

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA

LENGUAJES DE PROGRAMACIÓN 1

9na práctica (tipo b)
Segundo Semestre 2024

Indicaciones Generales:

Duración: 110 minutos.

NO SE PERMITE EL USO DE APUNTES DE CLASE, FOTOCOPIAS NI MATERIAL IMPRESO

- No se pueden emplear variables globales. Tampoco se podrán emplear las funciones de C que gesten memoria como malloc, realloc, memset, strdup, strtok o similares, igualmente no se puede emplear cualquier función contenida en las bibliotecas stdio.h, cstdio o similares y que puedan estar también definidas en otras bibliotecas.
- **EL PROYECTO DEBERÁ SER DESARROLLADO BAJO EL PARADIGMA DE PROGRAMACIÓN ORIENTADO A OBJETOS POR LO QUE SALVO EN LA SOBRECARGA DE LOS OPERADORES >> Y <<, NO SE PODRÁN DEFINIR FUNCIONES (NI PLANTILLAS DE FUNCIONES) INDEPENDIENTES QUE NO ESTÉN LIGADAS COMO MÉTODOS A ALGUNA DE LAS CLASES PLANTEADAS.**
- Deberá mantener en todo momento el encapsulamiento de todos los atributos de las clases, esto implica también que un método NO puede devolver un puntero a algún dato del objeto. También debe guardar los estándares en la definición y uso de todas las clases desarrolladas. No deberá repetir código en los diferentes métodos. No se considerará en la nota las clases que violen esto.
- **Cada método NO debe sobrepasar las 20 líneas de código aproximadamente.**
- En el archivo main.cpp deberá colocar un comentario en el que coloque claramente su nombre y código, **de no hacerlo se le descontará 0.5 puntos en la nota final.**
- El código comentado NO SE CALIFICARÁ. De igual manera NO SE CALIFICARÁ el código de un método si su llamado está comentado.
- Los programas que presenten errores de sintaxis o de concepto se calificarán en base al 40% de puntaje de la pregunta. Los que no muestren resultados o que estos no sean coherentes en base al 60%.
- Se tomará en cuenta en la calificación el uso de comentarios relevantes.
- **NO SE PODRÁ USAR CHAR* PARA MANEJAR LAS CADENAS SOLO PUEDE USAR STRING. DE LO CONTRARIO SE ANULARÁ LA SOLUCIÓN.**
- **TAMPOCO SE PODRÁ EMPLEAR LA CLÁUSULA protected NI LA CLÁUSULA friend, DE HACERLO SE NO SE LE CALIFICARÁN LAS CLASES INVOLUCRADAS.**

SE LES RECUERDA QUE, DE ACUERDO CON EL REGLAMENTO DISCIPLINARIO DE NUESTRA INSTITUCIÓN, CONSTITUYE UNA FALTA GRAVE COPIAR DEL TRABAJO REALIZADO POR OTRA PERSONA O COMETER PLAGIO.

NO SE HARÁN EXCEPCIONES ANTE CUALQUIER TRASGRESIÓN DE LAS INDICACIONES DADAS EN LA PRUEBA

- **Puntaje total:** 20 puntos.

INDICACIONES INICIALES

Cree un proyecto de C++ en NetBeans siguiendo estrictamente las indicaciones que a continuación se detallan:

- La unidad de trabajo será **t:** (Si lo coloca en otra unidad, no se calificará su laboratorio y se le asignará como nota cero)
- Cree allí una carpeta con el nombre **"CO_PA_PN_Lab09_2024_2"** donde **CO** indica: Código del alumno, **PA** indica: Primer Apellido del alumno y **PN** primer nombre (de no colocar este requerimiento se le descontará 3 puntos de la nota final). **Allí colocará los proyectos solicitados en la prueba.**

Cuestionario:

La finalidad principal de este laboratorio es la de reforzar los conceptos contenidos en el capítulo 9: "STL".

Descripción del Problema:

Análisis de Sentimiento en Comentarios de Restaurantes

El análisis de sentimiento es una técnica de procesamiento de lenguaje natural (NLP, por sus siglas en inglés) que se utiliza para identificar y clasificar las emociones expresadas en texto. En este caso, la tarea consiste en determinar si los comentarios de los clientes de un restaurante son positivos, negativos o neutros, basándose en las palabras contenidas en cada comentario.

Contexto del Problema

Eres un desarrollador contratado por una cadena de restaurantes para analizar los comentarios de los clientes y determinar la calidad del servicio en base a las emociones expresadas en los comentarios. Los comentarios serán proporcionados en un archivo de texto con el siguiente formato:

Sol y Sombra, "Excelente tarde en Sol y Sombra, El pollo a la brasa estuvo muy rico, supero nuestras expectativas. Desde nuestra llegada, fuimos recibidos por nuestro mozo, quien siempre estuvo al tanto para que nuestra estancia fuera excelente"

Penumbra, "Tuvimos una mala experiencia en este local, nada que comparar con el local Penumbra 2, todo estuvo frio y con la atencion de mala gana, no volveria a este local"

Cada comentario tiene un nombre del restaurante asociado (NOMBRE) y el texto del comentario (TEXTO), este último hasta el fin de cada línea.

¿Qué es un lexicon y cómo ayuda en el análisis de sentimiento?

Un lexicon es un conjunto de palabras o expresiones junto con información asociada, como su significado o su clasificación. En el caso del análisis de sentimiento, un léxico consiste en una lista de palabras que están clasificadas según su connotación emocional, como positivas, negativas o neutras.

Por ejemplo:

Palabras positivas: delicioso, excelente, increíble, bueno, perfecto, agradable

Palabras negativas: lento, mal, frío, terrible, pésimo, horrible

Palabras neutras: hola, local, todo.

Un léxico ayuda en el análisis de sentimiento porque permite asignar un peso emocional a cada palabra en un comentario. Al comparar las palabras del comentario con este léxico, se puede determinar la polaridad general del comentario (positivo, negativo o neutro), sumando las polaridades de las palabras que lo componen.

El lexicon para este ejercicio será proporcionado en un archivo de texto con el siguiente formato:

lexicon.csv

abandona,-2
abandonado,-2
...
confianza,2
confianza,1
...

Cada línea se compone por la palabra en minúsculas (PALABRA) y su polaridad (POLARIDAD).

Pregunta 1 (16 puntos):

Se solicita que desarrolle un proyecto "PREG01_LAB09" dentro de la carpeta correspondiente, DE NO COLOCAR ESTE REQUERIMIENTO SE LE DESCONTARÁ 2 PUNTOS DE LA NOTA FINAL, en la cual se declaren las clases descritas a continuación:

- ❖ Para manejar las palabras: La clase se denominará "Palabra" y deberá contener lo siguiente: 1) un atributo denominado **texto** (**string**) y 2) un atributo denominado **polaridad** (**int**).
- ❖ Para manejar las oraciones: La clase se denominará "Oracion" y deberá contener lo siguiente: 1) un atributo denominado **palabras**, este será un vector de palabras. 2) un atributo denominado **polaridad_total**(**int**). 3) un atributo **cantidad_palabras** (**int**).
- ❖ Para manejar todo el análisis: La clase se denominará "AnálisisSentimiento" y deberá contener lo siguiente: 1) un atributo **comentarios** (mapa: string y vector de Oracion): Almacena comentarios por restaurante. 2) un atributo **lexicon** (mapa: string y entero): Contiene palabras con su polaridad

"DEBE EMPLEAR OBLIGATORIAMENTE LOS NOMBRES DE LAS CLASES Y SUS ATRIBUTOS"

Con las clases indicadas debe realizar las siguientes operaciones:

- En la clase **AnálisisSentimiento**, debe implementar el método **cargar_lexicon**, que se encargará de la lectura del archivo **lexicon.csv** y lo almacenará en el atributo **lexicon**.
- En la clase **AnálisisSentimiento** debe implementar el método **cargar_comentarios_restaurantes**, que se encarga de la lectura del archivo "comentarios.csv" con la finalidad de cargar en el mapa **comentarios**.

Consideraciones:

- Debe tener en cuenta que el texto obtenido de **comentarios.csv** no está listo para usarse, debe realizar un **pre-procesamiento** simple con la finalidad de poder encontrar las palabras en el **lexicón**. Se recomienda crear un método donde corresponda para **procesar_oracion** que le ayude a eliminar caracteres no alfanuméricos, pasar a minúsculas y eliminar espacios extras. Para este proceso le recomiendo usar las funciones : **isalpha** y **tolower** ambas reciben un **char** y devuelven un **bool**.
- Luego de pre-procesar la oración, puede separar cada **Palabra** y continuar con el proceso, para esto le recomendamos utilizar los métodos de **string** **find**, **substr** y **erase**.
 - **find**: recibe un carácter y devuelve el índice de la primera aparición del carácter en la cadena.
 - **substr**: recibe un índice inicial y uno final, devuelve una cadena nueva con el resultado de extraer de la subcadena correspondiente.
 - **erase**: elimina los caracteres dentro de una cadena usando el índice inicial y final pasados como parámetros.
- En la clase **AnálisisSentimiento** implementar el método **imprimir_analisis**, que se encargue de realizar la impresión de un archivo debidamente tabulado (**sin usar el carácter '\t'**), que muestre el contenido del mapa **comentarios** correspondiente a los comentarios de los restaurantes. Solo debe imprimir el nombre del restaurante, incluir un estadístico de cantidad de Comentarios Positivos, Negativos y Neutros.

Consideraciones:

Para el desarrollo de esta pregunta debe considerar el siguiente código:

```
#include "AnálisisSentimiento.hpp"

using namespace std;

int main(int argc, char** argv) {
```

**NO PUEDE
CAMBIAR
ESTE CÓDIGO**

```
AnalisisSentimiento analisis;

 analisis.cargar_lexicon();
 analisis.cargar_comentarios_restaurantes();
 analisis.ordenar_comentarios();
 analisis.imprimir_analisis();

 return 0;
}
```

Pregunta 2 (4 puntos):

Se solicita que desarrolle un proyecto “**PREG02_LAB09**” dentro de la carpeta correspondiente, **DE NO COLOCAR ESTE REQUERIMIENTO SE LE DESCONTARÁ 2 PUNTOS DE LA NOTA FINAL**, en la cual se declarará el método descrito a continuación:

- En la clase **AnalisisSentimiento**, debe implementar el método **ordenar_comentarios**, que se encargará del ordenamiento del atributo **comentarios**. Estos deberán estar ordenados por la **polaridad_total** de las oraciones en orden decreciente.
- Modificar el método **imprimir_analisis**, para agregar lo siguiente por cada restaurante:
 - La oración con mayor polaridad.
 - La oración con menor polaridad.
 - Las primeras 5 oraciones con polaridad 0.

Al finalizar la práctica, **comprima** la carpeta de su proyecto empleando el programa Zip que viene por defecto en el Windows, **no se aceptarán los trabajos compactados con otros programas como RAR, WinRAR, 7zip o similares**.

Profesores del curso: Rony Cueva
 Andrés Melgar
 Miguel Guanira

Erasmó Gómez
Erick Huiza

San Miguel, 29 de Noviembre del 2024.