

UNIVERSIDADE SÃO JUDAS TADEU

GRUPO

Emilio Gabriel dos Santos Palacios	RA 823112595
Geovanna Camily da Rocha Guedes	RA:823130579
Guilherme Nunes Pereira	RA:823149439
Gustavo Baeza Calça	RA: 82319972
Gustavo Soares Rodrigues	RA:823133821

TEORIA DA COMPUTAÇÃO E COMPILADORES

PROJETO A3 - PDF ANALISADOR LÉXICO

Análise Léxica

Na ciência da computação, análise léxica é o processo de converter uma sequência de caracteres (como em um programa de computador ou página da web) em uma sequência de tokens (strings com um significado atribuído e, portanto, identificado).

A análise léxica pode ser dividida em duas etapas, a primeira chamada de escandimento que é uma simples varredura removendo comentários e espaços em branco, e a segunda etapa, a análise léxica propriamente dita onde o texto é quebrado em lexemas.

Podemos definir três termos relacionados a implementação de um analisador léxico:

- **Padrão:** é a forma que os lexemas de uma cadeia de caracteres podem assumir. No caso de palavras reservadas é a sequência de caracteres que formam a palavra reservada, no caso de identificadores são os caracteres que formam os nomes das variáveis e funções.
- **Lexema:** é uma sequência de caracteres reconhecidos por um padrão.
- **Token:** é um par constituído de um nome e um valor de atributo esse último opcional. O nome de um token é um símbolo que representa a unidade léxica. Por exemplo: palavras reservadas; identificadores; números, etc.

A tabela abaixo mostra os exemplos de uso dos termos durante a análise léxica.

Token	Padrão	Lexema	Descrição
<const, >	Sequência das palavras c, o, n, s, t	const	Palavra reservada
<while, >	Sequência das palavras w, h, i, l, e	while, While, WHILE	Palavra reservada
<if, >	Sequência das palavras i, f	If, IF, iF, If	Palavra reservada
<=, >	<, >, <=, >=, ==, !=	==, !=	
<numero, 18>	Dígitos numéricos	0.6, 18, 0.009	Constante numérica
<literal, "Olá">	Caracteres entre ""	"Olá Mundo"	Constante literal
<identificador, 1>	Nomes de variáveis, funções, parâmetros de funções.	nomeCliente, descricaoProduto, calcularPreco()	Nome de variável, nome de função
<=, >	=	=	Comando de atribuição
<{, >	{, }, [,]	{, }, [,]	Delimitadores de início e fim

Compilação X Interpretação

A compilação é o processo pelo qual o código fonte escrito em uma linguagem de alto nível é traduzido para código de máquina ou bytecode executável. Este processo é realizado por um compilador e geralmente envolve várias etapas: análise léxica, análise sintática, análise semântica, otimização e geração de código. O código compilado é então executado diretamente pelo hardware ou por uma máquina virtual.

A interpretação, por outro lado, envolve a execução do código fonte diretamente, sem a necessidade de traduzi-lo para código de máquina antecipadamente. Um intérprete lê e executa o código linha por linha ou bloco por bloco. Isso pode levar a uma execução mais lenta em comparação com o código compilado, pois o processo de tradução ocorre em tempo de execução.

Mecanismos Envolvidos

Autômatos Finitos:

A análise léxica geralmente utiliza autômatos finitos (AFDs ou AFNs) para a identificação de tokens. Um autômato finito é uma máquina de estado que lê uma sequência de caracteres e muda de estado com base na entrada atual e no estado corrente. Se, ao final da entrada, o autômato estiver em um estado de aceitação, a entrada é reconhecida como um token válido.

Expressões Regulares:

Expressões regulares são padrões utilizados para descrever conjuntos de cadeias de caracteres. Elas são amplamente utilizadas na definição de tokens. Ferramentas como Lex e Flex utilizam expressões regulares para gerar autômatos finitos que realizam a análise léxica.

Exemplo: Para identificar um identificador (variável) que começa com uma letra seguida por letras ou dígitos: `[a-zA-Z][a-zA-Z0-9]*`

Tabela de Símbolos

É uma estrutura de dados gerada pelo compilador com o objetivo de armazenar informações e melhorar o fluxo de caracteres sobre os nomes (identificadores de variáveis, de

parâmetros, de funções, de procedimentos, etc) definidos no programa fonte. Essa tabela associa atributos (tipo, escopo, limites no caso de vetores e número de parâmetros no caso de funções) aos nomes que foram definidos pelo programador. A construção dessa tabela, em geral, se dá durante a análise léxica, quando os identificadores são reconhecidos.

Quando um identificador é encontrado pela primeira vez, o analisador léxico armazena-o na tabela sem condições de associar atributos a esse identificador. Geralmente essa tarefa é executada durante as fases de análise sintática e semântica. Existem vários modos de se organizar e acessar as tabelas de símbolos sendo, os mais comuns: Listas lineares, árvores binárias e tabelas hash.