

CC5213 - Recuperación de Información Multimedia

Prof. de Cátedra: Juan Manuel Barrios

Estudiante: Andrés Calderón Guardia



Mini-Control 1

September 16, 2023

Pregunta 1

a) Solución:

Para cada sistema primero tenemos que calcular los AP de cada consulta:

Sistema 1				
Lugar	Q1	Precisión Q1	Q2	Precisión Q2
1	I39	1/1 = 1	I98	1/1 = 1
2	I81	2/2 = 1	I65	-
3	I07	-	I77	-
4	I24	-	I92	-
5	I98	-	I54	-
6	I51	-	I93	-
7	I40	-	I02	-
8	I10	-	I83	-
9	I35	-	I78	-
10	I68	-	I44	-

Para **Q1** no se encontraron dos valores así que su **AP** es:

$$AP_1 = \frac{1 + 1 + 0 + 0}{4} = 0.5$$

Y para **Q2** faltó uno, resultando en:

$$AP_2 = \frac{1 + 0}{2} = 0.5$$

Con esto finalmente calculamos el **MAP** para el primer sistema:

$$MAP_1 = \frac{0.5 + 0.5}{2} = \frac{1}{2} = 0.5$$

Repitiendo el procedimiento para el segundo sistema:

Sistema 2				
Lugar	Q1	Precisión Q1	Q2	Precisión Q2
1	I88	-	I39	-
2	I54	1/2 = 0.5	I42	1/2 = 0.5
3	I03	-	I51	-
4	I81	2/4 = 0.5	I98	2/4 = 0.5
5	I66	-	I32	-
6	I93	3/6 = 0.5	I13	-
7	I98	-	I81	-
8	I39	4/8 = 0.5	I21	-
9	I01	-	I09	-
10	I90	-	I54	-

En ambas consultas se encontraron todas las imágenes así que no hace falta considerar ceros:

$$AP_1 = \frac{0.5 + 0.5 + 0.5 + 0.5}{4} = \frac{2}{4} = \frac{1}{2} = 0.5$$

$$AP_2 = \frac{0.5 + 0.5}{2} = \frac{1}{2} = 0.5$$

Y ahora calculamos el **MAP** para el segundo sistema:

$$MAP_2 = \frac{0.5 + 0.5}{2} = \frac{1}{2} = 0.5$$

b) Solución:

Para el primer sistema el primer acierto estuvo en la primera posición para ambas consultas, lo que resulta en:

$$MRR_1 = \frac{1 + 1}{2} = 1$$

Y para el segundo sistema el primer acierto estuvo en la segunda posición para ambas consultas, resultando en:

$$MRR_2 = \frac{0.5 + 0.5}{2} = \frac{1}{2} = 0.5$$

c) Solución:

Sistema 1:

Hasta la primera posición se consigue un acierto en ambas consultas por lo que **precision@1** es 1.

Para la primera consulta se encuentran dos de las cuatro imágenes ($2/4 = 0.5$) en las primeras diez posiciones, y para la segunda se encuentra una de dos ($1/2 = 0.5$), lo cual promediándolo resulta en un **recall@10** = 0.5 para el sistema.

Finalmente para la primera consulta hay 4 imágenes que se espera encontrar, por lo que calculando el **recall@4** se obtienen dos aciertos ($2/4 = 0.5$) hasta la cuarta posición, y para la segunda consulta hay que obtener el **recall@2** pues hay dos imágenes a encontrar, y aquí solo se halla una hasta la segunda posición ($1/2 = 0.5$), por lo que promediando obtenemos **R-Precision** = 0.5 para el primer sistema.

Sistema 2:

En la primera posición no se halla ninguna imagen en las consultas, resultando en **precision@1** = 0 para este sistema.

Para ambas consultas se encuentran todas las imágenes antes de la décima posición, por lo que se tiene **recall@10** = 1 para el sistema.

Por último para este caso se repite lo mismo que en el primer sistema, en la primera consulta hay que calcular **recall@4** y para la segunda **recall@2**, en donde sus resultados son $2/4 = 0.5$ y $1/2 = 0.5$ respectivamente, promediando así **R-Precision** = 0.5 para el segundo sistema.

Pregunta 2

a) Solución:

i.

- Primero como tenemos un archivo *raw* significa que solo se tiene lo mínimo para almacenar el audio, así que no hay que considerar un uso adicional de memoria para otro tipo de información.
- Luego se menciona que tiene una profundidad de **s16le** lo cual indica que por sample se almacenan 2 bytes (16 bits).
- Además se nos indica un sample rate de 22.050, lo que significa que hay 22.050 samples por segundo.
- Finalmente la canción dura 4 minutos y 21 segundos.

Con esta información se tiene que la memoria en bytes de la canción es:

$$2 \frac{\text{bytes}}{\text{sample}} \cdot 22.050 \frac{\text{samples}}{\text{segundo}} \cdot (4 \cdot 60 + 21) \text{ segundos} = 44.100 \cdot 261 \text{ bytes} = 11.510.100 \text{ bytes}$$

ii.

Por el **Sampling Theorem (Nyquist-Shannon)** sabemos que para digitalizar sonido se necesita obtener samples a una frecuencia mayor al doble de la máxima frecuencia del audio, por lo que como esta canción está en 22.050 [Hz] solo se pueden escuchar frecuencias de hasta 11.025 [Hz].

b) Solución:

i.

Primero tenemos que calcular la cantidad de samples por ventana, hay 48.000 samples por segundo y las ventanas son de 50 [ms], es decir, que hay 20 ventanas por segundo, de modo que en cada una de estas ventanas hay 2.400 samples (48.000/20), y por último, para determinar la cantidad de coeficientes hay que notar que la transformada de Fourier es simétrica y estamos trabajando con valores reales, de modo que gracias a estas propiedades es que los coeficientes distintos solo van a ser la mitad de samples por ventana, lo que sería **1.200** para este caso.

ii.

Sabemos que hay 20 ventanas por segundo, así que si una ventana obtiene valores altos para un coeficiente que representa 200 [Hz] significa que para ese coeficiente hay una magnitud alta para frecuencias de $20 \cdot 200$ [Hz], es decir, **4.000 [Hz]**.

c) Solución:

i.

Como el oído humano no percibe mucha diferencia sobre altas frecuencias, la escala Mel ayuda al pasar la escala lineal a una logarítmica donde los valores altos poseen una baja diferencia pero las bajas no.

ii.

Como la primera columna suele contener información del volumen resulta útil eliminar esta columna pues nos interesa identificar canciones por la canción en sí y no por el volumen en el que se escuchan.

Pregunta 3

a) Solución:

Asumiremos que la codificación del color es de la forma YCbCr y que cada uno de estos 3 valores ocupa 2 bytes en memoria.

i.

- Cada frame tiene un tamaño de 1920x1080 píxeles, lo que da un total de 2.073.600 píxeles.
- La resolución de los canales de color es 4:2:0, es decir, que para el canal Y se usa el tamaño completo de cada frame, mientras que para Cb y Cr se ocupa un cuarto del total, es decir, 518.400 píxeles cada uno, por lo que en cada frame se están ocupando $2.073.600 \text{ píxeles} + 2 \cdot 518.400 \text{ píxeles} = 3.110.400 \text{ píxeles}$ entre los 3 canales.
- Finalmente se tienen 30 fps, por lo que en un segundo se ocupa 30 veces lo obtenido en el punto anterior.

Con esta información la cantidad de bytes que se ocupan en total en un segundo es:

$$2 \frac{\text{bytes}}{\text{píxel}} \cdot 3.110.400 \frac{\text{píxeles}}{\text{frame}} \cdot 30 \frac{\text{frames}}{\text{segundo}} = 186.624.000 \frac{\text{bytes}}{\text{segundo}}$$

ii.

- Para un tamaño de 1280x720 hay 921.600 píxeles.
- La resolución de los canales de color ahora es 4:4:4, por lo que los 3 canales ocupan el tamaño total de cada frame, es decir, $3 \cdot 921.600 \text{ píxeles} = 2.764.800 \text{ píxeles}$.
- Y ahora hay 60 frames por segundo.

Con esta información obtenemos los bytes necesarios:

$$2 \frac{\text{bytes}}{\text{píxel}} \cdot 2.764.800 \frac{\text{píxeles}}{\text{frame}} \cdot 60 \frac{\text{frames}}{\text{segundo}} = 331.776.000 \frac{\text{bytes}}{\text{segundo}}$$

iii.

- Tenemos un tamaño de 3180x2160 lo que da un total de 6.868.800 píxeles.
- Poseemos la misma resolución de los canales de color que en i. por lo que haciendo un cálculo análogo el total de píxeles será de 10.303.200.
- Y por último se tienen 30 frames por segundo.

Calculando los bytes:

$$2 \frac{\text{bytes}}{\text{píxel}} \cdot 10.303.200 \frac{\text{píxeles}}{\text{frame}} \cdot 30 \frac{\text{frames}}{\text{segundo}} = 618.192.000 \frac{\text{bytes}}{\text{segundo}}$$

De esto deducimos que i. es la configuración en donde se ocupan menos bytes por segundo.

b) Solución:

i.

Para este caso darían error los frames del 4 al 7, pues solamente estos 4, 5 y 7 dependen del 6 al ser bidireccionales.

ii.

Aquí ningún otro frame además del mismo 18 se vería afectado pues los B no se utilizan como referencia para ningún tipo.

iii.

Para este caso los frames desde el 11 al 21 tendrían el error, para el caso de los frames 17, 20 y 21 es porque estos son del tipo P y dependen del I que sufrió el error en 14, y el resto son de tipo que dependen directa o indirectamente de este I.

iv.

Para este caso solo el 22 y 23, el 23 por ser de tipo P depende del tipo I o P anterior a él, que resulta ser el I de 22.