

Gestión de Bases de Datos Geográficas

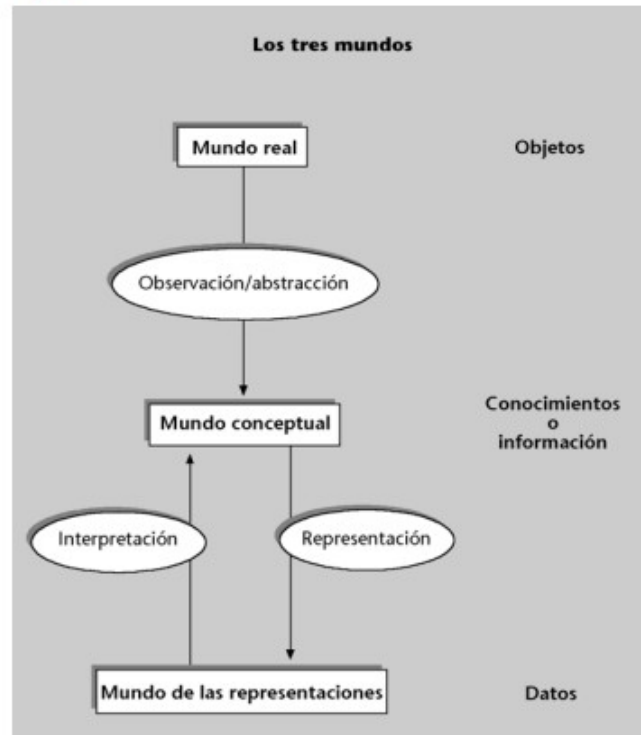
Clase 3



Los tres mundos: el real, el conceptual y el de las representaciones

- El mundo real con los objetos de nuestro interés.
- El mundo de las conceptualizaciones lógicas.
- El mundo de las representaciones informáticas.

Figura 1



La realidad: los objetos

Como informáticos que debemos analizar o construir un sistema de información (SI) determinado, necesitamos conocer el mundo real al que este SI debe hacer referencia o modelizar; así, nuestro mundo real podrá ser un hospital, una empresa distribuidora de productos alimenticios, la matriculación de los alumnos de una universidad, etc.

El **mundo real**, la parte de la realidad que nos interesa, es lo que percibimos con nuestros sentidos y está compuesto por objetos concretos, físicos o no.

Dado que trabajamos en el campo de los SI, los mundos reales que nos interesan son las organizaciones: empresas, instituciones, etc.

Ejemplos de objetos concretos

El enfermo Juan García, la cama 34 de la segunda planta, el almacén de Soria, el camión B-3452-AG, la alumna María Pi, la asignatura *Química I*, la enfermedad meningitis, la devolución de un pedido concreto, un determinado accidente de tráfico, etc. son algunos ejemplos de objetos que pertenecen al mundo real.

Las concepciones: la información

A partir de la observación del mundo real, los humanos somos capaces de deducir conocimientos e información. La observación de los objetos del mundo real nos conduce a su análisis y su síntesis; después, obtenemos abstracciones de ello, hacemos clasificaciones (podemos saber que dos objetos son de la misma clase a pesar de que sean diferentes), deducimos propiedades e interrelaciones, etc.

El conjunto de los conocimientos obtenidos a partir de la observación de un mundo real se denomina **mundo conceptual** o **mundo de las concepciones**. En la esfera de las concepciones construimos un modelo abstracto, conceptual, del mundo real, y esto nos ayuda a razonar y a expresarnos.

El **proceso de observación/abstracción** es básicamente un proceso para modelizar la estructura, las propiedades y el funcionamiento de la realidad.

De la observación obtenemos la información

La observación del campo de la matriculación en una universidad nos permite reconocer diferentes clases o tipos de objetos, como por ejemplo el *alumno* o la *asignatura*. Deducimos que todo alumno tendrá las propiedades (son abstracciones) *fecha de nacimiento*, *DNI*, *nombre*, etc., y de este modo obtenemos informaciones como las siguientes: el alumno de nombre Juan García tiene el DNI 34.567.854 y su año de nacimiento es 1979.

De hecho, existen diferencias entre *conocimiento e información*. La **información** es un conocimiento transmisible, es decir, que se puede representar. Los únicos conocimientos que nos interesarán aquí son, por lo tanto, las informaciones. !

Las representaciones: los datos

El mundo de las concepciones o de los conocimientos es un mundo mental. Sin embargo, para trabajar con estos conocimientos y comunicarlos, necesitamos proyectar los pensamientos al exterior, representándolos físicamente de alguna manera. Éste es el **mundo de las representaciones**.

Nosotros aquí nos ocuparemos de las representaciones informáticas, y hablaremos de datos, ficheros, bases de datos, registros, campos, *bytes*, discos, etc.

Damos el nombre de **datos** a las representaciones físicas de los conocimientos que tenemos de los objetos del mundo real. El paso de los conocimientos a los datos, o de una concepción a una representación informática, no es automático. Es un proceso humano: un proceso de diseño.

Podemos representar conocimientos...

... escribiendo a mano sobre un papel, grabando *bytes* en un disco magnético según un formato y una codificación determinados, etc.

La interpretación

Acabamos de ver el camino que nos conduce de la realidad a los conocimientos, y de éstos a los datos o las representaciones. Sin embargo, nos hará falta interpretar la representación. El proceso inverso al de representación se denomina **interpretación**.

¿Cómo se puede obtener conocimientos e información de una representación?

Hemos dicho que una información es un conocimiento que se puede representar, pero ahora, teniendo en cuenta el camino inverso, podremos decir que la **información** es el significado que le damos a los datos.

Introducción a las Bases de Datos

Antes de introducir las bases de datos geográficos y el almacenamiento de datos espaciales, es conveniente repasar los conceptos principales de almacenamiento de datos informáticos. De esta manera, podréis entender cómo se ha llegado al modelo de datos actual para los datos geográficos. Hay dos grandes familias de formatos de almacenes en los sistemas de información: los ficheros y las bases de datos.

Los ficheros se usan principalmente para informaciones con poco volumen que se suelen consultar en bloque. Las bases de datos se utilizan para grandes bloques de información que se quieren explotar y consultar de manera parcial.

Ficheros

Un fichero informático es un conjunto de información que se almacena en algún medio que permite al sistema operativo la lectura y escritura. Un fichero se identifica por un nombre y una extensión, que indica el formato. Los ficheros se organizan en carpetas dentro del sistema de ficheros que gestiona el sistema operativo.

Los ficheros almacenan información binaria en un orden y formato que permita a ciertas aplicaciones interpretar su contenido. La lectura y escritura del fichero pasa por funciones de bajo nivel del sistema operativo. Por ello la velocidad de acceso o la edición concurrente del fichero se ven limitados por las capacidades del sistema de ficheros del sistema operativo.

Edición concurrente

La edición concurrente se da cuando dos usuarios modifican un mismo conjunto de datos simultáneamente.

Bases de datos relacionales

Para definir qué es una base de datos relacional, debemos empezar por saber qué es una base de datos. Una **base de datos** es un conjunto de datos del mismo contexto almacenados y ordenados para que un sistema de información pueda seleccionar rápidamente una porción de éstos. Las bases de datos geográficas que veréis más adelante son las bases de datos que almacenan información para sistemas de información geográficos o SIG.

En contraposición con el resto de datos que podéis encontrar en un ordenador, las bases de datos se organizan y se gestionan mediante motores de bases de datos independientes del sistema de ficheros del sistema operativo denominados *sistemas de gestión de bases de datos* o *SGBD*.

Hay distintos modelos y tipos de bases de datos, pero actualmente hay uno que es el más extendido y el que usan todas las bases de datos geográficos: las bases de datos relacionales.

Las bases de datos relacionales son aquellas que están organizadas en tablas y relaciones entre tablas.

- Las **tablas** se organizan en campos y registros de manera similar a las columnas y filas de una hoja de cálculo.
- Cada uno de los **registros** de la tabla representa un objeto con las mismas propiedades.
- Cada una de las propiedades está almacenada en un **campo** de un tipo determinado.
- Toda tabla tiene una **clave primaria** formada por un campo o conjunto de campos. La clave primaria permite identificar inequívocamente cada uno de los registros según su valor.

No repasaremos el resto de tipos de bases de datos, ya que no se aplican a las bases de datos geográficos.

Ejemplo de tabla

En la figura 2 podéis ver representada una tabla con sus campos y sus registros.

Figura 2. Ejemplo de tabla de bases de datos

The diagram shows a table with four columns and six rows. A bracket labeled 'Campos' spans the top of the columns. A bracket labeled 'Registros' spans the left side of the rows. The table content is as follows:

DNI	Nombre	Apellidos	Teléfono
1234567X	Pep	Camps	897876765
2345671X	Maria	Camps	897876765
3456712X	Pau	Valls	-
4567123X	Pol	Prats	243562817
5671234X	Pep	Brou	-
6712345X	Marta	Miro	345654765

Esta tabla almacena datos de personas y tiene 4 campos (*DNI*, *Nombre*, *Apellidos* y *Teléfono*) de diferentes tipos. *DNI*, *Nombre* y *Apellidos* son cadenas de texto, mientras que *Teléfono* es un número. Cada uno de los registros de la tabla representa a una persona. Observad algunas características de los datos:

El campo *Nombre* tiene dos valores repetidos (el valor repetido es "Pep") y el campo *Apellidos* también (el valor repetido es "Camps").

El campo *Teléfono* tiene valores vacíos, ya que no todo el mundo tiene teléfono. Cuando un registro no tiene valor para un campo se dice que tiene el valor "NULL".

DNI es el único campo que tiene valores diferentes y no "NULL" para cada uno de los registros. *DNI* es el mejor (y único) candidato para ser la clave primaria de esta tabla.

otra tabla. Esta relación nos permite indicar que un registro de una tabla está relacionado con un registro de otra tabla. Una tabla también se puede relacionar con clave foránea con ella misma.

Podríamos añadir un campo a la tabla de personas representada en la figura 2 que fuera una clave foránea al padre de la persona. El campo contendría el valor del DNI del padre de la persona.