

Práctica 2.1

Introducción

El objetivo de esta práctica es aplicar una técnica para corregir mala ortografía. Esto puede llegar a ser algo muy complicado, ya que hay mucha gente que cuenta con una expresión escrita horrible, lo cual hace difícil entender qué es lo que realmente quiere expresar una persona que tiene mala ortografía.

A su vez, otro objetivo es analizar y comparar los histogramas obtenidos antes y después de implementar el diccionario para corregir las palabras mal escritas.

Desarrollo

La primera parte de esta práctica consistió en realizar un conteo de palabras contenidas en un archivo de texto. Al finalizar el conteo, se imprimió un histograma con las palabras y cuántas veces se encontró cada una de ellas en el archivo de texto.

A continuación, lo que se hizo fue definir en un archivo json un conjunto de palabras mal escritas con su equivalente, es decir, la palabra bien escrita a la que hace referencia. En pocas palabras, se creó un diccionario.

Ahora, con ayuda del diccionario, se corrigieron cada una de las palabras mal escritas en el archivo de texto. Nuevamente, imprimimos un histograma, ahora con las palabras corregidas y, de igual forma, el número de veces que se encontraron cada una de ellas en el archivo de texto.

Conclusión

Si analizamos qué fue lo que pasó con los histogramas, nos damos cuenta que, en el segundo histograma, el total de palabras diferentes encontradas en el archivo de texto disminuye, por lo que se eficientiza el conteo de las palabras. Esto se debe a que no importa cómo escribas de mal una palabra, siempre va a responder a una palabra (la palabra bien escrita). Mientras que antes de implementar el diccionario, si se escribe mal una palabra de dos o más formas diferentes, por cada que se escriba diferente, se tomará en cuenta como una nueva palabra.

Con esto concluimos que, efectivamente, es bastante complicado poder entender o descifrar lo que una persona está diciendo si tiene mala ortografía. Esto se debe a que existen muchísimas maneras diferentes de escribir mal una palabra, y es prácticamente imposible implementar, en un diccionario, todas las combinaciones posibles en las que se puede escribir mal cada palabra definida en dicho diccionario.