



Construção de Compiladores

Alterado por: prof. Rodrigo Pereira

Alterado por: prof. Marlon Oliveira

Elaborado por: Prof^a Christine Vieira

Análise Léxica

A entrada para um compilador é a cadeia de caracteres que representa o **programa fonte**.

O objetivo do bloco léxico é quebrar esta cadeia de caracteres separando as palavras que nela estão representadas.

Por exemplo, a cadeia de entrada poderia ser:

```
RESULTADO:=RESULTADO+15
```

variável RESULTADO
comando de atribuição
variável RESULTADO
operação aritmética
constante numérica

Assim, os 23 caracteres da cadeia de entrada são transformados em 5 novas entidades/**tokens**

Se a tabela de tokens for vista como um dicionário, então o processo léxico pode ser visto como o ato de agrupar letras em palavras e determinar suas localizações no dicionário.

Outras funções normalmente atribuídas ao bloco léxico são as de ignorar espaços em branco e comentários, além de detectar os erros léxicos.

A saída do analisador léxico é uma sequência de símbolos que é passada para a próxima fase, o analisador sintático.

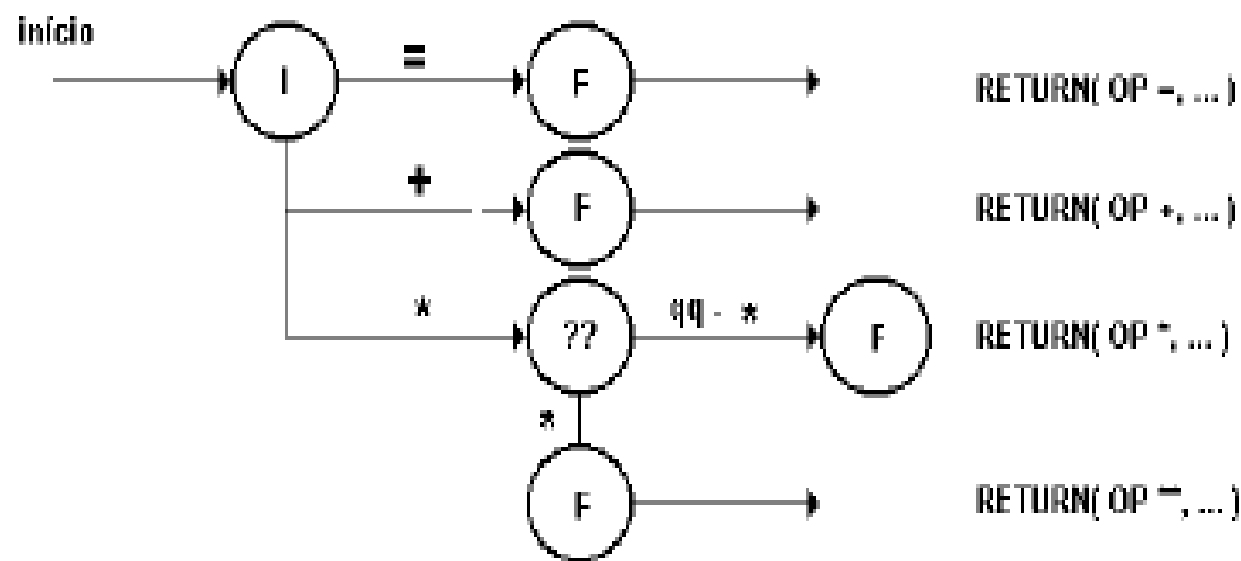
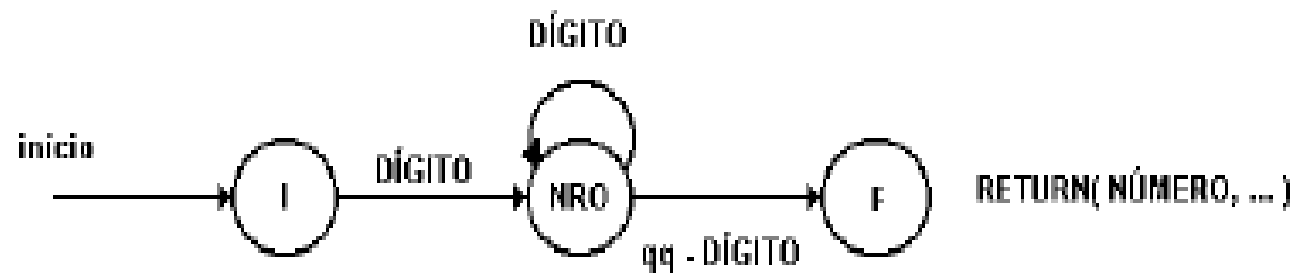
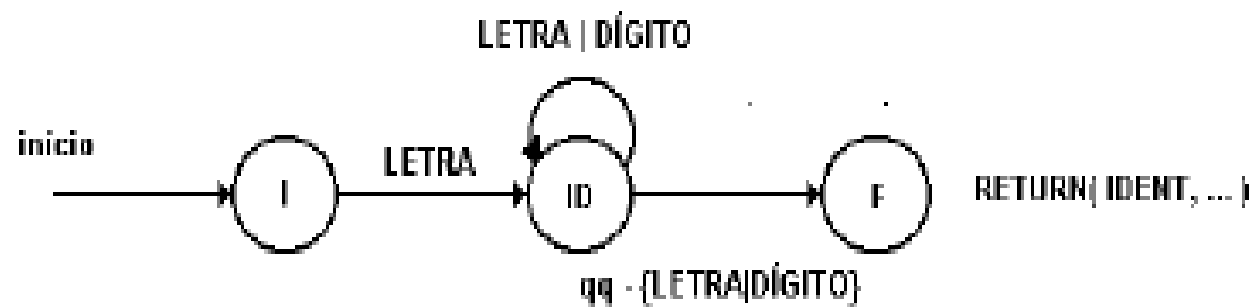
Os símbolos nessa sequência são geralmente representados por códigos (valores inteiros).

Os símbolos reconhecidos pelo léxico se enquadram em uma das seguintes classes:

- Identificadores (do usuário, da linguagem - palavras-chaves, funções embutidas);
- Números (inteiros, reais, complexos);
- Operadores (aritméticos, relacionais, lógicos);
- Parentisadores (abre/fecha aspas, abre/fecha parênteses, abre/fecha colchetes, abre/fecha chaves BEGIN/END);
- Sinais de pontuação (ponto, vírgula, ponto-e-vírgula)

DESCRIÇÃO DE SÍMBOLOS COM AUTÔMATOS FINITOS

Uma forma de descrever os símbolos de uma language seria com o uso de autômatos de estados finitos (que têm a vantagem de permitir uma tradução quase direta para código de programação).



Obter algoritmos a partir dos autômatos é uma tarefa bastante simples.

RECONHECIMENTO DE IDENTIFICADOR

```
procedimento identificador;  
{  
    enquanto caracter é LETRA ou DÍGITO faça {  
        símbolo = símbolo || caracter;  
        caracter = leia_caracter();  
    }  
  
    return( 1, símbolo);  
}
```

RECONHECIMENTO DE NÚMERO

```
procedimento número;  
{  
    enquanto character é DÍGITO {  
        símbolo = símbolo || character;  
        character = leia_character();  
    }  
  
    return( 2, símbolo );  
}
```

IMPLEMENTAÇÃO DO ANALISADOR LÉXICO

Para implementar o analisador léxico mais facilmente podemos agrupar os autômatos para reconhecimento de identificadores, números e operadores em um único autômato que seria o próprio analisador.

