

Práctica 1

Problema 1: Listas Ordenadas Análisis de Frecuencia de Palabras con Excepciones

Problema: Se te proporcionará un texto que contendrá múltiples líneas de palabras separadas por espacios y signos de puntuación comunes (coma, punto, signo de interrogación, signo de exclamación). Tu objetivo es analizar la frecuencia de cada palabra en el texto, ignorando las diferencias entre mayúsculas y minúsculas. Sin embargo, se te proporcionará también una lista de "palabras prohibidas". Estas palabras prohibidas no deben ser contadas en el análisis de frecuencia ni aparecer en el resultado final.

Entrada (texto.in y prohibidas.in):

- texto.in: Este archivo contendrá el texto a analizar. Puede tener múltiples líneas.
- prohibidas.in: Este archivo contendrá una lista de palabras prohibidas, una por línea.

La lectura de ambos archivos debe realizarse hasta el final del archivo.

Salida (Estándar):

Para cada palabra (no prohibida) encontrada en el texto, tu programa deberá imprimir en la salida estándar la palabra en minúsculas y su frecuencia de aparición, separadas por dos puntos y un espacio (:). Las palabras deben imprimirse en orden alfabético.

Ejemplo:

Entrada texto.in:

Este es un texto de prueba.

Este texto tiene varias palabras, algunas repetidas.

Prueba este texto!

Entrada prohibidas.in:

es

de

este

Salida Estándar:

algunas: 1

palabras: 1

prueba: 2

repetidas: 1

texto: 2

tiene: 1

un: 1

varias: 1

Consideraciones para la implementación:

- Lectura de archivos: Deberás leer el contenido de ambos archivos (texto.in y prohibidas.in).

- Procesamiento del texto:

1. Leer el texto línea por línea.

2. Dividir cada línea en palabras. Ten en cuenta que las palabras pueden estar separadas por múltiples espacios y pueden estar acompañadas de signos de puntuación. Deberás limpiar cada palabra eliminando la puntuación adjunta y convirtiéndola a minúsculas para realizar una comparación sin distinción de mayúsculas y minúsculas.

- Almacenamiento de palabras prohibidas: Almacena las palabras prohibidas de manera eficiente para facilitar la búsqueda (por ejemplo, en un conjunto o una lista).

- Conteo de frecuencia: Utiliza una estructura de datos adecuada (como un diccionario o un mapa) para almacenar la frecuencia de cada palabra encontrada en el texto que no esté en la lista de palabras prohibidas.

- Ordenamiento de la salida: Después de procesar todo el texto, deberás ordenar las palabras encontradas alfabéticamente antes de imprimir sus frecuencias.

- Formato de salida: Asegúrate de que la salida cumpla con el formato especificado (palabra en minúsculas, dos puntos y espacio, frecuencia).