EXPERIMENT NO: 02

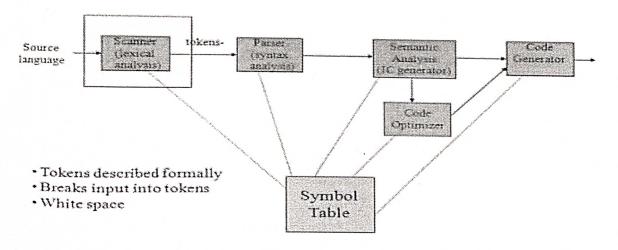
Title: Design a complete lexical analyzer for C language

Aim: Design a complete lexical analyzer for C language

Theory:

Lexical analysis is the process of converting a sequence of characters (such as in a computer program of web page) into a sequence of tokens (strings with an identified "meaning"). A program that perform lexical analysis may be called a lexer, tokenize or scanner.

Lexical Analysis - Scanning



TOKEN

A token is a structure representing a lexeme that explicitly indicates its categorization for the Purpose of parsing. A category of token is what in linguistics might be called a part-of- speech. Examples of token categories may include "identifier" and "integer literal", although the set of Token differ in different programming languages. The process of forming tokens from an input stream of characters is called tokenization. Consider this expression in the C programming language: Sum=3 + 2;

Tokenized and represented by the following table:

Lexeme	Token category		
Sum	"identifier"		
=	"assignment operator"		
3	"integer literal"		
+	"addition operator"		
2	"integer literal"		
:	"end of the statement"		

ALGORITHM:

- 1. Start the program
- 2. Include the header files.
- 3. Allocate memory for the variable by dynamic memory allocation function.
- 4. Use the file accessing functions to read the file.
- 5. Get the input file from the user.
- 6. Separate all the file contents as tokens and match it with the functions.
- 7. Define all the keywords in a separate file and name it as key.c
- 8. Define all the operators in a separate file and name it as open.c
- 9. Give the input program in a file and name it as input.c
- 10. Finally print the output after recognizing all the tokens.
- 11. Stop the program.

```
Program:
#include<stdio.h>
#include<conio.h>
#include<ctype.h>
#include<string.h>
void main()
FILE *fi, *fo, *fop, *fk;
int flag=0,i=1;
char c,t,a[15],ch[15],file[20];
clrscr();
printf("\n Enter the File Name:");
scanf("%s",&file);
fi=fopen(file,"r");
fo=fopen("inter.c","w");
fop=fopen("oper.c","r");
fk=fopen("key.c","r");
c=getc(fi);
while(!feof(fi))
if(isalpha(c)||isdigit(c)||(c=='['||c==']'||c=='.'==1))
```

```
fputc(c,fo);
else
Prepared by-Ramesh Chotiya GCEK, BHAWANIPATNA CSE Department
if(c='\n')
fprintf(fo,"\t$\t");
else fprintf(fo,"\t%c\t",c);
c=getc(fi);
fclose(fi);
fclose(fo);
fi=fopen("inter.c","r");
printf("\n Lexical Analysis");
fscanf(fi,"%s",a);
printf("\n Line: %d\n",i++);
while(!feof(fi))
if(strcmp(a,"\$")==0)
printf("\n Line: %d \n",i++);
fscanf(fi,"%s",a);
fscanf(fop,"%s",ch);
while(!feof(fop))
if(strcmp(ch,a)==0)
fscanf(fop,"%s",ch);
printf("\t\t%s\t:\t%s\n",a,ch);
flag=1;
} fscanf(fop,"%s",ch);
rewind(fop);
fscanf(fk,"%s",ch);
while(!feof(fk))
if(stremp(ch,a)==0)
fscanf(fk,"%k",ch);
printf("\t\t%s\t:\tKeyword\n",a);
flag=1;
Prepared by-Ramesh Chotiya GCEK, BHAWANIPATNA CSE Department
```

```
fscanf(fk,"%s",ch);
rewind(fk);
if(flag==0)
if(isdigit(a[0]))
printf("\t\t%s\t:\tConstant\n",a);
else
printf("\t\t%s\t:\tIdentifier\n",a);
flag=0;
fscanf(fi,"%s",a); }
getch();
Input Files:
Oper.c
(open para
) closepara
{ openbrace
} closebrace
< lesser
> greater
" doublequote ' singlequote
: colon
; semicolon
# preprocessor
= equal
== asign
% percentage
^ bitwise
& reference
* star
+ add
- sub
\ backslash
/ slash
Prepared by-Ramesh Chotiya GCEK, BHAWANIPATNA CSE Department
Key.C
int
void
main
char
if
for
while
```

```
else
printf
scanf
FILE
Include
stdio.h
conio.h
iostream.h
Input.c
#include "stdio.h"
#include "conio.h"
void main()
int a=10,b,c;
a=b*c;
getch();
Output:-
```

Mer the [1]	PEXICUL	PANTAN	S a significant of the				
ine : 1 ways	in a said to be a finer	range place our	week and one in price arrange of the consequent		-		
Sells de marie		the angle of the	prairaceasor	films at the high section of	rese educations, a first	and the second	A Price Value
Appropriate algorith	include.		keyword doublequate	种类的工作的现在形式的	chiedkielast.	- Salanda Salan	THE TANK
	gtdio.h		knyvord				
		*	doublequate				
ine ; 2			1944 1945 N. J. C. S.				
	Me. I daniela	in though a sum	preprocessor	And he divided		a Mizik	497
	include		AU VMU PHI STAR IS TO BE				100
	conio.h		daublequote keyvord	in the second	A due	The Court of	
		Free Arm November 1981	doublequote	MOVEMENT A CARRY LAS	E HOW	a salahan	e Bank fine
HOUSE AND ALL AND		Kathai .	WARRANT NO. 11 THE	MARK TEMPO	1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	1 1
ine : 3					1,30	1000	
	void nain		keyword keyword				
		1777	oponpara				
	>	\$	closepara				
ine : 4							
ine . a		1: 3:1	apenbrace				
			14 30 kg 184				
ina : 5 💮		1.29	* 34.36				
	int.	100	keyward identifier				
	egen and replacement	61 J. SW 210	Torius I per a de la			(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	
	12 10 3 5 5	1	constant			The Contract of	
			identifier identifier			A Charles	
Maria Maria Cara	Selection of the selection	OF SHIP FALLS	Thent Interest	中文的 经 经 经 计	A SA PARA	25. 人名英格兰	
	i kata	1	identifier				
		*	senicalan				
ine : 6							
1110 . 0			ident if ier				
	#	:	oqual				
		i da kara	identifier				
			identifier				
		•	seniculon				
ine : 2							
***************************************	getch	2 11 3 15	identifier				
	₹ 1	4	openuara				
		* ***	closepara seniculon				
			actifico for				
ine : II							
	· ·	•	clusebrace				-
ine : 7							
	W C. Carre	1 2 7 2 2 3 3 3 3 3	identifier			SUPPLY OF THE PARTY OF THE PART	1000