

Tecnológico de Costa Rica  
Escuela de Computación  
Curso: Introducción a los Sistemas Operativos

## Proyecto 1

El objetivo de este proyecto es comparar el rendimiento de realizar una tarea colaborativa utilizando múltiples hilos. Para ello se deberá desarrollar una versión multihilos del programa llamado *grep* que permite buscar palabras sobre varios archivos e imprimir el archivo en donde aparece.

En este proyecto se deben realizar múltiples pruebas para determinar la cantidad óptima de hilos para ejecutar este tipo de tarea sobre gran cantidad de archivos.

### Uso del programa

El programa *grep* recibe una lista de opciones, una expresión regular y una lista de archivos:

```
grep [opciones] [expresión regular] [archivos|directorio]
```

Por omisión el programa muestra la línea del archivo en donde se encuentra la expresión regular. Las opciones que se deben implementar son “-l” que solo muestra el nombre del archivo, y “-r” que busca recursivamente en directorios. El parámetro “archivos” puede ser uno o varios nombres de archivo, o bien, un directorio.

### Manejo de expresiones regulares

El lenguaje C por sí mismo no maneja expresiones regulares, sin embargo se puede usar la librería “regex.h” que cuenta con las funciones *regcomp* y *regex* que permiten realizar esta tarea. La función *regcomp* se debe ejecutar inicialmente y recibe la expresión regular ingresa por comando, luego la función *regex* se debe ejecutar sobre cada una de las líneas del archivo.

### Búsqueda recursiva en directorios

La opción “-r” del programa permitirá buscar en forma recursiva a través de todos los archivos de los directorios. Para implementar esto se deben utilizar las funciones *opendir*, *readdir*, y *closedir* con que cuenta Unix. La operación *readdir*

retornará un elemento del tipo *dirent* que internamente cuenta con una variable *d\_type* la cual indica si entrada es un archivo o un directorio.

## Manejo de múltiples hilos

El programa debe contar con múltiples hilos que se encarguen de inspeccionar los diferentes archivos. Debido a que la búsqueda puede abarcar una gran cantidad de archivos no sería eficiente crear un hilo por cada archivo. Es por ello que es necesario crear un conjunto compartido de hilos (pool de hilos) que incluya solo una cantidad limitada de hilos. Como parte del proyecto se deben realizar diferentes pruebas (con muchos archivos) para determinar un tamaño “óptimo” del pool de hilos. Se utilizará el tiempo de ejecución para determinar dicho tamaño óptimo

En este caso se utilizará un pool de hilos estático, es decir, el programa creará la totalidad de los hilos al inicio del programa y únicamente utilizará estos para realizar todo el trabajo. Tal como usted debe saber, a los hilos que terminen de buscar en un archivo se les deberá asignar otro archivo para realizar la búsqueda hasta terminar con todos los archivos de el/los directorio(s).

## Consideraciones generales

- Todo el desarrollo del proyecto debe realizarse en lenguaje C y sobre ambiente Unix.
- No es permitido el uso de otras librerías adicionales a “regex.h” ó “dirent.h” para programar el proyecto. Tampoco es válida la copia de porciones significativas de código desde Internet.
- Se deberá generar una documentación formal, en formato pdf, en donde se describan las diferentes etapas del desarrollo del proyecto, las decisiones de diseño que se tomaron, los mecanismos de programación utilizados, y los resultados de las diferentes pruebas al programa. Dicha documentación deberá incluir al menos las siguientes secciones:
  - Introducción
  - Descripción del problema (este enunciado)
  - Definición de estructuras de datos
  - Descripción detallada y explicación de los componentes principales del programa
  - Mecanismo de creación y comunicación de hilos
  - Pruebas de rendimiento
  - Conclusiones
- El proyecto puede realizarse en forma individual o en grupos de dos estudiantes.

- No se permite la copia de código entre grupos de estudiantes, tampoco es permitido utilizar librerías adicionales (desarrolladas por terceros).
- Se deberá crear un archivo zip que incluya: el código fuente de todos los programas desarrollados, binarios generados, y documentación en formato pdf.