

# DESARROLLO DE SISTEMAS DISTRIBUIDOS

## Proyecto Final

Elaborado por: Ukranio Coronilla

En este proyecto se debe elaborar un servicio que busca la frase más larga idéntica entre dos textos, cada uno almacenado en un archivo de texto plano.

Ejemplo:

Texto 1

—Desde luego, desde luego. Tome asiento, por favor —Ingelstad iba sin parar y se dejó caer alegremente en una de las sillas y le señaló otra, invitándole a que se sentara. Glokta tardó cierto tiempo en acomodarse. Primero se agachó con cautela y luego fue moviendo las caderas hasta que dio con una postura en la que la espalda no le sometiera a un martirio constante—. ¿Y qué es lo quería tratar conmigo?

Texto 2

Los cascos de los caballos golpeaban el arenoso sendero, las sillas crujían y los arneses tintineaban. El jinete iba sin parar de un lado para otro. El resto del mundo se desplazaba bajo ellos. El cielo del este se desangró, pasando del rojo a un rosa de matadero. El río surgió lentamente ante su vista, serpenteando entre los bosques otoñales que circundaban la base del empapado valle. Centelleante, como un ejército en marcha, corriendo rápido e implacable hacia el mar. Hacia Talins.

En este caso la frase más larga idéntica sería: iba sin parar

Los textos sobre los que se va a realizar la búsqueda se encuentran originalmente en el cliente el cual mediante una interfaz web podrá seleccionar los dos archivos para su envío hacia la aplicación distribuida, que se encargará de realizar el procesamiento y al finalizar le devuelva una respuesta al cliente.

Recuerde que para propósito de este curso es ver un sistema distribuido como un conjunto de computadoras que cooperan para brindar un servicio. Por consiguiente, el servicio es una aplicación distribuida con al menos tres computadoras que se distribuyen el procesamiento con objeto de brindar una respuesta al cliente en el menor tiempo posible y con el uso óptimo de recursos.

Todo el código deberá encontrarse completamente en C++ y orientado a objetos (salvo la interfaz web), y con sockets UDP.

Para mejorar la eficiencia se sugiere que todo el procesamiento de cadenas deberá hacerse en cadenas tipo C, en lugar de usar la clase String de C++.

Para evitar confusión en las búsquedas, su programa debe eliminar todos los caracteres extraños, signos de puntuación, eñes, etc... del texto original, de modo queden solo letras validas en ASCII.

La rúbrica para calificar será la siguiente:

Por cada rubro que se cumpla por completo se asignarán 2.5 puntos.

1.- La interfaz en el navegador incluye un botón para seleccionar los archivos que se van a subir. Cuando termine de realizarse la búsqueda, el navegador además de imprimir la oración común más larga imprime el tiempo que tardó la búsqueda con exactitud de segundos, contando desde que el usuario da clic en el botón de enviar el segundo archivo, y hasta la devolución de todos los resultados. Si la interfaz es en modo consola solo se otorgará un punto.

2.- La aplicación está distribuida y al menos tres computadoras realizan el cómputo. En caso de que no esté distribuido y una computadora haga todo el trabajo, solo se otorga medio punto.

3.- Trabajan todos los núcleos de las tres computadoras (en general cada CPU contiene varios núcleos y con ayuda del monitor de sistema podemos verificar que la mayor parte del tiempo todos los núcleos trabajan) de manera que el tiempo de búsqueda sea el más bajo posible con el hardware disponible (usar las computadoras más rápidas para el procesamiento), lo cual implica un buen balanceo de carga. En caso de que se usen todos los núcleos, pero no exista el balanceo de carga, se otorgará la mitad de la puntuación.

4.- La respuesta de la búsqueda es correcta y el tiempo de búsqueda es razonable (en promedio tarda lo mismo en los demás equipos). Si la respuesta no es la correcta, no habrá puntos en este rubro.

Con la intención de que puedan probar su código se dejan las ligas a los archivos A1.txt y A2.txt, cuya frase común más larga y que debiera devolver su programa, contiene 15 palabras y es la siguiente:

**“que produce el metal al raspar la tierra Un sonido que le era muy familiar”**

Para este proyecto solo se podrá programar en un intervalo de tiempo comprendido entre las 7:00 y las 18:00 y no habrá prórroga, debido a que la revisión lleva mucho tiempo y son muchos equipos. Para evitar inconvenientes se deben guardar todas las versiones de su aplicación, aunque no tenga todas las funcionalidades, de lo contrario no podrían tener una versión funcional que se pueda evaluar. Concluido el tiempo se irá calificando a cada equipo en el orden que fueron asignados, es decir primero al equipo 1, después el 2, etc... . No se podrá saltar un equipo en el proceso de evaluación.

En todos los equipos se tiene el derecho de asignar una calificación menor a uno de sus miembros en el caso de que no haya apoyado lo suficiente, lo cual deberá ser avalado por la mayoría de los integrantes. Favor de hacérmelo saber en el momento de la calificación.

Las calificaciones finales se entregarán en mi cubículo el día viernes 7 de junio, en un horario de 7:00 a 11:00 y de 12:00 a 15:00 o si aún hay alumnos formados, hasta que se termine de entregar a todos su calificación.