Practica 1

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.
   Respuesta:
   Comparar archivos, artículos es una tarea que puede llegar a ser dispendiosa y algo compleja. Es por esto que decidí elaboraron unos conjuntos de datos que comparan las palabras de dos noticias y la cantidad de veces que estas aparecen. Este es un punto de partida para luego mediante otras técnicas buscar similitudes entre las noticias y poder concluir si están trabajando sobre la misma temática.
2. Definir un título para el dataset. Elegir un título que sea descriptivo.
   Respuesta:
   Compara noticias.
3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).
   Respuesta:
   Se elaboraron tres documentos uno que compara las palabras comunes en las dos noticias, otro que busca en la segunda noticia, las palabras que aparecen en la primera y un último archivo que busca en la primera noticia las palabras que aparecen en la primera.
4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.
   Respuesta:

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.
   Respuesta:
   Se tomaron noticias del día de hoy
   Cada archivo tiene los siguientes campos
   Palabra: Texto (especifica la palabra encontrada)
   Contador1: Numérico (Cuenta las apariciones de la palabra en la noticia 1)
   Contador2: Numérico (Cuenta las apariciones de la palabra en la noticia 1)
6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).
   Respuesta:

Agradecimiento al diario el país de España porque de allí obtuvieron los links de las noticias.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.
   Respuesta:
   La motivación principal es la de realizar búsquedas en noticias a fin de identificar cuales hacen parte del mismo tema y contexto, y estos conjuntos de datos son un punto de partida, ya que encuentran las palabras comunes en ambas noticias.
8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:
   - Released Under CC0: Public Domain License
   - Released Under CC BY-NC-SA 4.0 License
   - Released Under CC BY-SA 4.0 License
   - Database released under Open Database License, individual contents under Database Contents License
   - Other (specified above)
   - Unknown License

   Respuesta:

   La licencia que se eligió fue CC BY-SA 4.0 License. Se escogió este tipo de licenciamiento por sus cláusulas que se adaptan al trabajo elaborado.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.
   Respuesta:
   Archivo disponible en el enlace de github
   github/comparanoticias
10. Dataset. Presentar el dataset en formato CSV
    Respuesta:
    Archivo disponible en el enlace de github
    github/comparanoticias