

5.10.2021

Seminar 2 - The scanning algorithm

Inputs: source code (<- code.txt) + lexical tokens (<- tokens.txt)

Outputs: - PIF(Program Internal Form) + ST(Symbol Table), if source is lexically correct

- lexical errors, otherwise

EXAMPLE: *B: Tiutin Andrada-Georgia*

Code.txt

```
program test;  
var a : integer;  
begin  
a := b+a+3;  
write("test");  
end.
```

PIF

Token	ST_pos
program	-1
id	0
;	-1
var	-1
id	1
:	-1
integer	-1
;	-1
begin	-1
id	1
:=	-1
id	2
+	-1
id	1
+	-1
const	3
;	-1
write	-1

(-1
const	4
)	-1
;	-1
end	-1
.	-1

ST (only identifiers and constants)

ST_pos	Symbol
0	test
1	a
2	b
3	3
4	"test"

Lexical error = unclassified token

Ex.: \$a – unrecognized alphabet symbol '\$'

1abc – invalid identifier

"mystring, mystring" - invalid string char

'a, a', 'abc', " - invalid char constant