

Preparação para o Analisador Léxico**Exercício:**

O objetivo deste exercício é construir uma função **proximo_token()** que realiza a análise léxica de um programa escrito na linguagem C. A função deve ler o programa em C de um arquivo e identificar os tokens da linguagem, retornando uma estrutura com as informações referentes ao token encontrado. As definições regulares para os tokens serão dadas abaixo. O arquivo de entrada deve ser fornecido, com a extensão **.c**, para o programa através da linha de comando.

Exemplo: Se o programa a ser analisado possuir o nome **ex03.c**, e o programa fonte sendo **lexico.py**, a execução deve ser:

C:\>python lexico.py ex03.c

A função **proximo_token()** deve fazer um controle das linhas e colunas do programa fonte e também fazer a eliminação dos delimitadores (espaços em branco, tabulação (\t), nova linha (\n) e retorno de carro (\r). Para cada token reconhecido devem ser mostrados os seguintes valores:

- token
- descrição
- linha
- coluna
- valor
- lexema

1. Definição dos Tokens que são Palavras Reservadas

As palavras chaves da linguagem são consideradas palavras reservadas, ou seja, não podem ser utilizadas como identificadores. Para simplificar a função de análise léxica, o reconhecimento de palavras chaves será feito com base na definição regular de identificadores, sendo que a determinação do token associado à palavra chave é feita por uma busca em uma tabela de palavras reservadas. A busca na tabela de palavras reservadas deve ser implementada utilizando um dicionário para tabela de palavras reservadas, lembrando que algumas palavras reservadas no C também são palavras reservadas no Python.

- Abaixo está a tabela com as palavras reservadas da linguagem C.

auto

break

case

char

const	continue	default	do
double	else	enum	extern
float	for	goto	if
int	long	register	return
short	signed	sizeof	static
struct	switch	typedef	union
unsigned	void	volatile	while

2. Definição dos tokens sem atributos

ATRIBUICAO	=
PONTO	.
ABRE_PAR	(
FECHA_PAR)
PONTO_VIRGULA	;
VIRGULA	,
SUBTRACAO	-
ADICAO	+
MULTIPLICACAO	*
DIVISAO	/
MODULO	%

3. Definição do token OP_RELACIONAL

OP_RELACIONAL → **< | <= | = | <> | >=**

Onde o atributo associado ao token **OP_RELACIONAL** deverá ser as constantes simbólicas:

LT	<
LE	<=
EQ	=
NE	!=
GT	>
GE	>=

4. Definição do token OP_LOGICO

Onde o atributo associado ao token **OP_LOGICO** deverá ser as constantes simbólicas:

AND	&
OR	
NOT	!

5. Definição do token IDENTIFICADOR

LETRA	[_A-Za-z]
DIGITO	[0-9]
IDENTIFICADOR	LETRA(LETRA DIGITO)*

Quando um IDENTIFICADOR é reconhecido, deve-se fazer uma busca na tabela de palavras reservados; se for encontrado, deve-se retornar o token associado com a palavra reservada; caso contrário, o token IDENTIFICADOR deve ser retornado

6. Definição do token NUMERO_INTEIRO e NUMERO_REAL

NUMERO_INTEIRO	DIGITOS+
OP_FRACAO	.DIGITOS*
OP_EXP	((E e) (+ - λ) DIGITO+)
OP_EXP	NUMERO_INTEIRO OP_FRACAO (OP_EXP λ)

O atributo para NUMERO_INTEIRO ou NUMERO_REAL é o valor da constante numérica que gerou o token.

7. Definição do token FRASE

FRASE	“ CHAR1* ”
--------------	-------------------

CHAR1 → {qualquer caractere ASCII menos os caracteres de retorno de carro (\r) e avanço de linha (\n) }

Quando a frase tiver “ (aspas) esta deve ser precedida pelo caracter ‘\’.
O atributo para FRASE é um apontador para a cadeia de caracteres.

8. Definição do token COMENTÁRIO

CHAR2	{qualquer caractere ASCII }
COMENTARIO	/* CHAR2* */

Para COMENTARIO deve ser retornado token correspondente, e o controle de linhas deve ser mantido dentro do comentário. COMENTARIO não possui atributo.

Observações sobre a Atividade Contínua:

- Esta Atividade Contínua poderá ser feita em grupo de no máximo 6 alunos.
- Os trabalhos serão entregues até a data indicada abaixo.
- Os trabalhos deverão ser entregues pelo Classroom.
- Entregar também um arquivo contendo o nome dos integrantes do grupo.
- Apresentar um breve relatório com observações, comentários e dificuldades encontradas para a execução do trabalho.

Data de Entrega do Trabalho: Dia 17 de Setembro de 2022.