Experiment 06 : Two Pass Assembler

Learning Objective: Student should be able to Apply 2 pass Assembler for X86 machine. **Tools:** Jdk1.8,Turbo C/C++, Python, Notepad++

Theory:

An assembler performs the following functions

1 Generate instructions

a. Evaluate the mnemonic in the operator field to produce its machine code.

b. Evaluate subfields- find value of each symbol, process literals & assign address.

2 Process pseudo ops: we can group these tables into passed or sequential scans over input associated with each task are one or more assembler modules.

Format of Databases:

a) POT (Pseudo Op-code Table):-

POT is a fixed length table i.e. the contents of these table are altered during the assembly process.

Pseudo Op-code (5 Bytes character)	Address of routine to process pseudo-op-code. (3 bytes= 24 bit address)
"DROPb"	P1 DROP
"ENDbb"	P1 END
"EQUbb"	P1 EQU
"START"	P1 START
"USING"	P1 USING

- The table will actually contain the physical addresses.
- POT is a predefined table.
- In PASS1 , POT is consulted to process some pseudo opcodes like-DS,DC,EQU
- In PASS2, POT is consulted to process some pseudo opcodes like DS,DC,USING,DROP

b) MOT (Mnemonic Op-code Table):-

MOT is a fixed length table i.e. the contents of these tables are altered during the assembly process.

Mnemonic Op-code (4 Bytes character)	Binary Op-code (1 Byte Hexadecimal)	Instruction Length (2 Bits binary)	Instruction Format (3 bits binary)	Not used in this design (3 bits)

"Abbb"	5A	10	001	
"AHbb"	4A	10	001	
"ALbb"	5E	10	001	
"ALRb"	1E	01	000	

b- Represents the char blanks.

Codes:-

Instruction Length

01= 1 Half word=2 Bytes

10= 2 Half word=4 Bytes

11= 3 Half word=6 Bytes

Instruction Format

000 = RR

001 = RX

010 = RS

011= SI

100= SS - MOT is a predefined table.

- In PASS1 , MOT is consulted to obtain the instruction length.(to Update LC) - In PASS2, MOT is consulted to obtain:
  a) Binary Op-code (to generate instruction)
  b) Instruction length (to update LC)
  c) Instruction Format (to assemble the instruction).

C) Symbol table (ST):

Symbol	Value	Length	Relocation
(8 Bytes charaters)	(4 Bytes Hexadecimal)	(1 Byte Hexadecimal)	(R/A) (1 Byte character)
"PRG1bbb"	0000	01	R
"FOURbbbb"	000C	04	R

- ST is used to keep a track on the symbol defined in the program.
- In pass1- whenever the symbol is defined an entry is made in the ST. - In pass2- Symbol table is used to generate the address of the symbol.

D) Literal Table (LT):

Literal	Value	Length	Relocation (R/A)
= F '5'	28	04	R

- LT is used to keep a track on the Literals encountered in the program.

- In pass1- whenever the literals are encountered an entry is made in the LT.
- In pass2- Literal table is used to generate the address of the literal.

E) Base Table (BT):

Register Availability (1 Byte Character)	Contents of Base register (3 bytes= 24 bit address hexadecimal)
1 'N'	-
2 'N'	-
. . .	-
15 'N'	00

- Code availability-
- Y- Register specified in USING pseudo-opcode.
- N--Register never specified in USING pseudo-opcode.
- BT is used to keep a track on the Register availability.
- In pass1- BT is not used.
- In pass2- In pass2, BT is consulted to find which register can be used as base registers along with their contents.

F) Location Counter (LC):

- LC is used to assign addresses to each instruction & address to the symbol defined in the program.
- LC is updated only in two cases:-

a) If it is an instruction then it is updated by instruction length.
b) If it is a data representation (DS, DC) then it is updated by length of data field

Pass 1: Purpose - To define symbols & literals

1) Determine length of machine instruction (MOTGET)
2) Keep track of location counter (LC)
3) Remember values of symbols until pass2 (STSTO)
4) Process some pseudo ops. EQU
5) Remember literals (LITSTO)

Pass 2: Purpose - To generate object program

1) Look up value of symbols (STGET)
2) Generate instruction (MOTGET2)
3) Generate data (for DC, DS)
4) Process pseudo ops. (POT, GET2)

Data Structures:

Pass 1: Database

1) Input source program
2) Location counter (LC) to keep the track of each instruction location
3) MOT (Machine OP table that gives mnemonic & length of instruction
4) POT (Pseudo op table) which indicate mnemonic and action to be taken for each pseudo-op
5) Literals table that is used to store each literals and its location 6) A copy of input to be used later by pass-2.

Pass 2: Database

1) Copy of source program from Pass1
2) Location counter
3) MOT which gives the length, mnemonic format op-code
4) POT which gives mnemonic & action to be taken
5) Symbol table from Pass1
6) Base table which indicates the register to be used or base register
7) A work space INST to hold the instruction & its parts
8) A work space PRINT LINE, to produce printed listing
9) A work space PUNCH CARD for converting instruction into format needed by loader 10) An output deck of assembled instructions needed by loader.

Algorithm:

Pass 1

1. Initialize LC to 0
2. Read instruction
3. Search for pseudo-op table and process it.
    a. If its a USING & DROP pseudo-op then pass it to pass2 assembler
    b. If its a DS & DC then Adjust LC and increment LC by L
    c. If its EQU then evaluate the operand field and add value of symbol in symbol table
    d. If its END then generates Literal Table and terminate pass1
4. Search for machine op table
5. Determine length of instruction from MOT
6. Process any literals and enter into literal table
7. Check for symbol in label field
    a. If yes assign current value of LC to Symbol in ST and increment LC by length
    b. If no increment LC by length
8. Write instruction to file for pass 2
9. Go to statement 2

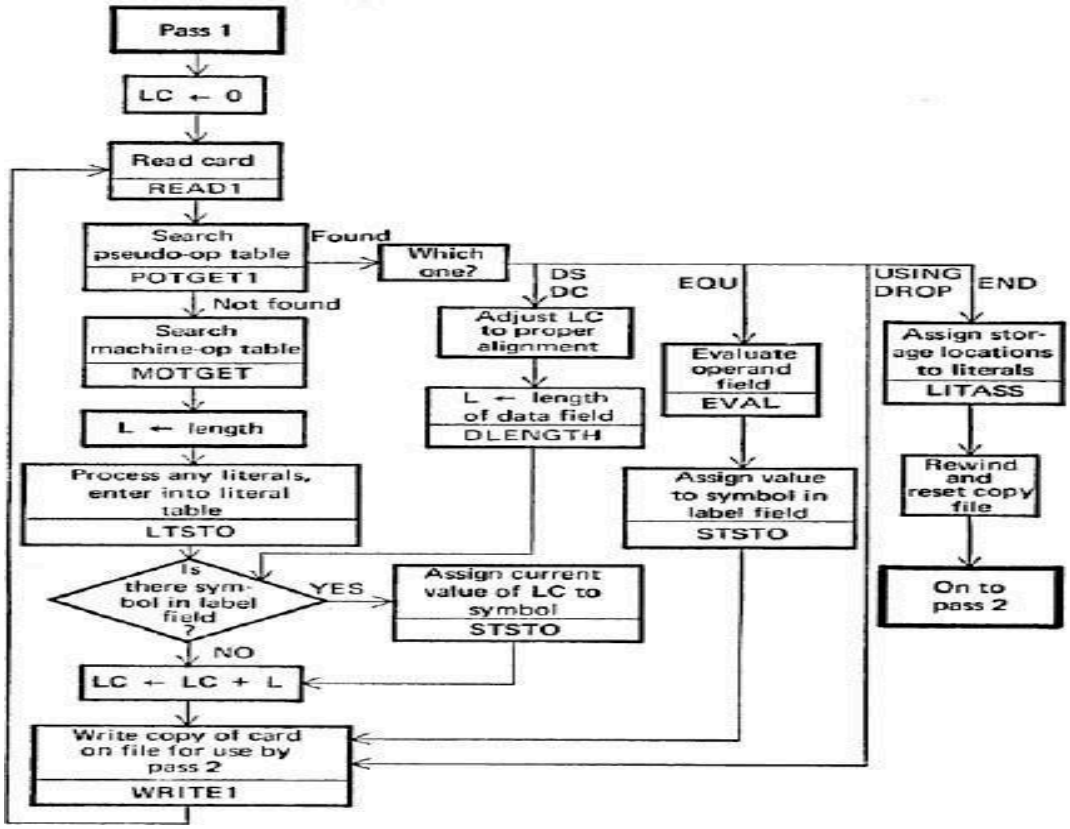
Pass 2

1. Initialize LC to 0.
2. Read instruction
3. Search for pseudo-op table and process it.
    a. If it's a USING then check for base register number and find the contents of the base register
    b. If it's a DROP then base register is not available

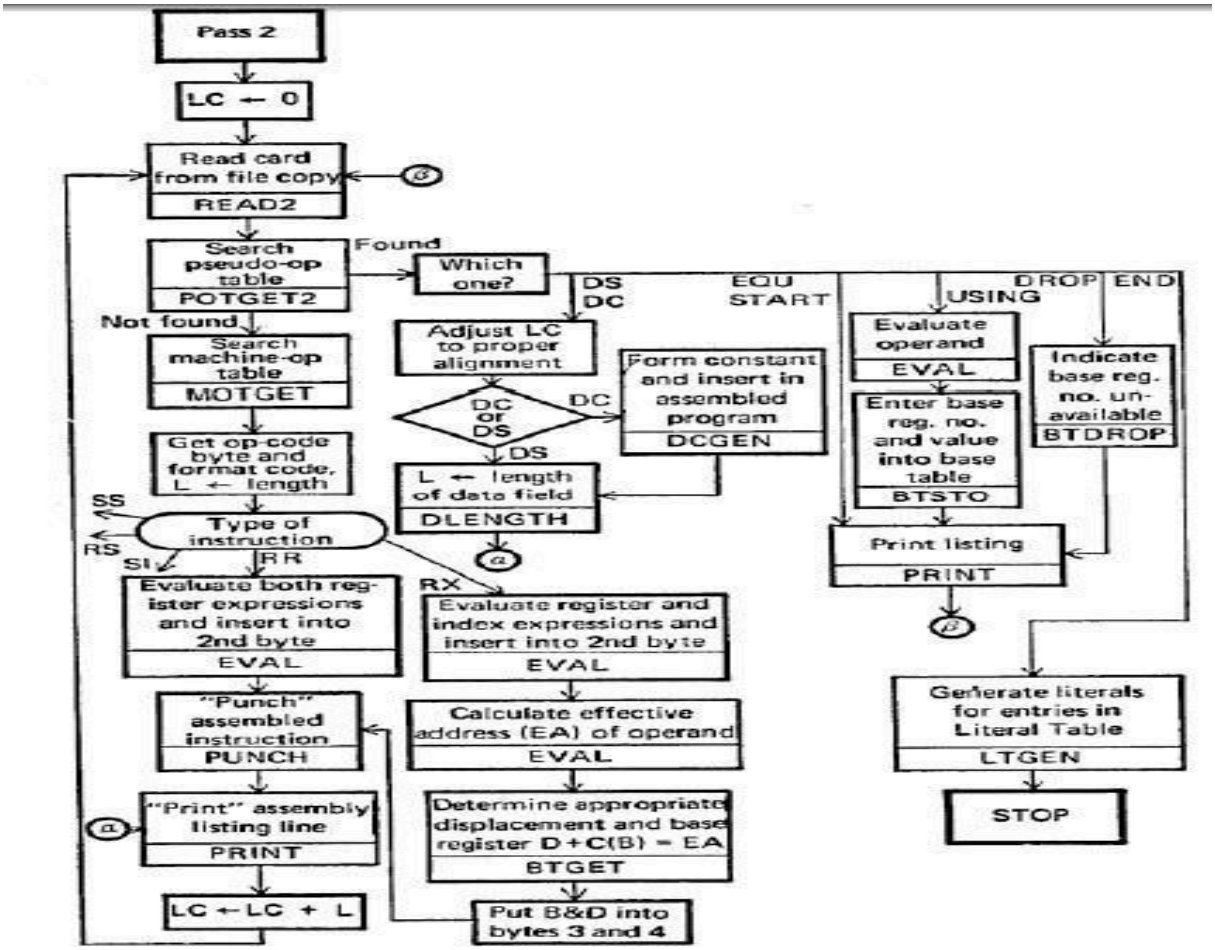
c. If it's a DS then find length of data field and print it
d. If DC then form constant and insert into machine code.
e. If its EQU and START then print listing
f. If its END then generates Literal Table and terminate pass1
g. Generate literals for entries in literal table
h. stop

4. Search for machine op table
5. Get op-code byte and format code
6. Set L = length
7. Check for type of instruction
    a. evaluate all operands and insert into second byte
    b. increment LC by length
    c. print listing
    d. Write instruction to file 8. Go to step 2

Flowchart:Pass1



Flowchart: Pass 2



Example:

```
JOHN START 0
           USING     *, 15
L          1, FIVE
           A         1, FOUR
           ST        1,
                     TEMP

FOUR       DC        F '4'

FIVE       DC        F '5'

TEMP       DS   1F

           END
```

Output: Display as per above format.

Application: To design 2-pass assembler for X86 processor.

Design:

```
SYMBOL_TABLE = {}  # Symbol table to store labels and their addresses
MEMORY = []  # Memory to store machine code
PC = 0  # Program counter

def pass_one(assembly_code):
    global PC
    tokens = assembly_code.split()
    for token in tokens:
        if token.endswith(":"):  # Label definition
            label = token[:-1]
            SYMBOL_TABLE[label] = PC
        else:  # Instruction
            MEMORY.append(token)
            PC += 1
def resolve_symbols():
    global MEMORY
    for i, instruction in enumerate(MEMORY):
        if instruction in SYMBOL_TABLE:
            MEMORY[i] = str(SYMBOL_TABLE[instruction])

def assemble(assembly_code):
    pass_one(assembly_code)
    resolve_symbols()
    print("Opcode Table:")
    print(SYMBOL_TABLE)
    print("Machine Code:")
    print(MEMORY)
assembly_code = """
START:  MOV A, B
        ADD C
        JMP END
LOOP:   SUB D
        JNZ LOOP
END:    HLT
"""
assemble(assembly_code)
```

Result and Discussion:

```
Opcode Table:
{'START': 0, 'LOOP': 7, 'END': 11}
Machine Code:
['MOV', 'A,', 'B', 'ADD', 'C', 'JMP', '11', 'SUB', 'D', 'JNZ', '7', 'HLT']
```

Learning Outcomes: The student should have the ability to
LO1: ***Describe*** the different database formats of 2-pass Assembler with the help of examples.
LO2: **Design** 2 pass Assembler for X86 machine.
LO3: ***Develop***2-pass Assembler for X86 machine.
LO4: Illustrate the working of 2-Pass Assembler.
Course Outcomes: Upon completion of the course students will be able to Describe the various data structures and passes of assembler design.

Conclusion:

For Faculty Use

Correction Parameters	Formative Assessment [40%]	Timely completion of Practical [40%]	Attendance / Learning Attitude [20%]	
Marks Obtained				