

DISSENY I CREACIÓ DE BASES DE DADES

TEMA 1: SISTEMES DE BASES DE DADES

Contextualitza les bases de dades en els sistemes d'informació (SI).

Les definicions depenen del punt de vista des d'on es mira: el document o la informàtica:

Sistema d'informació (informàtica)

- Es tracta del mecanisme que proporciona els mitjans d'emmagatzematge, creació i distribució d'informació per tal de donar suport a les operacions i funcions de gestió d'una organització (Willits)
- Conjunt d'elements ordenadament relacionats entre si d'acord amb certes regles, que aporten a l'organització a la què serveixen la informació necessària per al compliment dels seus fins, per la qual cosa haurà de recollir, processar, emmagatzemar dades provinents tant de la mateixa organització com de fonts externes, facilitant la seva recuperació, elaboració i presentació (Miguel)

Les dues definicions es diferencien en:

- Willits – posa l'accent en la funció de les bases de dades.
- Miguel – posa l'accent en què és un conjunt d'objectes relacionats entre si.

Components dels Sistemes d'Informació:

Hi ha el contingut, que són les dades com a l'objecte del SI, el hardware per emmagatzemar les dades amb un equip perifèric per capturar les dades, el software (Sistema operatiu + software de gestió + software de control de comunicacions + altres softwares), i un equip humà (administradors + usuaris)

En un SI només hi pot haver informació de qualitat, perquè d'això depèn la presa de decisions.

- Precisió: Grau d'aproximació entre la informació emmagatzemada i la real
- Oportunitat: Temps transcorregut entre el moment en què s'ha produït la informació i aquesta es posa a disposició de l'usuari
- Completesa: Que la informació sigui completa per poder complir els seus fins
- Significació: Que la informació sigui significativa: tingui el màxim contingut semàntic possible i sigui comprensible
- Segura (Protegir la informació del deteriorament – físic – i d'accessos no autoritzats – restriccions amb drets d'usuaris–)
 - ◆ Confidencialitat
 - ◆ Integritat
 - ◆ Disponibilitat

En els contextes del SI, les bases de dades són un component del SID, però també un sistema de bases de dades. Un SI pot ser més ampli que una base de dades. Els sistemes de bases de dades (SBDD) es pot considerar un SI.

És important diferenciar entre SBDD i el concepte de BDD.

Bases de dades:

Les Bases de dades (BDD) són un component essencial dels sistemes d'informació, ja que habitualment el contingut (les dades) dels SI està estructurat en bases de dades.

- Definicions:

Definició funcional d'una base de dades:

- Es un sistema d'informació que manté registres sobre un altre sistema del món real i ha de representar amb la màxima fidelitat possible aquella part del món que intenta modelar (Codina: 1993, 69)
- És un conjunt de dades sobre una part seleccionada del món real que es vol fer servir per a propòsits particulars (Fidel, 1987: 5)

Es la representació d'un món real que té en compte dos elements:

- ♦ L'objecte de la base de dades (*empresa* segons la terminologia de Fidel)
- ♦ La finalitat (*entorn* segons la terminologia de Fidel): les característiques de l'entorn (les necessitats de l'organització) configuraran l'estructura de la base de dades.

La descripció és fonamenta pel què representa i no pel què és, ja que és un intent de representar el món real. Una BDD és una decisió arbitrària, perquè es fa perquè algú ho vol – decisió personal – amb unes intencions determinades. La funció i els objectius han d'estar molt ben definits.

Té en compte:

Objecte (documents)	Empresa
Finalitat	Entorn (necessitats) condicionat pels objectius de l'organització)
	FIDEL

Definició descriptiva d'una base de dades

- Col·lecció de dades interrelacionades, mantinguda en fitxers informàtics, que és accessible per múltiples usuaris i per múltiples aplicacions.
- Característiques de les bases de dades:
 - Les dades estan interrelacionades i estructurades segons un model, pensat i decidit. No és un dipòsit de dades. Fora de la informàtica, no existeix, ja que només són similis, com el catàleg manual d'una biblioteca. Les dades entraran si han estat prèviament modulades.
 - La redundància lògica de les dades ha de ser controlada. Que les dades no és repeteixin. És un control lògic. Introducció una sola vegada de les dades, malgrat que l'explotació de la dada sigui múltiple. Evita errors i ajuda a la consistència de les dades.
 - Atenen a múltiples usuaris i a múltiples aplicacions. Contrast amb el sistema de fitxers (un sol fitxer per aplicació). Explotació de les dades diverses vegades i de forma diferent segons l'usuari.
 - Independència física i lògica de les dades i el tractament (és la característica més important). L'emmagatzematge és diferent a l'explotació. Les dades s'estructuren en funció de l'explotació posterior.
 - Són autodocumentades: la definició o descripció de les dades està integrada amb les dades. Cada dada té la seva etiqueta i la seva interrelació.
 - Asseguren el manteniment de la seguretat (confidencialitat, disponibilitat i integritat) de les dades
- Arquitectura funcional:

La distinció entre la representació física i lògica de les dades implica que l'esquelet funcional de les bases de dades presenta una arquitectura a tres nivells:

- Nivell conceptual: Visió global de les dades. Inclou la descripció de tots les dades i les seves interrelacions
 - Nivell extern: És la visió que un usuari en particular té de la base de dades. Només hi ha de quedar reflectits aquelles dades i interrelacions que necessiti el corresponent usuari. S'ha de pensar en els drets d'accés i el disseny del sistema de gestió per visualitzar les dades.
 - Nivell intern (físic): Mètode d'accés físic. Fa referència a l'emmagatzematge de les dades i als mètodes d'accés. S'adapta a un SGBD determinat (depèn d'ell).
- Avantatges de les bases de dades:

Respecte a la gestió de les dades:

- Major eficàcia en la recollida, validació i introducció de les dades al sistema En un SGBD, les dades es centralitzen en un sistema i es poden exportar en múltiples casos. En un sistema tradicional de gestió de dades, les dades es repeteixen en múltiples llistats.
- Reducció de l'espai d'emmagatzematge S'entra una única vegada.
- Independència de les dades respecte als tractaments En una base de dades quan canvien el tractament fet fins ara no cal canviar les dades, d'altra manera si cal canviar les dades. La independència de les dades respecte el seu tractament és el que dóna avantatges a la base de dades. Un canvi en els tractaments no suposa un canvi de disseny de la BDD (ni l'estructura ni les dades que tenim)

Respecte als resultats:

- Coherència dels resultats Reducció d'errors, perquè una única entrada per dada.
- Major valor informatiu Les interrelacions entre les dades poden donar més visions. Un registre d'una BDD et dóna tota la informació possible, en un fitxer no és així.
- Respecte als usuaris:
 - ◆ Accés més ràpid i homogeni. Una única interfície de consulta. Els fitxers s'han d'obrir per separat.
 - ◆ Facilitat per atendre demandes canviants. Si es necessita un nou tractament de les dades, demanes una presentació diferent.
 - ◆ Facilitats per compartir les dades entre diversos usuaris

Les bases de dades estan fetes per tractar grans volums d'informació.

- Inconvenients de les bases de dades:
 - Instal·lació costosa pel software, sobretot per disseny i manteniment.
 - Personal especialitzat: un informàtic, un analista de la informació especialitzat. És un dels papers del documentalista. Fins ara hi havia híbrids d'informàtics i documentalistes.
 - Implantació llarga i laboriosa
 - Falta de rendibilitat a curt termini
 - Grau d'estandardització poc elevat. Hi ha sistemes de gestió no compatibles. No és pot assumir el cost de migrar dades d'un sistema a un altre.

Per les grans biblioteques hi ha un estàndard, VTLIS. Les biblioteques petites no ho poden assumir.

TEMA 2: METODOLOGIA PER AL DISSENY DE LES BASES DE DADES

La metodologia per dissenyar les bases de dades és convencional. Tots els manuals segueixen la mateixa metodologia.

Les premisses metodològiques per al disseny de les bases de dades són:

- Cal una planificació estratègica que analitzi al detall el sistema objecte que es vol modelar
- El disseny lògic del sistema d'informació és el pas anterior a la decisió sobre la infraestructura tècnica que l'implementarà.
- Els objectius i la definició de requeriments són les directrius que primen sobre la definició de base de la base de dades. Els SGBD s'han d'ajustar als requeriments establerts en la fase de disseny

En definitiva:

- Cal partir dels aspectes coneguts – l'entorn de l'empresa
- Anar dels aspectes generals als concrets — sempre s'ha de dissenyar en global, tenint en compte l'organització
- Anar dels aspectes lògics als físics – és prioritari el disseny intel·lectual

El disseny de la base de dades és el més important, perquè si el programa informàtic està ben dissenyat sempre es poden migrar les dades. Sempre hem de partir dels objectius (de les necessitats). S'ha de planificat al detall tots els components de les necessitats.

El cicle de vida de la creació de la base de dades:

- Anàlisi
- Anàlisi de l'entitat
- Anàlisi de la informació
- Estudi de necessitats
- Disseny
- Definició i objectius
- Elaboració de l'esquema conceptual
- Descripció funcional de la base de dades
- Implantació

Les fases posteriors sempre condicionen les fases anteriors. S'ha de tornar enrera. Quan dissenyes perfils i perfecciones coses de la fase d'anàlisi.

Fase d'anàlisi

- Anàlisi de l'entitat (*entorn*) : Cal recollir informació sobre el tipus d'organisme, l'activitat, els objectius, funcions, estructura i organització del treball i l'estratègia de l'organització quant a les tecnologies de la informació (vital als arxius)
- Anàlisi de la informació (*empresa*) : Recollir informació sobre les classes d'informació o els tipus d'entitats sobre els quals el sistema d'informació ha de mantenir registres.
- Estudi de necessitats : Establir les necessitats per tipus d'usuaris del sistema d'informació. Les necessitats condicionaran els objectius.

Fase de disseny

- Definició i objectius
 - ◆ Funcions i objectius
 - ◆ Tipus de bases de dades

◆ Idioma

Hem de tenir clares les funcions i els objectius, saber amb quins sistemes farem la gestió, etc. Tipus d'accés: local o extern... les coses més generals.

- Elaboració de l'esquema conceptual (*nivell conceptual*)
 - ◆ Determinació d'entitats
 - ◆ Determinació de les relacions entre les entitats
 - ◆ Diccionari de dades
 - ◆ Determinació d'usuaris i drets d'accés

Com estructurarem la informació?? Hem de conèixer les entitats que hi estan implicades.

Per entendre el concepte d'*entitat*, quan parlem de bases de dades, parlem d'objectes o activitats. Ex. Futbolistes, equips, partits de futbol (activitat)

Les entitats tenen *relacions entre elles* i hem de saber quines relacions hi ha. Hem de fer el *diccionari de dades* (camps) i definir els paràmetres sobre els camps. Determinar els *usuaris* i els *drets d'accés*. Usuaris de la base de dades (informàtics, bibliotecaris i usuaris de consulta)

- Descripció funcional (*nivell extern*)
 - ◆ Mètodes d'introducció de dades
 - ◆ Interfície de consulta
 - ◆ Formats de visualització
 - ◆ Serveis i productes que genera el sistema (llistat bibliogràfic, etc.)
 - ◆ Sistema de distribució (en CD-ROM, on-line, etc.)

Decidir quin tractament que es dona a les dades.

Fase d'implantació

- Selecció dels instruments
 - ◆ Descripció de l'oferta
 - ◆ Avaluació de propostes

Després del disseny hem de buscar un programa informàtic. A vegades és més important l'actitud de l'entitat davant les TIC. En els SIDs es té molt present el software utilitzat en organitzacions similars.

- Instal·lació i elaboració de l'aplicació

Quan hem escollit un producte, cal elaborar la base de dades. La tasca del disseny és tasca dels documentalistes.

- Fase de proves
- Elaboració del llibre d'estil (implicant a tots els que hi ha treballat, si cal els usuaris)
 - ◆ Concepció del projecte d'automatització
 - ◆ Disseny tècnic de l'aplicació
 - ◆ Normes de descripció i manual d'estil (s'ha de posar per escrit el que s'ha decidit i tota la fase de disseny)
 - ◆ Manual de manteniment (com reindexar, com fer còpia de seguretat, com compactar, etc.)
- Formació d'usuaris
- Control de qualitat (manteniment i avaluació)

- Promoció

Eines per fer el disseny:

- L'entitat (entorn)
- Models abstractes (teòric) de disseny de bases de dades (ajuden a estructurar les dades; s'estudia l'entorn i s'aplica el model)
- Eines de normalització de representació de la informació (això ens condicionarà els camps, la puntuació, etc.)

TEMA 3: MODELS DE DADES

Models en funció de l'arquitectura funcional de la base de dades. Els models abstractes (conceptuals) ens permetran la construcció de les bases de dades.

Model de dades Un model de dades és un conjunt d'eines conceptuals que permeten descriure a diferents nivells d'abstracció l'estructura d'una base de dades.

[model de dades = conjunt de conceptes, regles i convencions que ens permeten descriure i manipular (consultar i actualitzar) les dades d'un cert món real que desitgem emmagatzemar a la base de dades]

Model extern = com ho veu l'usuari. *[punt de vista de casa usuari en particular – permeten representar les dades que necessita cada usuari, sobretot amb estructures pròpies del llenguatge de programació]*

Model intern = com s'emmagatzemen les dades. *[orientats a la màquina – punt de vista de la màquina]*

Els models de dades conceptuals faciliten la descripció global del conjunt d'informació del món real que es vol emmagatzemar en una base de dades. Permeten representar i descriure les dades i les seves interrelacions. El model de dades sempre s'aplica en el món real.

És una representació de la realitat a través d'un procés d'abstracció. Ens permet analitzar el món real per poder convertir-lo en una base de dades. Això ho traduïm a un esquema conceptual.

Tradueixen el món real en un esquema conceptual. Una cosa és el model conceptual i l'altre l'esquema de la base de dades.

Tipus de models de dades:

- Models de dades pròpiament **conceptuals** (independents de qualsevol tecnologia)
 - ◆ Faciliten la descripció del sistema objecte amb independència de la màquina
 - ◆ Són models d'anàlisi no d'implantació
 - ◆ Els seus conceptes són propers al món real
 - ◆ Alt nivell d'abstracció

Exemple: Model Entitat-Relació

- Models de dades **lògics**
 - ◆ Descriuen les dades segons l'estructura i el model suportat per un SGBD determinat.
 - ◆ Són models més enfocats a la implantació

Exemples: Relacional, Xarxa, Jeràrquic

Quan es dissenya una base de dades s'usa un model conceptual que s'ha de passar a un model lògic (usant tecnologies). Un bon disseny passa per aquests dos processos.

Model conceptual Entitat–Relació:

Fou creat per Chen (1976 i 1977)

Es centra en l'estructura lògica i abstracta de les dades com a representació del món real. L'objectiu del model és identificar els objectes i les relacions entre els objectes i poder-los representar de forma gràfica i entenedora a través del diagrama Entitat–Relació.

[model E/R visió unificada de les dades, centrant-se a l'estructura lògica i abstracte de les dades, com la representació del món real, amb independència de la consideració del tipus físic]

Entitats Són cadascun dels objectes del món real sobre els quals volem emmagatzemar informació a la base de dades.

- Poden ser persones, llocs, coses, conceptes o successos, reals o abstractes
- Es designa amb un nom
- Es representa amb un rectangle

Un **tipus d'entitat** és una categoria abstracta que designa les classes d'objectes. *[Estructura genèrica]*

Ex. AUTOR

Una **ocurrència d'entitat** és un objecte concret del tipus d'entitat. *[cadascuna de les realitzacions concretes d'aquest tipus d'entitat]*

Ex. Cervantes

Es divideixen en entitats regulars i entitats dèbils:

- **Entitats regulars:** les ocurrències del tipus d'entitat tenen existència pròpia.

Ex. Documents

- **Entitats dèbils:** les ocurrències del tipus d'entitat dèbil depenen de l'existència del tipus d'entitat regular de la qual depenen. Si s'elimina l'ocurrència del tipus d'entitat regular desapareixen les ocurrències de l'entitat dèbil.

Ex. Exemplars (núm.) Si no tenim cap document, no tindrem cap exemplar.

Atribut És cadascuna de les propietats o característiques d'un tipus d'entitat.

- Són definibles pel creador de la base de dades

Ex. l'entitat DOCUMENT té com atributs: AUTOR, TÍTOL, EDITORIAL, ANY, ISBN

- És un concepte abstracte
- Es representa gràficament amb una elipsi, o, representació simplificada:

El **valor de l'atribut** correspon a cadascuna de les dades associades a una ocurrència d'entitat.

Ex. El valor de l'atribut AUTOR de l'ocurrència d'entitat *El Quijote* és Cervantes

El **Domini** és el conjunt de valors que admet un atribut. Moltes vegades és la informació de l'atribut, els valors que hi poden contenir i els que no. [*el domini té un nom i una existència pròpia amb independència de qualsevol entitat o atribut*]

Ex. El domini de l'atribut DATA DE PRÉSTEC és DATES

Tipus:

- Atribut Identificador Principal (AIP): del conjunt d'atributs d'un tipus d'entitat se n'ha de definir un que identifiqui de forma unívoca cada ocurrència
- Atributs simples i compostos
 - ◆ Simples: contenen valors no subdividits en parts. Ex. Editorial
 - ◆ Compostos: contenen valors subdividits en parts. Ex. Autor
- Atributs singulars i múltiples (univalorats i multivalorats)
 - ◆ Singulars: admeten un valor
 - ◆ Múltiples: admeten més d'un valor
- Atributs nuls: un valor nul es quan una ocurrència d'entitat no té un valor per un atribut
- Atributs derivats: resultat d'operacions fetes amb el valor d'altres atributs. Ex. Data de devolució

Interrelació Són associacions entre dues entitats.[*associació o correspondència entre entitats*]

- Sol ser un verb ja que es refereix a una acció
- Representació gràfica: Rombe
- **Tipus de correspondència:** Expressa el número màxim d'ocurrències d'una entitat a les que una altra entitat pot estar relacionada.
- La **raó de cardinalitat** expressa per a cada entitat el número mínim i màxim d'ocurrències de cada tipus d'entitat que participa en la interrelació.

Possibilitats: (0,1), (1,1), (0,n), (1,n)

Un a un:

Una ocurrència de capital ha d'estar vinculada a 1 com a mínim i a una com a màxim ocurrència de nació.

Una ocurrència de nació ha d'estar vinculada com a mínim a una i com a màxim a una ocurrència de capital

Un a molts:

Una ocurrència d'article ha d'estar vinculada com a mínim a una ocurrència de revista (Cada article per força es publica a una revista) i com a màxim a una revista (se suposa que un article no es pot publicar en més d'una revista).

Una ocurrència de revista ha d'estar vinculada com a mínim a una ocurrència d'article (una revista té com a mínim un article, no en pot tenir 0). Com a màxim pot estar vinculada a molts articles.

Molts a molts:

Una ocurrència de llibre ha d'estar vinculada a cap autor (un llibre pot no ser escrit per ningú) o a molts autors

(pot estar escrit per més d'un autor)

Una ocurrència d'autor ha d'estar vinculada a un llibre com a mínim (si no hagués escrit cap llibre no estaria a la base de dades) i pot estar vinculada a més.

Les interrelacions poden ser reflexives, és a dir, una entitat pot estar relacionada amb si mateixa.

O dues interrelacions entre les mateixes entitats:

- **INTERRELACIÓ ÉS_UN**: és el tipus d'interrelació que es dona entre un supertipus i un subtipus.

La característica més important és l'herència: els atributs d'un supertipus sempre són heretats per un subtipus. Les interrelacions que afecten a tots s'associen al supertipus i les específiques s'associen a cada subtipus.

Les interrelacions també poden tenir atributs.

EXEMPLE

Característiques de l'entorn i l'empresa:

L'entitat es dedica a la venda d'articles i funciona de la següent manera:

- Els clients realitzen demandes dels articles que volen comprar
- Cada venedor recull les demandes dels clients
- Cada demanda pot estar formada per un o més articles
- Es treballa amb diversos proveïdors. Cada proveïdor té més d'un article. Cada article, però, es compra a un únic proveïdor.
- Una vegada confeccionada la demanda se li comunica al client. Si el client l'admet s'emet la factura corresponent.
- Alguns empleats supervisen altres empleats

Identificar les entitats

- Els **clients** realitzen **demandes** dels **articles** que volen comprar
- Cada *venedor* recull les *demandes* dels clients
- Cada *demanda* pot estar formada per un o més *articles*
- Es treballa amb diversos **proveïdors**. Cada *proveïdor* té més d'un *article*. Cada *article*, però, es compra a un únic *proveïdor*.
- Una vegada confeccionada la *demanda* se li comunica al *client*. Si el client l'admet s'emet la **factura** corresponent.
- Alguns **empleats** supervisen altres *empleats*.

TEMA 4: MODELS LÒGICS DE DADES

Després del model conceptual, s'ha de convertir en model lògic, segons el suport tecnològic. Descriuen les dades segons l'estructura i el model suportat per un SGBD determinat. Models enfocats a la implantació

Evolució dels models de dades

Model jeràrquic: (dècada 1960)

Model en xarxa (model CODASYL) (dècada 1960)

Model relacional (dècada 1970)

Model orientat a l'objecte (dècada 1990)

Hi ha diferents models lògics (jeràrquic, en xarxa, relacional), que funcionen segons una estructura compartida a tots ells. Cada model és una millora de l'anterior (el model en xarxa és una millora del jeràrquic).

L'objecte de treball són les dades i el procés que hi ha en les dades (cada dada porta incorporat un procés). El procés és decisió del dissenyador.

El model orientat a l'objecte en una base de dades de factures d'un arxiu, cada factura té la data de quan ha de morir.

- **El model jeràrquic**

Sorgeix als anys 60's.

Estructura del model jeràrquic: estructura en forma d'arbre composta de nodes, que representen les entitats, i enllaçats per línies que representen les interrelacions.

Permet representar perfectament la jerarquia: un element pot tenir diversos elements dependents (fills) però només pot dependre d'un (pare). Permet tenir diversos nivells de la jerarquia

Les dades es connecten entre si mitjançant punters físics, entenent com a tal una direcció física que identifica on es pot trobar un registre sobre el disc

Inconvenients del model jeràrquic

- No permet representar la relació molts a un, ni la relació molts a molts
- La necessitat de recórrer pels punters físics per recuperar la informació fa que la realització de consultes sigui complexa.

[el món real no està format de monojerarquies, per tant, no ens serveix per representar-lo. La realitat és més complexa]

[la consulta és molt lenta, perquè al fer-la s'ha de passar per tots els nivells a fi d'arribar a les dades]

- **El model en xarxa**

Es comença a investigar el 1965 i es va normalitzar segons la norma CODASYL de 1973.

Estructura: conjunt de registres connectats entre si mitjançant enllaços.

Un element pot tenir elements dependents i a l'hora pot dependre d'un o més elements. Els enllaços són associacions exactes entre dos registres i es realitzen mitjançant punters físics

Avantatges i inconvenients:

- És més flexible que el model jeràrquic perquè permet representar les múltiples relacions pares fills
- Les relacions es fan també mitjançant punters físics, el que representa també un inconvenient ja que les interrelacions s'han definir quan es dissenya la base de dades i no en el moment de l'explotació

• El model relacional

Fou definit per CODD, matemàtic de l'Institut d'Investigació d'IBM de Califòrnia el 1970. Durant la dècada dels setanta es va desenvolupar la teoria i es produïren prototips [*basat en la teoria de les relacions*]. En la dècada dels vuitanta es creen els primers productes comercials que compleixen els requisits d'un model relacional. El model es va imposar i pràcticament es va convertir en un estàndard, per diverses raons:

- ◆ Les dades es presenten de forma comprensible
- ◆ Les dades queden separats de l'estructura física

Les característiques fonamentals del model són les següents:

- ◇ L'estructura de les dades és independent del mitjà físic.
- ◇ Incorpora regles d'integritat que proporcionen una base sòlida per la consistència de les dades
- ◇ Manipula les relacions en forma orientada a conjunts a través del llenguatge basats en l'àlgebra relacional

[*La base de dades relacional és un conjunt de taules que es poden relacionar.*]

Estructura de les dades Elements:

- Relació: es representa mitjançant una *taula* de dos dimensions que conté files i columnes de dades. Una taula és [*sempre*] la representació gràfica d'una relació [*no necessàriament*] o d'una entitat.
- Atribut: correspon a una *columna* d'una taula, que representa una propietat de la mateixa.
- Domini: correspon al conjunt de valors admesos per un atribut. [*conjunt d'un atribut. Definició concreta. No es pot barrejar més d'un concepte dins d'un atribut*]

Ex. Domini = tesarurs (tots els temes han d'estar dins del tesaurus).

Atribut títol normes bibliogràfiques

- Tupla: una fila de la relació o taula. És el registre. Són les ocurrences de l'entitat. [*contenen els valors que pren cadascun dels atributs per cada element de la relació*]
- Clau primària: És un atribut que identifica de forma unívoca cada fila. Cada fila del model relacional ha de ser única: no hi pot haver dos files amb els mateixos valors.
- Clau aliena: és un atribut d'una taula que a la vegada és clau primària d'una altra taula. S'utilitza per establir relacions entre les taules.

El valor atribut ha de ser igual a les entitats.

Restriccions d'integritat: assegurin la consistència de les dades:

- Integritat de l'entitat o integritat de la clau: cap atribut que és clau primària pot tenir valors nuls. Els valors de les claus primàries han de ser únics.
- Integritat referencial: els valors de les claus alienes han de ser els mateixos que els valors de les claus primàries corresponents. No es pot introduir un valor en una clau aliena que no existeixi com a valor en la clau primària. Les claus alienes només existeixen si hi ha claus primàries.
- Dependències funcionals: conjunt de regles de normalització que ajuden a estructurar les relacions evitant possibles redundàncies i anomalies.

Àlgebra relacional Conjunt d'operacions que permeten relacionar les dades de les diferents taules per crear

noves relacions (o taules). [*explotació de les dades*]

Tipus d'operacions:

- ◆ Selecció: subconjunt de les files de les taules. [*registres d'una entitat*]
- ◆ Projectió: subconjunt de les columnes d'una taula
- ◆ Unió: crear una taula combinant les dades d'altres relacions [*mostrar relacions*]

Pas del model Entitat-Relació al model relacional

- Transformació d'entitats i atributs:
 - Una Entitat amb atributs = a una taula o relació amb atributs
 - L'atribut identificador principal de l'entitat = clau primària de la taula
 - Els altres atributs = columnes o atributs de la relació
- Transformació d'interrelacions
- Interrelació molts a molts (N:M)
 - Un tipus d'interrelació N:M es transforma en una relació o taula
 - Els atributs d'aquesta nova taula seran:
 - ◆ Els atributs identificadors principals de les entitats que associa
 - ◆ Els atributs propis que pugui tenir

[*és la manera de solucionar la repetició de les dades*]

- La clau primària de la relació és la concatenació dels atributs identificadors principals de les entitats que associa
- Cadascun d'aquests atributs que formen la clau primària és una clau aliena que referència a l'entitat de la qual és clau primària
- Interrelació molts a molts (N:M).
- Interrelació 1:N. Possibilitats
 - Transformació de la interrelació en una taula: només se sol fer quan la interrelació té atributs propis
 - Propagació de clau: cada entitat es transforma en una taula. La interrelació no es transforma en cap entitat. Es propaga la clau primària de l'entitat que té com a cardinalitat màxima 1 a l'entitat que té n. I aquest atribut es converteix en clau aliena.
- Interrelació 1:1. Possibilitats
 - Crear una relació
 - Propagació de la clau: és el comportament més habitual. En aquest cas la propagació de la clau es pot fer en les dues direccions.

TEMA 5: SISTEMES DE RECUPERACIÓ D'INFORMACIÓ

◆ Els models lògics de dades en la gestió documental

En el cas de la Gestió Documental el model Entitat-Relació és l'aparell instrumental necessari per analitzar el món real i representar-lo en un esquema de base de dades. El model relacional s'aplica dins dels Sistemes de Gestió Documental per a controlar les operacions de gestió (adquisició i préstec) però no és l'adequat per a la representació dels documents.

[Dades " informació Tot el que implica processos es fa amb models relacionals, mentre que el que implica descripció es fa amb el model textual (sistema de gestió documental). No és el mateix recuperar informació que recuperar dades].

Raons:

- És un model excessiu per representar els documents, els quals formen un conjunt homogeni que es pot representar mitjançant una sola entitat.

Excepcions: Gestió d'arxius i descripció de documents administratius; i descripció bibliogràfica de nivell analític.

- Les restriccions imposades per les regles d'integritat no permeten processar informació de gran quantitat de text poc estructurat
 - ♦ els documents tenen atributs que no poden ser tractats de forma atòmica (títol, resum, text íntegre)
 - ♦ Els documents contenen informació no estructurada i estadísticament difícil de predir (títol, resum, text íntegre)
 - ♦ La recuperació d'informació textual té característiques pròpies i diferencials de la recuperació de dades i els seus reptes no són sempre satisfets per les prestacions del model relacional.

♦ **La recuperació de dades versus la recuperació d'informació**

Blair diferencia la Recuperació d'Informació o recuperació documental (RI) i la recuperació de dades en 4 àrees:

- ♦ La forma i el mètode d'elaborar una pregunta

RD: Quin és el salari d'aquesta persona? [*recuperació de dades – pregunta concreta (dades)*]

RI: Quants documents hi ha sobre el tema? = factor d'indeterminació [*recuperació d'informació = recuperació documental*]

[*RI pregunta més genèrica. La resposta ens dona documents on podem trobar la informació. Resposta que no és de si/no/valor concret. Resposta no és directa, sinó obtenció de documents per resoldre-la*]

- ♦ La relació entre la pregunta formal al sistema i la satisfacció del usuari

RD: relació directa entre una pregunta ben formulada i una resposta correcta: recuperació determinista. [*resposta positiva*]

RI: relació probable entre una pregunta ben formulada i una resposta satisfactòria: recuperació probabilística.

Ex. Quants llibres hi ha del Quim Monzó?

[*Quan es refereix al contingut és probabilístic, ja que depèn de molts factors indeterminats. Factors indeterminats determinats pel punt d'accés dels documents*]

- ♦ Criteris d'èxit en la recuperació [= utilitat del que s'ha recuperat per a l'interessat]

RD: el criteri d'èxit és la correcció: l'avaluació és objectiva.

[criteri d'èxit = utilitat recuperació per un usuari / cercador]

RI: el criteri d'èxit és la utilitat i es medeix per la taxa de precisió i la taxa d'exhaustivitat: l'avaluació és subjectiva

[rellevància dependrà de l'interès de la persona]

◆ Velocitat de la resposta

RD: depèn de la velocitat de l'accés físic

RI: depèn del número de decisions lògiques que pren l'usuari en el transcurs de la cerca

◆ **El model textual**

Desenvolupat a partir dels anys 70 per tal d'atendre les necessitats de biblioteques i centres de documentació.

Components del model:

· Estructura: registre textual

Codi	020103
Títol de la sèrie	Expedients de personal
Dependència	Administració de personal
Identificador	3
Descripció	Fernandez Rodriguez, J
Data – inici	1987/03/5
Data – final	1989/95/6
Nº transferència	89/1
Any d'eliminació	
Suport	Paper
Dipòsit	Administratiu
Matèries	Arquitectes Pla de Palau Moll de la Fusta Ponts

[Entitat que es compona de registres, que tenen atributs que els defineixen (camps)]

· Fitxer mestre / lineal: Format pel conjunt de registres introduïts a la base de dades. La disposició dels registres és seqüencial (seguint l'ordre d'altres)

[Correspondència similar al catàleg manual de fitxes]

· Fitxer invers: Relació alfabètica de totes les paraules dels camps indexats. El fitxer invers és transparent a l'usuari i es pot consultar. Funció: recuperar la informació a partir del llenguatge d'interrogació.

TERME	Núm. Reg.	Camp	Seqüència
-------	-----------	------	-----------

Ajuntament	23	650	3
	54	23	5
Documentació	23	101	1

La innovació és el fitxer invers, que sempre té: una llista alfabètica de termes, un núm. De registre en què apareix i una seqüència (si és la primera paraula, segona paraula, etc.). Utilitzem terme i no paraula, ja que el terme és la funció de les tècniques d'indexació, es determina què és el terme i què no:

- Terme sencer (ex. Pla, Pere)
- Paraula per paraula: cadena de caràcters entre espai i espai o signes de puntuació definits per cada programa.

Pots marcar manualment què va a parar al fitxer invers (dels resums, títols, etc.) Tot depen del programa.

- Fitxer de paraules buides: Llista de termes sense significació semàntica (articles, conjuncions, pronomes, preposicions...)

Funció: actuar de filtre per impedir que aquestes paraules formin part del fitxer invers:

- Evita un creixement desmesurat del fitxer invers
- Augmenta la rapidesa de la cerca d'informació

Característiques

- Camps de longitud il·limitada
- Permet múltiples entrades per camp
- Incorpora un llenguatge d'interrogació basat en l'àlgebra de Boole per recuperar la informació
- Gestiona un conjunt d'informació poc estructurada i de gran quantitat de text

◊ **Comparativa entre el model relacional i el model textual**

Els models no són purs i sempre tenen prestacions d'uns i altres, però sempre tenen un precedent, un model base.

Ex. ORACLE model relacional amb prestacions de model textual

TEMA 6: Elements per al disseny d'una base de dades documental

Definició d'entitats:

- ◊ Habitualment és una ja que els documents són un conjunt homogeni d'objectes.
- ◊ Es crea una única estructura de dades per a representar tots els documents i aquesta estructura es subdivideix en diferents models de registre segons els tipus de documents

Definició d'atributs i camps:

- ◊ Els formats estàndards de descripció automatitzada (MARC) defineixen l'estructura de la base de dades.
- ◊ Cal tenir en compte tres àmbits:
 - Camps de control: data d'introducció, analista
 - Camps de descripció bibliogràfica: (seguint normativa ISBD, AACR)
 - Camps de descripció del contingut

El **diccionari de dades** és una aplicació informàtica destinada a mantenir descripcions de totes les dades utilitzades en una base de dades. Conté la següent informació:

- ◊ Estructura de dades
- ◊ Controls d'integritat
- ◊ Drets d'accés i autoritzacions d'usuaris
- ◊ Formats de presentació de dades
- ◊ Informació sobre l'emmagatzematge de les dades

Per extensió, el diccionari de dades s'ha definit com un document escrit que conté la descripció dels camps. És, per tant, una eina útil de disseny.

Paràmetres:

- Etiqueta: etiqueta numèrica que identifica un camp
- Nom: Nom únic que identifica un camp
- Tipus de dades: tipus de dades que admet un camp com a valor:
 - Alfanumèric, Alfabètic, Numèric, Data, Lògic, Gràfic, Altres
- Longitud d'un camp (en el model textual normalment no s'ha de definir, mentre que el model relacional és molt estricte amb aquesta característica)
- Repetibilitat: Camp singular o múltiple
- Subcamps: Possibilitat de dividir el camp en parts anomenades subcamps
- Control d'integritat:
 - Valor únic: el valor del camp no pot ser repetit en cap altre registre
 - Obligatori: el camp ha de contenir obligatòriament un registre
- Domini: definició del contingut d'un camp
- Tractament documental: indica si el contingut d'un camp està subjecte a una normativa de descripció o a una llista de valors determinada
- Indexació: defineix si el camp s'indexa o no. Tècniques:
 - Paraula a paraula
 - Camp sencer
- Missatges d'ajuda

Els paràmetres d'un diccionari de dades depenen en part de les prestacions ofertes per un determinat Sistema de Gestió Documental: per exemple, no tots els sistemes de gestió documental permeten definir subcamps.

Els paràmetres mínims que ha de tenir un diccionari de dades són els següents:

- ◊ Nom de camp
- ◊ Repetibilitat
- ◊ Domini (perquè és una decisió conceptual)
- ◊ Indexació

TEMA 7 : SISTEMES DE GESTIÓ DOCUMENTAL

Estructura i prestacions:

- ◊ Definició de bases de dades
(agrupa les funcions que permeten aplicar les especificacions del diccionari de dades)
 - Definició de l'estructura de camps i les seves propietats
(alguns programes incorporen sistemes estàndards de descripció)
 - Definició dels models de registre
 - Definició dels formats de visualització
 - Definició dels informes de sortida
 - Possibilitat de relacionar bases de dades

- Possibilitat de modificar les bases de dades
- Permeten operacions de càlcul entre camps encara que molt limitades

◊ Manteniment

- Introducció de dades
- Modificació de dades
- Supressió de dades
- Modificacions col·lectives de dades
- Associar fitxers externs al registre
- Detecció de duplicats
- Validació d'entrades a través del control de fitxers d'autoritats
- Validació d'entrades a través del control de tesaurus
- Corrector ortogràfic
- Facilitats d'edició: entrada automàtica de dades

◊ Indexació i recuperació

Indexació

- Definició d'índexs
 - Camps indexats
 - Tècniques d'indexació: paraula a paraula, camp sencer, ocurrències, per segments
 - Índex bàsic
 - Agrupar diversos camps en un índex
- Construcció del fitxer invers: diferit o immediat
- Construcció del fitxer de paraules buides
- Creació de go words: llista de paraules que seran les úniques vàlides per a la indexació

Recuperació:

- Interfície de consulta: elaboració del quadre de diàleg que permet a l'usuari consultar la base de dades
 - Consulta simple
 - Consulta avançada
- Sistemes de recuperació de la informació
 - Recuperació booleana
 - Recuperació per patrons
 - Recuperació semàntica
- Historial de cerques
- Ordenació dels resultats: per ordre d'entrada, per un camp, per càlcul de rellevància

◊ Sortida i intercanvi

- Disseny de formats de sortida
 - Seleccionar els registres en relació a una condició establerta
 - Dissenyar el format de visualització: escollir els camps, escollir la tipografia i l'ordre de presentació
 - Dissenyar el format de pàgina per les sortides impreses
 - Definir l'ordre de sortida dels registres
- Disseny de formats d'importació: instruccions que permeten interpretar la informació d'un fitxer per incorporar-lo a la base de dades
- Possibilitats de distribució:
 - Impressió
 - Consulta des del web
 - Distribució en CD-ROM

- ◊ Gestió electrònica de documents
 - Vinculació del document digital a la descripció textual
 - Mètodes de vinculació:
 - Incrustació d'imatges (definició d'un camp que admet com a valor una imatge)
 - Vinculació del registre amb imatges digitals
 - Vinculació del registre amb qualsevol document digital
- ◊ Administració de bases de dades
 - configuració general: ubicació de fitxers
 - Sistema de seguretat: definició d'usuaris i drets d'accés
 - Còpies de seguretat
 - Optimització de l'espai
 - Menús del programa: possibilitat de definir menús de pantalla per facilitar l'accés a certes funcions del programa

AVALUACIÓ DE SISTEMES DE GESTIÓ DOCUMENTAL

- ◊ Passos per l'adquisició d'un programa informàtic:
 - Obtenir una llista de programes comercialitzats
 - Estudiar les característiques dels programes
 - Testar el programa

Diversos programes:

- ◆ SGBD clàssic Inmagic (*information retrieval systems*): incorporen gestió electrònica de documents (GED)
- ◆ sistemes d'indexació (*motors d'indexació*). Ex. Excalibur. Generació avançada de SGBD. Prestacions per indexació de documents digitals.
- ◆ *Electronic document management* = Sistema de gestió electrònica de documents, portals corporatius, programes de workflow. SG documents, però BD controlen el procés de producció digital.

Aplicacions documentals diferents, també aplicacions informàtiques per gestionar centres de documentació concrets:

- SG bibliotecària (VTLS)
- SG arxius
- SG museus

.

Avaluació de productes a partir d'una taula de prestacions Indicadors

- ◊ Indicadors de qualitat
 - Cost i condicions de pagament
 - Capacitat del sistema
 - Escalabilitat: capacitat del producte a adaptar-se a les novetats tecnològiques.
 - Flexibilitat: facilitat del programa per a adaptar-se a diferents programes. Ex. Possibilitat de modificar l'estructura de camps.
 - Parametritzable: dissenyable
 - Amigabilitat: facilitat d'ús

Ex. Pantalla d'ajuda

- Migrabilitat: enregistrament de dades de format estàndard per canviar

registres (ISO 2709). Possibilitat d'entrar al Catàleg Col·lectiu.

· Rendiment: testar el programa amb diferents registres.

Només es pot saber a partir d'altres instal·lacions del programa, perquè només es coneix amb altes quantitats de volum.

◇ Anàlisi de les característiques generals del programa

· Equipament informàtic

- Tipus de SGBD: relacional, documental, motor d'indexació
- Productes que incorpora: passarel·la web, tesaurus, edició en CD-ROM
- Configuració requerida de hardware
- Entorn operatiu
- Limitacions del programa: quantitat màxima de bases de dades, de registres

· Comercialització i suport tècnic

- Solvència de l'empresa distribuïdora
- Territorialitat
- Amplitud de mercat
- Versió de demo disponible
- Suport tècnic en la instal·lació i en la postvenda
- Formació inicial
- Existència de manuals comprensibles
- Cost del programa
- Garantia
- Cost del manteniment anual

· Interfície d'usuari

- Facilitat d'utilització
- Ajuda contextual
- Idioma

TEMA 8 : Distribució de bases de dades

S'ha de diferenciar entre el disseny, la producció i la distribució. La distribució és completament diferent a la seva producció.

Hi ha 4 possibilitats:

- ◆ Consulta local: Es refereix a la possibilitat de distribuir la base de dades en el centre on s'ha creat a través del mateix SGBD de producció de la base de dades

Avantatges:

- Simplificació del procés ja que no cal fer interfícies de consulta diferents
- Es pot oferir ajuda personalitzada als usuaris

Inconvenients:

◇ Aspectes de seguretat per evitar que els usuaris manipulin les dades.

◇ El cost de disposar de diversos ordinadors i mantenir diverses llicències en xarxa del SGBD

◇ El desplaçament de l'usuari

- ◆ Edició impresa

◇ Aplicacions:

- Publicació de novetats i noves adquisicions

- Edició de bibliografies de temàtica molt especialitzada
- ◊ Estructura: habitualment conté dos apartats
 - Llistat de referències bibliogràfiques
 - Índexs que remeten a les referències bibliogràfiques
- ◊ Es construeixen automàticament amb el SGD (les prestacions que inclouen els SGD poden ser en aquest aspecte molt diferents i cal estudiar-les amb deteniment: formats d'impressió, llistats jeràrquics, encapçalaments, construcció d'índexs)
- ◊ Avantatges: l'usuari no necessita equipaments ni programes informàtics per consultar el contingut
- ◊ Està en desús pels inconvenients que suposa: dificultat d'actualitzar el contingut i costos de producció i edició.
- ◆ Edició en suport òptic: Implica incorporar en el disc la base de dades i el programa de recuperació de la informació (la major part de SGBD tenen mòduls de consulta que es poden utilitzar per a la distribució de bases de dades en CD-ROM)

Avantatges:

- ◊ Facilitat d'edició perquè no cal elaborar una interfície de consulta diferent
- ◊ Facilitat a l'usuari perquè aprofita totes les prestacions d'indexació i del llenguatge d'interrogació de la base de dades
- ◊ Cost: les facilitats d'autoedició han abaratit molt els costos. Per l'usuari final moltes vegades és gratuït.

Inconvenients:

- ◊ Actualització de les dades
- ◊ Costos derivats de la distribució per correu, les llicències del software, l'edició del disc
- ◆ Publicació en el web: És el sistema de distribució actualment preferit i amb més futur

Avantatges:

- ◊ Només cal disposar d'un navegador per accedir a la base de dades. La consulta es fa un formulari html que és més intuïtiu i fàcil de manejar
- ◊ Les bases de dades poden estar permanentment actualitzades
- ◊ Es pot accedir des de qualsevol part del món.

Inconvenient:

- ◊ L'únic inconvenient des del punt de vista del productor és el cost de les CGI dels SGD que són encara ara molt cares.

Elements que intervenen en el procés:

- ◊ Navegador (p.e. Netscape o Internet Explorer): la funció del navegador és proveir d'un entorn per realitzar les consultes i obtenir els resultats.
- ◊ Servidor http (p.e. Internet Information Server, Apache): la funció del servidor és emmagatzemar la base de dades i proveir la informació demanada pel client.
- ◊ Programa CGI (*Common Gateway Interface*) (p.e. Knosys Internet, Wxis, WebPublisher): actua de passarel·la entre la base de dades i la interfície de consulta.
- ◊ Interfície de consulta: defineix el sistema de consulta i visualització dels resultats.
- ◊ Base de dades: es troba en el format de SGD en què s'ha produït.

El Protocol CGI

- ◊ És el que permet consultar bases de dades des d'un formulari html
- ◊ És un protocol que actua com un sistema de comunicació o passarel·la entre un programa extern i el web.
- ◊ Permet obtenir documents html dinàmics (que es generin en un moment determinat ja que depenen de la consulta que fa l'usuari)
- ◊ És executat i interpretat per un servidor Web per contestar peticions d'usuari
- ◊ El protocol CGI és capaç d'interpretar dades des d'un formulari html, d'executar-les i de transferir noves dades a l'usuari en format html.
- ◊ S'utilitza per: consulta de bases de dades, comptadors, motors de cerca, formularis, generadors d'e-mail automàtic...

Passos que es produeixen en un procés de consulta d'una base de dades al web:

- El navegador obre una connexió amb el servidor httpd. Mostra un formulari de consulta de la base de dades en html (interfície de consulta).
- L'usuari introdueix les condicions de cerca i n'ordena l'execució.
- El servidor executa l'aplicació CGI, inicia les variables de l'entorn (els paràmetres introduïts per l'usuari al formulari i altres paràmetres ocults). Aquestes variables s'emmagatzemen en un registre virtual.

Cada variable conté una etiqueta i un valor: l'etiqueta conté el nom del camp d'entrada i el valor conté la dada introduïda per l'usuari:

Per ex. Name=autor; value=fuentes quintana

- El programa CGI examina les variables d'entorn i efectua la consulta a la base de dades
- L'aplicació genera els resultats.
- L'aplicació envia els resultats al servidor.
- El servidor envia la informació al navegador en format html.

Tema 9: interfícies de consulta

El disseny de la interfície de consulta és una de les parts més importants del disseny conceptual ja que serveix d'intermediari entre l'usuari i els registres de la base de dades.

La interfície d'usuari està formada pels següents elements:

- ◊ Pàgina de consulta
- ◊ Pàgina dels resultats
- ◊ Visualització dels documents
- ◊ Consulta als índexs
- ◊ Informació general i ajudes
- ◊ Pàgina de consulta: formulari en html que indica a l'usuari les opcions de cerca

Elements:

- ◊ Títol
- ◊ Requadres de cerca:
 - Els camps sobre els quals es pot fer una consulta
 - La possibilitat d'utilitzar operadors booleans o altres operadors
 - Sistema de consulta: Requadres de text, Selectors de camp, Llistes desplegable
 - Possibilitat de consultar els índexs
- ◊ Botons d'execució i d'esborrar
- ◊ Missatges d'ajuda

◇ Navegació

Estructura dels requadres de cerca:

- ◇ Consulta als camps bàsics Apareixen en primer lloc els requadres corresponents a la consulta dels camps bàsics: autor, títol, matèries
- ◇ Limitadors: camps secundaris o complementaris Apareixen en segon lloc aquells camps que permeten delimitar la cerca: data de publicació, format, etc.
- ◇ Especificacions de visualització Apareix en últim lloc opcions relacionades amb el format de visualització: el número de registres per pàgina, l'ordre dels registres, el format de visualització...

Les interfícies solen disposar de dues pàgines de consulta

◇ Consulta simple

- Pensat per la majoria d'usuaris que cerquen informació sobre un sol camp sense necessitat de combinar elements
- Característica principal: simplicitat
- Disposa de:
 - Un requadre de text que cerca la informació en qualsevol camp
 - Pot també disposar de dos requadres més de text: cerca per autor, i cerca per matèria.
 - L'operador que s'utilitza per defecte hauria de ser la intersecció (AND)

◇ Consulta avançada

- Pensada per usuaris que fan consultes combinant termes amb operadors booleans o altres operadors
- Inclou diverses línies amb requadres de text perquè es puguin combinar termes
- Inclou la possibilitat d'escollir l'operador booleà

◇ Consulta als índexs:

- ◇ La pàgina de consulta ha de tenir la possibilitat de visualitzar els índexs de cada camp.
- ◇ La consulta als índexs és necessària pels camps que tenen una indexació sencera, ja que per l'usuari és molt difícil saber exactament quina és l'expressió indexada.
- ◇ Funcionament per activar les pàgines d'índex:
 - Desplegar els índexs per cada camp prement el botó corresponent. Si no s'indica el contrari se situa al principi de la llista.
 - S'activa l'índex de forma automàtica després d'executar una consulta: l'índex es desplega a partir de la paraula que s'ha escrit.
- ◇ Estructura de la pàgina d'índex:
 - Requadre de selecció (s'hauria de permetre la selecció de més d'un terme)
 - Terme
 - Número de documents associats
 - Opció d'executar la consulta o de tornar a la pàgina de consulta
- ◇ Pàgina dels resultats: llista dels resultats d'una pregunta formulada per l'usuari. Inclou només la informació bàsica dels documents.

Elements:

- ◇ Informació sobre la cerca realitzada: equació de cerca, nombre total de resultats obtinguts, nombre de resultats que es visualitzen
- ◇ Llistat amb la descripció bàsica dels documents: en forma de taula o en forma de paràgraf. Inclou una casella per seleccionar els documents.
- ◇ Criteris d'ordenació: el criteri d'ordenació principal pot ser seleccionat per l'usuari,

mentre que els criteris complementaris s'aplicaran de forma automàtica (ponderació, data, autor)

◇ Sortida de la informació

- Impressió col·lectiva dels documents
- Exportació col·lectiva dels documents
 - Format ASCII
 - Format estructurat (compatible amb un gestor bibliogràfic)

◇ Visualització del document: format de visualització del document complet.

Elements:

◇ Visualització del document amb un format per defecte

◇ Possibilitat d'escollir entre diferents formats

◇ Enllaç amb el text complet de l'article si està disponible

◇ Sortides del document:

- Impressió del document
- Exportació del document a un fitxer
 - ASCII
 - RTF
 - Format estructurat (que sigui compatible amb un Gestor Bibliogràfic)

TEMA 10 : QUALITAT DE LA BASE DE DADES

La literatura estableix indicadors (taules) per avaluar si les bases de dades són de qualitat.

Scoug és el primer decàleg a partir del qual s'ha anat treballant. És una llista de referència assumit per les biblioteques americanes.

Equipe ha fet una cosa similar per promoure l'avaluació de les bases de dades sobre qualificadors.

Els indicadors *Scoug* per avaluar la qualitat de les bases de dades són:

- ◆ consistència en l'indexació
- ◆ cobertura temàtica (exhaustiu en allò que es proposa)
- ◆ abast temporal (anàlisi de l'actualització)
- ◆ exactitud (que no hi hagi errors)
- ◆ accessibilitat (facilitat d'ús i accés)
- ◆ integració (amb altres bases de dades del mateix distribuïdor)

En l'avaluació s'analitzen les dades, el programa informàtic i el nivell administratiu. Es tracta d'analitzar quins són els indicadors que hauríem de mirar, en funció del que sigui correcte i del que pot produir un error.

◇ Nivell base de dades:

◇ Fonts:

- grau de cobertura exhaustivitat de la base de dades.
- Especialització temàtica (interessa que sigui molt alt)
- accés als documents originals han de ser accessibles via xarxa, o sinó facilitar on està la referència o el lloc on està.

◇ Errors de precisió. Són molt importants, però són fàcils d'arreglar. Provoquen molts problemes de recuperació.

Ex. Ortogràfics, sintaxi, semàntica, mecanografia.

- ◇ Errors d'actualització i presentació de les dades:
 - Presència de registres duplicats. Hi ha programes que detecten els duplicats abans d'entrar-los, però n'hi ha que no.
 - Grau d'actualització en relació a tots els registres de la base de dades. Proporció de documents actuals en proporció de registres de la base de dades. Anàlisi documents de l'últim any, però també del transcurs dels anys, a fi que no hi hagi buits significatius. Pressuposa que el nivell màxim de la base de dades ha de ser igual als documents científics publicats en un any.
- ◇ Errors de consistència. Molt importants!
 - Consistència en la descripció bibliogràfica. Grau de compliment respecte a les normes de descripció – sistematització de la base de dades.
 - Consistència en els punts d'accés. Opció d'informació dels punts de forma sistemàtica. Llistes de validació, regles pels camps, etc.
 - Consistència en l'anàlisi del contingut. Que hi hagi igual proporció de descriptors amb els documents. Dificil d'aconseguir. Una proporció de 70–80% de descriptors és considera correcte.
- ◇ Nivell del sistema de recuperació:
- ◇ Estructura. Es pretén que cada camp processi informació homogènia. Un sol concepte. Unir informació de 2 camps es pot fer, però separar informació d'una camp és molt complicat.
- ◇ Punts d'accés. Existència d'un índex general i diferents índexs de camps (concrets). Possibilitat d'accedir a diferents índexs de camps.
- ◇ Formats de visualització. Possibilitat existència de més d'un format per adequar-se a més d'un usuari.
- ◇ Consistència de la base de dades al llarg del temps. Cal renovar-la, ja que té una vida.

Dependrà del disseny de la base de dades

Altres indicadors que fan referència al programa:

- ◇ Potencialitat del llenguatge d'interrogació.
- ◇ Ordenació dels resultats per diferents criteris
- ◇ Amigabilitat. (intuïtat + grau de fiabilitat)
- ◇ Sistema d'ajuda del programa informàtic.
 - Nivell administratiu:
- ◇ Documentació (correcte, llegible i en diferents idiomes)
- ◇ Atenció a l'usuari
- ◇ Preus i sistema de facturació

Indicadors a valorar quan s'ha d'adquirir una base de dades.

Sistemes de recuperació d'informació = literatura anglosaxona. Model textual = literatura espanyola.

SGD : model adequat a la recuperació d'informació.

(Miguel, 1997: 14)

Informàtics

No informàtics

Àrea de dades

Àrea informàtica

Usuaris

Sistema operatiu

Gestión de datos – SGBD

Control de las comunicaciones

Tratamientos específicos

Equip perifèric

Unitat central de procés

Dades no estructurades

Dades estructurades

Factual

Referencial

Administrador

Suport lògic – *software*

Equip físic – *hardware*

Contingut – *dades*

SBDD " SI

SBDD

"

SI

- Dades estructurades en forma de BDD.
- Continent
- Software (que és un sistema de gestió de bases de dades)
- Administrador

BDD

=

Conjunt de dades (no parlem de tecnologia

Model intern (físic)

Nivell intern

Nivell extern

Model de dades conceptual (o model de dades global)

Nivell conceptual

Model extern

(model de dades per a cada usuari)

Models abstractes

Atributs

Atribut identificador principal

DOCUMENT

DOCUMENT

Escriu

AUTOR

◊ Molts a molts: n:m

◊ Un a molts: 1:n

◊ Un a un: 1:1

◊ Molts a molts: n:m

◊ Un a molts: 1:n

◊ Un a un: 1:1

N:M

N:M

N:M

escriu

publica

PERSONA

LLIBRE

N:M

consta

TEMA

1,1

Data préstec

Data retorn

0,1

0,1

És_un

ARTICLE

LLIBRE

DOCUMENT

DNI

Codi Llibre

N:M

Presta

USUARI

LLIBRE

Identificació d'entitats i atributs:

Interrelacions:

1:N

1:N

N:M

1:1

1:N

1:N

Es supervisat per

recull

genera

té

té

fa

EMPLEAT

EMPLEAT

FACTURA

PROVEÏDOR

ARTICLE

COMANDA

CLIENT

Diagrama Entitat/Relació:

EDITORIAL: Nom, adreça,

LLIBRE: Codi, títol, editorial

LLIBRE: Codi, títol, idioma,

AUTOR: Nom, Institució, Nacionalitat

ESCRIU: Codi, Nom

Model relacional

Model Entitat relació

Taula Client

Id. Client

Nom

Adreça

Ciutat

Entitat Client

Id. Client

Nom

Adreça

Ciutat

COMANDA: Núm. comanda, data, total,

FACTURA: Núm. factura, data, núm. comanda

N:M

escriu

publica

PERSONA

LLIBRE

– Serveis d'informació i documentació

– Ofimàtica

Àmbit

– Model textual:

. Longitud variable

. Camps repetibles

– Fitxer pla: no permeten la relació de

taules

– Fitxer invers

– Possibilitats de recuperació a través

de la lògica booleana. (Rec. *probabilista*)

– Taula

. Camps de longitud fixa

. Una ocurrència

– Relacions entre taules: capacitat

d'herència

– No té fitxer invers

– Eines de recuperació limitades

(recuperació *determinista*)

– Informació poc estructurada

– Dades permanents

– Informació molt estructurada

– Les dades canvien amb freqüència

Objecte

Estructura

Model textual

Model relacional

Pàgines html

Consulta

Servidor httpd

Programa CGI

Ordinador client:

Navegador

Interfície de consulta

Base de dades