

Descrição Geral da Linguagem Simple C

Felipe Gomes da Silva
Luis Henrique Salomão Lobato

Setembro, 2025

1 Linguagem Regular - Analisador Léxico

1.1 Tokens Reconhecidos

O analisador léxico da linguagem Simple C toma a decisão de aceitar ou não determinada sentença com base nos tokens destacados na Tabela 1

Tabela 1: Tabela de Tokens do Simple C

Nome do Token	Lexema(s)	Descrição
Palavras-chave: Tipos de Dados		
KEYWORD_INT	int	Palavra-chave para tipo inteiro.
KEYWORD_FLOAT	float	Palavra-chave para tipo ponto flutuante.
KEYWORD_CHAR	char	Palavra-chave para tipo caractere.
KEYWORD_STRING	string	Palavra-chave para tipo string (cadeia de caracteres).
KEYWORD_VOID	void	Palavra-chave para tipo vazio/nulo.
KEYWORD_BOOL	bool	Palavra-chave para tipo booleano.
Palavras-chave: Controle de Fluxo e Comandos		
KEYWORD_IF	if	Inicia uma estrutura condicional.
KEYWORD_ELSE	else	Bloco alternativo de uma estrutura condicional.
KEYWORD_FOR	for	Inicia um laço de repetição 'for'.
KEYWORD_WHILE	while	Inicia um laço de repetição 'while'.
KEYWORD_DO	do	Inicia um laço de repetição 'do-while'.
KEYWORD_SWITCH	switch	Inicia uma estrutura de seleção múltipla.
KEYWORD_CASE	case	Define um rótulo dentro de um 'switch'.
KEYWORD_DEFAULT	default	Define o rótulo padrão de um 'switch'.
KEYWORD_BREAK	break	Interrompe a execução de um laço ou 'switch'.
KEYWORD_CONTINUE	continue	Pula para a próxima iteração de um laço.
KEYWORD_RETURN	return	Retorna um valor de uma função.
Identificadores e Literais		
IDENTIFICADOR	var, _x, ...	Nome de variável, função, etc.
INT	123, 42	Valor literal inteiro.
FLOAT	3.14, 0.5	Valor literal de ponto flutuante.
CHAR	'a', '\n'	Valor literal de caractere.
STRING	"hello"	Valor literal de string.
BOOLEAN_LITERAL	true, false	Valor literal booleano.
Operadores		
OP_SOMA	+	Operador de adição.
OP_SUB	-	Operador de subtração.
OP_MULT	*	Operador de multiplicação.
OP_DIV	/	Operador de divisão.
OP_MOD	%	Operador de módulo (resto da divisão).
OP_ATRIBUICAO	=	Operador de atribuição simples.
OP_INC_ATRIBUICAO	+=	Operador de atribuição com adição.
OP_DEC_ATRIBUICAO	-=	Operador de atribuição com subtração.
OP_MULT_ATRIBUICAO	*=	Operador de atribuição com multiplicação.

Tabela 1: Tabela de Tokens do Simple C (Continuação)

Nome do Token	Lexema(s)	Descrição
OP_DIV_ATRIBUICAO	/=	Operador de atribuição com divisão.
OP_INC	++	Operador de incremento.
OP_DEC	-	Operador de decremento.
OP_IGUAL	==	Operador relacional de igualdade.
OP_DIFERENTE	!=	Operador relacional de desigualdade.
OP_MENOR	<	Operador relacional menor que.
OP_MAIOR	>	Operador relacional maior que.
OP_MENOR_IGUAL	<=	Operador relacional menor ou igual que.
OP_MAIOR_IGUAL	>=	Operador relacional maior ou igual que.
OP_AND	&&	Operador lógico E (AND).
OP_OR		Operador lógico OU (OR).
OP_NOT	!	Operador lógico de negação (NOT).
Pontuadores e Delimitadores		
PONTO_VIRGULA	;	Finalizador de instrução.
DOIS_PONTOS	:	Usado em casos de 'switch'.
VIRGULA	,	Separador de elementos (ex: em listas).
ABRE_PARENTESES	(Abre lista de parâmetros ou expressão.
FECHA_PARENTESES)	Fecha lista de parâmetros ou expressão.
ABRE_CHAVES	{	Abre um bloco de código.
FECHA_CHAVES	}	Fecha um bloco de código.
ABRE_COLCHETES	[Abre a declaração/acesso de um array.
FECHA_COLCHETES]	Fecha a declaração/acesso de um array.

1.2 Descrição

A linguagem regular estruturada para a aceitação de cadeias pelo analisador léxico da linguagem Simple-C pode ser descrita através das expressões regulares e definições presentes na Tabela 2. Deste modo, é garantido que variáveis não sejam declaradas com números como primeiro caractere da cadeia, além de outros erros léxicos devidamente discutidos.

Tabela 2: Expressões regulares do analisador léxico.

Expressão Regular	Descrição	Tipo
<code>[a-zA-Z_][a-zA-Z0-9_]*</code>	Reconhece identificadores válidos que começam com uma letra ou um sublinhado.	Identificador
<code>0 [+-]?[1-9][0-9]*</code>	Reconhece números inteiros.	Literal
<code>[+-]?([0-9]+\.[0-9]* \.[0-9]+)</code>	Reconhece números de ponto flutuante.	Literal
<code>//.*</code>	Comentário de linha única, que começa com <code>//</code> e vai até o final da linha.	Comentário
<code>/*</code>	Inicia um comentário de bloco.	Comentário
<code><IN_COMMENT>"*/"</code>	Fecha um comentário de bloco.	Comentário
<code>"(\. \^[^\n])+"</code>	Conteúdo de uma string, incluindo caracteres escapados.	String
<code>'(\. \^[^\n])+'</code>	Reconhece literais de caracteres.	Caractere
<code>==, !=, <=, >=, etc.</code>	Operadores de comparação, lógicos e aritméticos.	Operador
<code>;, <, >, {, }</code>	Símbolos de pontuação e agrupamento.	Símbolo
<code>{INT}{ID}</code>	Captura identificadores que começam com um número, que é um erro léxico.	Erro
<code>.</code>	Captura qualquer caractere que não corresponda a nenhuma regra.	Erro
<code>[+-]?0[0-9]+</code>	Captura números inteiros com zeros à esquerda, que é um erro léxico.	Erro

2 Linguagem Livre de Contexto - Analisador Sintático

2.1 Gramática da Linguagem Livre de Contexto

A partir da definição dos tokens na Seção 1.1 com a Tabela 1, a linguagem Simple-C pode ser gerada pela gramática livre de contexto definida pela tupla na Equação 1.

Neste contexto, Σ é o conjunto de variáveis não terminais (alfabeto), P é o conjunto de regras de produção, S é o símbolo inicial da derivação e A é o conjunto de variáveis não terminais.

$$G = (\Sigma, P, S, A) \quad (1)$$

Conjunto de Símbolos Terminais (Σ) Este é o conjunto de símbolos terminais (tokens) que formam as sentenças da linguagem.

$$\Sigma = \left\{ \begin{array}{lll} \text{IDENTIFICADOR,} & \text{PONTO_VIRGULA,} & \text{OP_ATRIBUICAO,} \\ \text{OP_SOMA,} & \text{OP_SUB,} & \text{OP_MULT,} \\ \text{OP_DIV,} & \text{OP_INC,} & \text{OP_DEC,} \\ \text{KEYWORD_INT,} & \text{KEYWORD_FLOAT,} & \text{KEYWORD_STRING,} \\ \text{KEYWORD_BOOL,} & \text{INT,} & \text{FLOAT,} \\ \text{STRING,} & \text{CHAR,} & \text{BOOLEAN_LITERAL,} \\ \text{ABRE_PARENTESSES,} & \text{FECHA_PARENTESSES,} & \text{ABRE_CHAVES,} \\ \text{FECHA_CHAVES,} & \text{KEYWORD_IF,} & \text{KEYWORD_ELSE,} \\ \text{KEYWORD_FOR,} & \text{KEYWORD_WHILE,} & \text{KEYWORD_DO,} \\ \text{OP_IGUAL,} & \text{OP_DIFERENTE,} & \text{OP_MENOR,} \\ \text{OP_MAIOR,} & \text{OP_MENOR_IGUAL,} & \text{OP_MAIOR_IGUAL,} \end{array} \right\}$$

Conjunto de Regras de Produção (P) : O conjunto das regras de produção é definido por:

$\langle \text{Programa} \rangle \rightarrow \langle \text{ListaDeComandos} \rangle$

$\langle \text{ListaDeComandos} \rangle \rightarrow \langle \text{Comando} \rangle \mid \langle \text{Comando} \rangle \langle \text{ListaDeComandos} \rangle$

$\langle \text{Comando} \rangle \rightarrow \langle \text{Declaracao} \rangle \mid \langle \text{Atribuicao} \rangle \mid \langle \text{EstruturaDeControle} \rangle \mid \langle \text{ChamadaDeFuncao} \rangle \mid \langle \text{Bloco} \rangle$

$\langle \text{Declaracao} \rangle \rightarrow \langle \text{Tipo} \rangle \text{ IDENTIFICADOR } \langle \text{AtribuicaoOpcional} \rangle \text{ PONTO_VIRGULA}$

$\langle \text{Atribuicao} \rangle \rightarrow \text{IDENTIFICADOR } \text{OP_ATRIBUICAO } \langle \text{Expressao} \rangle \text{ PONTO_VIRGULA}$

$\langle \text{AtribuicaoOpcional} \rangle \rightarrow \text{OP_ATRIBUICAO } \langle \text{Expressao} \rangle \mid$

$\langle \text{Tipo} \rangle \rightarrow \text{KEYWORD_INT} \mid \text{KEYWORD_FLOAT} \mid \text{KEYWORD_STRING} \mid \text{KEYWORD_BOOL}$

$\langle \text{Literais} \rangle \rightarrow \text{INT} \mid \text{FLOAT} \mid \text{STRING} \mid \text{CHAR} \mid \text{BOOLEAN_LITERAL}$

$\langle \text{Expressao} \rangle \rightarrow \langle \text{Termo} \rangle \mid \langle \text{Expressao} \rangle \text{ OP_SOMA } \langle \text{Termo} \rangle \mid \langle \text{Expressao} \rangle \text{ OP_SUB } \langle \text{Termo} \rangle$

$\langle \text{Termo} \rangle \rightarrow \langle \text{Fator} \rangle \mid \langle \text{Termo} \rangle \text{ OP_MULT } \langle \text{Fator} \rangle \mid \langle \text{Termo} \rangle \text{ OP_DIV } \langle \text{Fator} \rangle$

$\langle \text{Fator} \rangle \rightarrow \text{IDENTIFICADOR} \mid \langle \text{Literais} \rangle \mid \text{ABRE_PARENTESSES } \langle \text{Expressao} \rangle \text{ FECHA_PARENTESSES} \mid \text{OP_INC} \mid \text{OP_DEC}$

$\langle \text{EstruturaDeControle} \rangle \rightarrow \langle \text{If} \rangle \mid \langle \text{For} \rangle \mid \langle \text{While} \rangle$

$\langle \text{If} \rangle \rightarrow \text{KEYWORD_IF ABRE_PARENTESSES} \langle \text{Condicao} \rangle \text{FECHA_PARENTESSES} \langle \text{Comando} \rangle$
 $\langle \text{ElseOpcional} \rangle$

$\langle \text{ElseOpcional} \rangle \rightarrow \text{KEYWORD_ELSE} \langle \text{Comando} \rangle \mid \epsilon$

$\langle \text{Condicao} \rangle \rightarrow \langle \text{Expressao} \rangle \langle \text{OperadorRelacional} \rangle \langle \text{Expressao} \rangle$

$\langle \text{OperadorRelacional} \rangle \rightarrow \text{OP_IGUAL} \mid \text{OP_DIFERENTE} \mid \text{OP_MENOR} \mid \text{OP_MAIOR} \mid$
 $\text{OP_MENOR_IGUAL} \mid \text{OP_MAIOR_IGUAL}$

$\langle \text{Bloco} \rangle \rightarrow \text{ABRE_CHAVES} \langle \text{ListaDeComandos} \rangle \text{FECHA_CHAVES}$

Conjunto de Não-Terminais (A) O conjunto de símbolos não-terminais é definido por:

$$A = \left\{ \begin{array}{lll} \langle \text{Programa} \rangle, & \langle \text{ListaDeComandos} \rangle, \langle \text{Comando} \rangle, \\ \langle \text{Declaracao} \rangle, & \langle \text{Atribuicao} \rangle, & \langle \text{AtribuicaoOpcional} \rangle, \\ \langle \text{Tipo} \rangle, & \langle \text{Literais} \rangle, & \langle \text{Expressao} \rangle, \\ \langle \text{Termo} \rangle, & \langle \text{Fator} \rangle, & \langle \text{EstruturaDeControle} \rangle, \\ \langle \text{If} \rangle, & \langle \text{ElseOpcional} \rangle, & \langle \text{Condicao} \rangle, \\ \langle \text{OperadorRelacional} \rangle, & \langle \text{Bloco} \rangle, & \langle \text{ChamadaDeFuncao} \rangle, \\ \langle \text{For} \rangle, & \langle \text{While} \rangle, & \langle \text{DoWhile} \rangle \end{array} \right\}$$

Símbolo Inicial (S) O símbolo inicial da gramática é:

$$S = \langle \text{Programa} \rangle$$