



estudios de postgrado
en computación



Bases de datos avanzadas

Universidad de Los Andes

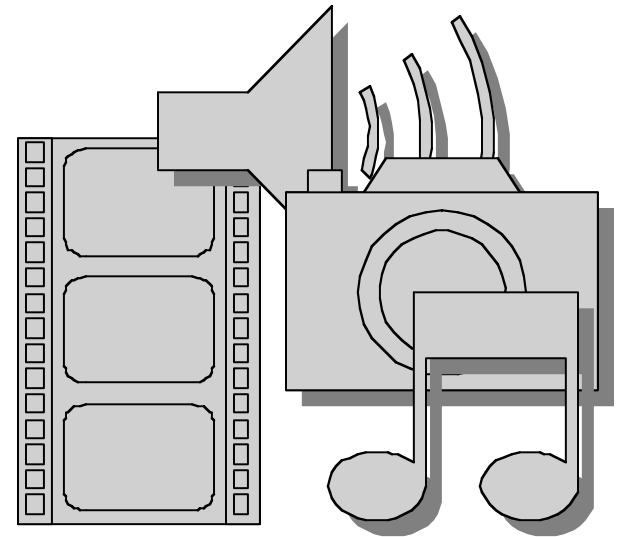
Postgrado en Computación

Prof. Isabel M. Besembel Carrera

***Unidad IV. Sesión 24. Bases de datos
multimedia.***

Introducción

- Avances de la computación en software y hardware
- Multimedia:
 - ✓ presentar información en más de una forma:
 - texto,
 - gráficos,
 - animación,
 - video y sonido.
- Aplicaciones:
 - ✓ Entretenimiento,
 - ✓ información pública,
 - ✓ publicidad y educación



- Almacenamiento y recuperación de datos multimedia.
 - ✓ Manejador de almacenamiento multimedia expandible y transparente para cada tipo de dato multimedia. Modelo general que capture los diferentes tipos de dispositivos.
 - ✓ Grandes capacidades de almacenamiento con jerarquías multinivel (principal, disco, “rocola” de discos).
 - ✓ Recuperación en tiempo real que requieren buen rendimiento para ambientes sensibles a datos multimedia. Imágenes como un todo (compresión), pero audio y video por bloques sincronizados.
 - ✓ Uso efectivo del almacenamiento y procesamiento paralelo y distribuido en todas las jerarquías.



Modelado, estructuración y barrido

- Reconocimiento, agrupamiento, tipos y relaciones para capturar y recuperar la información multimedia.
 - ✓ Diferencia entre recuperación en base de datos y en recuperación de información.
- Modelos hipertexto e hipermedia
 - ✓ sugieren un método de navegación para recuperar la información.
 - ✓ 3era generación (2000 - 2xxx): integración de las herramientas documentales en un sistema de información para las empresas
- Recuperación basada en contenido
 - ✓ Entendimiento del lenguaje natural, velocidad de procesamiento, visión y modelado de usuarios



Indexación por contenido

- Documento audiovisual: película, reportaje, emisión TV, etc.
- Estructuraciones posibles:
 - ✓ Temporal: no están presentes todos los niveles. Actualmente DAV y SDAV se indexan en tiempo real por los documentalistas especializados (documentación textual)
 - Documentos (DAV)
 - Subdocumentos (SDAV): de 1 hora a varios minutos
 - Escenas: de varios minutos a varios segundos
 - Planos (PAV): varios segundos
- SAV = DAV o SDAV o Escena o PAV
- STRATE-IA: modelo para la estructuración de las imágenes como secuencias audiovisuales (SAV) y modelos subyacentes para la interpretación de las mismas



Interacción y presentación

- **Especificación de consultas**
 - ✓ Clases de usuarios (novatos-expertos)
 - ✓ Enfoque de la consulta,
 - ✓ filtrado y redireccionamiento de la búsqueda.
- **No está restringida a la especificación de atributos**
 - ✓ Nuevas formas de interacción (escritura manual, etc.)
- **Transformación de la información**
 - ✓ Representaciones alternativas,
 - ✓ extracción,
 - ✓ simulación de fenómenos naturales



Enfoques de desarrollo

➤ Extensión de sistemas existentes

- ✓ sobre SMBDR: MDBMS.
- ✓ SMBDO-R: TITAN+ y DASDBS
- ✓ sobre SMBDOO: ORION manejador de inf. multimedia
- ✓ sobre recuperación de información: Manchester MIS y I3R.

➤ Desarrollo de nuevos sistemas

- ✓ nuevos sistemas de documentos multimedia (MINOS, MULTOS): modelado de la estructura de objetos complejos, recuperación basada en contenido y diferenciación de la representación interna y externa.
- ✓ sistemas hipermedia:
 - Intermedia, Harmony, is-News MIS

➤ KYDONIA

- ✓ Próxima generación de sistemas multimedia.
- ✓ Soporta todos los requerimientos de datos multimedia.
- ✓ Orientado por objetos con extensiones y optimizaciones apropiadas a los datos multimedia.
- ✓ Manejador de objeto: **maneja objetos complejos persistentes y transientes.**
- ✓ Manejador de almacenamiento OO en capas: **lógica, física y de dispositivos.**
- ✓ Módulo de interfaz: **Núcleo OVM que es un LPANOO. Asegura el acceso uniforme a todos los objetos y funciones del sistema**

- Recuperación de texto: por claves o booleana.
 - Archivos invertidos con tesauros (árbol_B+).
 - Archivos firma:
 - ✓ versión hash utilizando patrones de bits superpuestos y usando la operación “or”
- | | |
|-------|-----------------|
| data | 001 000 110 010 |
| base | 000 010 101 001 |
| firma | 001 010 111 011 |
- Agrupamiento de documentos por similaridad:
 - ✓ documento-a-documento (coseno) o
 - ✓ documento-a-grupo (grafo $O(n^2)$, iterativos ($O(n \lg n)$)).



Indexación multimedia

➤ Tipos de consulta:

- ✓ Igual: $\{O_1, \dots, O_n\}$ y Q encontrar los objetos en una distancia d de Q , $d(O_a, O_b) =$ distancia euclidiana (suma de la diferencia de cuadrados).
- ✓ Subpatrones: Dados $\{O_1, \dots, O_n\}$, d y un subobjeto Q encontrar los objetos cuyas partes igualan a Q .
- ✓ Vecino-más-cercano: árbol_R y el algoritmo “podar las ramas”
- ✓ Todos los pares: árbol_R y el algoritmo producto espacial.

➤ GEMINI:

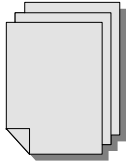
- ✓ Índice-F: Pasar Q a un punto $F(Q)$

Usar un SAM y d para recuperar los puntos

recuperar los objetos, calcular su d de Q y eliminar los que no la cumplen (false hits)

Documentos multimedia

El concepto de documento



- Diploma, carta, relación u otro escrito que ilustra acerca de algún hecho, principalmente los históricos.
- Cualquier otra cosa que sirve para ilustrar o comprobar algo.

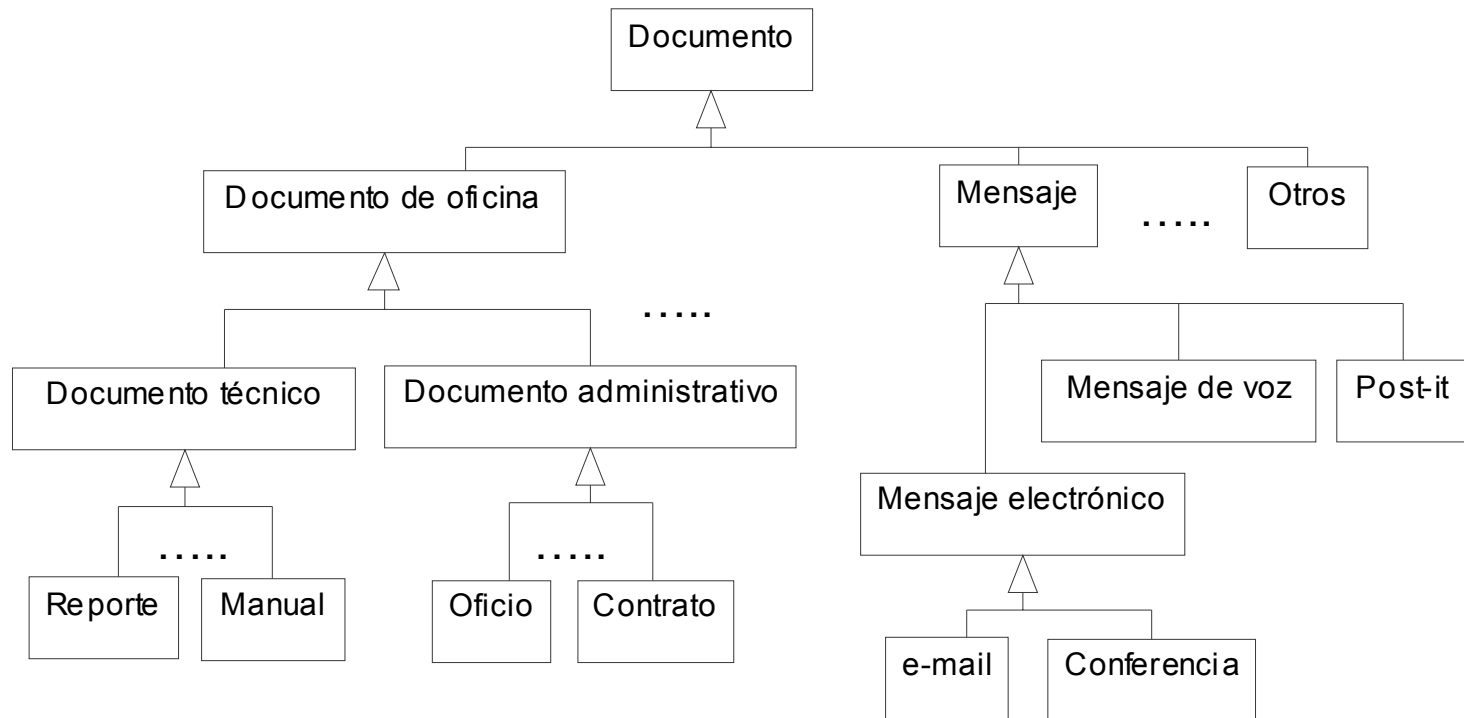
Documentación

- Conjunto de operaciones, métodos, etc., que facilita la recopilación, almacenamiento, búsqueda y circulación de documentos.

Tipos de documentos

- Cartas
- Informes
- Monografías
- Tesis
- Libros
- Artículos
- Páginas web, etc.

Tipos de documentos



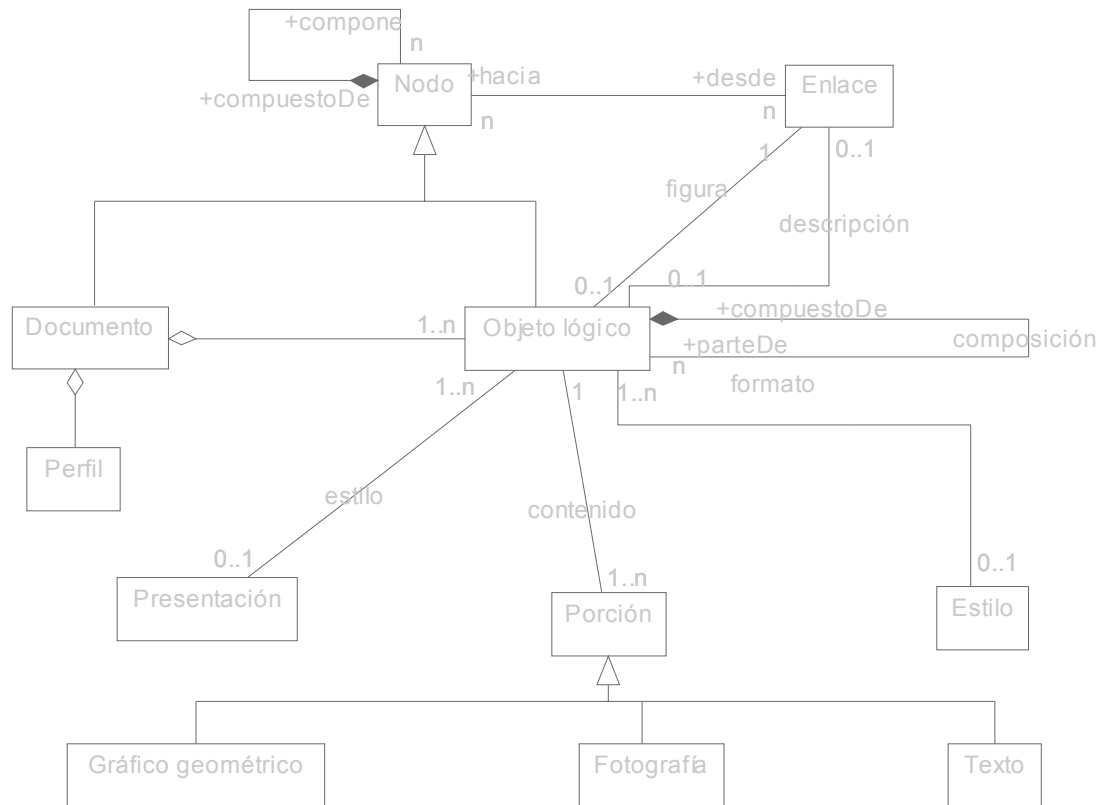


Bases de documentos

- Aplicaciones concebidas desde los inicios como hiperdocumentos y el sistema hipertexto (o hipermedia)
- Ejemplos:
 - ✓ Enciclopedias
 - ✓ Hipermapas o mapas geográficos con hiperenlaces
 - ✓ Sistemas de gestión de información urbana
- La constitución de bases de documentos con enlaces en el interior de cada documento y entre los documentos conducen a una estructura de hiperdocumento
- Ejemplos:
 - ✓ Bases de documentos multimedia
 - ✓ Documentos multiversión y multiautores



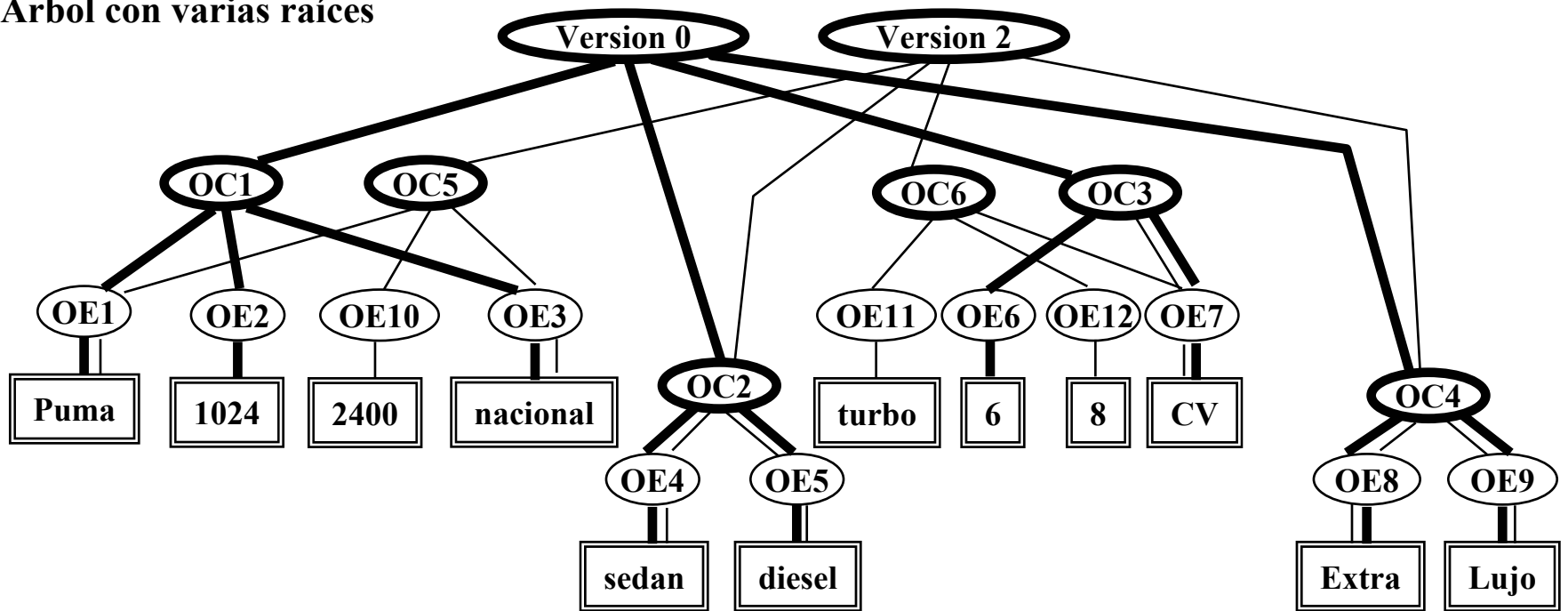
Modelo



Documentos multiversión

- Familia de documentos que tienen muchos elementos comunes
 - ✓ Documentación técnica de un producto industrial
 - ✓ Documentos multiautor (trabajo en grupo)

Árbol con varias raíces





Documentos multiautores

- Documentos técnicos bajo control de calidad son
 - ✓ Voluminosos (páginas de texto con figuras, esquemas, etc.)
 - ✓ Complejos: necesitan la colaboración de varios especialistas
- Se dividen en varias secciones homogéneas donde cada parte tiene:
 - ✓ Redactor para la primera versión
 - ✓ Varios verificadores (al menos 3) que deben controlar cada versión y proponer las correcciones o modificaciones
 - ✓ Redactor en jefe para comparar las diferentes versiones y resolver los conflictos
 - ✓ Las diferentes partes se compendian y se verifican

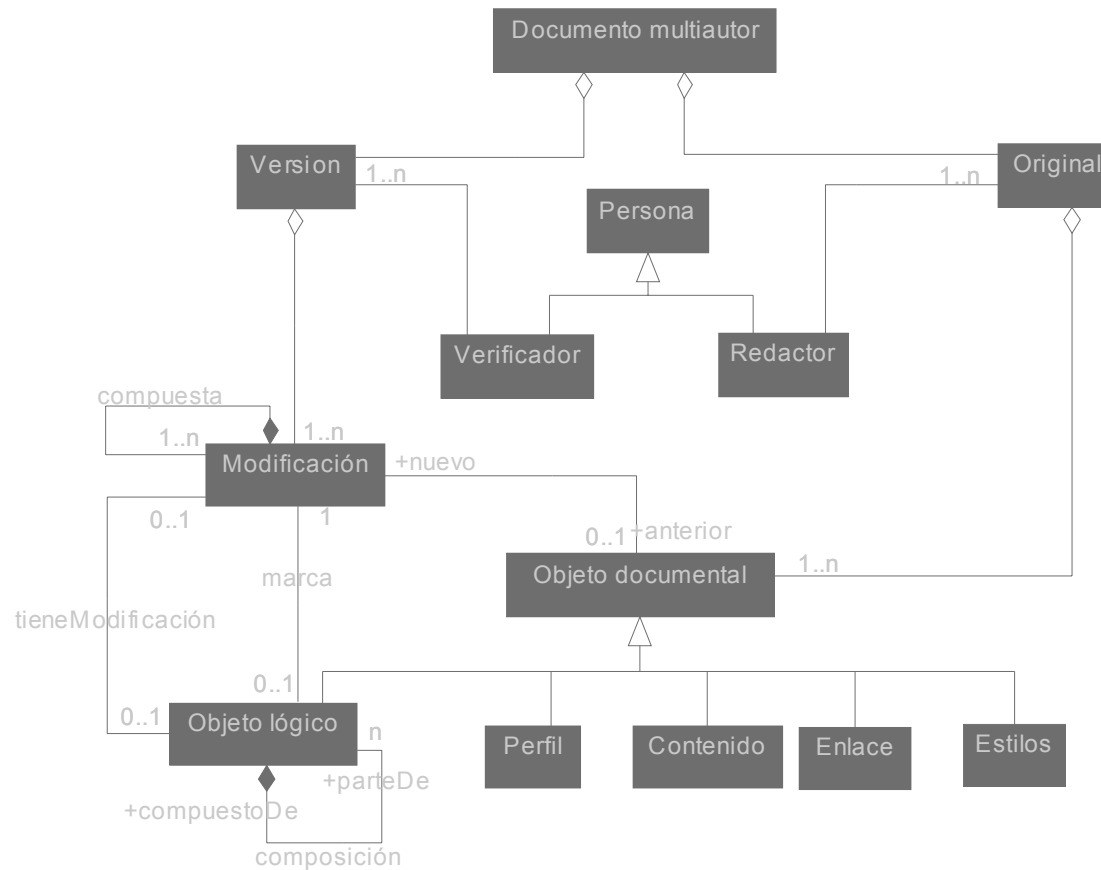


Documentos multiautores

- Caracterización de una modificación
 - ✓ Tipología:
 - Editorial: falta de ortografía, error en la presentación, etc
 - Técnico: error en un concepto, cálculo erróneo, etc
 - ✓ Importancia: menor, mayor, etc
 - ✓ Alcance: número de objetos documentales involucrados
 - Local: un único objeto
 - Extendido de orden n: n objetos involucrados
 - ✓ Estructura: compuesta o elemental
- Un objeto documental puede ser:
 - El perfil o uno de sus componentes
 - Un objeto lógico simple o compuesto
 - Una porción del contenido
 - Estilo: formato o presentación
 - Enlaces



Modelo



Conclusiones

- **Es un área de investigación que aún está en desarrollo.**
- **Su mejor soporte está basado en la orientación por objetos.**
- **Sus métodos de acceso se basan en los métodos de acceso espaciales y en la transformada rápida de Fourier**
 - ✓ Transformación en DCT (Discrete Cosine Transformation) para imágenes
 - Inspirada en la transformada de Fourier
 - Los 64 valores de cada bloque se transforman en 64 coeficientes DCT sin pérdida, esto es la transformación inversa obtiene los 64 valores originales
 - El primer coeficiente llamado DC representa el color medio del bloque
 - Los coeficientes siguientes denominados AC_i representan las frecuencias cada vez más elevadas, que se corresponden con los detalles cada vez más finos, en las dos direcciones