

TEMA 5: SUPORTS I MÈTODES D'ACCÉS A FITXERS

(Informàtica II: assignatura troncal de la diplomatura en Biblioteconomia i Documentació)

0. INTRODUCCIÓ

Tenint en compte que l'única memòria directament accessible per a la unitat de control és la **memòria principal**, tot allò que l'ordinador gestione haurà de residir en aquesta memòria. Tanmateix, aquesta memòria és cara i la seua disponibilitat és limitada. Si, a més a més, afegim que és volàtil (és a dir, que s'esborra quan es talla el subministrament d'energia), podem concloure que no és adequada per a mantenir volums grans d'informació de manera permanent.

Tot això justifica la connexió a l'ordinador de memòries **perifèriques**: no són volàtils, són més barates i es troben a l'abast en més quantitat.

La informació, a les memòries perifèriques, s'organitza en arxius. En aquest tema estudiarem els tipus de suports perifèrics i els mètodes d'accés a les dades emmagatzemades en forma d'arxius. En la classe següent, veurem algunes maneres d'organitzar la informació, cosa que facilita la gestió de les dades emmagatzemades i acurta el temps de procés.

En realitat, tots estem familiaritzats amb el treball amb arxius en la vida quotidiana. Per això hem trobat convenient iniciar aquest tema establint una comparació entre la gestió manual i la gestió d'arxius.

1. GESTIÓ NO AUTOMATITZADA DE LA INFORMACIÓ

1.1. SUPORTS DE LA INFORMACIÓ EN UNA GESTIÓ MANUAL

Sovint, les dades que gestionem no es troben aïllades, sinó que formen un conjunt relacionat lògicament. Posem per cas que som en una biblioteca i fem la recerca d'un llibre concret. D'entrada, anirem a un arxivador on tal vegada hi haurà una fitxa per a cadascun dels llibres; en la fitxa trobarem dades al voltant de les característiques del llibre. En termes informàtics, aquesta fitxa s'anomena **registre**; al conjunt de fitxes

del mateix tipus, **fitxer** o **arxiu**, i a cada un dels conceptes inclosos en un registre, **camp**. En el nostre cas de la biblioteca, potser tinguem els camps següents: *títol, autor, any, editorial i prestatge*. Quan dissenyem la fitxa, haurem d'indicar, per a cada un dels camps, el nom i la llargària.

Una vegada decidida l'estructura de la informació que volem gestionar, cal un suport físic on emmagatzemar-la i guardar-la. En l'exemple que acabem de veure, les dades de cada llibre s'escriuran sobre una fitxa de cartolina. Les fitxes de tots els llibres es guardaran en un arxivador.

1.2. QUINES GESTIONS HEM DE FER AMB UN FITXER MANUAL?

Com acabem de veure, el primer que hem de fer és **crear-lo**, és a dir, preparar el format de la fitxa amb les seues dades per enviar-la a la impremta. Haurem d'obrir l'arxiu per efectuar-ne el **manteniment**. Aquesta gestió es fa a nivell de registre, és a dir, donarem una **alta** (emplenar la fitxa), una **baixa** (suprimir-ne una), rectificarem una fitxa, és a dir, **modificarem** el valor d'algun o alguns dels camps, o podem **consultar** un registre concret. Així mateix, podem generar un **informe**, és a dir, obtenir una llista de tot o part del contingut del fitxer. Una vegada acabat el treball amb l'arxiu, hem de **tancar-lo**.

Una altra gestió que es pot fer en un fitxer és **reorganitzar-lo**, cosa que consisteix en optimitzar la col·locació física dels registres dins del fitxer, com per exemple, per eliminar els esborraments produïts per baixes o modificacions, o per mantenir-lo físicament emmagatzemat en un ordre determinat. Aquest segon tipus de reorganització s'anomena **classificació**.

1.3. RELACIÓ ENTRE EL SUPORT FÍSIC I EL MÈTODE D'ACCÉS

La manera d'accedir a les dades es troba condicionada pel suport que les conté. En la cultura egípcia s'escrivia en tires llargues de paper enrotllades sobre uns eixos. Aquest suport tenia, però, molts entrebancs:

- Ha de desenrotllar-se amb molta cura per accedir a la informació.
- Si és molt llarg, s'haurà d'anar enrotllant el principi en un segon eix, la qual cosa entrebanca l'accés a la informació que abans hem passat.

- Si cal fer alguna modificació en les dades escrites, cal copiar tot el paper sencer.

No obstant això, l'avantatge és la seua simplicitat: tan sols és necessari un full de paper continu i un parell de llistons. Si l'ús que es fa del paper és, per exemple, la seua lectura sistemàtica des del començament fins al final, aquesta senzilla tècnica compleix perfectament la seua missió.

D'altra banda, si quan accedim a la informació emmagatzemada no desitgem llegir-la tota, sinó que només estem interessats en una part concreta, l'organització d'un llibre sembla més adequada. Facilitar aquesta forma d'accés va implicar el desenvolupament d'una nova tecnologia: la confecció de llibres.

La informació en un suport com un llibre pot ser llegida des del començament fins al final (de manera seqüencial), tal com habitualment fem amb una novel·la, però també ens permet accedir directament a una pàgina concreta sense haver d'obrir ni tan sols les anteriors.

El suport paper només ens permet un accés seqüencial, mentre que el suport llibre ens permet un accés seqüencial i també un accés directe.

1.4. ORGANITZACIÓ I MÈTODE D'ACCÉS

El nou suport llibre ens permet l'accés directe a una pàgina, encara que això també té una mancança d'utilitat. Normalment no es desitja accedir a «allò que hi ha» en una pàgina determinada, sinó accedir a un contingut (siga on siga emmagatzemat en el llibre). Per permetre un **accés directe** a la informació, una tècnica utilitzada és la creació d'un **índex**, el qual conté dos tipus de dades; el primer descriu els continguts i el segon és un **punter** cap la pàgina del llibre on comença aquesta temàtica. És més ràpid llegir seqüencialment l'índex, fins a trobar la posició de la matèria en la qual estem interessats, i accedir-hi directament, que fer la recerca a través del llibre.

Una altra organització de les dades que s'utilitza freqüentment per facilitar la recerca en un llibre consisteix a tenir-les en un ordre determinat. Si pensem en un diccionari, quin procés seguiríem per fer la recerca del significat d'una paraula, si aquestes hi aparegueren sense cap ordre? Hauríem d'anar llegint paraula a paraula (seqüencialment) des de la primera fins a trobar la nostra paraula.

Tanmateix, si el tenim ordenat, podem accedir-hi més o menys directament; caldrà només obrir el llibre per qualsevol pàgina, llegir-ne una i moure'ns cap enrere i cap endavant, i repetir el procés fins a arribar a la paraula desitjada.

Aquest procés de consulta resulta molt eficient i no requereix disposar de cap informació afegida, com en el cas de l'índex. La limitació més important que té és que les dades han d'estar emmagatzemades físicament per ordre; per donar d'alta una nova paraula, cal fer un buit al lloc que li correspon.

Tornem a l'exemple de la biblioteca que hem esmentat abans. Estem interessats a fer la recerca d'un llibre del qual coneixem el títol, l'autor o la matèria. Podríem emmagatzemar les fitxes (registres) ordenant-les per qualsevol d'aquests conceptes, però només per un. Una manera freqüent de resoldre aquest problema en les biblioteques informatitzades és mantenir diversos arxivadors en què es troben les mateixes fitxes, però ordenades d'acord amb criteris diferents. Triplicar la informació és una solució adequada per a la recerca realitzada per persones. Per a la recerca amb l'ordinador en proposarem alguna variació, però la solució és similar conceptualment.

Per disposar d'accés directe a la informació, hem d'emmagatzemar-la sobre suport d'accés directe, però, a més, hem d'organitzar-la de manera que sapiguem localitzar-la posteriorment. Hem vist en aquests exemples que totes les organitzacions tenen uns entrebancs o altres (l'índex consumeix espai addicional; en el diccionari cal obrir un buit per incloure-hi una paraula nova). Hem de seleccionar aquella organització que s'ajuste tan bé com es pugui a les necessitats de cada tractament de les dades.

Hem volgut fer aquestes reflexions al voltant de les tasques que nosaltres habitualment fem en la recuperació d'informació fora de l'entorn de l'ordinador perquè ens són tan familiars que, fins i tot, no en som conscients i la veritat és que són molt semblants a les consideracions que farem en un entorn informatitzat.

2. GESTIÓ AUTOMATITZADA DE LA INFORMACIÓ

2.1. SUPORTS DE LA INFORMACIÓ EN INFORMÀTICA

Així com en el tractament manual empràvem, per emmagatzemar-hi la informació, un pergamí, unes fitxes de cartolina o les pàgines d'un llibre, per emmagatzemar la informació en l'entorn informàtic necessitem un suport que siga accessible a l'ordinador. La memòria principal és l'única accessible per a la unitat de control. Per manipular una dada concreta que resideix en un arxiu, es transfereix una petita porció d'informació, la qual conté la dada, des del perifèric fins a la memòria principal. Aquesta zona de la memòria principal s'anomena **memòria intermèdia (buffer)**. La porció del suport extern transferida s'anomena **bloc**.

El suport d'emmagatzematge és molt més lent que la unitat de control. Evidentment, uns suports són molt més lents que altres, però, per seguir els raonaments posteriors, ara suposem que en un ordinador coherent una operació de lectura (un accés a suport) tarda 1.000 vegades més que una instrucció de la unitat de control. Per evitar temps d'espera de la unitat de control, el suport disposa d'un controlador. La unitat de control dóna l'ordre de lectura al controlador i queda alliberada per atendre altres programes. El suport es troba fragmentat en blocs. El controlador deixa la informació del bloc complet al *buffer*. Per al controlador, un bloc és una seqüència de bits. Es despreocupa de l'estructura interna de la informació. Tanmateix, en el fitxer hi ha registres, tants com entitats del món real (en el fitxer de clients, per exemple, hi ha tants registres com clients té l'empresa). Així mateix, el registre està estructurat en **camps**. Un bloc conté un o diversos registres de fitxers.

Observem que el concepte de *registre* que acabem d'introduir correspon al de l'estructura de dades de registre. Aquesta estructura es troba composta per una sèrie de camps; cadascun dels camps té un únic nom en el fitxer (selector de camp) i s'implementarà amb l'estructura de dades adequada a la informació, la qual contindrà: cadena, enter, lògic, etc.

Quan el bloc sol·licitat ha arribat al *buffer*, la unitat de control localitza el registre que li interessa; dins del registre localitza el camp que desitja modificar, el modifica i torna a copiar el *buffer* complet al bloc on estava en el suport.

En la fig. 3.1 es representa la memòria principal d'un ordinador, ocupada, en primer lloc, pel sistema operatiu, els programes que estan en execució i dues àrees de *buffer*. El controlador envia un bloc al *buffer* de la memòria principal.

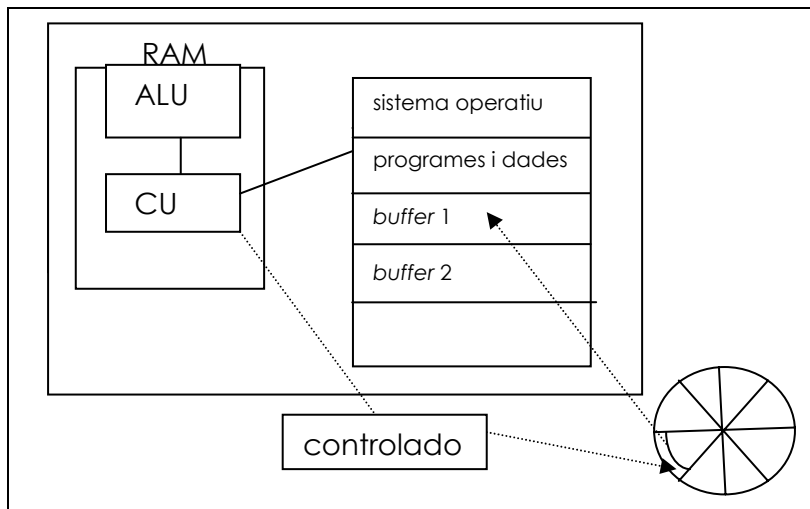


Fig. 3.1. Transferència d'un bloc des d'un suport d'emmagatzematge a la memòria principal.

La paraula *buffer* vol dir «amortidor»; la missió d'aquesta àrea de memòria és que la lentitud del suport físic tinga poca repercussió al voltant del funcionament del sistema. La CU (*control unit* / unitat de control) dóna l'ordre perquè es carregue el *buffer* i continue treballant en una altra aplicació. Quan les dades d'un registre ja es troben a la memòria principal, la CU les manipula. La grandària d'un *buffer* coincideix amb la grandària del bloc o múltiples del bloc. En un bloc solen haver-hi diversos registres. El més normal és que un registre es trobe inclòs íntegrament en un bloc perquè la CU el tinga sencer al *buffer* quan necessite treballar amb aquest registre.

2.2. GESTIONS QUE CAL FER EN UN FITXER INFORMÀTIC

D'entrada, el primer que hem de fer per poder utilitzar un fitxer automatitzat és **crear-lo**, és a dir, definir-ne els camps, indicant-ne el nom i el tipus de cadascun dels camps (l'equivalent a preparar el format de la fitxa per enviar a la impremta). Quan definim el tipus de dada, queda determinat el **domini** de valors, és a dir, hi especifiquem el conjunt de valors possibles per al camp, amb la qual cosa s'eviten errors a l'hora d'introduir les dades. Així, en l'exemple de la biblioteca amb el qual presentarem la gestió manual, el camp *any* probablement estarà

comprés entre l'any 1900 i l'any actual; per a una determinada fitxa, el camp *editorial* contindrà una de les editorials possibles, definides prèviament.

Després de crear el fitxer, podem **obrir-lo** per fer el **manteniment** de l'arxiu; aquesta gestió es fa a nivell de registre, és a dir, donem una **alta** (afegim un registre al fitxer), una **baixa** (suprimim un registre), **modifiquem** el valor d'algun o alguns dels camps d'un determinat registre i **consultem** la informació que conté un registre. Podem, així mateix, generar un **informe**, és a dir, obtenir una llista de tot o part del contingut del fitxer. Quan hàgem acabat de treballar amb un fitxer, haurem de **tancar-lo**.

Obrir un fitxer consisteix a avisar al sistema que treballarem amb ell, és a dir, se sol·licita que prepare la zona de memòria on transferir les dades. Quan tanquem el fitxer, es força, al *buffer*, l'escriptura de la informació que conté la memòria si ha estat modificada i, tot seguit, s'allibera la memòria. Per tot això, és important el tancament de l'arxiu quan acabem d'utilitzar-lo; així garantim que totes les modificacions han estat emmagatzemades al suport i, d'altra banda, no consumim memòria principal per més temps del que és necessari.

Així com en la gestió manual, podem trobar necessari **reorganitzar** un fitxer, és a dir, optimitzar la col·locació física dels registres; per exemple, per aprofitar els buits que han pogut produir-se per les successives baixes o modificacions, o també per mantenir-lo físicament emmagatzemat per un ordre determinat: **classificat**.

2.3. RELACIÓ ENTRE EL SUPORT I EL MÈTODE D'ACCÉS

Els suports més freqüents per a l'emmagatzematge de la informació a l'ordinador utilitzen tecnologia magnètica. Aquests suports són molt semblants, tant en la composició com en el funcionament, a les cintes magnètiques i als discos convencionals emprats en música. Així, per escoltar la tercera cançó en una cinta de casset, hem de «passar» les dues anteriors (com a màxim podem aspirar a tenir dues velocitats de rebobinatge). Al contrari, el braç d'un tocadiscos es pot posicionar en qualsevol part de la superfície i escoltar les cançons des d'aquesta posició només, o també podem escoltar un disc seqüencialment: des del començament fins al final.

En una cinta magnètica cal llegir els metres de cinta que hi ha davant del registre que desitgem recuperar i, per tant, només permet gestionar la informació de manera **seqüencial** (és a dir, llegint tots els registres en l'ordre que es troben emmagatzemats des del primer fins que es troba el registre desitjat). Una cinta és un suport seqüencial i tan sols admet aquest mètode d'accés. Per contra, un disc és un suport d'accés directe i permet, o bé llegir-lo de manera seqüencial, o bé de manera **directa**. El mètode d'accés a un fitxer informàtic en un suport directe pot ser seqüencial o directe, però quan es troba emmagatzemat en suport seqüencial, només pot ser seqüencial.

2.4. ORGANITZACIÓ DE FITXERS I MÈTODES D'ACCÉS

Així com ens passava amb la gestió manual, normalment no estem interessats a accedir a una posició del disc, sinó a un contingut concret. En la gestió manual, proposàvem dues solucions que consistien a organitzar la informació en un determinat ordre que ens permetia fer la recerca com en un diccionari, o bé construir un índex que ens permetia un accés directe al contingut. Per poder fer la recerca d'una dada concreta a l'ordinador, també cal organitzar la informació. En la classe següent presentarem algunes organitzacions de fitxers més usuals: la seqüencial ordenada, l'aleatòria i la indexada. Independentment de com es trobe organitzat el fitxer, es podrà llegir en seqüència o en accés directe (sempre que el suport ho permeti).

Organitzar les dades porta l'afegit d'una gestió addicional i, per tant, més temps de procés pel manteniment i, freqüentment, un consum addicional de memòria. Proposarem variants que minimitzen un o l'altre problema, però la millora en una gestió sempre repercuteix en l'altra. No existeix la «millor» organització.

La tria dependrà fonamentalment de les **expectatives d'ús**. Si es preveu que es faran moltes altes, o aquestes es produiran en moments crítics per al sistema (moments de molta càrrega de treball, o que afecten la imatge de la companyia o parts vitals del seu funcionament), s'haurà de triar una organització on aquesta gestió concreta, és a dir, les altes, consumisquen pocs recursos i, a més, siguin ràpides. Si el sistema ha de donar un servei ininterromput, totes les hores del dia, tots els dies de l'any, no serveix una organització que necessite reorganitzacions freqüents, ja que no es disposa de temps per a fer-les.

Així les coses, la tria de l'organització sembla una tasca molt complexa, perquè hi ha moltes variants possibles, però és molt sistemàtica. Hem vist que els processos que cal fer en un fitxer són sempre els mateixos: altes, baixes, modificacions, consultes d'un registre i consultes globals. Per a cada organització proposada, caldrà examinar l'eficiència de cada un d'aquests cinc processos i, a més, tres qüestions: la necessitat de reorganització, el consum addicional de memòria que realitza i la facilitat de desenvolupar el sistema. I això és tot el que hi ha. Si comparem l'eficiència de cadascun d'aquests aspectes amb la manera amb què esperem que s'utilitzaran les dades, obtindrem l'adequació o la inadequació d'una organització adreçada a resoldre el problema concret.

Si, per facilitar una gestió de manteniment, permetem una petita desorganització, el fitxer es pot degenerar i els processos funcionaran més lentament. Com a conseqüència, l'eficiència es deteriorarà i serà necessària la reorganització. Abans de decidir-se per una o altra organització, caldrà avaluar la freqüència de reorganització i el consum de recursos que suposa aquest procés.

Pel que fa al consum addicional de memòria secundària, ja hem comentat que és relativament barata i, si cada vegada s'abarateix més, aquest criteri resulta menys condicionant. Consumir entre un 30 i un 50 per cent més, o fins i tot el doble d'espai de l'estrictament necessari, és cosa freqüent i fins i tot pot estar justificat si això millora els temps de resposta. No obstant això, per molt barat que siga l'emmagatzematge, no podem plantejar-nos consumir 100 vegades més, ni tan sols 10 vegades més, del que necessitem.

Finalment, la facilitat de desenvolupar el programa que tracte les dades i la seua fiabilitat també condiciona l'elecció de l'organització. Així, si el sistema proporciona eines **depurades** (és a dir, que sabem que no produiran error en el tractament) i **optimitzades** (o siga, que han estat preparades per consumir els mínims recursos) per fer funcionar una determinada organització, serà més fàcil desenvolupar l'aplicació i pot resultar adequat emprar aquesta organització de dades, fins i tot encara que no siga l'òptima per al tractament del problema. Aquesta és la circumstància que esdevé amb l'organització indexada, la qual veurem a la classe següent. Aquesta, la indexada, és una organització fàcilment reutilitzable, accelera l'accés per tants camps com es desitja, i molts sistemes incorporen la gestió de la generació dels índex, de

manera que el seu ús està molt estès, encara que consumeix bastants recursos i entrebanca els processos de manteniment dels fitxers.

En la fig. 3.2 es recullen, a mena de resum, els vuit criteris d'eficiència que hem d'avaluar en cada organització de dades. En la fig. 3.3 figuren les característiques que hem d'analitzar en l'entorn que desitgem automatitzar, per obrar en conseqüència i seleccionar el mètode d'accés més adient. Recorrerem cadascun dels punts en el tauler, posant com a exemple el catàleg dels fons d'una biblioteca imaginària concreta. D'entrada, hem d'assabentar-nos del tipus de dades que es mantindran i quin volum inicial tenen. Per exemple, indicarem quines són les dades que cal incloure per a cada llibre: títol, autor, editorial, signatura, etc. També indicarem el tipus de dada corresponent a cada camp, per poder avaluar així el consum de bytes necessari per a cada llibre. Ara per ara, els fons contenen 10.000 llibres.

El pas següent que cal tenir en compte és l'evolució previsible del volum i el nombre d'altres, baixes, modificacions i consultes que rebrà el sistema. Continuant amb l'exemple del catàleg, posem per cas que esperem enriquir la col·lecció amb 2.000 obres per any, la qual cosa vol dir que es produiran 2.000 altres en el catàleg. El personal encarregat de la catalogació coneix amb escreix el seu treball i està previst que rebrà un curs de formació sobre el nou sistema. Les modificacions serien com a conseqüència d'errors de catalogació, o bé produïdes per canvis de criteri en la catalogació esmentada; per tant, és previsible que hi haja poques modificacions. Pel que fa a les baixes, d'entrada, no hauria d'haver-ne cap, però tenint en compte el perfil dels nostres usuaris i el volum de treball amb els llibres, l'experiència demostra que el 0,1% de les obres es perd o es deteriora al llarg de l'any. Aquest, per tant, és el volum de baixes. Es reben al voltant de 85 consultes diàries, en què se sol·licita una obra concreta en què l'usuari sap el títol i/o l'autor. Així mateix, es produeixen una mitjana de 24 consultes diàries que cerquen informació disponible sobre una o diverses matèries. Algunes (8-10) vegades a l'any, el Patronat i altres benefactors de la institució sol·liciten informació exhaustiva dels fons... En aquestes circumstàncies, l'organització dels fitxers que hem triat ha de proporcionar consultes molt eficaces per títol i autor i, a més a més, per matèries. No importa que la gestió de les baixes i modificacions siga ràpida o lenta. Cal parar alguna atenció a l'eficiència de les altres.

Normalment, les operacions que es fan amb més freqüència són les que han de ser més eficients en l'organització triada, però també cal tenir en

compte la prioritat relativa de les operacions. Per entendre millor aquest punt, prendrem l'exemple d'un altre entorn. Si els clients d'una empresa poden pagar per finestreta, cal generar el rebut a l'instant; no se'ls pot tenir esperant. Al contrari, per exemple, una ordre de cobrament per enviar al banc, no importa si es retarda 10 minuts més o menys.

Tornant a la biblioteca, la pregunta que pot demanar el Consell de Direcció quan es troba en una roda de premsa davant dels mitjans de comunicació, serà més prioritària que l'alta d'un llibre, encara que la pregunta fóra molt més infreqüent que aquest procés.

És important tenir en compte que, a banda de la freqüència d'un procés, també tenim la distribució en el temps. No és el mateix que la compra dels documents nous es produïska homogèniament al llarg de l'any, que, per contra, s'apleguen tots el dia 30 de desembre i hagen d'estar registrats en el sistema el dia 1 de gener. Encara que el volum de treball emprat per les noves altes no fóra importantíssim respecte del total efectuat pel sistema, pel fet de concentrar-se en un moment determinat, cal poder resoldre'l ben aviat. La resta del sistema funciona bé, però ací hi ha un embós, una obstrucció i, si no se soluciona, la resta no pot continuar endavant. A aquest tipus de situacions se sol anomenar *colls de botella*. Prou sovint, en la vida real, es concentren determinades activitats en el temps. Pensem en la gestió dels alumnes d'un centre educatiu. No és gens significatiu parlar de 20.000 matrícules cada any. Les matrícules de tot l'any es produeixen, per exemple, de l'1 al 5 de setembre; per tant, són 20.000 altes en 5 dies. El sistema que s'implante ha de poder assumir tot aquest treball en aquest breu període, encara que en altres èpoques de l'any estiga menys ocupat.

El darrer punt que proposa analitzar la fig. 3.3 és la **continuitat del servei**. Tant els processos de reorganització del fitxer com la realització de còpies de seguretat, consumeixen temps de procés i necessiten que no es produïsqen modificacions en el continguts dels fitxers. En el cas que hem pres com a exemple, les còpies i reorganitzacions podran fer-se després del tancament de la biblioteca, o els cap de setmana. Al contrari, un aparell que estiga subministrant oxigen a un malalt, en una unitat de cures intensives d'acord amb unes constants vitals, no pot desconnectar-se «un momentet» per reorganitzar les dades. L'organització del fitxer ha de garantir la **no aturada** del sistema.

- | |
|--|
| <ol style="list-style-type: none">1. Temps per a llegir un registre concret.2. Temps per a llegir el següent registre físic o tots els registres. |
|--|

3. Temps per a inserir un nou registre.
4. Temps per a modificar un registre existent.
5. Temps per a esborrar un registre.
6. Temps i freqüència de la reorganització.
7. Requisits d'espai.
8. Simplicitat en la programació. Reutilització.

Fig. 3.2. Criteris d'eficiència d'una organització de fitxers.

1. Tipus d'informació i el seu volum actual.
2. Augment o disminució del volum en el temps.
3. Freqüència (absoluta) de cadascuna de les operacions que cal fer amb les dades.
4. Prioritat (relativa) de cada operació per a l'organització.
5. Distribució en el temps de cada operació (colls de botella).
6. Continuitat del servei: reorganització i còpies de seguretat.

Fig. 3.3. Anàlisi dels patrons d'ús d'una aplicació.

3. TIPUS DE SUPORTS

A continuació, farem una revisió dels tipus de suports que s'utilitzen més freqüentment i els classificarem atenent a conceptes diferents.

3.1. TIPUS DE SUPORT SEGONS LA CONNEXIÓ A L'ORDINADOR

La memòria d'emmagatzematge es pot classificar en secundària i terciària segons estiga contínuament accessible a l'ordinador, o bé siga necessària la intervenció de l'operador per col·locar-la al dispositiu de lectura/escriptura abans que l'ordinador pugui llegir-la. Quan el suport es troba connectat a l'ordinador, es diu que el suport **està en línia**. Encara que el dispositiu de lectura siga lent, sempre serà més ràpid que quan s'ha de muntar manualment. D'altra banda, la memòria terciària està fora de línia i és imprevisiblement lenta, però és més barata. El cost sempre és un condicionant a l'hora de fer-ne una tria.

En la fig. 3.4 es representen els costos d'emmagatzematge en línia i fora de línia. Calculem el preu per unitat d'informació, diguem-ne el preu per byte. Un suport té una capacitat en bytes i un preu, però cada suport en línia requereix un dispositiu de lectura/escriptura, mentre que diferents suports (n) comparteixen el mateix dispositiu quan estan fora de línia. El

nombre de dispositius que es poden connectar a l'ordinador sol ser limitat, condiona el preu de l'ordinador i, per tant, s'ha de repercutir com a part del preu de la informació. Els dispositius de lectura/escriptura tenen una **durada limitada**; per tant, l'empresa ha de tenir en compte la seua amortització. A més a més, tenen un **manteniment**, que freqüentment està contractat amb una companyia i això suposa una quantitat fixa al mes. Els dispositius cada vegada són més petits, però una empresa ha de tenir en compte els costos de **lloguer de l'espai** físic que ocupen les seues instal·lacions. En el cas del dispositiu en línia, cal assignar aquests quatre costos íntegrament al preu de la informació emmagatzemada en el suport, mentre que en el cas de fora de línia, el preu es reparteix entre els n suports. En la memòria terciària cal considerar l'espai que ocupa el suport, però aquest factor és prou més inferior que el que comporta el preu del dispositiu. D'altra banda, en la memòria terciària cal tenir en compte el cost corresponent al temps consumit per l'operador, quan busca el suport al magatzem i, després, el temps de muntatge. Així mateix, hi podem incloure el cost corresponent al temps extra d'espera de l'usuari, o la insatisfacció generada en l'usuari a causa de la gestió manual.

<p>Cost en línia = preu suport + dispositiu (ports + amortització + manteniment + espai dispositiu)</p> <p>Cost fora de línia = preu suport + espai suport + operador + insatisfacció usuari + dispositiu (ports + amortització) + manteniment + espai dispositiu)/n</p>
--

Fig. 3.4. Costos de l'emmagatzematge de la informació.

Les tendències d'evolució dels costos són els següents:

- Disminució en el preu de suports i dispositius.
- Disminució de la grandària dels suports i dispositius, i augment de la densitat de gravació.
- Augments grans en els costos del personal.
- Augment (proporcionalment menor) dels preus de lloguer d'instal·lacions.
- Augment en el cost de manteniment.
- Augment de les expectatives i exigències dels usuaris.

Tot això comporta que cada vegada més s'organitze la informació en línia, encara que grans volums d'informació de dades que tenen poca freqüència d'ús continuen organitzant-se a l'emmagatzematge terciari. Per exemple, les còpies de seguretat de les dades, o les dades històriques que cal guardar per imperatiu legal (per exemple, les dades econòmiques cal guardar-les durant cinc anys). Així mateix, durant dos anys, cal conservar el diari del tractament de dades que poden aportar-hi perfils de la personalitat d'individus).

Hi ha aparells que permeten que diferents suports que comparteixen el dispositiu de lectura puguin ser considerats en línia. Per exemple, hi ha grans biblioteques de suports informàtics gestionades per robots, els quals s'encarreguen de seleccionar i instal·lar el suport més adient a la unitat llegidora. En relació amb això, hi ha un dispositiu comercial anomenat *Jukebox*, que gestiona una col·lecció reduïda de suports (diguem-ne entre 25 i 100) i un braç mecànic que carrega i descarrega un disc cada vegada que actua. És molt semblant al tocadiscos que hi ha en alguns bars o llocs públics, on el consumidor hi introdueix una moneda, selecciona una cançó i tot seguit pot escoltar-la. L'ús d'aquests aparells no és molt extens i, per tant, el seu preu, a hores d'ara, és molt elevat.

3.2. TIPUS DE SUPORT SEGONS EL MÈTODE D'ACCÉS

Com hem comentat abans, hi ha suports que permeten només l'accés seqüencial, mentre que n'hi ha altres que suporten l'accés directe i també el seqüencial. Els suports seqüencials solen tenir format de cinta, mentre que els d'accés directe solen ser discos giratoris. Els primers, normalment, són molt més barats, encara que més lents i més exposats a trencaments, a causa de les tensions que suporten les cintes quan s'aturen, es posen en marxa i quan rebobinen. Tanmateix, solen tenir gran capacitat i ocupar espais més reduïts. Sovint els tractaments de dades necessiten accés directe i, per això, els dispositius seqüencials s'empren per emmagatzemar-hi arxius històrics i còpies de seguretat.

3.3. TIPUS DE SUPORT SEGONS LA TECNOLOGIA DE GRAVACIÓ

3.3.1. SUPORTS MAGNÈTICS

Els suports magnètics han estat els més àmpliament emprats per a l'emmagatzematge de la informació automatitzada en les dècades

dels anys vuitanta i noranta, i, fins i tot, a hores d'ara també. Estan formats per un substrat (una base de material inert) recobert de partícules de material magnètic molt petites. Cada partícula té dos possibles estats: pot ser magnetitzada o bé a l'esquerra o bé a la dreta. Cada partícula és una unitat d'informació. El suport sol estar protegit per una carcassa. En el mercat hi ha cintes i discos. Els disquets i els discos durs normalment són magnètics.

Els capçals del dispositiu tenen dues posicions: una de lectura i una altra d'escriptura. En lectura, el capçal detecta l'orientació del camp magnètic. En l'escriptura, el capçal és capaç de canviar la polaritat de cada partícula. Aquest tipus de tecnologia és, per això, **regravable**, és a dir, es pot modificar la informació escrita en el suport. El suport es deteriora amb l'ús, és a dir, després de moltes llegides i/o escriptures, és previsible que el suport estiga inservible. La informació emmagatzemada queda afectada per la presència de camps magnètics i també per les temperatures superiors a 40-50°C (si deixem un disquet dins del cotxe aparcat al carrer, al més d'agost, possiblement no podrem accedir a la informació que contenia). Els suports magnètics també es deterioren amb el transcórrer del temps; tenen una expectativa de vida al voltant de 10 anys i, com hem esmentat abans, les cintes són més propenses als trencaments per l'ús que els discos.

Els discos magnètics es graven sobre la superfície, a mena de pistes concèntriques. El disc es formata en sectors, tal com es mostra en la fig. 3.5

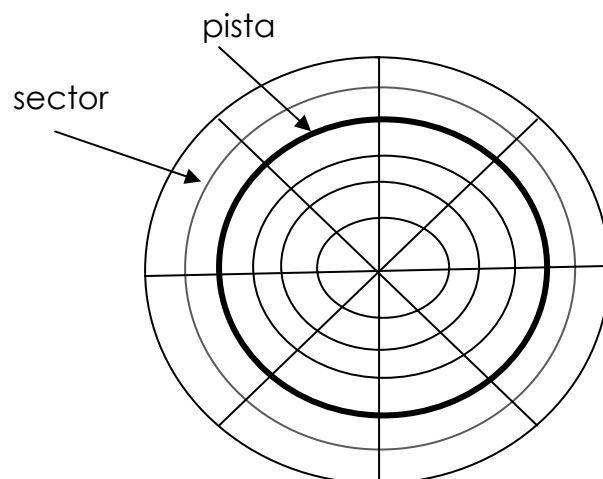


Fig. 3.5. Distribució de la informació en un suport magnètic giratori.

Tots els sectors tenen la mateixa capacitat d'emmagatzematge (hi poden albergar el mateix nombre de bits). Percebem que, amb aquesta distribució, la informació es troba més condensada a les pistes

interiors que a les de l'exterior. Alguns sistemes permeten seleccionar la grandària del sector. En altres, però, és un valor fix.

Cada suport dedica la pista més exterior per a mantenir un directori del contingut del suport. Aquest directori rep el nom de *taula d'assignació de fitxers*, FAT (*file allocation table*) o TOC (*table of contents*). Quan es crea el fitxer en un suport, se li assignen una sèrie de sectors consecutius i aquesta assignació s'apunta en el directori. Si posteriorment el fitxer creix, se li assignaran noves extensions, la qual cosa també queda reflectida en el directori. Quan es desitja accedir a un registre d'un fitxer ubicat en un suport, el sistema va primer al directori del suport a localitzar el lloc on és el fitxer dins del disc.

Per tal d'augmentar la capacitat del suport, freqüentment, s'apilen uns quants plats (discos) que giren sobre un eix dins de la mateixa carcassa. S'organitza a mena d'una pinta de cabells composta per capçals llegidors, un per cada superfície del disc, tots muntats sobre un mateix braç. El braç es desplaça sobre una serra dentada, on cada dent de la serra correspon a una pista d'un disc. Com es veu en la fig. 3.6, s'accedeix a la mateixa pista de cadascun dels plats al mateix temps, sense cap moviment del braç dels capçals llegidors. En la figura no s'han representat capçals en la primera i en l'última cara.

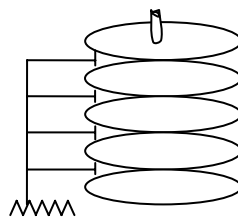


Fig. 3.6. Disc compost per cinc plats.

3.3.2. SUPORTS ÒPTICS

La tecnologia òptica grava la informació sobre el suport, mitjançant un raig làser molt concentrat i calorífic, que fa cremar una porció molt menuda de la superfície del disc. Els dos estats possibles per a la unitat d'informació en aquesta tecnologia són cremat o no cremat. Per llegir, es fa incidir sobre la superfície un feix de llum de menor intensitat (de manera que no siga danyada). Si la superfície estava intacta, és a dir, si no s'hi havia gravat res abans, el feix eixirà reflectit amb el mateix angle d'incidència i un mesurador de llum detectarà una quantitat de llum

similar a l'emesa. Si, al contrari, la superfície ja estava cremada, l'angle reflectit variarà i el detector no percebrà la llum. En la fig. 3.7 es representa un esquema dels processos d'escriptura i de lectura en un suport òptic.

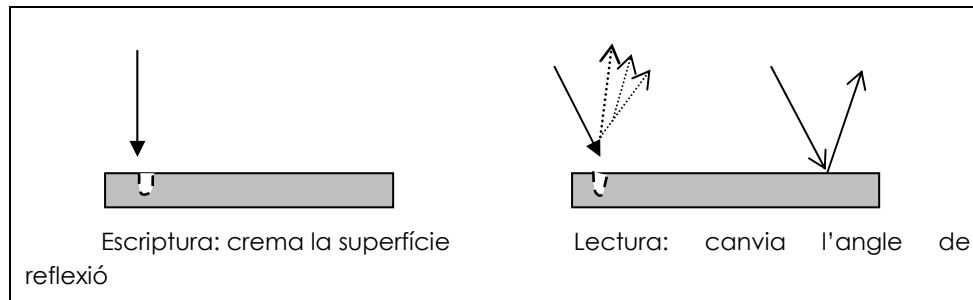


Fig. 3.7. Gravació i lectura en un suport òptic.

Els suports d'emmagatzematge òptic són de gravació única, és a dir, no es pot modificar la informació que contenen. Aquesta informació no s'altera per la presència de camps magnètics. Els perifèrics amb aquesta tecnologia presenten molta resistència a la humitat i a la temperatura. A més a més, no hi ha desplaçament dels caps llegidors sobre la superfície. Tot això fa pensar que no hi haurà deterioració amb l'ús de perifèrics amb aquesta tecnologia.

El substrat emprat determina les expectatives de durada del suport. En suports durs, el procés de gravació requereix la generació d'un negatiu (màster) amb el qual s'estampen els discos. La generació del negatiu necessita un làser d'alta potència, amb un preu que és car i, fins i tot, perillós. Per això les unitats de gravació són escasses. L'estampació es realitza en cases especialitzades. Fins a l'any 1992, s'havia d'enviar la informació a gravar fora d'Espanya. El làser de la unitat lectora és més accessible als usuaris. Per tot això aquesta tecnologia va rebre el nom de ROM (*read only memory*), que significa «memòria només de lectura».

Com a alternativa al costós procés d'estampació, va sorgir un suport òptic sobre substrat orgànic, que requereix un feix de menor intensitat, per la qual cosa és molt més accessible i permet la gravació per l'usuari. Al començament de la dècada dels anys noranta aquests tipus de suports estaven protegits en una carcassa com els disquets. En aquest cas, no hi ha negatiu ni tampoc procés d'estampació. La gravació es fa directament.

Però la vertadera alternativa als discos d'estampació va sorgir quan es va adoptar per a l'entorn informàtic els discos òptics compactes que procedien de la música, cosa que va esdevenir amb l'aparició del CD-ROM. En l'entorn musical hi ha estàndards ben definits pel que fa a la grandària i al format de gravació, la qual cosa permet garantir la compatibilitat, la reutilització i la continuïtat en el mercat. La música és un mercat obert al gran públic. Tot això permet abaratir els preus i incentivar la investigació de millores. En pocs anys s'han produït caigudes molt cridaneres de preus tant en els suports com en els dispositius de lectura. També han augmentat espectacularment les velocitats d'accés a les dades (la velocitat inicial d'accés era de 150 Mb/s i, ara per ara, han eixit lectors 8x, 16x i, fins i tot, 52x). A més a més, s'han fet gravadores molt barates i se n'ha generalitzat la utilització. Aquests tipus de suports s'anomenen popularment *CD-ROM gravable* o *enregistrable*, la qual cosa és un anacronisme (disc compacte de memòria només de lectura gravable) i, encara ho és més, si s'anomena CD-ROM regravable. Aquests suports hereten els estàndards dels CD-ROM, però no són tecnologia òptica, sinó magnéticoòptica. Els CD-ROM, CD-ROM gravables i CD-ROM regravables tenen externament l'aspecte dels discos d'estampació, no porten carcassa i tenen 4,75 polzades.

Els discos d'estampació continuen sent una alternativa que cal tenir en compte quan cal fer un nombre de còpies molt gran, perquè el preu en aquestes circumstàncies és menor que si ho grava l'usuari i, a més a més, són més resistents al transcórrer del temps i a condicions ambientals extremes.

El DVD (*digital versatile disc*) és la generació de suports òptics que segueixen al CD-ROM. Així com va passar amb el CD-ROM, hem de distingir entre els discos DVD de vídeo i els DVD-ROM, que són els de dades. El DVD-ROM proporciona una velocitat de transferència al voltant de 9 vegades més gran que el CD-ROM (entre 1.321 Mb/s per a lectors 1x, a 21 Mb/s per als de 16x). La capacitat depèn de la grandària (hi ha suports de 12 i de 8 centímetres), del format de gravació (CLV o CAV, que s'expliquen a continuació) i del nombre de cares útils. Amb aquestes variables, la capacitat es troba compresa entre 4 i 15 gigabytes. El fet que hi haja aquesta versatilitat implica que cap dels formats és realment estàndard i, per tant, la seua compatibilitat i expectativa de supervivència en el mercat és menor que la dels CD-ROM. El primer gravador de DVD-ROM estigué disponible en el mercat per l'any 1997. Els gravadors de DVD de vídeo s'han començat a

comercialitzar per l'any 2000 i tenen la mancança de no poder enregistrar pel lícules xifrades.

3.3.3. FORMATS DE GRAVACIÓ

Els suports giratoris informàtics sempre s'han gravat en pistes concèntriques, com hem representat en el cas dels discos magnètics. En aquest format d'enregistrament, el disc gira sobre l'eix amb una velocitat angular constant CAV (*constant angular velocity*). Aquest format permet posicionar el capçal llegidor exactament en una pista i en un sector. Els suports òptics generats exclusivament per l'entorn informàtic són gravats en aquest format. Els estàndards de música i, per tant, els CD-ROM de dades, ho graven en espiral. El disc fa voltes sobre l'eix, que són més ràpides si el capçal es troba col·locat prop de l'eix i més lentes a mesura que es desplaça cap a l'exterior, de manera que l'únic que hi ha constant és la velocitat lineal del capçal CLV (*constant linear velocity*). En format CLV s'aprofita millor la superfície de gravació, ja que s'hi manté una densitat uniforme (pensem que les pistes exteriors estaven desaprofitades en el format CAV). No obstant això, el format CLV no garanteix el lloc de posicionament del capçal llegidor. L'accés directe als discos CLV és, d'alguna manera, simulat. De fet, no és possible aturar la imatge en moviment quan aquesta ha estat enregistrada en un suport amb format CLV, i sí que és possible aturar-la quan ha estat gravada en un suport amb format CAV. Un mateix suport pot, en principi, gravar-se en qualsevol dels dos formats. Tot depèn del dispositiu de lectura/gravació. En la fig. 3.8 es representen ambdós formats.



Fig. 3.8. Formats de gravació.

3.3.4. SUPORTS MAGNETICOÒPTICS

La tecnologia magnetoòptica aprofita la propietat que tenen alguns materials de romandre inalterats davant de la presència de camps magnètics a baixes temperatures i, no obstant això, polaritzar-se a partir d'una temperatura determinada, anomenada *temperatura de Curie* del material. Aquesta propietat s'anomena *coercitivitat*.

En la gravació de dades sobre un suport magnetoòptic, un raig làser d'intensitat mitjana, però de molt alta resolució (molt estret), escalfa un punt molt concret del suport fins a la temperatura de Curie; llavors, un camp magnètic enregistra la informació i el raig làser s'atura; com a conseqüència, el punt incidit es gela. En situació freda la informació és inalterable, però si es torna a aplicar el raig làser fins a assolir la temperatura adequada, es pot modificar la informació emmagatzemada i aplicar-hi el camp magnètic contrari.

La lectura d'aquest tipus de suport es basa en l'*efecte Kerr*, que consisteix en el canvi que experimenta un feix de llum polaritzada quan s'enfronta a un camp magnètic. Un detector examina la polarització del feix reflectit.

En aquesta tecnologia, l'expectativa de durada de la informació emmagatzemada també és molt alta, ja que també és un feix lluminós el que fa la lectura sense tocar la superfície. Així mateix, és inalterable per la presència de camps magnètics i per rangs molt elevats de temperatura, ja que s'elegeixen substrats per als quals les temperatures de Curie estiguen al voltant de 150°C.

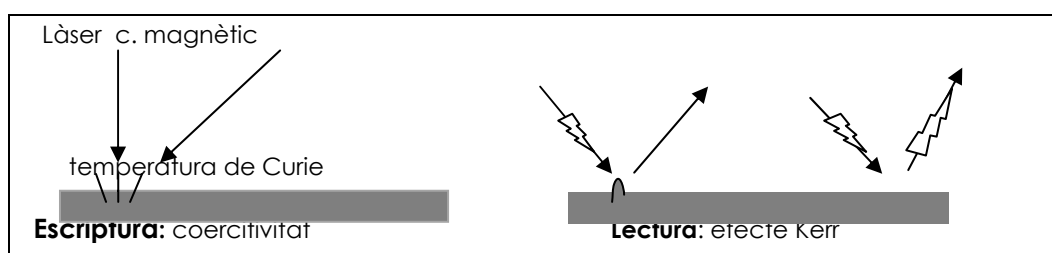


Fig. 3.9. Gravació i lectura amb tecnologia magnetoòptica.

Els CD-ROM regravables, com hem comentat abans, són tecnologia magnetoòptica. La següent generació de suports serà el DVD-RAM.

3.3.5. ALTRES TECNOLOGIES D'EMMAGATZEMATGE

En la dècada dels anys setanta, els perifèrics més utilitzats per emmagatzemar-hi informació estaven basats en tecnologia mecànica. Els suports més comuns eren targetes o cintes de paper perforat. Però es feien malbé amb l'ús, eren incòmodes de transportar i es deterioraven amb la humitat. Aquests suports només s'empren en l'actualitat per comunicar dues màquines en entorns de manufacturació assistida per ordinador (CAM).

S'han proposat suports d'emmagatzematge basats en memòries associatives, però estan menys esteses que les tecnologies abans esmentades.

Quan cal emmagatzemar-hi grans quantitats de documents, es poden emprar microfitxes i holografies. Aquests suports tenen l'avantatge que ocupen molt poc volum. Les microfitxes han estat àmpliament emprades per emmagatzemar-hi documents. Les holografies tenen l'avantatge que, una vegada generades, és molt barat fer-ne còpies a partir del negatiu. Són costoses de copiar si no es disposa del negatiu. Per això, pot ser una bona opció per a la difusió massiva d'informació, sobretot quan es desitja protegir la propietat intel·lectual. Tanmateix, ara per ara, és massa cara la generació del negatiu.

En ambdues tecnologies, la informació emmagatzemada no és accessible a l'ordinador. Els documents en cada microfitxa o holografia es troben catalogats en una base de dades. L'usuari consulta la base de dades, decideix els documents rellevants i el sistema informàtic gestiona de manera automàtica el visionament de la imatge sol·licitada, però el sistema informàtic realment no té accés a la informació del document, sinó a la seua catalogació.

4. TEMPS D'ACCÉS

Tal com hem vist en l'apartat 2, en un ordinador coherent el temps d'accés a la informació que conté un suport perifèric és al voltant de 1.000 vegades més gran que el d'una operació o que el de l'accés a una posició de la memòria principal. Per això, en les aplicacions que fan un ús intensiu de dades (**bases de dades i sistemes de recuperació de la informació**), el temps de procés es troba **condicionat** per les operacions d'**entrada/eixida** als perifèrics, és a dir, el temps de resposta d'un programa que accedeix a fer recerca d'informació en suports depèn

del nombre de vegades que haja d'accedir al disc, i no de les operacions que faça amb les dades.

Per exemple, si en cada accés es llegeix un registre:

- a. Cal llegir 1.000 registres i fer 2 operacions amb cadascun: el procés tardarà $1.002 * 1.000$ unitats de temps.
- b. Cal llegir 1.000 registres i fer 10 operacions amb cadascun: el procés tardarà $1.010 * 1.000$ unitats de temps.
- c. Si en el cas a i b, després de llegir cada registre, es fa la primera operació i, tot seguit, es tornen a llegir per fer la segona, el temps d'accés serà $1.001 * 1.000 * 2$.

Els temps de a i b són pràcticament iguals. El temps de c es quasi el doble que el de a.

L'organització del treball proposat en a és millor que la de c, ja que aconseguix els mateixos resultats en la meitat de temps. En general, intentarem organitzar els programes i també les dades dins del fitxer, de manera que s'hi facen el menor nombre d'accessos.

4.1. ACCÉS A UN SECTOR D'UNA PISTA

Més avall, en la fig. 3.10 a), reproduïm el braç del capçal lector que accedeix a cada una de les pistes quan es desplaça sobre una serra que marca posicions fixes. En la fig. 3.10 b) ressaltem una petita perforació a la superfície del disc, que servirà per identificar els sectors. El plat està girant a velocitat angular constant. Sota el plat, hi ha un emissor de llum i, per damunt, un detector. Quan el detector veu llum, vol dir que es troba sobre la marca d'inici. Com que la velocitat de rotació és constant, en temps constants es trobarà sobre els sectors 1, 2, etc.

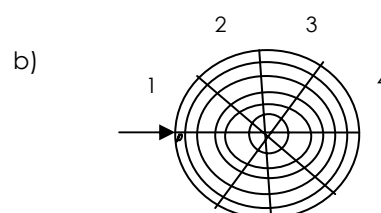
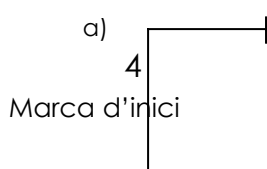




Fig. 3.10. Localització d'una pista i un sector del disc.

Quan el controlador rep l'ordre de portar a la memòria un sector del disc, ha de moure el braç del capçal a la pista indicada, fer girar el disc fins que el detector de llum trobe la marca d'inici i girar fins que hagen passat els sectors anteriors. En aquest moment, començarà a transferir la informació. En la fig. 3.11 es descompon el temps per a accedir a una posició concreta del disc, des de qualsevol altra posició (temps aleatori d'accés), en el temps d'arrencada, temps de latència i el de transferència de dades. S'hi ha posat en negreta el nom com sol figurar cada paràmetre en el fullets comercials dels suports. El temps de transferència bruta és allò que es tarda a llegir tot el disc, des del primer sector fins a l'últim.

t_1 = col·locació del capçal lector en pista = posada en marxa, acceleració, posicionament, desacceleració, ajust = **t. arrencada**.

t_2 = localització del sector 0 i recerca del sector requerit = **t. latència rotacional**.

t_3 = càrrega de dades = proporcional a la grandària del sector = **temps de càrrega** =

t. transferència.

Temps aleatori d'accés = $t_1 + t_2 + t_3$.

Temps de transferència brut = temps d'accés a tot el disc seqüencialment = $t_1 * \text{nombre de pistes} + t_3 * \text{nombre de sectors al disc}$.

Fig. 3.11. Paràmetres freqüents de temps d'accés a disc.

El sistema assigna sectors consecutius a cada extensió del fitxer. Quan el disc conté diversos plats llegits per un braç de capçals, com en la fig. 3.7, una pista de cada plat es troba accessible sense moure el braç, de manera que formen un cilindre d'informació, i probablement el fitxer s'escriurà en una pista de cada plat.

Una lectura seqüencial sol·licita «llegir el registre següent», per a la qual cosa necessitem consumir el temps t_2 . Una lectura en accés directe sempre necessita localitzar el sector concret. La lectura seqüencial és més ràpida que la directa. El temps de transferència, t_3 , serà més gran com més gran siga la grandària del sector.

En un sector hi ha diversos registres i tots fan cap a la memòria quan un és sol·licitat. Si els registres que se sol·licitaran es poden predir en un ordre determinat, és avantatjós emmagatzemar-los en aquest ordre, perquè quan sol·licitem el «següent», ja estarà en memòria principal i pot convenir que els blocs siguin relativament grans. Si l'ordre d'accés no es pot predir, possiblement convinga que la grandària del bloc siga menor perquè no viatgen dades innecessàries.

Quan triem una organització per a les dades d'un fitxer, el criteri fonamental serà aconseguir fer el mínim nombre d'accessos a disc, llegir en seqüència i evitar el moviment dels capçalers llegidors.

4.2. ACCÉS A UN REGISTRE CONCRET D'UN FITXER

Hem vist com el sistema accedeix a les dades contingudes en un sector concret. En aquest apartat intentarem entendre com el sistema respon a la pregunta següent: quin sector del disc ha de portar el controlador a la memòria per operar amb el registre núm. 37 del fitxer d'EMPLEATS?

El directori del suport indica en quina pista i sector es troba cada extensió del fitxer. Si el registre té un nombre fix de caràcters, en cada sector cap un nombre determinat i enter de registres. El nombre de registres que caben en un sector s'anomena **factor de bloqueig**. Això pot produir un desaprofitament en cada sector. El sistema calcula el nombre d'ordre del sector on hi ha el fitxer que conté el registre desitjat, ja que si sabem el nombre de sectors per pista i també la pista i el sector d'inici, és possible calcular la pista i el sector concret per portar-hi la memòria.

Tots aquests càlculs els fa el sistema operatiu en el moment que se sol·licita accedir a un registre. Això té l'avantatge que el programa només ha de demanar el número del registre que desitja que conté en un fitxer, on el programa coneix l'organització del fitxer esmentat, independentment d'on es trobe ubicat el fitxer dins del disc o que es porte a un altre disc, fins i tot encara que tinga una grandària diferent de sector. Tot això proporciona **independència** entre el programa i el dispositiu d'emmagatzematge. Tanmateix, ens interessa entendre aquest procés per organitzar les dades de manera que l'ús siga tan ràpid com es puga.

Exercici 3.1. Un fitxer es troba emmagatzemat en un suport formatat a 1.024 bytes per sector. El registre es troba definit de la manera següent:

Tipus de registre emprat

Nom = cadena [20]
Cognoms = cadena [40]
DNI: cadena [9]
sexe: [d, h]
sou, complements: real
antiguitat: enter curt
END

Quina és la grandària del registre? Quin és el factor de bloqueig? Quin és el desaprofitament per bloc?

Exercici 3.2. Una pista té capacitat per a 20.500 bytes. Quants sectors contindrà si es formata a 1.024 bytes per sector? Quin és el desaprofitament per pista quan es formata?

Exercici 3.3. Calculeu la pista i el sector en el disc del registre núm. 37 d'un fitxer de 30 caràcters de llargària sobre un suport amb les característiques següents:

- 8 sectors per pista (SP).
- 512 caràcters per sector (CS) (aquestes dades són fixes per al suport, després de formatar).
- La primera extensió del fitxer comença a la pista 16 (PI), sector 6 (SI) i ocupa 20 sectors (aquesta dada figura a la FAT del disc).
- El fitxer té una segona extensió d'uns altres 20 sectors a partir de la pista 50, sector 3.

Calculeu el nombre del caràcter del bloc on comença el registre.

Exercici 3.4. Emprant les dades de l'exercici anterior, indiqueu en quin byte del *buffer* comença el registre 37.

Exercici 3.5. Emprant les dades de l'exercici anterior, quina seria la pista i el sector al disc del registre núm. 179 del fitxer?

4.3. REGISTRES DE LLARGÀRIA VARIABLE

En l'apartat anterior hem vist que el sistema requereix que la llargària del registre siga fixa i, a més, coneguda (és a dir, amb un nombre determinat de bytes) per a poder fer el càlcul de la pista i el sector adequadament. Si la llargària del registre és variable, l'aplicació no pot sol·licitar al sistema que accedisca a un registre concret segons la seua posició i, per això, no pot proporcionar un accés calculat com acabem de fer. Addicionalment, si el registre és de llargària fixa, fer una

modificació suposa portar el sector a la memòria, modificar el valor contingut en un o diversos camps i reescriure el sector al seu lloc. Si el registre és de llargària variable, la modificació pot implicar un augment de la grandària i, per tant, que el registre ja no càpiga al bloc on estava. Com a conseqüència, en aquest cas, caldrà reescriure el bloc sense aquest registre i portar a la memòria un altre bloc, que s'escriurà on es puga. El tractament de registres de grandària variable és més complicat. Per això, sempre que siga possible, es procura considerar-lo de llargària fixa, encara que això suppose més consum de memòria que l'estrictament necessari.

A continuació, es presenten les solucions que regeneren aquests tipus de registres.

4.3.1. CAMPS DE LLARGÀRIA VARIABLE

La llargària d'aquests camps depèn de la informació que continguem. Per exemple, podem tenir un camp nom, en què la seua llargària s'adapte a la llargària del nom concret que emmagatzeme.

Solucions:

- Fixar una grandària acceptable a partir d'una previsió de grandàries i dels requisits de precisió de l'entorn, **completar amb blancs** fins al final del camp i convertir-lo així en un camp de llargària fixa.
- **Marca de fi de camp:** per delimitar la llargària del camp, es pot utilitzar un **caràcter especial de fi de camp**. Això complicarà el programa.
- **Màscara:** un camp addicional indica la llargària del contingut de cada camp de grandària variable. Aquest camp addicional s'anomena **màscara de camp**.

4.3.2. CAMPS NO APLICABLES O OPCIONALS

Poden haver-hi camps que no s'emplenen depenent del contingut d'altres camps. Aquests camps buits ocupen un espai inútil. Per exemple, en un fitxer d'un hospital, un camp sexe pot condicionar l'aparició d'altres camps com per exemple *data_del_darrer_embaràs*, etc.

Solucions:

- Si hi ha pocs camps opcionals, es pot forçar que siguem sempre presents en tots els registres i els que no s'utilitzen, **emplenar-los amb blancs** i tractar-los com a registres de llargària fixa.

- **Caràcter de fi de camp:** com a mínim, generarà tants caràcters com camps opcionals. Si el nombre de camps opcionals és molt gran, pot ser preferible l'opció següent.
- Parelles de valors (camp-valor).

4.3.3. GRUPS REPETITIUS

Aquests són grups de camps que es poden repetir un nombre de vegades no definit. Per exemple, *fills* és un grup repetitiu en un fitxer d'empleats que continga el nom i l'edat de cadascun dels fills de cada empleat.

Solucions:

- Si s'hi troba definit el nombre màxim de vegades que es pot repetir, o la semàntica del món real permet menysprear els que excedeixen un determinat valor, es pot **reservar un nombre considerable de camps** per cada registre, amb la consegüent pèrdua d'espai. Es tractaran com a registres de llargària fixa.
- **Caràcter de fi de grup repetitiu.** Cal indicar que el caràcter següent ja correspon a un altre camp.
- Organitzar el grup repetitiu **en un altre fitxer** on hi ha tants registres com vegades es repeteix el grup. Així es converteix el problema d'un fitxer de llargària variable en dos fitxers de llargària fixa, sense cap pèrdua de generalitat, ni pèrdues d'espai. L'inconvenient d'aquesta opció és que les dades del grup repetitiu queden físicament separades del grup a l'emmagatzematge secundari, i llegir-los tots dos suposa fer diversos accessos al disc.
- Una **màscara** que indica el nombre de vegades que es repeteixen els camps esmentats.

4.3.4. REGISTRES DE DIFERENTS TIPUS

Aquest cas es podria evitar amb alguns fitxers que continguin els registres de cadascun dels tipus. Però tenir diversos tipus de registres en un mateix fitxer pot tenir alguns avantatges. Així que si cada fitxer obert consumeix un *buffer* de memòria, pot ser interessant tenir menys fitxers oberts. A més a més, sovint els fitxers que especifiquen el domini són menuts. Per exemple, *sexe {home, dona}*, *dies {dilluns, dimarts...}*. Si cada domini es troba en un fitxer, cal fer-hi un accés a cadascun; possiblement en caben uns quants en un sòl *buffer*.

Solucions:

- Una **màscara** que indique, amb un codi, de quin tipus de registre es tracta, i d'aquesta manera saber quina n'és l'estructura interna.

Hem de tenir en compte que, en els casos de marca i màscara, es requereix doble lectura: la primera per transferir les dades del sector al *buffer* sobre la variable «cadena de caràcters» en la qual s'investigue quin és el format real del registre. La segona lectura és sobre el *buffer*, cosa que no suposa accés al disc, ja que porta codis llegits a les variables de memòria, les quals són comprensibles pel gestor (normalment, els noms dels camps de cadascun dels registres).

Sempre que ha estat possible hem intentat convertir els registres a llargària fixa. Les raons per les quals cal mantenir una gestió de registres de llargària variable poden ser, d'una banda, que el consum de memòria que comportaria la gestió de registres fixos fóra francament molt més gran que la memòria que realment ocupen les dades; de l'altra banda, que els temps d'accés a les dades, atès que es troben fraccionades, siguin tan elevats que superen les reduccions que aporten la gestió més senzilla del registres fixos. El primer cas esdevé quan la grandària màxima del registre (aquell que no suposa restriccions en el món real) és molt més gran que la grandària de la mitjana dels registres, la qual cosa suposa un desaprofitament de memòria inacceptable. El segon cas té lloc quan els patrons d'ús de les dades són de tal manera que la majoria de les vegades que s'accedeix a un registre, s'accedeix a totes les seues dades. Totes dues situacions es fan més sovint en els entorns documentals que en els tractaments de dades típics de gestió.

5. SOLUCIÓ ALS EXERCICIS PROPOSATS

3.1. La grandària del registre és $(20 + 40 + 9 + 1 + 4 + 4 + 2) = \mathbf{80 \text{ bytes}}$. El factor de bloqueig és el nombre de registres que caben al sector $FB = \text{part_entera}(1.024/80) = \mathbf{12}$. El desaprofitament per bloc és $1.024 - 12 * 80 = \mathbf{64 \text{ bytes}}$.

3.2. El nombre de sectors per pista és $\text{part_entera}(20.500 / 1.024) = \mathbf{20 \text{ sectors per pista}}$. El desaprofitament per pista és $20.500 - 1.024 * 20 = \mathbf{20 \text{ bytes}}$.

3.3. **Factor de bloqueig** (nre. registres per bloc) = $\text{Int}(CS/L) = \text{Int}(512/30) = \mathbf{17}$.

- El registre es trobarà al sector número SF:
 $SF = \text{Int}(R/FB) + 1 = \text{Int}(37/17) + 1 = \mathbf{3 \text{ del fitxer}}$.
- Ubicació al disc
 $P = \text{pista del disc} = PI + \text{Int}[(SF + SI) / SP] = 16 + \text{Int}[(3 + 6) / 18] = \mathbf{17}$.
 $S = \text{sector del disc} = \text{mod}8(SF + SI) = \mathbf{1}$.

- Portar a la memòria principal el bloc que hi ha a la pista 17, sector 1.
- **D = desplaçament** = (caràcter del bloc on comença un registre concret)
 $D = (R \text{ interès} - 1r \text{ registre del sector } R1) * \text{nre. caràcters} + 1.$
 $R1 = (SF-1) * FB + 1 = 2 \text{ sectors} * 17 \text{ reg./sector} + 1 = 35.$
D = (37-35) * 30 + 1 = 61.

BIBLIOGRAFIA

PEÑA, Rosalía, BAEZA, Ricardo i RODRÍGUEZ, José V. *Gestión digital de la información. De bits a bibliotecas digitales y la web*. Ed. Ra-Ma. 2002.

ELMASRI, Ramez i NAVATHE, Shamkant. «Sistemas de bases de datos. Conceptos fundamentales». Ed. Addison-Wesley. 1997.