



Recuperación de Información Multimedia

Descriptores de Contenido

CC5213 – Recuperación de Información Multimedia

Departamento de Ciencias de la Computación

Universidad de Chile

Juan Manuel Barrios – <https://juan.cl/mir/> – 2019



Descripción del Contenido

- Descripción del Contenido: Generación de datos estructurados que resume el contenido de cada documento multimedia
 - Descriptor de contenido, Vector característico, Feature vector, Fingerprint
- Para efectos prácticos, ese dato generado (descriptor) reemplaza al documento multimedia
- Es necesario además definir la función de comparación entre descriptores
- Se espera que la (di-)similitud entre descriptores corresponde a la (di-)similitud entre documentos



Fases de Sistemas de Recuperación

- Fase Offline: Preparación de servicio
 - Obtener documentos de referencia **R**
 - Calcular descriptores para todo **R**
 - Almacenar descriptores **R** y construir índice **I**
- Fase Online: Atender consultas (paralelo)
 - Recibir documento de consulta **Q**
 - Calcular descriptores de **Q**
 - Comparar descriptores de **Q** con **R** usando el índice **I**
 - Generar respuesta al usuario
- Usualmente se prefiere una fase offline lenta para lograr fase online lo más rápida posible

Comparación de Descriptores

■ Función de Similitud

- Entrega un número con el grado de “parecido” entre dos elementos

■ Función de Disimilitud (o distancia)

- Entrega un número con el grado de “diferencia” entre dos elementos



$$f\left(\begin{pmatrix} \vdots \end{pmatrix}, \begin{pmatrix} \vdots \end{pmatrix}\right)$$

Si f es Similitud

$$f(x,y) \sim 1 \leftrightarrow x \text{ e } y \text{ muy parecidos}$$

$$f(x,y) \sim 0 \leftrightarrow x \text{ e } y \text{ no se parecen}$$

Si f es Disimilitud

$$f(x,y) \sim 0 \leftrightarrow x \text{ e } y \text{ muy parecidos}$$

$$f(x,y) \sim \infty \leftrightarrow x \text{ e } y \text{ no se parecen}$$



Nivel de Descriptores

- Descriptores de bajo nivel o “**low-level**”:
 - Se calculados por medio de estadísticas y algoritmos de análisis del contenido de un documento
 - Ej.: histogramas, frecuencias de palabras, orientaciones de bordes, etc.
- Descriptores de alto nivel o “**high-level**”:
 - Se calculados por personas (directa o indirectamente)
 - Contienen una componente “semántica” o de interpretación del contenido
 - Ej.: etiquetas, tags, clicks, likes, etc.
- Descriptores de nivel medio o “**mid-level**”:
 - Agregación estadística de descriptores de bajo nivel, usualmente usando datos de entrenamiento y clasificadores
 - Ej.: frecuencias de patrones relevantes y comunes

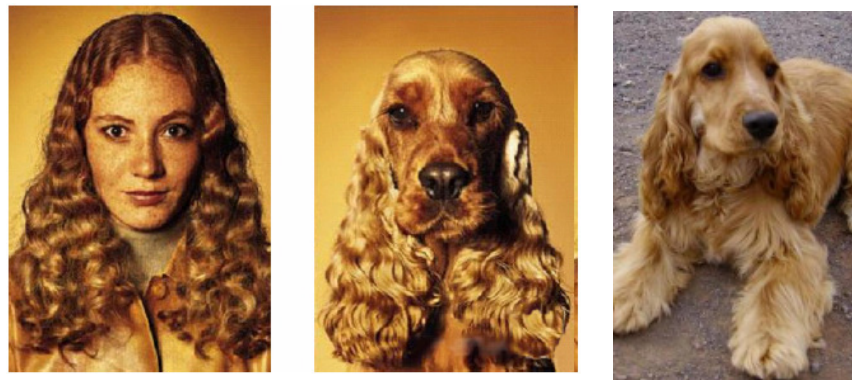


Semantic Gap

- Brecha Semántica: *“La diferencia entre la información que extrae un computador de documentos multimedia y la interpretación que un usuario puede dar a los mismos”*
- Diferencia entre la interpretación dada por un software y por un ser humano para un mismo documento

Concepto de Similitud

- Similitud es un concepto inherentemente subjetivo y difícil de modelar
- La similitud definida por personas puede no cumplir la desigualdad triangular



• — cerca — • — cerca — •
• — muy lejos — •



Formato de Descriptores

- Vector de dimensionalidad fija:

$$X=(x_1,\dots,x_n)$$

- ☐ Usualmente se comparan con distancias L_p
- ☐ Cada coordenada es usualmente un float (32 bits)
- ☐ A veces cada coordenada es un entero 8bit o 16 bits
- ☐ A veces los vectores son normalizados (largo 1)

- Cadena binaria:

$$X=10010101011011$$

- ☐ Se comparan con distancia de Hamming

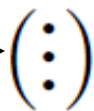
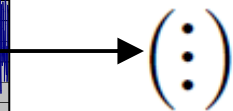
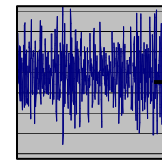
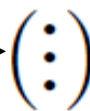
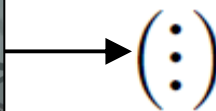
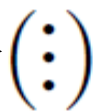
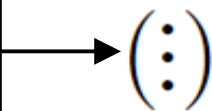
- Signatures:

$$X= (w_1,\dots, w_n) (c_1,\dots, c_n)$$

- ☐ Donde w_i escalares y c_i vectores u otros descriptores (señalan qué representa cada peso w_i)

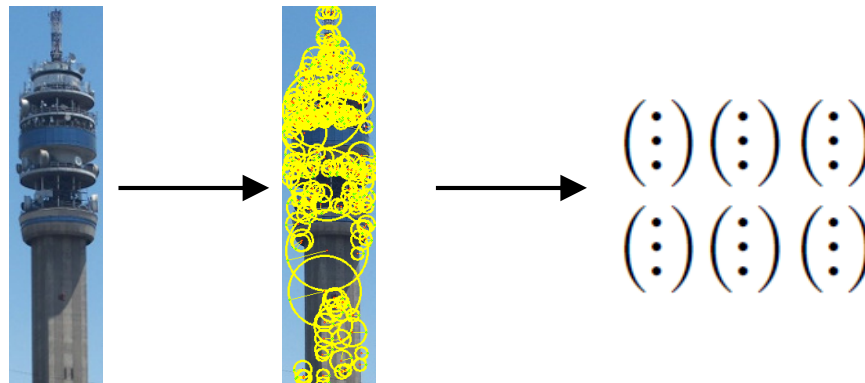
Tipos de Descriptores

- Descriptor Global: Representa la totalidad del documento multimedia
 - Una imagen, un audio, un video, un texto se representa por un único descriptor
 - Útil para buscar documentos parecidos en general, pero falla para ubicar pequeños patrones comunes



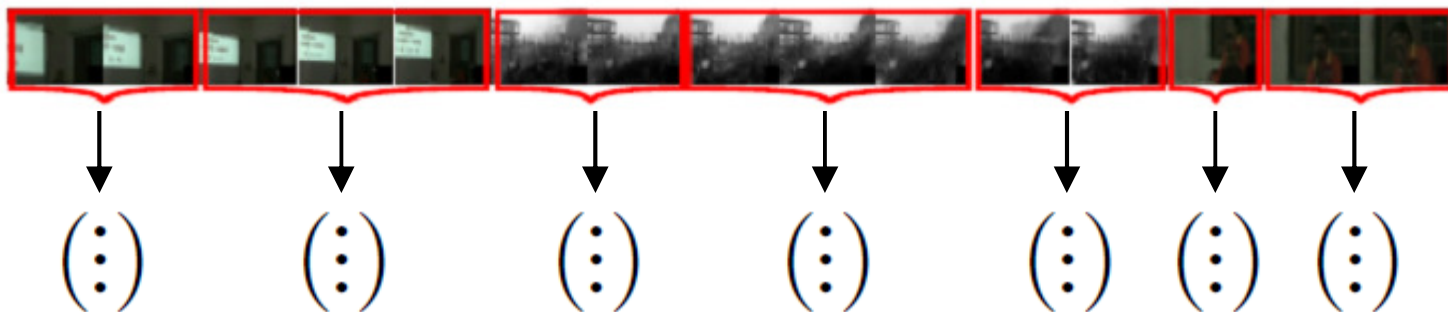
Tipos de Descriptores

- Descriptor Local: Representa una zona espacial del documento
 - Una imagen se representa por múltiples descriptores
 - Tiene un costo muy alto comparado con descriptores globales (del orden de 1000 veces)
 - Útil para calcular similitud parcial



Tipos de Descriptores

- Descriptor de Ventana (temporal):
Representa un periodo de tiempo del documento
 - Un audio, un video se representa por múltiples descriptores
 - Pueden ser ventanas de tamaño fijo o variable





Papers

- **Smeulders et al.** “Content-Based Image Retrieval at the End of the Early Years”. 2000.
- **Manjunath et al.** “Color and Texture Descriptors”. 2001.