

Les dades: conceptes introductoris

Índex

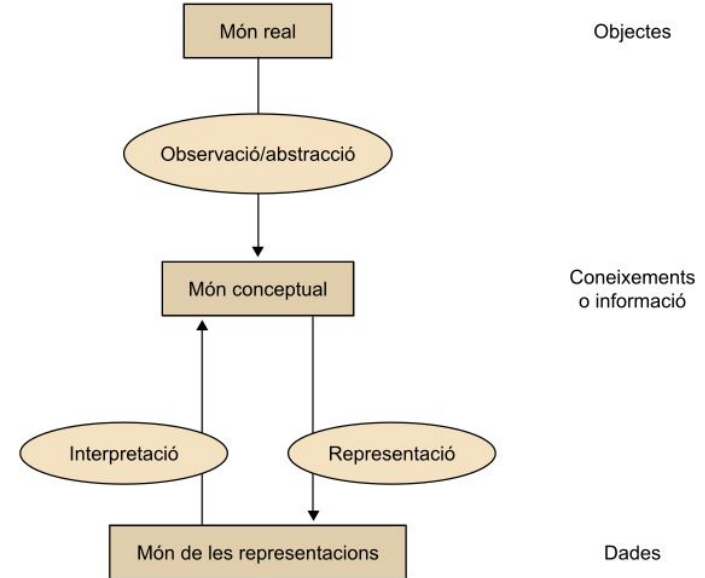
1. Els tres mons: el real, el conceptual i el de les representacions
2. El món conceptual: entitats i atributs
3. El món de les representacions
4. La memòria persistent

1. Els tres mons: el real, el conceptual i el de les representacions

Els tres mons

Per a tenir un marc on situar els termes i els conceptes que explicarem, distingirem tres àmbits diferents:

- El món **real** amb els **objectes** del nostre interès.
- El món de les **conceptualitzacions** lògiques.
- El món de les **representacions** informàtiques.



La realitat: els objectes

El món **real**, la part de la realitat que ens interessa, és el que percebem amb els nostres sentits i és compost per objectes concrets, físics o no.

Exemples d'objectes concrets:

- El malalt Joan Garcia
- El llit 34 de la segona planta
- El magatzem de Sòria
- El camió B-3452-AG
- L'alumna Maria Pi
- L'assignatura Química I
- La devolució d'una comanda concreta

Les concepcions: la informació

El conjunt dels coneixements obtinguts observant un món real, l'anomenem **món conceptual** o **món de les concepcions**.

En l'esfera de les concepcions construïm un model abstracte, conceptual, del món real, i això ens ajuda a raonar i a expressar-nos.

El procés d'**observació/abstracció** és bàsicament un procés per a modelitzar l'estructura, les propietats i el funcionament de la realitat.

La **informació** és un **coneixement transmissible**, és a dir, que es pot representar.

Les representacions: les dades

Per a poder treballar amb coneixements (que pertanyen al món mental) i poder-los comunicar, necessitem projectar els pensaments a l'exterior representant-los físicament d'alguna manera. Aquest és el món de les **representacions**.

Donem el nom de **dades** a les representacions físiques dels coneixements que tenim dels objectes del món real.

El pas dels coneixements a les dades, el pas d'una concepció a una representació informàtica, no és automàtic. És un procés humà, un procés de disseny.

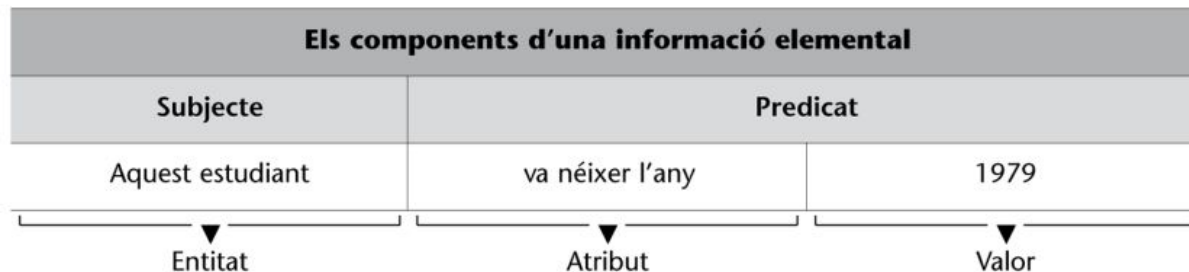
Acabem de veure el camí que ens porta de la realitat als coneixements, i d'aquests a les dades o representacions. Però ens farà falta poder interpretar la representació. El procés invers al de representació, l'anomenem **interpretació**.

2. El món conceptual: entitats i atributs

Entitats, atributs i valors

En termes lingüístics: una **informació** (un coneixement elemental) es pot expressar amb un **subjecte** (l'estudiant concret) i un **predicat** (“va néixer el 1979”). El predicat és format pel **verb** i el **complement**.

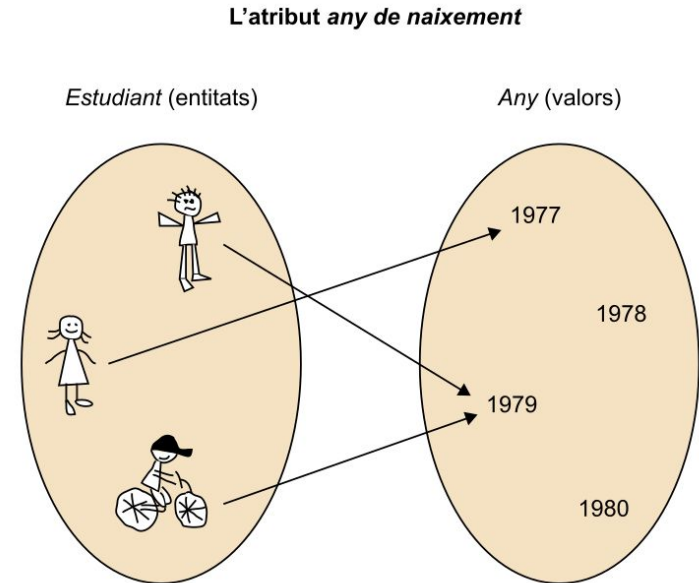
Des d'un punt de vista informàtic: anomenem **entitats** els objectes que conceptualitzem com a distingibles els uns dels altres (és a dir, que són identificables) i dels quals ens interessen algunes propietats. El *predicat* és la propietat descrita, i les seves dues parts, *verb* i *complement*, les anomenem **atribut** (any de naixement) i **valor** (1979), respectivament.



El marc de la teoria de conjunts

Fent servir la teoria de conjunts, podem veure la correspondència entre els estudiants i els anys com una **aplicació** (en el sentit de les matemàtiques) del conjunt dels estudiants sobre el conjunt dels anys:

- Cada estudiant té un sol any de naixement.
- Diferents estudiants poden tenir el mateix any de naixement.



Dominis, identificadors i claus

El conjunt de tots els valors vàlids que pot arribar a tenir un atribut, rep el nom de **domini** de l'atribut.

Pot passar que el valor d'un atribut determinat d'alguna entitat individual sigui desconegut o no existeixi. Llavors direm que el domini d'aquest atribut **accepta el valor nul**.

Els atributs que concebem com a aplicacions injectives¹ s'anomenen **identificadors**.

Tot atribut o conjunt d'atributs que permet identificar les entitats individuals rep el nom de **clau**.

Com que l'atribut és una aplicació entre conjunts, a cada entitat li pot correspondre com a màxim un sol valor. En conseqüència **un atribut no podrà ser multivalor** (o multivaluat).

¹ Una aplicació injectiva és aquella en la qual a cada element del conjunt imatge li correspon un element del conjunt origen com a màxim.

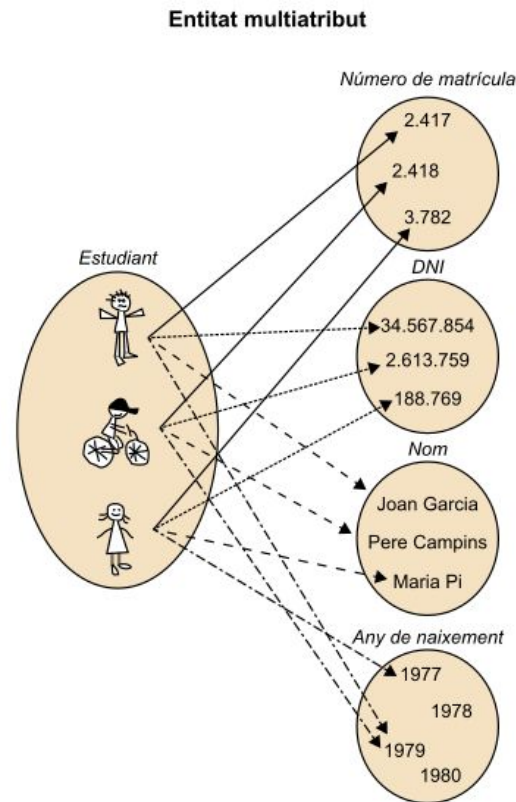
3. El món de les representacions

La representació tabular

La representació informàtica d'una informació elemental s'anomena **dada**.




La figura de l'esquerra és una **representació gràfica**, no informatitzada, de la informació dels estudiants, fent servir la representació clàssica de les aplicacions en la teoria de conjunts, per una entitat multiatribut.

Amb tantes fletxes i conjunts, no resulta gaire còmode per a ser processada....



La representació tabular

Resulta molt més senzilla una representació **tabular** amb una fila per a cada entitat individual i una columna per a cada atribut.

Representació tabular de la informació de la figura 5				
<i>Estudiant</i>	<i>Número de matrícula</i>	<i>DNI</i>	<i>Any de naixement</i>	<i>Nom</i>
	2.417	34.567.854	1979	Joan Garcia
	3.782	188.769	1977	Maria Pi
	2.418	2.613.759	1979	Pere Campins

Fitxers, registres i camps

Un **fitxer de dades** és una representació informàtica equivalent a la representació tabular:

- La representació d'una entitat, l'equivalent a una *fila* de la taula, rep el nom de **registre**.
- La representació del *valor* d'un atribut d'una entitat s'anomena **camp**.

El conjunt de camps constitueix el registre, i el conjunt de registres constitueix el fitxer.

Fitxer d'estudiants				
<i>número de matrícula</i>	<i>DNI</i>	<i>any de naixement</i>	<i>nom</i>	} Nom o capçalera dels camps
2.417	34.567.854	1979	Joan Garcia	
3.782	188.769	1977	Maria Pi	} Registres
2.418	2.613.759	1979	Pere Campins	

Fitxers interrelacionats

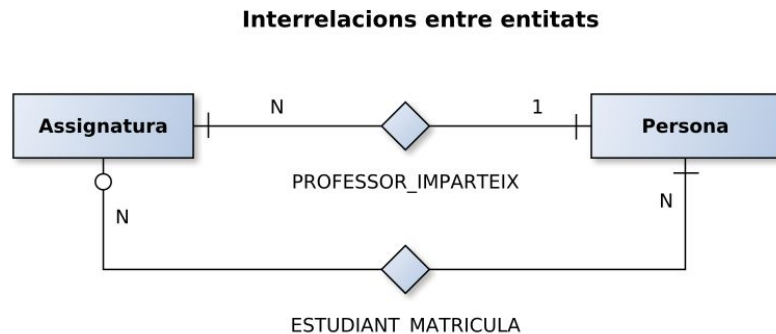
De moment, en aquest mòdul introductori donarem el nom de **base de dades** (BD) a un conjunt de fitxers de dades interrelacionats.

Suposem que els tipus d'objectes del nostre interès són *estudiants*, *assignatures* i *professors*.

Podrem representar aquestes entitats mitjançant tres fitxers, un per a cada entitat, amb els camps corresponents als atributs. Però hi falta la informació que permet interrelacionar les entitats entre si.

1) Tot estudiant pot cursar més d'una assignatura i tota assignatura pot ser cursada per molts estudiants.

2) Tota assignatura és donada per un sol professor, però cada professor pot donar diverses assignatures.



Els programaris especialitzats en aquests conjunts complexos de dades reben el nom de **Database Management Systems** (DBMS) o Sistemes de gestió de BD (SGBD).

Accés a les dades

Hi ha dues formes bàsiques d'accés a les dades:

- Accés **seqüencial**: l'accés a un registre pressuposa l'accés previ a tots els registres anteriors.
- Accés **directe**: l'accés a un registre no requereix l'accés previ a tots els registres anteriors.

Adicionalment, distingim entre:

- Accés per **valor**: l'accés per valor ens porta al registre en funció del valor d'algun dels seus atributs, sense tenir en compte la posició que ocupa el registre.
- Accés per **posició**: l'accés per posició ens porta a un lloc - una posició - on hi ha un registre de dades, sense tenir en compte el contingut.

Nivell lògic i nivell físic

Al món de les representacions informàtiques, s'acostuma a distingir dos punts de vista o nivells: el **nivell físic**, quan és necessari considerar la realització física, i el **nivell lògic**, quan no cal conèixer-la.

- **Nivell lògic:** per exemple, el programador d'aplicacions que treballa amb un llenguatge d'alt nivell com el C, el C++, el Java, etc., pot veure o imaginar que els fitxers són formats simplement per registres, l'un darrere l'altre, i que contenen camps amb lletres i números. No veu, ni li cal conèixer, la realització física que potser constarà d'encadenaments de registres físics (cadascun amb diversos registres lògics), marques separadores entre camps, compressió de dades, índexs, etc.
- **Nivell físic:** entrem al nivell físic quan hem de considerar la realització física. Un programador de programari bàsic (SGBD, SO, etc.), un dissenyador físic d'una BD, un tècnic de sistemes que administra una BD, etc., han d'entrar al nivell físic.

4. La memòria persistent

Justificació de la utilització de la memòria persistent

La necessitat d'emmagatzemar les dades ens obliga a utilitzar memòries **no volàtils** amb suports permanents, com ara els discs magnètics, els discos òptics, memòries flash, cintes, etc. A més de la no-volatilitat, aquesta memòria també té altres propietats: **gran capacitat** i el **baix preu per byte**.

El principal inconvenient d'aquests perifèrics d'emmagatzematge persistent és el **temps d'accés**, que és significativament més lent que el de la memòria interna.

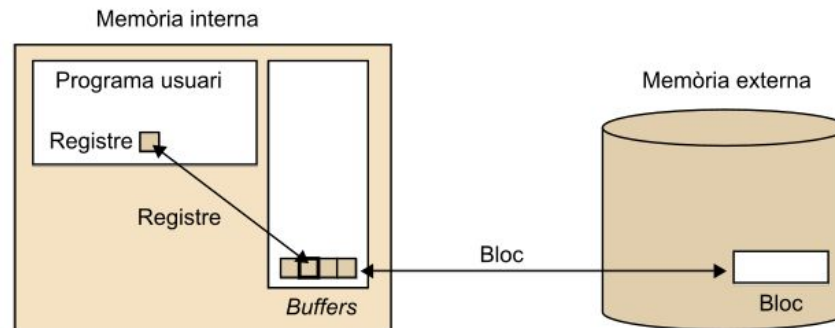
Esquema d'E/S

El **bloc** és allò que es llegeix o s'escriu de cop en una sola operació física d'E/S (entrada/sortida).

En el món de les BD es fa servir sovint el terme **pàgina** com a sinònim de bloc.

De la mateixa manera que la unitat de transferència entre la memòria persistent i els buffers és el bloc, la unitat de transferència entre aquestes i el programa d'usuari és el **registre**.

Esquema bàsic de l'E/S



Temps d'accés

El temps necessari per a completar una operació física de lectura o escriptura d'un bloc a una memòria persistent mòbil consta de dues parts:

- El temps d'**accés** és el temps necessari perquè el mecanisme es col·loqui a l'inici del bloc que s'ha de llegir o escriure.
- El temps de **transferència** és el temps necessari per a llegir o escriure el bloc.

Universitat Oberta
de Catalunya

Uoc

 UOC.universitat
 @UOCuniversitat
 UOCuniversitat
