

Compiladores

Proyecto de Curso

Generalidades

El proyecto a desarrollar durante este curso consiste en la implementación de un compilador para el lenguaje Mini-Pascal, un subconjunto del lenguaje Pascal, cuya sintaxis se define más adelante en este documento.

El código objeto a generar será una maquina virtual de arquitectura RISC, denominada MIPS R2000, para lo cual se utilizará un simulador PCSPIM con capacidad de correr en un ambiente Windows. Se utilizarán los generadores JFlex y CUP para su implementación y el lenguaje a utilizar será Java.

Mini-Pascal es un lenguaje *case insensitive*, es decir que no se hace diferencia entre mayúsculas y minúsculas.

El proyecto se desarrollará en grupos de dos alumnos y en dos fases, una en cada curso de la clase de Compiladores, éstas se describen a continuación:

Fase No.1

La primera fase del proyecto consiste en crear el analizador léxico y sintáctico. El proyecto deberá crear un AST, el cual debe poderse visualizar de alguna forma por el usuario. El proyecto a presentar debe considerar:

- Los comentarios en Mini-Pascal se encierran entre { y }. Es importante apuntar que los comentarios no pueden ser anidados y pueden extenderse sobre varias líneas. El compilador deberá ser capaz de ignorar los comentarios y los caracteres de tabulación como ser: nueva línea, tabulación y espacio blanco.
- Se manejarán variables de tipo integer, char, boolean y records
- Una constante de carácter (**constchar**) y una constante de string (**conststr**) se escriben entre comillas simples. Ejemplo: 'a' y 'Número de cuenta'
- Deberán implementarse las funciones *write* y *read* para lectura y escritura de variables. Se deberá adaptar la gramática para incluir estas funciones.
 - o La función *read* recibe como parámetro una variable donde se almacenará el valor leído. Este valor podrá ser de tipo integer o char, no podrán leerse datos de tipo boolean.
 - o La función *write* recibe uno o dos parámetros. El primer parámetro es un **conststr**. El segundo parámetro (si existe) será una variable de tipo integer, char o string.
- Se deberán reconocer operaciones aritméticas conforme a la precedencia de operadores comúnmente conocida.

- Se deberán reconocer expresiones booleanas conforme a la precedencia lógica ya conocida. Los operadores serán los mismos utilizados por Pascal (<>, =, >, <, >=, <=, and, or, not, etc.)
- Los identificadores deberán empezar con una letra, seguidos de cero o más dígitos o letras. Por letra entendemos caracteres de la a-z, A-Z o “_”.
- Las funciones podrán recibir y retornar cualquiera de los tipos previamente indicados, y los parámetros se pasarán por valor. La cantidad de parámetros a recibir es ilimitada.
- Se debe permitir la recursividad.
- El programa deberá contener un cuerpo principal (main) y no se utilizarán los procedimientos anidados.
- Debe ser capaz de reconocer y recuperarse de errores léxicos, errores en comentarios y errores sintácticos. Se deberá imprimir la línea y columna donde fueron encontrados cada uno de éstos y una descripción clara y útil del mismo.
- Se implementará el ciclo while, for, repeat y la instrucción if.
- Se entregarán todos los programas fuente; así como los archivos .class o .jar. El programa final deberá consistir en un solo programa que reciba como parámetros un archivo. Deberán entregarse además, 3 programas fuente correctos y 3 con error.
- Los operadores de suma son +, -
- Los operadores de multiplicación son *, /, div, mod
- El lexema para la asignación es :=
- Otros símbolos deberán ser considerados como errores léxicos y reportados apropiadamente.
- La evaluación del proyecto se basa en la detección y recuperación de errores y la creación del AST. Con respecto a la detección de errores, se deberá imprimir la línea y columna donde fueron encontrados cada uno de éstos y una descripción útil del mismo.
- Se entregarán todos los programas fuente; así como los archivos .class o .jar. El programa final deberá consistir en un solo programa que reciba como parámetros un archivo .pas. Deberán entregarse además, 3 archivos fuente correctos y 3 con error.