

Primeiro trabalho de Compiladores

Profº Leonardo Takuno

Analizador Léxico

O objetivo do trabalho é construir uma rotina (**AnaLex**) que realiza a análise léxica de um programa escrito na linguagem PORTUGOL. A rotina deve ler o programa em PORTUGOL de um arquivo e identificar os átomos da linguagem. Retornando uma estrutura com as informações referentes ao átomo encontrado. As definições regulares para os átomos serão dadas abaixo. O arquivo de entrada deve ser fornecido, com a extensão **.PTL**, para o programa através da linha de comando.

Exemplo: Se o programa a ser analisado possuir o nome **TESTE.PTL**, e o programa fonte sendo **Lexico.C**, a execução deve ser:

```
C:\>Lexico TESTE.PTL
```

A rotina **AnaLex** deve fazer um controle das linhas do programa fonte e também fazer a eliminação dos delimitadores (espaços em branco, tabulação, nova linha e retorno de carro). Para cada átomo reconhecido devem ser mostrados os seguintes valores:

```
{Número da Linha do Átomo, Átomo, Lexeme }
```

Definição dos Átomos que são Palavras Reservadas

As palavras chaves da linguagem são consideradas palavras reservadas, ou seja, não podem ser utilizadas como identificadores. Para simplificar a rotina de análise léxica, o reconhecimento de palavras chaves será feito com base na definição regular de identificadores, sendo que a determinação do átomo 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 uma busca binária na tabela de palavras reservadas.

Átomos Retornados	Descrição
ALGORITMO	Início do algoritmo
ATE	Inicializa a variável de controle do PARA
CADEIA	Tipo de dado para sequência de caracteres
CHARACTER	Tipo de dado para definir um caracter
ENQUANTO	Determina um laço com condição no início
ENTAO	usado no bloco de estrutura de seleção
FACA	usado no bloco de estrutura de repetição
FIM	Fim de um bloco ou do programa
FUNCAO	Determina uma função no algoritmo
INICIO	Início do algoritmo
INTEIRO	Tipo de dado para números inteiros
PARA	Determina um laço com variável de controle
PASSO	Incremento da variável de controle na repetição
PROCEDIMENTO	Determina um procedimento no algoritmo
REAL	Tipo de Dado para números reais (ponto flutuante)
REF	Operador para de referência para variáveis
RETORNE	Retorna de uma função
SE	Determina uma estrutura de seleção

SENAO	Caso contrário da instrução de seleção
VARIAVEIS	Início do bloco de declaração de variáveis

Obs: As palavras Reservadas podem ser informadas com caracteres maiúsculos ou minúsculos, ou seja, aLgoriTmo o átomo retornado deve ser ALGORITMO

Definição dos átomos sem atributos

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

Definição do átomo OP_RELACIONAL

OP_RELACIONAL → < | <= | = | > | >=

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

ME → <

MEI → <=

IG → =

DI → >

MA → >

MAI → >=

Definição do átomo OP_LOGICO

OP_LOGICO → & | \$ | !

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

E → &

OU → \$

NEG → !

Definição do átomo 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 átomo associado com a palavra chave; caso contrário, o átomo **IDENTIFICADOR** deve ser retornado

Definição do átomo NUMERO_INTEIRO e NUMERO_REAL

NUMERO_INTEIRO → DIGITOS+

OP_FRACAO → .DIGITOS*

OP_EXP → ((E|e) (+ | - | λ) DIGITO+)

NUMERO_REAL → NUMERO_INTEIRO OP_FRACAO (OP_EXP | λ)

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

Definição do átomo FRASE

CHAR1 → {qualquer caractere ASCII menos os caracteres de *retorno de carro e avanço de linha* }

FRASE → “ CHAR1* ”

Quando a frase tiver “ (aspas) esta deve ser precedida pelo caracter ‘\’.

O atributo para **FRASE** é um apontador para a cadeia de caracteres.

Definição do átomo COMENTÁRIO

CHAR2 → {qualquer caractere ASCII }

COMENTARIO → /* CHAR2* */

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

Obs: O programa deve estar bem documentado. O trabalho poderá ser feito em grupo de no máximo 2 alunos, qualquer tentativa de fraude será punida com a **nota zero**. Os trabalhos serão recolhidos na data abaixo. O trabalho deverá ser entregue em disquete com etiqueta dos integrantes do grupo, além de um breve relatório sobre o funcionamento, dificuldades encontradas e demais comentários.

Data de Entrega do Trabalho: Dia 30 de Março de 2018