

7. Estrategias:

- Crear una función que identifique el alfabeto del texto y que identifique los idiomas que tienen alfabeto propio
- Para los idiomas con alfabeto compartido, crear una lista con las x palabras comunes de los idiomas y compararlos con el texto y la lista que más concuerde es la del idioma
- Con la lista de “stop words” identificar con el texto con que idioma concuerdan la mayor cantidad de palabras

9. Estrategias:

- Sacar las palabras que empiezan en mayúsculas, ahí van a estar los nombres propios
- “Tratar” de descartar las palabras que no sean nombres y dejar nada más los nombres junto con su contador de cuantas veces están

10.

- tomar la lista de personajes generada en el punto anterior, y retornar los 2 más frecuentes

11. Estrategias:

- Identificar las palabras que tengan un nombre propio
- Revisar las preposiciones que están junto a este, para identificar si es nombre o lugar

12 Estrategias:

- Si la obra cuenta con fechas, extraerlas del texto y hacer un promedio de las fechas para un aproximado de el tiempo en que esta escrita
- Si no dice la fecha, usar una lista con palabras comunes según el contexto histórico usar palabras “filtro” que solo existan en la actualidad pero no en el pasado

**Implementacion:**

7.

Al principio se pensó en crear una función que identificarse los idiomas que tuviesen alfabeto propio y hacer las comparaciones para identificar idiomas utilizando una lista con las palabras más comunes de cada idioma, sin embargo a la hora de implementar las soluciones, me di cuenta que la creación de listas de palabras era innecesario, puesto que cada idioma por similar que sea a otro, tiene una serie de caracteres especiales únicos que lo diferencia de otro(dígase la ñ en el español, las tildes dobles y triples en el francés etc.), así que bastaba con crear una serie de diccionarios que asociasen dichos caracteres especiales con su respectivo idioma, de modo que solo hacia falta verificar la frecuencia con la que aparecían dichos caracteres en el texto, y así identificar su idioma.

Reto: encontrar la manera mas optima de identificar el idioma en que esta escrito un texto determinado

Análisis: La solución es funcional, y no presenta ninguna complicación a la hora de identificar el idioma en el texto, además de ser más óptima que la estrategia que se pensaba implementar al principio.

punto 9

Se implementó la solución planteada inicialmente, de distinguir nombres propios, a partir de una función creada para el desarrollo de otro punto (revisar solución punto 11 ), y para contar la cantidad de veces que aparecen se creó un diccionario en el cual se almacenaron los nombres identificados en el texto y se contó cuántas veces se repite cada nombre.

Retos:

- Diferenciar los nombres propios de otras palabras que inician en mayúscula
- identificar nombres iguales pero que pasasen por diferentes debido a la presencia de caracteres especiales (';',',')

Análisis:

sin librería: se tiene una buena lista de nombres y lugares, sin embargo al tratar de diferenciar dichos elementos haba varias fallas en la ejecución, puesto que no terminaban de separarse completamente.

con librería: en base a la misma lista de nombres y lugares se utilizo la libreria spacy que permite la identificación de lugares y asi diferenciarlos de los nombres

punto 10

se implementó la solución planteada inicialmente, y se tomaron los 3 personajes mas frecuentes de la lista generada en el punto anterior

Retos:

- No hubo mayor reto puesto que con la realización del punto anterior, ya teníamos la solución de este punto.

Análisis: No hubo mayor complicación puesto que con la realización del punto anterior, ya teníamos la solución de este punto.

punto 11.

Se implementó la solución planteada inicialmente, identificando los nombres propios a partir de si la palabra iniciaba con una letra mayúscula y se creó una función de diferenciación de lugares y nombres, a partir de identificar el adverbio de lugar o sujeto anterior a nombre propio encontrado en el diccionario.

Retos:

- Diferenciar nombres de personas de nombres de lugares, puesto que hay muchos lugares que tienen nombres de personas.

Análisis: revisar análisis punto número 9 (ambos códigos siguen los mismos principios)

punto 12.

En este caso si se implementó la estrategia planteada inicialmente, extrayendo las fechas que estuviesen en el texto y sacando un promedio de todas para identificar la época en que transcurre el texto, si la fecha es menor que 1800 (la edad contemporánea inicia a partir del 1800) el texto es anterior a la edad contemporánea, si es mayor a 1800 y menor que 2023 pertenece a la edad contemporánea y si es mayor a 2023 es futurista y posterior a la edad contemporánea, sin embargo en caso de que el texto no cuente con ningún tipo de fecha numérica, se crea una lista de palabras claves que correspondan con elementos característicos de cada época, sin embargo esta solución no es funcional en el caso de las obras futuristas, debido a la misma naturaleza hipotética del tratar de narrar hechos futuros.

Retos: tratar de identificar el tiempo en los casos en los que no se cuenta con una fecha explícitamente.

Análisis: la solución es perfectamente funcional en los casos en los que se cuenta con una fecha, sin embargo en los casos en los que se trate de una obra futurista y que no cuente con una fecha, solamente será capaz de identificar que no corresponde a una época anterior a la edad contemporánea