Introducción:
Este informe presenta un código que realiza el procesamiento de stemming en un archivo de texto sin utilizar la biblioteca NLTK y sin el uso de la estructura Counter para contar palabras. El stemming se logra aplicando reglas específicas para simplificar las palabras reduciéndolas a su forma raíz.

Pasos del Código:

Importación de Bibliotecas:
El código comienza importando la biblioteca re para manejar expresiones regulares. Esta biblioteca se utiliza para tokenizar el texto en palabras.

Definición de Función de Stemming:
Se define una función llamada simple_stem que toma una palabra como entrada y aplica reglas específicas de stemming para reducir la palabra a su forma raíz.

Definición de Función de Procesamiento de Lote:
Se define una función llamada process_batch que toma un lote de texto y un diccionario word_counts como argumentos. Dentro de esta función, el lote de texto se tokeniza en palabras utilizando una expresión regular. Luego, se aplica el stemming a cada palabra utilizando la función simple_stem, y se actualiza el diccionario word_counts con las palabras stemmeadas y su frecuencia.

Función Principal:
La función main es donde se realiza el procesamiento principal. Se inicializa un diccionario word_counts vacío y se establece el tamaño del lote.

Lectura del Archivo por Lotes:
El archivo de texto es leído en lotes utilizando un bucle while. En cada iteración, se lee un lote de texto del archivo. Luego, el lote de texto se pasa a la función process_batch junto con el diccionario word_counts.

Ordenamiento y Muestra de Palabras Más Comunes:
Al final del procesamiento, las palabras y sus frecuencias se ordenan en orden descendente. Se imprime un número determinado de palabras más comunes junto con sus frecuencias.

Ejecución Principal:
La ejecución principal del código ocurre en el bloque if __name__ == "__main__":. En este bloque, se llama a la función main().

Conclusiones:
El código demuestra cómo implementar el proceso de stemming en un archivo de texto utilizando reglas específicas de reducción de palabras. Aunque este método de stemming es simple y puede no ser tan preciso como los algoritmos más avanzados, proporciona una forma

básica de normalizar las palabras en un archivo de texto. El uso de un diccionario word_counts para realizar un seguimiento de las frecuencias de las palabras sin el uso de Counter muestra una alternativa para contar palabras. El código puede ser un punto de partida para aquellos que desean comprender el proceso de stemming y explorar sus propias reglas personalizadas para el procesamiento de texto.