

Universidade Federal de Goiás
Curso de Bacharelado em Engenharia da Computação
Compiladores 2019-2
Compilador para a Linguagem Cafezino
Etapa I - Análise Léxica

Adaptado de Thierson Couto Rosa

1 Objetivo

Este projeto tem como objetivo a implementação de um compilador didático para a linguagem *Cafezinho*. O presente texto descreve a Etapa I do projeto que consiste na especificação do analisador léxico para a linguagem.

A Seção 2 descreve a gramática da linguagem *Cafezinho*. A Seção 3 descreve o trabalho de análise léxica. A seção 4 descreve os detalhes de como a Etapa I do projeto deve ser entregue.

2 Gramática para a linguagem *Cafezinho*

A seguir, é apresentada a gramática para a linguagem *Cafezinho*. Os não-terminais da gramática iniciam-se com letra maiúscula e o símbolo inicial é *Programa*. Os terminais aparecem em negrito quando são formados por palavras ou por sequência de caracteres. Exemplo: **int**, **+**, **-** **()**, **<=**, etc.

Gramática:

Programa → *DeclFuncVar DeclProg*

DeclFuncVar → *Tipo id DeclVar ; DeclFuncVar*
| *Tipo id [intconst] DeclVar ; DeclFuncVar*
| *Tipo id DeclFunc DeclFuncVar*
| ϵ

DeclProg → **programa** *Bloco*

DeclVar → **,** *id DeclVar*
| **,** *id [intconst] DeclVar*
| ϵ

<i>DeclFunc</i>	→ (<i>ListaParametros</i>) <i>Bloco</i>
<i>ListaParametros</i>	→ ε <i>ListaParametrosCont</i>
<i>ListaParametrosCont</i>	→ <i>Tipo id</i> <i>Tipo id</i> [] <i>Tipo id</i> , <i>ListaParametrosCont</i> <i>Tipo id</i> [] , <i>ListaParametrosCont</i>
<i>Bloco</i>	→ { <i>ListaDeclVar</i> <i>ListaComando</i> } { <i>ListaDeclVar</i> }
<i>ListaDeclVar</i>	→ ε <i>Tipo id DeclVar</i> ; <i>ListaDeclVar</i> <i>Tipo id</i> [intconst] <i>DeclVar</i> ; <i>ListaDeclVar</i>
<i>Tipo</i>	→ int car
<i>ListaComando</i>	→ <i>Comando</i> <i>Comando</i> <i>ListaComando</i>
<i>Comando</i>	→ ; <i>Expr</i> ; retorne <i>Expr</i> ; leia <i>LValueExpr</i> ; escreva <i>Expr</i> ; escreva “cadeiaCaracteres” ; novalinha ; se (<i>Expr</i>) entao <i>Comando</i> se (<i>Expr</i>) entao <i>Comando</i> senao <i>Comando</i> enquanto (<i>Expr</i>) execute <i>Comando</i> <i>Bloco</i>
<i>Expr</i>	→ <i>AssignExpr</i>
<i>AssignExpr</i>	→ <i>CondExpr</i> <i>LValueExpr</i> = <i>AssignExpr</i>
<i>CondExpr</i>	→ <i>OrExpr</i> <i>OrExpr</i> ? <i>Expr</i> : <i>CondExpr</i>
<i>OrExpr</i>	→ <i>OrExpr</i> ou <i>AndExpr</i> <i>AndExpr</i>

<i>AndExpr</i>	→	<i>AndExpr</i> e <i>EqExpr</i> <i>EqExpr</i>
<i>EqExpr</i>	→	<i>EqExpr</i> == <i>DesigExpr</i> <i>EqExpr</i> != <i>DesigExpr</i> <i>DesigExpr</i>
<i>DesigExpr</i>	→	<i>DesigExpr</i> < <i>AddExpr</i> <i>DesigExpr</i> > <i>AddExpr</i> <i>DesigExpr</i> >= <i>AddExpr</i> <i>DesigExpr</i> <= <i>AddExpr</i> <i>AddExpr</i>
<i>AddExpr</i>	→	<i>AddExpr</i> + <i>MulExpr</i> <i>AddExpr</i> - <i>MulExpr</i> <i>MulExpr</i>
<i>MulExpr</i>	→	<i>MulExpr</i> * <i>UnExpr</i> <i>MulExpr</i> / <i>UnExpr</i> <i>MulExpr</i> % <i>UnExpr</i> <i>UnExpr</i>
<i>UnExpr</i>	→	- <i>PrimExpr</i> ! <i>PrimExpr</i> <i>PrimExpr</i>
<i>LValueExpr</i>	→	id [<i>Expr</i>] id
<i>PrimExpr</i>	→	id (<i>ListExpr</i>) id () id [<i>Expr</i>] id carconst intconst (<i>Expr</i>)
<i>ListExpr</i>	→	<i>AssignExpr</i> <i>ListExpr</i> , <i>AssignExpr</i>

3 Analisador Léxico

A programa desenvolvido em python deve ser capaz de reconhecer os *tokens* da gramática da linguagem *Cafezinho* especificada na Seção 2. Deverá também processar e descartar comentários. Os comentários podem ocupar mais de uma linha do programa fonte. Um comentário em *Cafezinho* inicia com o par de símbolos “/*” e termina com o par de símbolos “*/”. O analisador léxico deverá reportar erro, caso um comentário não termine. As palavras

reservadas de *Cafezinho* são: **programa**, **car**, **int**, **retorne**, **leia**, **escreva**, **novalinha**, **se**, **entao**, **senao**, **enquanto**, **execute**. No caso em que o *token* é uma constante *string* (**constString**), ou um identificador (**id**), a função deve gerar como lexema, o texto que forma a *string* ou o identificador (ex.: “x1”, “cont2”).

O analisador léxico deve reportar através de mensagem impressa erros léxicos encontrados no arquivo de entrada. São exemplos de erros léxicos: caracteres inválidos na linguagem, comentários que não terminam, cadeia de caracteres que não terminam ou que ocupam mais de uma linha no arquivo de entrada.

Caso o arquivo contenha um erro léxico, o programa deverá imprimir uma linha contendo exatamente o seguinte: **ERRO:**, seguido por um espaço em branco e, por uma das seguintes mensagens de erro: **CARACTERE INVÁLIDO** ou **COMENTÁRIO NAO TERMINA** ou **CADEIA DE CARACTERES OCUPA MAIS DE UMA LINHA**. Após a mensagem de erro o programa deve imprimir, na mesma linha, o número da linha do programa fonte onde o erro foi encontrado.

4 Informações Sobre a Implementação e a Entrega

O trabalho poderá ser feito em dupla e deve ser entregue até o dia 03/11/2019 via tarefa criada no sistema SIGAA. Devem ser entregues:

- O código python de entrada para o gerador de analisador léxico utilizado (arquivo com extensão “.py”).

Importante:

Cópias idênticas ou modificadas dos códigos ou de partes dos códigos são consideradas plágios. **Plágio é crime**. O aluno que cometer plágio em seu trabalho receberá nota zero no trabalho.