

COMPILADORES E INTÉRPRETES

Proyecto
Aspectos Léxicos de MINIJAVA
Segundo Cuatrimestre de 2022

1 Introducción

Este documento describe parte de la sintaxis de MINIJAVA. Como se menciona en el documento de especificación del proyecto, MINIJAVA es un lenguaje que puede verse como una simplificación de Java, donde por una parte no se consideran elementos avanzados como Genericidad, Excepciones o Hilos, y por otra parte la estructura de su sintaxis es más estricta. En este documento, se presentan los elementos léxicos correspondientes a la sintaxis del lenguaje, donde en particular se identificarán las palabras reservadas, forma de los identificadores y literales del lenguaje.

2 Componentes Léxicos

El primer paso para compilar un programa de MINIJAVA radica en convertir una entrada de caracteres en una entrada de tokens, donde cada token se corresponderá con un lexema de MINIJAVA. Los lexemas de MINIJAVA son palabras clave, nombres de tipos primitivos, literales, símbolos de puntuación, operadores e identificadores. Cada uno de estos elementos se describe en las siguientes secciones.

2.1 Espacios en Blanco

Los espacios en blanco en MINIJAVA, al igual que en Java, simplemente hacen más fácil la lectura del programa por un humano y **no** son tokens. Un espacio en blanco es un espacio, tab o un salto de línea.

2.2 Palabras Clave

MINIJAVA tiene las siguientes palabras clave:

class	interface	extends	implements	
public	private	static		
void	boolean	char	int	
if	else	while	return	var
this	new	null	true	false

Note que varias de las palabras clave originales de Java no están presentes en MINIJAVA. Un compilador de MINIJAVA podría considerar esas palabras clave no presentes como prohibidas en MINIJAVA.

2.3 Identificadores

En MINIJAVA hay dos tipos de identificadores: los identificadores de clases y los identificadores de métodos y variables. Un identificador de clase es una letra mayúscula seguida de cero o más letras (mayúsculas o minúsculas), dígitos y underscores. Un identificador de método y variable es una letra minúscula seguida de cero o más letras (mayúsculas o minúsculas), dígitos y underscores.

En los documentos con la gramática de MINIJAVA los identificadores de clase son denotados como **idClase**, mientras que los identificadores de metodos y variables son denotados como **IdMetVar**. Ningún identificador puede ser una palabra reservada.

2.4 Comentarios

Los comentarios son ignorados por el analizador léxico, análogamente a los espacios en blanco, simplemente hacen más fácil la lectura del programa por un humano y **no** son tokens. MINIJAVA, al igual que Java, tiene dos estilos de comentarios:

- `/* comentario */` Comentarios multi-línea: Todos los caracteres desde `/*` hasta `*/` son ignorados.
- `// comentario` Comentario simple: Todos los caracteres de desde `//` hasta el final de la línea son ignorados.

2.5 Literales

En MINIJAVA hay cinco tipos de literales, los cuales se corresponden a los tipos primitivos del lenguaje.

2.5.1 Literales Enteros

Un literal entero es una secuencia de uno o mas dígitos. El literal entero no puede contener mas de 9 dígitos y su valor corresponde a su interpretación estándar en base 10. En el documento con la gramática de MINIJAVA los literales enteros son denotados como **intLiteral**.

2.5.2 Literales Caracteres

Un literal valido caracter tener la forma:

- `'x'` donde x es cualquier caracter excepto la barra invertida (`\`) o la comilla simple (`'`).
- `'\x'` donde x es cualquier caracter.

En el documento con la gramática de MINIJAVA los literales caracter son denotados como **charLiteral**.

2.5.3 Literales String

Un literal string se representa mediante una comilla doble (`"`) seguida de una secuencia de caracteres y finaliza con otra comilla doble (`"`). Al igual que en Java la secuencia de caracteres no puede ser interrumpida por un salto de linea. También como en Java, es posible que en la secuencia de caracteres aparezca una comilla doble siempre y cuando se la anteceda por el caracter `\`. Por esto último la comilla doble que marca el final de un string correcto nunca estará antecedido por ese caracter. El valor del literal corresponde a la cadena de caracteres entre las comillas.

En el documento con la gramática de MINIJAVA los literales string son denotados como **stringLiteral**.

2.5.4 Literales Booleanos

Los literales booleanos se representan mediante las palabras reservadas **true** y **false**.

2.5.5 Literal Nulo

El literal nulo se representa mediante la palabra reservada **null**.

2.6 Puntuación

MINIJAVA cuenta con los siguientes símbolos de puntuación, utilizados para estructurar los programas:
() {} ; . .

2.7 Operadores

El lenguaje MINIJAVA cuenta con los siguientes operadores que serán utilizados en sus expresiones:

>	<	!
==	>=	<=
+	-	*
&&		%

2.8 Asignación

Al igual que en Java, en MINIJAVA el simbolo de asignación básica es el `=`. Además MINIJAVA es posible realizar las asignaciones compuestas `+=` y `-=` que se corresponden con las de Java.

2.9 Otros caracteres

Cualquier otra entrada que no conforme con las reglas anteriormente presentadas debe generar un error.