

UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação
Bacharelado em Ciência da Computação

Especificações dos Tokens

Hellena Almeida Canuto
João Vitor Santos Tavares
Wagner Anthony de Medeiros Silva

Maceió - AL, 17/12/2021

1. Linguagem de implementação

Os analisadores da linguagem Linkin Park foram implementados em Python (Versão 3.9.0).

2. Enumeração das categorias de Tokens

```
from enum import Enum

class Token(Enum):

    # Identifiers

    ID = 1

    # Operations

    OP_SUM = 2
    OP_SUB = 3
    OP_DIV = 4
    OP_MUL = 5
    OP_GREATER = 6
    OP_LESS = 7
    OP_EQUAL = 8
    OP_EQUALG = 9
    OP_EQUALL = 10
    OP_AND = 11
    OP_OR = 12
    OP_NOT = 13
    OP_NOTUN = 14
    OP_ATR = 15
    OP_MOD = 16
    OP_CONCAT = 27

    # Reserved words

    RW_FUNCTION = 17
    RW_RETURN = 18
    RW_WHILE = 19
    RW_FOR = 20
    RW_IF = 21
    RW_ELSE = 22
    RW_FLOAT = 23
    RW_INT = 24
    RW_CHAR = 25
    RW_STR = 26
    RW_BOOL = 28
    RW_TRUE = 29
```

```
RW_FALSE = 30
RW_NULL = 31
RW_OPEN = 32
RW_CLOSE = 33
RW_SCAN = 34
RW_PRINT = 35
RW_PRINTNL = 36
RW_MAIN = 37
RW_EMPTY = 38
BOOL_VALUE = 39
CTE_INT = 40
CTE_FLOAT = 41
CTE_CHAR = 42
CTE_STR = 43
```

```
# Delimiters
```

```
DEL_OPEN = 44
DEL_CLOSE = 45
DEL_OPENP = 46
DEL_CLOSEP = 47
DEL_ENDBRA = 48
DEL_OPENBRA = 49
DEL_COMMA = 50
DEL_SEMI = 51
```

```
# Errors
```

```
ER_UNK= 52
ER_ID = 53
ER_NUM = 54
ER_RW = 55
ER_CHAR = 56
```

```
# Others
```

```
EOF = 57
COMMENT = 58
```

3. Expressões Regulares Auxiliares

- `small_letters = '[:lower:]'`
- `capital_letters = '[:upper:]'`
- `digits = '[:digit:]'`

- alphanumeric = '[:alphanum:]'
- symbol = '[:punct:]{-}[\']'
- alpha_num_symb = ' [{alphanumeric}{symbol}]'
- sequential_char = '\ {alpha_num_symb}+[:space:]+[:word:]+\ '

4. Expressões Regulares dos Lexemas

4.1. Identificadores ERRADO

- ID - [{letra_maiuscula}{alfanumérico}][{1,16}]

4.2. Palavras Reservadas

RW_FUNCTION	'Function'
RW_RETURN	'Return'
RW_IF	'If'
RW_ELSE	'Else'
RW_WHILE	'While'
RW_FOR	'For'
RW_INT	'Int'
RW_FLOAT	'Float'
RW_CHAR	'Char'
RW_STR	'Str'
RW_BOOL	'Bool'
RW_INPUT	'Input'
RW_PRINT	'Print'
RW_TRUE	'True'
RW_FALSE	'False'
RW_NULL	'Null'
RW_EMPTY	'Empty'

RW_OPEN	'Open'
RW_CLOSE	'Close'
RW_MAIN	'Main'

DEL_BOOL	'True False'
CTE_FLOAT	[:digit:]+
CTE_INT	
CTE_CHAR	
CTE_STR	

4.3. Operadores

OP_ATR	'='
OP_EQUAL	'==' '!='
OP_SUM	'+'
OP_SUB	'_'
OP_MUL	'*'
OP_DIV	'/'
OP_MOD	'%'
OP_GREATER	'>'
OP_LESS	'<'
OP_EQUALG	'>='
OP_EQUALL	'<='
OP_AND	'And'
OP_OR	'Or'
OP_NOT	'!'
OP_NOTUN	'~'

OP_CONCAT	'&'
-----------	-----

4.4. Delimitadores

DEL_OPEN	'Open'
DEL_CLOSE	'Close'
DEL_OPENP	'('
DEL_CLOSEP	')'
DEL_ENDBRA	'['
DEL_OPENBRA	']'
DEL_COMMA	','
DEL_SEMI	','

4.5 Erros Léxicos

ER_UNK	Símbolo desconhecido
ER_ID	Identificador mal formado
ER_NUM	Número mal formado
ER_RW	Palavra reservada mal formada
ER_CHAR	Caracter ou cadeia mal formado