

COMPILADORES

PROYECTO-PRIMERA PARTE

CONSTRUCCIÓN DE UN ANALIZADOR LEXICOGRÁFICO

I. Objetivo general

Aplicar las principales técnicas de diseño de lenguajes de programación y de sus compiladores.

II. Objetivos específicos

- Construir un analizador lexicográfico para un lenguaje de programación, que tome en cuenta los aspectos básicos pero fundamentales involucrados en esta fase de compilación.
- Que el estudiante enfoque el problema de la construcción de compiladores y de los ambientes en que estos trabajan, usando principios de la programación de sistemas y de la ingeniería de software.
- Que el estudiante encuentre un sitio natural para la aplicación del conocimiento adquirido en otras áreas de las ciencias de la computación, tales como: programación, estructura de datos, sistemas operativos y teoría matemática de la computación.
- Que el estudiante conozca un generador automático de analizadores lexicográficos (Flex).

II. Descripción

Se construirá un analizador lexicográfico para un lenguaje de programación que será asignado por el profesor, así como una interfaz con el usuario, que se constituirá al final del proyecto de compiladores, en un sencillo pero funcional ambiente de desarrollo para el programador.

Se generará un analizador lexicográfico para este mismo lenguaje de programación usando la herramienta Flex.

IV. Requerimientos

Una vez asignado el lenguaje de programación, el estudiante:

- a. Seleccionará al menos 15 palabras reservadas del lenguaje
- b. Especificará la expresión regular para definir identificadores
- c. Especificará expresiones regulares para números enteros y números reales
- d. Seleccionará una lista de operadores y caracteres especiales (al menos 15 entre ambos)
- e. Indicará como se construyen comentarios en el lenguaje

PARTE 1: (60%)

El estudiante procederá a construir el analizador lexicográfico para estas componentes de léxico, usando las técnicas vistas en clase, con los siguientes requerimientos específicos:

1. Se debe sustituir la búsqueda lineal del analizador lexicográfico, por una búsqueda binaria.
2. Se deben omitir los comentarios bien formados (los comentarios no son tokenizados)
3. Se escribirá además una interfaz para el usuario con las siguientes funciones básicas: abrir un archivo, crear y/o modificar texto (editar), copiar texto, pegar texto, guardar texto y ejecutar el analizador lexicográfico sobre el archivo abierto. La salida de la ejecución debe ser el programa fuente (texto) tokenizado. Esta salida se debe mostrar en una ventana “hija” dentro de la interfaz, en el formato:

```
lexeme -> token
lexeme -> token
.....
lexeme -> token
```

4. La interfaz para el usuario debe ofrecer la posibilidad de configurar los siguientes parámetros para el analizador lexicográfico:

- Tamaño máximo del buffer de líneas (MAXLINEA)
- Máximo número de dígitos en los enteros (MAXDIGIT)
- Máxima longitud de los identificadores (MAXID)
- Ubicación física del archivo scanner.exe. Este archivo contiene el código ejecutable del analizador lexicográfico.

Estos parámetros serán almacenados en un archivo de texto denominado “param.txt”, con el formato:

```
parámetro;valor
parámetro;valor
.....
parámetro;valor
```

Este archivo de parámetros debe ser leído por el analizador lexicográfico para tomar en cuenta estos valores.

DEBE DE ENTREGAR DEBIDAMENTE PARAMETRIZADO SU PRODUCTO, CON VALORES POR DEFAULT.

5. Una opción del menú de la interfaz debe ser la “Ayuda”. En esta opción se accederá a un archivo llamado “ayuda.pdf”, que contendrá la siguiente información para el usuario:

- a. Breve historia y aplicaciones del lenguaje asignado
- b. Palabras reservadas seleccionadas
- c. Expresión regular para los identificadores
- d. Expresiones regulares para números enteros y números reales
- e. Lista de operadores y caracteres especiales seleccionados
- f. Forma de construcción de comentarios en el lenguaje
- g. Un ejemplo de programa para escribir “Hola mundo” en el lenguaje
- h. Cualquier otra observación (o limitación –o aclaración) que considere necesaria

6. Se requiere que el analizador lexicográfico sea escrito en C o C++. **Ningún otro lenguaje es elegible**

7. Se requiere que la interfaz con el usuario sea escrita en Visual C++, Visual Basic, C# o Java. **Ningún otro lenguaje es elegible**

PARTE 2: (40%)

Generar un analizador lexicográfico para el lenguaje asignado usando Flex.

V. Entregables

Un CD con una portada especificando el nombre completo de los miembros del grupo, el nombre del grupo y el lenguaje asignado. **NO LO ENTREGUE EN BOLSA DE PAPEL MANILA NI ENTREGUE DUPLICADOS O COPIAS DEL CD.**

El CD contendrá:

1. En la raíz:

- Un archivo ejecutable llamado IDE.EXE (o su equivalente) que arranca la interfaz con el usuario
- Dos programas ejemplos en el lenguaje asignado llamados: ejemplo1.txt y ejemplo2.txt
- Los archivos:”param.txt” y “ayuda.pdf”

2. En una carpeta denominada “scanner”:

- Una subcarpeta llamada “fuentes” conteniendo el código fuente completo correspondiente al scanner. Puede contener todas las subcarpetas que sean necesarias.
- Una subcarpeta llamada “objeto” conteniendo el código objeto correspondiente al scanner. Puede contener todas las subcarpetas que sean necesarias.

3. En una carpeta denominada “ide”:

- Una subcarpeta llamada “fuentes” conteniendo el código fuente completo correspondiente a la interfaz con el usuario. Puede contener todas las subcarpetas que sean necesarias.
- Una subcarpeta llamada “objeto” conteniendo el código objeto correspondiente a la interfaz con el usuario. Puede contener todas las subcarpetas que sean necesarias.

4. Una carpeta denominada "flex", con los siguientes archivos:

- El archivo con extensión .l
- lex.yy.c
- scanner.exe
- Los archivos ejemplo1.txt y ejemplo2.txt (otra vez)

POR FAVOR: No incluya nada más en el medio. Revise que su medio no esté defectuoso y que esté libre de virus.

VI. Notas importantes

- Las únicas plataformas que se pueden presuponer disponibles son Windows 7 y Windows 8.
- El producto será probado por el profesor desde la interfaz con el usuario. **NO SUPONGA QUE LOS ARCHIVOS CON LOS QUE SE PROBARA EL PROGRAMA ESTAN EN ALGUN SITIO PREDETERMINADO. NO SUPONGA QUE EL PROGRAMA EJECUTABLE DEBE ESTAR EN ALGUN SITIO PREDETERMINADO.**
- Lo que usualmente se hace es copiar el contenido del CD a disco duro y probarlo desde ese lugar.

El profesor fijará hora y fecha para la entrega del producto. **Por ninguna razón se recibirá una segunda entrega del CD, ni se recibirá el CD más allá de la fecha y hora asignada. No se aceptará ningún otro medio de entrega que no sea CD.**