

Predictor de texto

Trabajo de fin de curso, Santiago Calligari.

Introducción

Este programa implementa un predictor de texto para completar frases con base en el estilo de un autor específico. El sistema funciona en dos etapas:

1. Preprocesamiento (C)

- **Entrada:** Nombre del autor.
- **Proceso:**
 - Se verifica si existe un directorio para el autor en RUTA_DIRECTORIO_TRABAJO.
 - Si no existe, se crea un archivo .list con los nombres de los archivos en el directorio del autor.
 - Se crea un directorio Entradas/{Artista}.
 - Se normalizan los archivos del autor, reemplazando caracteres especiales y eliminando puntuación.
 - Se ejecuta el script RUTA_SCRIPT_PYTHON.

2. Entrenamiento y predicción (Python)

- **Entrada:** Archivo de texto con frases del autor.
- **Proceso:**
 - Se crea un diccionario de diccionarios, donde cada palabra tiene un subdiccionario con las siguientes claves:
 - siguientes: Diccionario con las palabras que siguen a la actual y su frecuencia.
 - anteriores: Diccionario con las palabras que preceden a la actual y su frecuencia.
 - Se recorre el archivo de frases, actualizando el diccionario con las palabras y sus relaciones.
 - Se abre el archivo de frases a completar y se busca el caracter _.
 - Se obtienen las palabras anterior y posterior a _.
 - Se consultan los diccionarios siguientes y anteriores para obtener las palabras más probables.
 - Se selecciona la palabra con mayor frecuencia y se reemplaza en la frase.
 - Se repite el proceso hasta completar todas las frases.

Detalles de la implementación

- **Normalización de texto:** Se utiliza una función para convertir el texto a minúsculas, eliminar acentos y puntuación, y reemplazar caracteres especiales.
- **Diccionario de palabras:** Se implementa un defaultdict(dict) para almacenar las palabras y sus relaciones.
- **Predicción:** Se calcula la frecuencia de las palabras que siguen a la anterior y preceden a la posterior, y se selecciona la más probable.