
CC52V

Bases de datos multimedia

Prof. Benjamin Bustos

Capítulo 1 *Introducción*

1.1 Motivación

- La modernización del Registro Civil contempla el registro computacional de la huella digital de cada chileno
- Problema: Determinar si una persona está registrada o no en el sistema usando su huella digital

Se utilizará un lector de huellas digitales para escanear y digitalizar la huella (en una **imagen**). Las huellas digitales estarán almacenadas en una BD.



1.1 Motivación

■ Dificultades:

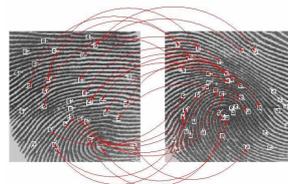
- ¿Pueden ser dos lecturas de la huella digital (incluso de la misma persona) exactamente iguales?
 - **R:** es improbable que suceda eso
 - Sólo basta con que 1 bit de ambas imágenes sea distinto para que dejen de ser iguales
 - Si la imagen ocupa n bits: Espacio total imágenes $O(2^n)$
 - Búsqueda exacta no es útil para resolver el problema
 - Sin embargo, ambas huellas deberían ser parecidas...
 - **Búsqueda por similitud**

3

1.1 Motivación

■ Más dificultades

- ¿Cómo se pueden comparar dos huellas digitales?
 - **R:** fijarse en características importantes de la huella
 - **Modelo de similitud (dependiente de la aplicación)**



4

1.1 Motivación

- Aún más dificultades:
 - ¿Cómo se pueden buscar las huellas digitales más parecidas en la BD?
 - Solución ingenua: búsqueda secuencial
 - Lector estándar USB: 40.000 huellas/s
 - Habitantes en Chile: ca. 16 millones
 - Tiempo de búsqueda: más de 6 minutos, ¡muy lento!
 - Suponer que el cliente no es Chile sino China...
 - Algoritmos y estructuras de datos eficientes que apoyen la búsqueda por similitud

5

1.1 Motivación

- Programa del curso
 - Búsqueda por similitud en BDs multimedia
 - Modelos de similitud para datos multimedia
 - Algoritmos de búsqueda por similitud
 - Índices para BDs multimedia
 - Tópicos avanzados
- Presentación prototipo

6

1.2 ¿Por qué bases de datos multimedia?

■ Multimedia

- Definición de la Real Academia Española:
“multimedia. (Del ingl. multimedia). 1. adj. Que utiliza conjunta y simultáneamente diversos medios, como imágenes, sonidos y texto, en la transmisión de una información.”
- Tendencia irreversible en los últimos años
- Mejoramiento de la calidad de la información
- Mayor ancho de banda

7

1.2 ¿Por qué bases de datos multimedia?

■ ¿Por qué tecnología de BD para multimedia?

- Altos requerimientos de espacio
- Acceso consistente en paralelo, multiusuario
- Necesidad de facilitar la búsqueda con estos tipos de datos
- Características de RDBMS (deseables en MMDBMS):
 - Independencia física y lógica de los datos
 - Optimización de consultas
 - Estructuras de datos para acceso eficiente
 - Transacciones (control de concurrencia, recuperación, restricciones de integridad...)
 - Seguridad de datos

8

1.2 ¿Por qué bases de datos multimedia?

- ¿Qué se entiende por BD multimedia?
 - Concepto colectivo:
 - Enciclopedia en CD-ROM
 - Sistema de Video-on-demand
 - Sistema CAD basado en RDBMS
 - BLOBs (“Binary Large Objects”) en BD relacionales
 - Definición de BD Multimedia para este curso
 - BD con alta capacidad y de alta performance
 - Para tipos de datos multimedia y alfanuméricos
 - Que soporte grandes volúmenes de información

9

1.3 ¿Qué es un medio?

- Definición (RAE): “medio, dia. (Del lat. medĭus). 15. m. Espacio físico en que se desarrolla un fenómeno determinado.”
- Comúnmente:
 - Medio traspa (transmite) mensajes
 - Un emisor genera una señal en el medio
 - Un receptor interpreta la señal

10

1.3 ¿Qué es un medio?

- Clasificación de los medios según los sentidos humanos
 - Visuales
 - Lingüísticos
 - Texto
 - Pictograma
 - No lingüísticos
 - Gráfico
 - Foto
 - Acústicos
 - Lenguaje
 - Música
 - Ruido
 - Táctiles
 - Gustativos
 - Olfativos

11

1.4 Dependencia temporal de los medios

- Medios dependientes del tiempo (inglés: *continuous media*)
 - Música
 - Lenguaje
 - Video
- Medios independientes del tiempo (inglés: *still media*)
 - Imagen
 - Gráfico
 - Texto

12

1.4 Dependencia temporal de los medios

- Características de los medios dependientes del tiempo:
 - Gran cantidad de información por objeto multimedia
 - Mayor *data rate* (transmisión por unidad de tiempo)
 - Requerimientos temporales

13

1.4 Dependencia temporal de los medios

- Problemas de los medios dependientes del tiempo:
 - Requerimientos de espacio
 - Ancho de banda mínimo del medio de almacenaje (e.g., disco duro)
 - Ancho de banda mínimo del medio de transmisión (e.g., network)
 - Reproducción a la velocidad correcta
 - Latencia (tiempo que necesita el sistema para el procesamiento de la información)
 - *Jitter* (variaciones en la latencia)
 - *Reliability* (confiabilidad)
 - Sincronización (e.g., video y audio)

14

1.5 Principios de las BD multimedia

- Modelos de datos
 - Son la característica principal de un DBMS
 - Determinan los siguientes aspectos:
 - ¿Qué tipos de objetos serán modelados y almacenados?
 - ¿Qué tipos de operaciones sobre los objetos serán soportadas?
 - Requisitos de integridad
 - Un sistema administrador de bases de datos multimedia (MMDBMS) implementa un *modelo de datos multimedia*

15

1.5 Principios de las BD multimedia

- Garantías de un DBMS
 - Persistencia
 - Independencia física y lógica de los datos
 - Sistema multiusuario
 - Recovery
 - Etc...
- Diferencia fundamental entre DBMS tradicional y MMDBMS: modelo de datos utilizado

16

1.5 Principios de las BD multimedia

- Usar DBMS tradicional con datos multimedia
 - Modelar datos multimedia a través de:
 - Referencias externas
 - Almacenaje de información no interpretada
 - Almacenar secuencia de bits + metadatos
 - Procedimientos externos
 - Problemas
 - No hay garantía de la persistencia de los datos multimedia (sólo los metadatos)
 - No hay independencia física de los datos
 - Aplicaciones deben saber que tipo de dato está almacenado
 - No hay soporte directo para datos dependientes del tiempo

17

1.6 Tipos de datos multimedia

- Texto
 - A menudo simplemente: strings de caracteres
 - Representación en un MMDBMS debería prever también información estructural (título, autor, capítulo, etc.)
 - Es el medio menos intensivo en uso de espacio
 - Requerimientos importantes de un sistema administrador de documentos
 - Búsqueda por texto completo
 - Tolerancia a faltas de ortografía (e.g., fallas del OCR)
 - Administración de sinónimos, catálogo
 - Búsqueda de documentos similares

18

1.6 Tipos de datos multimedia

- Imagen (estática)
 - Fuente: cámara digital, escáner
 - Muchos formatos distintos
 - Muchas formas de compresión (lossless, lossy)
 - En la medida de lo posible, formato de la imagen debe ser transparente para el usuario
- Una imagen es conceptualmente una matriz de puntos (píxeles)

19

1.6 Tipos de datos multimedia

- Bitmap
 - 1 bit por pixel
 - Funciones de corte y superposición muy eficientes a través de operadores booleanos
- Imágenes en colores
 - Profundidad del color: número de bits por pixel que codifican su color
 - La profundidad del color en una imagen es constante, pero puede variar entre distintas imágenes en una BD
 - Codificación a través de paleta de colores o descripción directa en un espacio de colores (e.g., RGB, CYMK)
 - Metadatos almacenan, ancho, alto y profundidad del color o paleta indexada

20

1.6 Tipos de datos multimedia

- Espacio para almacenar imágenes
 - Varía fuertemente dependiendo de la aplicación y del método de compresión utilizado
 - Imágenes en blanco y negro comprimidas: un par de Kbytes
 - Fotos satelitales: > 100 Mbytes
- Operaciones básicas permitidas por MMDBS
 - Selección de partes de imágenes
 - Búsqueda en BD de imágenes

21

1.6 Tipos de datos multimedia

- Gráficos vectoriales
 - Descripciones abstractas del contenido de una imagen
 - Se componen de formas geométricas básicas como líneas, círculos, polígonos
 - Ventaja principal: “editabilidad”
 - Se puede trabajar directamente con las formas básicas
 - Casi imposible de hacer con imágenes normales
 - Típicamente ocupan poco espacio
 - Problema hasta el momento: no hay estandarización

22

1.6 Tipos de datos multimedia

- **Objetos 3D**
 - Modelos tridimensionales de objetos geométricos
 - Estructura: típicamente mallas de triángulos
 - Con o sin orientación (interior, exterior)
 - Cerrados, abiertos
 - Propiedades adicionales: textura, color, etc.
 - Requerimientos de espacio variables
 - Número de polígonos utilizados
 - Resolución deseada
 - Caso particular: objetos CAD
 - Extensión de gráficos vectoriales a 3D
 - Muchas aplicaciones prácticas

23

1.6 Tipos de datos multimedia

- **Otros tipos:**
 - Moléculas
 - Utilización en docencia, investigación, marketing, etc.
 - Investigación en búsqueda de geometría parcialmente similar (propiedades químicas)
 - Investigación en búsqueda de geometría complementaria (protein docking)
 - Grafos
 - Estructura de datos importante para la representación de muchos conceptos del pensamiento humano
 - Investigación en similitud parcial y global
 - Time series
 - Valor de alguna característica periódicamente medida
 - Investigación en búsqueda de trazados similares

24

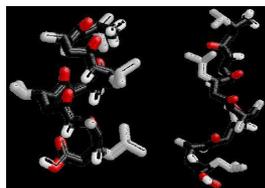
1.6 Tipos de datos multimedia

- No estudiaremos en este curso
 - Problemas con datos dependientes del tiempo
 - Audio
 - Video

25

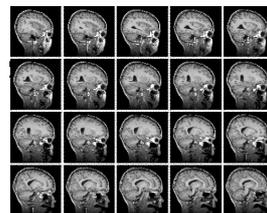
1.7 Aplicaciones

- BD multimedia: muchas aplicaciones prácticas



Biología
molecular

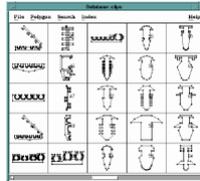
Medicina



Geografía



Industria
manufacturera



Y muchas otras...

26