integer\_4** — 32-бітові цілі числа.
* **Boolean** — логічні змінні (істинно/хибно).
* **True** та **False** — логічні константи.
* *// ... //* — коментар.
* **"..."** — рядкова константа.
* **;** — кінець оператора.
* **(** та **)** — дужки для виразів.
* **,** — розділювач між деклараціями змінних.

Крім того, в мові використовуються:

Латинські літери (A-Z, a-z).

Арабські цифри (0-9).

Символи табуляції, пробіли та символи переходу на нову стрічку.