

DOCUMENTAÇÃO ANALISADOR LÉXICO

Gabriel Benfatti de Campos gabriel-benfatti.campos@unesp.br

Murilo Tolentino Rufino murilo.t.rufino@unesp.br

1. Introdução

Um analisador léxico tem como objetivo identificar erros léxicos, ou seja, verificar se a grafia das palavras se adequam à linguagem. Este processo é chamado de tokenização. Neste modelo o analisador verifica números, palavras, identificadores e símbolos.

Com a utilização da ferramenta Flex (fast lex) podemos fazer a tokenização usando técnicas de expressões regulares, que traduzem automatos finitos. Um exemplo que podemos citar é a expressão regular que verifica o identificador:

`ID [a-z][a-z0-9_]*`

Isso quer dizer que um identificador é caracterizado por letras minúsculas e números e o caracter '_' e deve começar com uma letra.

OBS: As funções de palavras reservadas e tipos de dados não foram desenvolvidas conforme as orientações da professora.

2. Descrição do analisador

Primeiramente os arquivos interpretados devem conter a extensão 'exemplo.mg'.

O analisador tem uma variável 'line' que conta as linhas do código fonte para poder apresentar em qual linha um erro foi verificado, por exemplo. Esse contador é incrementador ao se verificar o caractere "\n".

Em seguida há a definição de algumas expressões, são elas: ID (identificador), DIGIT (número), WORD (palavra), WHITESPC (espaço), ENTER ('\n'), TAB ('\t') e EQUAL('=').

Por fim há a verificação dos símbolos. Estes são divididos em 3 subclasses visando facilitar a análise sintática que virá depois. As subclasses são: 1- Símbolos Aritméticos [+ - * / % ++ -- += -=], 2- Símbolos Lógicos [NOT ! AND & OR | < <= > >= == !=], 3- Símbolos Diversos [' () [] ; , -> { }].

Além disso o analisador entende comentários representados por "#" até o final da linha.

3. Erros

Há três mensagens de erro possíveis. Uma quando o arquivo código fonte não é encontrado no diretório. Uma quando o identificador não cumpre os requisitos (começar com uma letra). E uma quando o analisador se depara com um símbolo não especificado.

4. Exemplo

O arquivo exemp.mg contém alguns comandos da linguagem C e apresenta um erro de identificador na linha 4.