Module 10

Compiler Basics

Module Ten

- Compiler Basics - Part Two
- In this presentation, we are going to talk about :
- Lexical Analysis

Overview

- Previously we talked about:
- Compiler Basic Functions
- Language Definition - Grammar

Now: Lexical Analysis

Lexical Analysis

- Find the **TOKENs**
- Scan the input source program - Scanner
- Recognize: tokens
  - keywords
  - operators
  - Identifiers
  - integers
  - real numbers
  - character strings

A program

```
PROGRAM Stats

VAR
   sum, sumSQ, I, value, mean, variance : INTEGER

BEGIN
   sum    := 0;
   sumSQ  := 0;
   FOR I  := 1 TO 100 DO
        BEGIN
               READ ( value );
               sum    := sum + value;
               sumSQ  := sumSQ + value * value
        END;
   mean     := sum DIV 100;
   variance := sumSQ DIV 100 - mean * mean;
   WRITE ( mean, variance )
END.
```

A program as seen by the Scanner

- as seen by the Scanner:

```
PROGRAM Stats VAR sum,sumSQ,I,value,mean,variance:
INTEGER BEGIN sum:=0;sumSQ:=0;FOR I:=1 TO 100 DO BEGIN
READ(value);sum:=sum+value;sumSQ:=sumSQ+value*valueEND;
mean:=sumDIV100;variance:=sumSQDIV100-mean*mean;
WRITE(mean,variance)END.
```

- A stream of characters

What are the Tokens

- What are some of the tokens?

```
PROGRAM
Stats
VAR
sum
,
sumSQ
,
I
,
Variance
:
INTEGER
```

```
sum
:=
0
;
sumSQ
:=
0
;
```

```
FOR
I
:=
1
TO
100
DO
BEGIN
```

Lexical Analysis

- Convert TOKEN to a code
- Usually an integer plus an optional value
  - BEGIN - 3
  - END. - 5
  - READ - 8
  - WRITE - 9
  - **;** - 12
  - **)** - 21
  - **id** - 22 sumsq
  - **int** - 23 100
- More efficient for the Parser to process

Lexical Analysis

- Different for each programming language
- Is white space significant ?
- Are there Reserved words ? What are they ?
- Identifiers are entered into Symbol Table
- Multi-character tokens

  READ vs R E A D

  Keywords

  Variable names

  Best handled by the scanner. That is its job.

Lexical Analysis

- Integers and Floating point numbers
- Constants can be character strings
- Comments
- Read the source code, write the listing file
- Called by the Parser
- Automatic construction of Scanner

Issues

- Some issues
- FORTRAN
  - `DO I = 1,15`
  - `DO I = 115`
- Reserved words - Key words

```
IF THEN THEN THEN = ELSE ELSE ELSE = THEN
```

- Quoted strings

Summary

- Lexical Analysis
  - Find and identify the TOKENS

Next: Syntax Analysis