

## Lecture # 04

### IDENTIFIERS:

- starts with alpha or -
  - First char can be followed by alphabets, digits, -
  - Only - is not valid ID limitation
- \* Ags Language ki specification ye hai ki
- $$\Sigma = \{ A, D, - \}$$

$$R \cdot E = (-)^+ (A + D) (A + D + -)^* / \\ (A) (A + D + -)^*$$

if limitation is changed.

- Consecutive - are not allowed.

$$R \cdot E = A (A + D + - (A + D))^* (- + \lambda) / \\ (-) (A + D) (A + D + - (A + D))^* \\ (- + \lambda).$$

## Tasks

- language mai identifier k ye rules. batane.
- R.E's for identifier.
- Implementation team register se.

For char const:

A: ' " \ .

always with \

B: n, r, t, b, 0  
with or without \

C: always without \ : +, \*, @ ...

- Const ki R.E batani  
hai specification k mutabiq.  
(Implementation whi hai)

⇒ Program to validate all R.Es

⇒ lexical ki implementation k ye  
tokens ko ek file mai save  
kawana hai

Token.txt (CP, VP, Inot#)

⇒ String specification batani k  
single line hogi ya double line

## Constants (4 classes now)

- 1 INT CONST
  - 2 FLOAT / DOUBLE CONST
  - 3 CHAR CONST
  - 4 STRING CONST

\* INT const.

$$R.C = (+ + - + \wedge) (D)^+$$

\* FLOAT / DOUBLE Const

$$R.B = (+ + - + \wedge) (D)^* \cdot (D)^+ \quad \text{and one } (optional)$$

\* Chas const

$$R.B. : \quad '[(A+B)+B+C]'$$

\* STRING CONST.

"(anychar)\*" "abc" invalid

R.E = "[\A+\B]+\B+\C]\*"

- Language to language checking varies each time

Source file

```
class A-B-C :: a+b-55
while (a<=b++ + 5886)
if-else (a<=!+ = 56
Int ch = "abc + = 56-7,d
return a-b+ //56=abc
ch = b+c
```

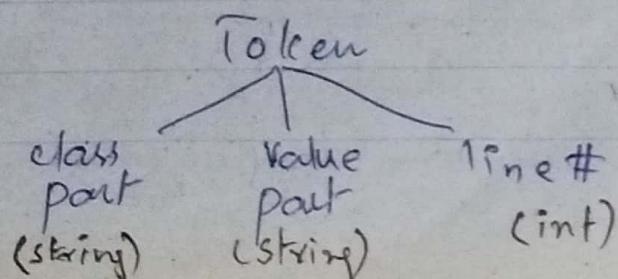
→ LA →

Tokens

(CP, VP, line)
(class, -, 1)
(ID, A-B-C, 1)
(:, -, 1)
(:, -, 1)
(ID, a, 1)
(PM, +, 1)

Input: "class :: b+c"  
temp:

- (structure)
- Token is container that holds the information about valid word



- \* Token banne ke bad source file khatam ho jati
- \* age & phase sof token use karte hain.
- \* LA, SA, Sent in save phases mai error reporting / error detection hoti hai.
- \* Token save kame ke liye ap dynamic data structure use karte hain.
- \* ek Token mai multiple information hoti ye sab mila ke ek entity banti hai. Tu esliye hum Token ki ek class banate hain tak ek es token ke bare mai jo cheezain hain wo ek jaga ek cheer ke liye aya (means object bana de).

class Token {

String CP

String VP

int line #

}

\* Jese ek token aya ek object  
bina us token ka sare informa-  
tion us ki obj mai save  
hoyega, then wo obj ek data  
structure mai kardo.

Input = "class ... b+c"

temp =  
1

- Input file se word part ke  
tak mai ata hai jab tak  
space encounter na hoye.
- Hai char k. & byte hoti  
\* agar CP .. VP same nota hai  
tu syntax token ka  
\*: (CP, VP, line)  
(class, class, 1)  
(class, -, 1)
- \* Punctuator one se word break  
nota hai
- \* Jab punctuator temp mai  
ayega tu q k punctuator  
ek 1 length ke hole isi  
me wo word me break  
kardo.

Specification pe bat ha

- Operator (+) single length k thi  
note tu isi lyce jab + aye ga  
temp mai tu wo next char  
ko read karega agar next  
char se ( ++, +, += ) m se tu  
concatenate karega warna  
break karega kuch he.

\v = line terminator

temp = c55 (line break)

↙  
Valid hai.

(ID, C55, 1)

⇒ line change char ajs rakhia  
hai tu iska token banne ga.  
mazhi hei k rakhne line  
change char rakhya na  
lekin line change pe line#  
increment karega.

2<sup>g</sup> temp = <  
possibility

↙  
<=

Agi = kawa tu concatenate hot  
break hoga ye ga maa

direct break hogaya ga  
⇒ (RO, <=, 2)

⇒ temp =  
⇒ (=, -, 2)

(Int-concat, 5, 2)

Single β allowed nhi tu wo next  
char lena con read karega age  
& hoga du concatenation valid  
Token generate hogi waha X  
wo invalid kardega.

(SS, -, 2),

⇒ (ID, Int, 4)  
(ID, ch, 4)

Int ch"

\* " jab phle bar ata tu phla  
word break hojata hai  
" k bad ek sare break  
check hata dete or closing.  
" one pe break kartे.

(str-const, abc + = 56-7, 4)

\* Tokan mai value part mai  
(") hata dete hai

e.g. (exception)

strmy = "abc\"56-7"

\* Starting " se read karke  
takhe " tu ya tu \" mila ga  
ya " aga \ mila tu next  
part ko string ka part  
bana denge or " one pe  
break kadega

$\Rightarrow$  (str-const, abc \"56-7, 4)

\* words end hamara " pe  
he hata hai

\* Multi lines string handling :

" \_\_\_\_\_ " +

" \_\_\_\_\_ " +

" \_\_\_\_\_ " +

Is scenario mai compiler mai  
hari change mai zarurat wo  
phile us taarek he karega

Agg

15 mai " " search karna ha,  
end of code tak search hoga.

- Agg multi string nahi support karte. Tu agg end " nahi aye line breaker agaya tu break hojaye ga.  
Then phr lexical ab error ayega  
agg token
- Lexical error lea token generate karne hoga.
- Error Reporting karne  
 $\Rightarrow$  temp = "abc \\" 56-7,d  
 $\Rightarrow (\text{invalid}, "abc \\ 56-7,d, 4)$   
Lexeme

- Agg single line string support karne kahin tu ek or bat k agar read karte we \ ayega tu bhi break hojaye ga.

Possibilities

/ / =  
// //

temp = /

- Agar // aagaya tu sab ignore hoga sare checks ignore honge loop chalta raha ga chota ignore hote raha ga jab tak line break na aye.
- Agar /\* aagya thi yaad karte raho jab tak \* na aye thi aagya \* k bad / aagya tu break hoga maini thi chalta raha ga.

## Lecture # 05

Continue...

Char ch = '+' ++ = '\n' + = 5 || n + '\*' = 2

\* character or float pe discussion  
hui

\* f character ko kise break  
karte hui;

token:

(10, Char, 1), (1D, ch, 1), (=, -, 1)

\* temp = ''

\* ('') phela jab aye ga to wo  
word - break hua dega.

⇒ first ('') ke bad ya ek char  
aye ga yo 2 means length ya 3  
hoga ya 4.                              eg 't' = 3  
  '\n' = 4

⇒ ('') quote k bad ending ('')  
wahi se change balle ye  
det ('') quote k bad ek char  
lazmi pahne ga agi \ howa  
tu next char lazmi pahne  
ga warna first char pahne  
break hua dega.

Requirement = character of word having length k  
will be valid or not  
(char-cont, t, l)

(inc-dec, t, l)  
(cont, l)

temp = 1

next two word part k by word  
break koudega chhavi wo 2.  
waqt jo bhi hai phr check  
leenga valid word or not.

(cont compend, t, l)  
(int-cont, 5, l)

temp = 111

(invalid "111")  
new

The next word dekhai ga

if

1 t = 5

2 abc...

character multi line whi hoga kha

line change character ka case m  
mostly word break houdega.

\* multi lines comment mai line  
break ka word ko break nahi  
karta.

\* DOT :

① 959.99.556 → lena ekawali  
a.b55.c  
a99.556.cbsa  
55b.sbc.abc

lenan ekawali  
ayege syntan  
euro aye ge.

⇒ 9.9

↳ is Scienctific mai ek  
character q, ye ek  
pure word hai

⇒ a.b

↳ break karega word  
ko

⇒ dot ko word ka part kab  
& bana chahiye.

① ⇒ temp = 959

Two checks k dot character  
ka part banege

→ check temp contain only digits

→ dot k bad wala

character digit hai

ja nahi bagdisi ko

concatenate kya baghais

hai  
agr checker (True) too dot  
concatenate ho jaye.

$$\text{temp} = 959.99$$

• phr dot aya second dot word  
breaker hai q k checker  
findil whi nowA

• agr • phela character ho  
tu.

$$\text{temp} = .$$

• check k , k bad digit  
hai agr hai tu check kare  
ge next <sup>char</sup> digit hai agr hai  
wo tw. word continue kare  
ga.

⇒ (float-const, 959.99, 1)

(float-const, -556, 1)

② ⇒ a.b55.c i + n = a

↳ word break kare  
ga q k temp mai digit whi

• alphabets or digits kehni thi word  
break into range  
toe (10, 9, 2)

next

temp = \*

→ token (., -12)

q k next char digit nahi

→ (ID, 655, 2)

(', 1, 2)

(ID, 0, 2)

③ → 99.556 : cb5a 4 tokens

ck  
word

recent debtain sf

④ 55b, 5bc : abc

word

word

word

error

error

word

Implementation &  
coding structure.

1. 5 RE regex implement  
karni

2. Language w words kesi data  
structure mai daina

3. if else k checks fis pe  
word break hogi.



isStr-const() return bool  
isInt-const() return bool  
isFloat-const() return bool

main()  
{

str Input = Read-Input

while (length of Input) {

str temp = breakword(Input, index)

function space, line break  
char, comments ko return

int category

breakword ek word

break kar ke take dega

• keyword alph se start hoga.

• id alpha or -

• float & digits / .

• int C digits

• String " "

• character ' '

• word break jo ho k aye  
ga us k first character  
se decide hogya k his  
boundary se pass karwao

if switch (temp[0])

{  
Case '-' : if (isID(temp)) {

tok = new Token

tok.CP = 'ID'

tok.VP = temp

• live ~~at~~ =

TS.add(tok)

} Else {

tok = new Token()

tok.CP = 'Lexem Error'

tok.VP = temp

TS.Add(tok)

}

Case 'alph':

esa koi keyword nahi jo valid  
identifier na ho to is hi  
the ID check hogi then  
keyword. Tum agar ID true  
hoga to the int token k  
mila ID ka token bana den  
bulk phle keyword ka check  
hoga

if (isID(temp)) {

str = isKW(temp)

LA submission

2<sup>nd</sup> - Sep - 2022

1st Week - Word

if (str == "") breaker

{ . C.P = "ID" bana hona  
} chariye.

else { + text file mai  
output hona

• CP = str chariye.

}

3 else {

error

}

float

Case = '.' ; dot itself

① photo temp ki length check  
karna.

② 1 # word breaker mila hai  
deafault mai jake error generate  
hoga.

# Syntax Analyzer

- totally CFG pe depend

## Overview of CFG.

- CFG language ko model karke  
hai in 4 cheezon:-
  - Terminals : anything jo CP mei  
- Non terminal : mai cost nahi hoga NT
  - Starting NT : hoi.
  - Production Rule.
- Terminal jo extend nahi karke
- Non terminal jo extend karke

e.g

$$S \rightarrow a b c$$

↳ production rule.

- Starting terminal hoi to
- Khud call hota hoi

⇒ Syntax Analyzer token input  
lelia hoi

Token mai

CP VP time

↓ ↗ ye consider nahi karte  
Syntax  
Analyzer mai use hota

2 k data type honi chaiye  
konsi honi ye matter nahi,  
AM hona chaiye konsa ye  
matter nahi karla.

Terminals → Tokens, CP

for example,

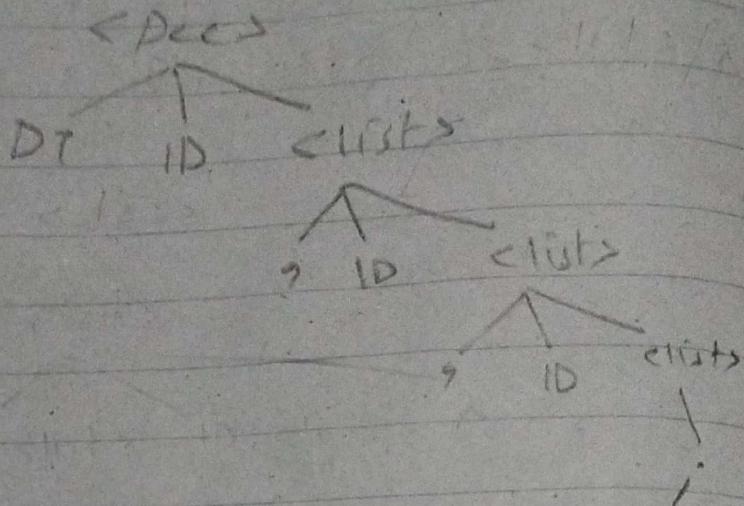
• int a,b,c;

Assignment ko agr abhi ignore  
karde tu DT k bad ID  
thim c/i list or k ek se ziyada ha  
< Dec > → DT ID < list > banayega  
< list > → ; / , ID < list >

• CFG ki strength recursion

• PDA stack ko use karta  
nhi / means recursion he  
nhi

int a, b, c; d;  
DT ID, ID, ID; ID;



Simple declaration CFG

<Decls> → DT ID <list>  
<list> → ; / , ID <list>

- tree should be complete
- input should be completely parse.
- Agar ye dons conditions true  
hoga tu syntan echi hai  
rules hamare honge.

Assignment simple grammar ke  
declare or assign any type of  
constant, any type of var

int a,b,c,d=5,x=d; Valid

float b=x,d=c=5=x,z; Invalid

CF by initialize karne ke  
no bhi sakhti ya nahi  
anytype of constant or value  
bhi.

parse tree bhi banan