

Finger Exercise 2: Un Uso Inesperado para Un Compresor

Tiempo Estimado: 3hs

(5 puntos)

En este ejercicio vamos a usar un compresor de datos para buscar imágenes similares a otras. Puede usarse cualquier compresor y/o lenguaje pero recomendamos tener cuidado con el tamaño del bloque que use el mismo.

Recomendamos usar Python como lenguaje de programación para este ejercicio pero puede usarse cualquier otro lenguaje.

Recomendamos usar LZMA, Block Sorting (con bloque grande), LZW o PPMC como métodos de compresión. Es sencillo encontrar bibliotecas Python soportando estos métodos de compresión. No es recomendable usar gzip.

El objetivo es usar un compresor para calcular la distancia normalizada de compresión entre archivos y encontrar cual es la imagen más parecida a un conjunto de imágenes dadas dentro de un set de imágenes.

Recordemos que la distancia normalizada de compresión se calculaba de la siguiente forma:

Donde $C(X+Y)$ es la longitud en bits o bytes del resultado de comprimir X concatenado con Y . $C(X)$ y $C(Y)$ denotan la longitud en bits o bytes de comprimir X e Y respectivamente.

Podemos comprobar que el compresor elegido funciona bien calculando $NCD(X,X)$ con diferentes imágenes y debería dar un número muy cercano a cero. Cuanto más chico el número mejor.

En el archivo 'data.zip' van a encontrar el set de datos. Cuando lo descompriman deberían encontrar dos carpetas una con el nombre 'images' y otra con el nombre 'test'. El objetivo es por cada imagen dentro de 'test' encontrar cuales son las 5 (cinco) imágenes más parecidas en 'images' e indicar el NCD entre ambas.

El alumno que encuentre primero las imágenes más similares a las imágenes solicitadas y publique sus resultados en el grupo de Facebook de la materia se llevará los preciados 5 puntos.

A divertirse!