

FECHAS IMPORTANTES

Entrega parcial 1: **Viernes 3 Noviembre** (vía webcursos)

Entrega FINAL: Viernes 17 Noviembre (vía webcursos)

El nombre del archivo que se sube a webcursos debe llamarse **Apellido1_Apellido2_EntregaFinal.c**. Donde Apellido1 y Apellido2 es el apellido de cada uno de los integrantes del grupo. Si el archivo no tiene ese nombre, se **descuenta 1 punto de la nota de la tarea**.

GOOGLITO

En esta tarea cada grupo construirá el programa *Googlito* que simulará un sistema simple de búsqueda del tipo *Google*. En vez de buscar en Internet, la búsqueda se realizará sobre un conjunto de documentos almacenados en el mismo computador en el que se ejecuta *Googlito*. Para esto, se deberá realizar un programa que le solicite al usuario **una palabra** a buscar (la *consulta*), y el programa deberá mostrar una lista con los 3 documentos (almacenados) más *relevantes* (ordenados de mayor a menor relevancia) ó “ceranos” a dicha consulta.

El algoritmo para construir *Googlito* tiene los siguientes cuatro pasos:

- 1) Recolectar los documentos de texto puro sobre los que se realizarán las búsquedas.
- 2) Crear los vectores de frecuencia:
 - a) Leer las palabras de cada documento almacenado para crear el vocabulario.
 - b) Convertir cada documento almacenado en un vector de *frecuencias de palabras* (qué palabra aparece cuántas veces en el documento), y generar luego la matriz de frecuencias para todos los documentos en todas las palabras del vocabulario.
- 3) Comparar la consulta con la matriz de frecuencias creada previamente:
 - a) Encontrar los 3 documentos donde la palabra consultada aparezca con mayor frecuencia.
- 4) Mostrar los resultados de la búsqueda:
 - a) Ordenar los 3 documentos según la frecuencia con la que aparece la palabra consultada.

En la entrega 1, cada grupo realizó los pasos 1 y 2, terminando con la construcción de la matriz de frecuencias. En la entrega final, cada grupo deberá agregar el código de los pasos **3 y 4**.

Por ejemplo:

Paso 3. Encontrar frecuencias máximas: Supón que la consulta es “pasa” y que la matriz de frecuencias de 4 documentos es la siguiente:

		Doc0	Doc1	Doc2	Doc3
		[0]	[1]	[2]	[3]
hola ->	[0]	1	0	0	1
Juan ->	[1]	1	1	0	2

pasa -> [2]	1	2	1	0
entro -> [3]	0	1	0	1
lento -> [4]	0	1	1	1
tiempo -> [5]	0	0	1	0

Luego, el vector de *frecuencia de palabras*, que identifica los documentos con las máximas frecuencias para la consulta “pasa”es:

[0]	[1]	[2]
1	0	2

En la primera posición se encuentra el 1, que identifica a Doc1.txt con la máxima frecuencia (2). Luego en la segunda posición se encuentra el 0, que identifica a Doc0.txt, con la segunda máxima frecuencia. Finalmente en la tercera posición se encuentra el 2, que identifica al documento Doc2.txt con la tercera mayor frecuencia.

Paso 4. Mostrar resultados: Se despliega por pantalla la identificación y las 4 primeras palabras de los 3 documentos donde la palabra “pasa” aparece con mayor frecuencia. En el caso de este ejemplo:

Doc1.txt
pasa juan y entro

Doc0.txt
hola que tal juan

Doc2.txt
el tiempo pasa lento

ENTREGA FINAL

La *Entrega Final* consiste en dejar operativo el sistema *Googlito*. Para eso, deben construir las siguientes funciones:

```
void MayoresFrecuencias(string consulta[], char
Vocabulario[MaxPal][LargoMaxPal], int Matriz[MaxPal][MaxDoc], int
Frecuencias[])
```

esta función toma como argumento de entrada la palabra consultada (consulta) y almacena en el arreglo Frecuencias la identificación de las columnas de Matriz con los mayores valores de

frecuencia para la palabra consultada. Por ejemplo, si los documentos con los mayores valores de frecuencia para la palabra consultada son Doc1.txt, Doc0.txt y Doc2.txt, entonces el arreglo Frecuencias contiene la siguiente información:

[0]	[1]	[2]
1	0	2

Void MostrarResultados(int Frecuencias[])

Esta función toma como argumento de entrada el arreglo Frecuencias y despliega por pantalla el nombre del documento donde la palabra consultada aparece con mayor frecuencia junto con las primeras 4 palabras de dicho documento.

La estructura del programa de la *Entrega Final* con las funciones principales, se debería ver como se muestra a continuación:

```
#include <stdio.h>
#include <cs50.h>
#include <string.h>

// El programa supone que los archivos de texto recopilados ya están
// en el directorio "documentos"

const int MaxDoc=50;           // Máximo Núm. de documentos
const int MaxPal=1000;        // Máximo núm. de palabras en Vocabulario
const int LargoMaxPal=20;     // Largo máximo de una palabra
const int LargoMinPal=4;      // Largo mínimo de una palabra
const int Listado=3;          // Número de documentos a mostrar
int NumPal;                   // Núm. de palabras actuales del Vocabulario

// Función correspondiente a entrega 1
void GenerarMatrizDatos(char Vocabulario[MaxPal][LargoMaxPal],int
Matriz[MaxPal][MaxDoc]){

    CrearVocabulario(Vocabulario);
    CrearMatrizFrecuencias(Vocabulario,Matriz);
}

int main(void)
{
    string entrada;
    char Vocabulario[MaxPal][LargoMaxPal];
```

```
int Matriz[MaxPal][MaxDoc];  
//almacena la identificación de los 3 documentos relevantes  
int Frecuencias[3];  
  
    NumPal=0; // num de palabras actuales del Vocabulario  
    GenerarMatrizDatos(Vocabulario, Matriz);  
    printf("Ingrese Consulta: ");  
    entrada = GetString();  
    MayoresFrecuencias(consulta,  
                        Vocabulario,Matriz,Frecuencias);  
    MostrarResultados(Frecuencias);  
}
```