



## EDUCACIÓ INTERDISCIPLINÀRIA

### Les bases del coneixement

#### *Conceptes sobre bases de dades per a neòfits*

per Jordi Xargay

Descobrir quina és l'essència del coneixement, com identificar els conceptes fonamentals i quines relacions s'estableixen entre aquests conceptes són tasques que han preocupat històricament tant filòsofs com lingüistes. L'evolució de la tecnologia que tracta la informació (informàtica<sup>1</sup>), en els darrers trenta anys, ha proporcionat el camp on aplicar conceptes clàssics sobre el coneixement, que fins al moment entraven només dins àmbits purament teòrics. La possibilitat d'emmagatzemar i gestionar quantitats molt grans d'informació, inimaginables fins al moment, ha revifat la problemàtica sobre la relació entre coneixement, conceptualització i lingüística. Aquesta part de la ciència té ara la possibilitat de proporcionar coneixement a partir del tractament de la informació.

Actualment, informació i coneixement són conceptes que es relacionen habitualment, fins i tot, poden arribar a tenir el grau (fictici) de sinònims. L'expansió de l'ús de conceptes com ara Tecnologies de la Informació ha portat a l'englobament de sectors de l'anteriorment coneguda com Informàtica, que sobrepassen la informació per a introduir-se, fins a un fites gens trivials, en terreny del coneixement. Aquest seria el cas de programes de gestió de bases de dades, que busquen l'emmagatzematge, l'estructuració i la consulta d'informació per a extreure'n coneixement.

Quan ens plantegem la transformació d'informació en un estructura informàtica consultable, haurem de tenir presents diversos aspectes: quina informació ens aporta coneixement, quin coneixement ens és realment útil i com sistematitzar el tractament de la informació per a tenir una font de coneixement versàtil i pràctica. Per a establir les bases, primerament haurem de diferenciar informació i coneixement. Aparentment, no sembla que hi pugui haver confusió:

---

1 Provenent dels termes francesos relatius al tractament automàtic de la informació (INFORMAtion automaTIQUE)



---

**Informació<sup>2</sup>:** Contingut d'una o més dades, fent abstracció de la representació concreta que adopta.

**Coneixement:** Funció i acte de la vida psíquica la característica dels quals és llur caràcter representatiu i, en el cas del coneixement humà, objectiu.

Podem comprovar que el coneixement implica una activitat psíquica i veurem que no és tan fàcilment estandaritzable com ens podríem pensar. És per això que, a l'hora d'introduir-nos en les bases de dades, no podem començar a tractar informació sense tenir clar el coneixement que en volem obtenir. Si no ho fem d'aquesta manera estem abocats a una molt costosa reestructuració posterior de la feina.

En el món informàtic, haurem de traduir aquests conceptes tan genèrics a un nivell pràctic: com a informació entendrem les dades que utilitzem, desprovistes de sentit profund; mentre que el coneixement serà el sentit que tenen aquestes dades, per a poder extreure'n relacions o conclusions. Per a saber quin és aquest sentit és preferible tenir una experiència en l'àmbit d'aplicació. Per a traduir aquestes relacions en estructures correctament formades caldrà un coneixement informàtic. Per tant, la cooperació d'un expert informàtic i un expert en la matèria sobre la qual versarà un projecte, serà la situació més habitual.

Per a dur a terme un projecte a un nivell no gaire elevat, ens seran suficients els gestors de bases de dades que ens proporcionen els paquets ofimàtics genèrics (MS Office, OpenOffice, etc.). Aquests gestors ens permeten emmagatzemar i consultar informació. L'agrupament des del punt de vista de la practicitat, dins de paquets ofimàtics (principalment el Microsoft Office, amb l'Access com a gestor de BBDD<sup>3</sup>), ha comportat que aquesta assimilació desemboqués en la consideració de la base de dades com una eina més per a contenir paraules, perdent de vista que el més important són els conceptes. Aquest desenfocament porta associat, principalment en neòfits, errors en l'anàlisi i el disseny de bases de dades, que posteriorment són molt difícils de solucionar. Per exemple, podem caure en la temptació de considerar tota la informació relacionada amb un tema com a important, i preveure la creació de consultes irrelevantes.

Les BBDD agrupen conceptes o tipus d'objectes en les anomenades **Taules**. Aquestes taules contenen tota la informació relacionada amb un concepte singular (per exemple, llibres d'una biblioteca). Una taula inclou informació sobre molts objectes del tipus definit per la taula (per exemple, emmagatzema informació sobre molts llibres). Els objectes del tipus de la taula se'ls

---

<sup>2</sup> Definicions extretes del diccionari en línia de la Gran Enciclopèdia Catalana (<http://www.grec.net/home/cel/dicc.htm>)

<sup>3</sup> Bases de Dades



anomenada **Registres**. Aquests registres representen tota la informació associada a un objecte a tractar. Cadascun dels aspectes informatius d'un objecte (per exemple, títol, editorial, autor, ...) són les cèl·lules d'informació que s'anomenen **Camps**. Les consultes d'informació sempre van orientades a demanar quins registres d'una taula determinada compleixen la condició de tenir certa informació en un camp determinat.

Crear una BDD pot arribar a ser un projecte per si sol, en funció de la mida, la complexitat de les dades i l'ambició en els objectius. Per a desenvolupar-lo haurem de seguir els passos habituals en tot projecte informàtic. A no ser que tinguem molta experiència en la creació de BBDD, ens trobarem sovint davant d'aspectes conceptuals inesperats i haurem de pensar i repensar l'estructura del projecte diverses vegades. Aquesta és una situació no gens estranya quan comencem a introduir-nos en les BBDD.

Saber el *què?*, l'*on?* i el *com?* de la majoria de conceptes és, actualment, feina fàcil. La quantitat d'informació que ens proporciona la Xarxa<sup>4</sup> és un exemple clar de com es pot obtenir informació massiva sobre qualsevol tema. Però també ens mostra que aquesta informació no es tradueix fàcilment en coneixement. Cal saber què preguntar a un cercador d'Internet per a obtenir informació que ens aportï coneixement útil. Aquest exemple transportat a la nostra temàtica ens indica que l'esforç s'ha de centrar, doncs, en el *per què?* i el *per a què?* per a treure profit de les eines informàtiques. Aquesta fase, en un projecte informàtic, se l'anomena **anàlisi**, i representa un pas vital en la vida del programa que es vulgui desenvolupar. Està calculat que un error detectat en la fase d'anàlisi comporta la despesa d'1 unitat d'esforç (temps, recursos,...) per a ser solucionat. El mateix error, si no és en les fases de desenvolupament i és detectat a la darrera fase (la fase de proves del programa), cal la despesa de 40 unitats d'esforç solucionar-lo. Aquesta diferència ens mostra quin és el principal perill en la creació de projectes informàtics: els errors en l'anàlisi dels requeriments inicials del problema.

Un altre dels aspectes que no són evidents a primer cop d'ull i que poden portar a la creació d'una BBDD que es quedarà ràpidament obsoleta o a una eina útil i versàtil, és la mida i l'abast del camp d'aplicació. Sovint, per la metodologia tradicional de l'ensenyament de qualsevol matèria, es considera que és millor començar amb un problema petit i senzill i anar-lo complicant. En aquest aspecte, qualsevol pedagog hi posaria la mà al foc. Es considera que obtenir un resultat correcte, per petit que sigui, en la fase d'aprenentatge és motivador i genera

---

4 La generalització d'Internet (1993: obertura de la primera pàgina comercial) provinent de la interconnexió de nodes a través de tot el món ha desembocat en el concepte de Xarxa amb majúscula: la xarxa que interconnecta xarxes menors fins a tenir un abast mundial.



una bona dinàmica de treball. En l'aprenentatge del disseny i gestió de BBDD, però, és aconsellable anar una mica més enllà. Podem considerar que la persona que s'introdueix en un món com aquest és una persona adulta, que ja ha tingut experiències en l'aprenentatge d'altres aspectes formatius i que ho fa per interès personal. Encara avui no és habitual que es demani aquesta mena de formació per a la vida professional. Per això no cal patir excessivament per la tan temuda desmotivació. De fet, ampliar les mires, de forma realista i sensata, comportarà una major satisfacció al final del projecte. És evident, però, que complicar la problemàtica farà que l'inici del projecte sigui més lent, però ja se sap que qui vol lluir ha de patir.

Però res millor que un exemple per a entendre tots els aspectes comentats en aquest article. Suposem que som un ornitòleg aficionat a la fotografia. És molt provable que hàgim ja compaginat les dues facetes de la nostra vida i tinguem un fons documental en forma de fotografies d'ocells. En poc temps, però la quantitat d'informació (fotografies i notes preses en el lloc i moment de fer-les) sobrepassarà la nostra capacitat de recerca. Aquesta és una situació propícia per a introduir-nos en la creació d'una BBDD.

Cada informació que pretenem emmagatzemar ocuparà espai. L'objectiu fonamental és tenir tota la informació que ens proporcioni coneixement utilitzant el mínim espai possible. Per això haurem de triar acuradament com estructurarem els nostres objectes. Per a cada objecte haurem de considerar una taula. Començarem per descriure la informació que podem obtenir de les nostres fotografies i relacionar-la amb la altres informacions que podem obtenir de la nostra faceta professional o personal. Els objectes principals són: els ocells i les fotografies. Anem a veure quina informació tenim de cadascun:

Sobre els ocells (Taula Ocells): nom, nom científic, família, hàbitat, migració, alimentació...

Sobre les fotografies (Taula Fotografies): dia, hora, lloc, obertura, exposició, contingut...

A la taula d'ocells tindrem registres que contindran informació específica d'ocells. Aquesta seria la part d'informació que la faceta d'ornitòleg pot aportar. A la taula de fotografies tindrem registres que contindran la informació relativa a cadascuna de les fotografies que hem fet. Però no n'hi ha prou amb la informació més evident, necessitem introduir altres dades per a poder utilitzar les taules amb eficàcia. Anem a veure quines i per què:

**Codi:** Un concepte fonamental de l'organització estructurada de qualsevol col·lecció de valors és que cada element diferenciat ha de poder ser identificat sense possibilitat d'error. Per això s'afegeix un camp anomenat codi, habitualment amb un valor numèric per la seva major simplicitat. Si volguéssim podríem fer servir la codificació tradicional en els ocells (el nom



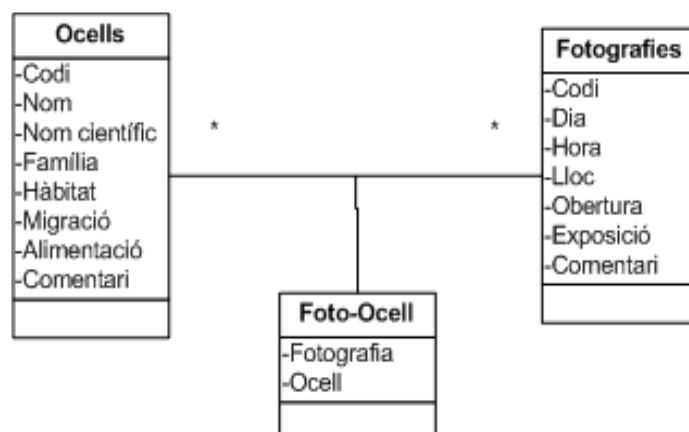
científic) però al ser llarg i en llatí podria comportar-nos errors tipogràfics que repercutirien en el coneixement que volem obtenir.

**Carpeta:** Tenir la informació pròpia de cada fotografia no és suficient, haurem de poder trobar-la per a veure-la. Lògicament, les fotografies les haurem de tenir agrupades d'alguna manera. Una possible agrupació seria per rodets o carpetes (en funció de si són analògiques o digitals). D'aquesta manera podrem utilitzar un camp que contindrà l'identificador de rodet o carpeta. El rodet l'haurem d'etiquetar físicament amb el mateix valor de **carpeta**. La carpeta que contingui les fotografies digitals haurà de coincidir amb el valor d'aquest camp. A més a més, per a major rapidesa, haurem de saber a quin fotograma correspon (en cas de fotografia analògica) o com es diu el fitxer que la conté (en cas de fotografia digital). Per a saber-ho haurem de considerar un camp anomenat **fotografia**.

Un altre aspecte que hem de tenir en compte és el fonament de la relació entre la taula d'ocells i la de fotografies. Cal que sapiguem quin o quins ocells apareixen a cada fotografia. La primera reacció podria ser la d'incloure al camp **contingut** de la taula de fotografies el nom dels ocells. Però tenim un problema. Cada camp només pot contenir un valor i en una fotografia poden haver-hi més d'un ocell. Si volem solucionar aquest problema tenim dues opcions:

Crear diversos camps dins la taula de fotografies per a poder contenir els diversos ocells que hi puguin haver. Però, quants camps? És impossible de preveure si farem fotografies on apareguin dos, tres o quatre ocells. Si ens decidim per una quantitat, és possible que algun dia no sigui suficient, per a una fotografia concreta. A més, podem veure que estem malgastant espai, en cas de fer fotografies amb un sol ocell.

Una altra possibilitat és considerar que cada registre de la taula fotografia conté un ocell, i si en una n'apareixen més, creem un altre registre (una mena de duplicat de la fotografia) amb el codi del segon ocell. En cas que n'hi hagi un tercer, repetiríem el procés. És fàcil comprovar que estem repetint informació sobre la fotografia (dia, hora, obertura,...), i per tant, malgastant espai de memòria. Per a solucionar-ho utilitzarem una estructura molt pròpia de les BBDD: les taules de relació.





Una fotografia contindrà un o més ocells, i un ocell apareixerà en una o més fotografies. Aquesta situació es defineix, en informàtica, com una relació *molts a molts* i es representa en un diagrama com el següent:

Per representar aquestes relacions amb el mínim d'espai de memòria utilitzarem la taula de relació Foto-Ocell, que contindrà els codis dels ocells i les fotografies que els contenen. Un exemple d'aquesta estructura seria el representat a continuació (elideixo alguns camps per a major comoditat):

Taula Ocells		
Codi	Nom	Nom científic
o1	Agró blanc	Egretta alba
o2	Corb marí gros	Phalacrocorax carbo

Taula Fotografies				
Codi	Dia	Hora	Carpeta	Fotografia
f1	22/10/05	11:35	20051022A	10
f2	22/10/05	11:40	20051022D	img14

Taula Foto-Ocell	
Ocell	Fotografia
o1	f1
o2	f1
o2	f2

Al camp **carpeta**, he afegit un sufix (A o D) per saber si la fotografia és analògica o digital.

D'aquesta manera, per saber en quines fotografies apareix l'agró blanc haurem de procedir de la següent manera:

- Saber a quin codi correspon l'agró blanc (o1).
- Consultar aquest codi a la taula Foto-Ocell, dins el camp Ocell.
- Obtenir els codis de Fotografia per a cada registre que contingui el codi de l'ocell al camp Ocell (o1).
- Consultar a la taula de fotografies quin rodet o carpeta la conté i el fotograma o nom del fitxer de la fotografia (20051022A i 10, respectivament).

Qualsevol gestor de BBDD ens permet automatitzar d'una manera fàcil aquestes tasques, per a obtenir-ne un resultat molt ràpid i eficient.



---

Cal destacar que a les taules que representen objectes és habitual utilitzar un camp amb una llargada indefinida de text (**comentari**) per a poder fer comentaris sobre els objectes que representa, en aquest cas, sobre les fotografies o els ocells.

Amb aquesta estructura ja podem començar a veure que el nostre projecte ja té cara i ulls. Aquest és el moment de no deixar-se portar per l'eufòria mecanogràfica, i prendre'ns un temps de reflexió abans d'introduir les dades que tenim fins el moment. És el moment propici per a obrir l'abast del nostre projecte, perquè ja hem reflexionat sobre les relacions entre objectes i tenim una estructura força correcta. Podem analitzar altres aspectes que puguin estar relacionats o puguin ser semblants a l'àmbit que hem tractat fins el moment, per a fer que la nostra BBDD sigui més versàtil, sense por d'equivocar-nos massa. Un exemple perfectament compatible amb el que hem dissenyat seria si, a més d'ornitòleg i fotògraf aficionat, també estiguéssim interessats per altres ciències naturals com ara la geologia o la mineralogia. Bàsicament, la relació que pot tenir qualsevol d'aquestes ciències amb la fotografia és la mateixa que té l'ornitologia: en fem fotografies per a tenir-ne un fons documental. Per això, no seria gens complicat ampliar la nostra BBDD per a cobrir també aquestes ciències. Suposem que volem afegir la nostra faceta mineralògica. Haurem de crear una taula (Taula Minerals), on desarem registres que estaran formats per diversos camps (codi, nom, composició, color...) i una taula de relació entre les fotografies i els minerals (Taula Foto-Mineral) que contindrà els índexs de la fotografia i el mineral representat. En aquest instant podem comprovar quina és la potència de les taules de relació: podem afegir informació sobre conceptes relacionats amb un mínim esforç i una mínima despesa d'espai.

En resum, hem vist que la creació d'una base de dades no és tan simple com pot semblar en un principi, ja que la necessitat de consultar les dades combinada amb la voluntat d'emmagatzemar-les en el menor espai possible fan que ens veiem obligats a la utilització d'estructures i tècniques pròpies de professionals de la informàtica, situació gens evident en un principi. Però tot l'esforç compensa, ja que una BBDD ben estructurada permet contenir moltíssima informació, que és fàcilment consultable i que podem adaptar a noves facetes de coneixement amb un esforç molt petit.

Desembre de 2005