



## Examen final – Convocatoria de febrero de 2005 ADQUISICIÓN Y TRATAMIENTO DE DATOS II

Nombre, apellidos y DNI

---

### Ejercicio de diseño de bases de datos

Supongamos que se nos ha encargado el diseño de un sistema de recuperación de información (SRI) mediante el que se pueda acceder a una base de datos documental. Tras analizar detenidamente el problema, enumeramos los requisitos que ha de cumplir el SRI:

- Hemos de mantener un registro de **documentos**, cada uno de los cuales viene identificado por un número de registro.
- Todos los documentos tienen título e incluyen una URL mediante la que se puede acceder al documento en sí.
- Los documentos aparecen indexados en la base de datos en función de los **términos** (palabras) que aparecen en ellos.
- En la base de datos se ha de mantener la frecuencia de cada término en cada documento.
- Para facilitar la actualización del índice, junto con los datos de cada documento, se almacena el tamaño en bytes del documento, la fecha de última actualización del documento en el índice y el valor de una función hash (*checksum*) que se utilizará para comprobar si el documento actual es el que ya está indexado en la base de datos.
- Aparte de poder realizar búsquedas por palabras clave (al estilo de un buscador como Google o Altavista), el sistema de recuperación de información también ha de permitir al usuario navegar por la base de datos documental. Para ello, se han de mantener los documentos clasificados por **temas** (al estilo de un directorio como Yahoo! o dmoz.org).
- Los temas se organizan de la forma tradicional formando una taxonomía (un tema puede tener varios subtemas y ha de estar situado en una posición concreta dentro de la jerarquía de temas).
- Además, también se mantienen relaciones entre temas afines para facilitar la navegación del usuario por la base de datos documental (por ejemplo, los algoritmos de generación de números pseudoaleatorios usados en simulación están obviamente relacionados con la Estadística, aunque probablemente no aparezcan dentro del tema “Estadística” en nuestra clasificación oficial por temas).

Por cuestiones de eficiencia, la base de datos del sistema de recuperación de información almacenará de forma redundante los siguientes datos (por ejemplo, para permitir la ordenación de los resultados obtenidos para una consulta):

- Para cada término, el número de documentos distintos en los que aparece y el número total de veces que aparece en todos los documentos de la base de datos.
- Para cada documento, su tamaño (en palabras), el número de palabras diferentes que incluye y la frecuencia de la palabra que más se repite en el documento.



# Examen final – Convocatoria de febrero de 2005 ADQUISICIÓN Y TRATAMIENTO DE DATOS II

Nombre, apellidos y DNI

---

Diseñe una base de datos relacional para el problema planteado en la página anterior:

- Identifique las entidades y los atributos correspondientes a cada una de las entidades, indicando explícitamente cuáles serán las claves primarias de las distintas entidades.
- Analice las relaciones existentes entre las distintas entidades y determine su cardinalidad.
- Dibuje un diagrama entidad/relación que represente el modelo de datos definido por las entidades y las relaciones identificadas en los apartados anteriores.
- A partir del diagrama, obtenga el conjunto de tablas que nos permita implementar nuestro modelo en un sistema gestor de bases de datos relacionales, indicando las claves primarias de todas y cada una de las tablas de nuestra base de datos.

## Ejercicio de HTML

Escriba el documento HTML necesario para generar la página de inicio de nuestro sistema de recuperación de información:



NOTA: Puede realizar cuantas suposiciones considere oportunas acerca de las URLs donde se encuentran las imágenes de la página mostrada y acerca de los destinos correspondientes a los distintos enlaces.