

Edició de discos CD-ROM

JOSEP ÀNGEL BORRÀS*

CARME CANTOS*

Resum: Descripció de les fases d'edició de discos CD-ROM amb una anàlisi de les principals problemàtiques que incideixen en el desenvolupament de les aplicacions: preparació de les dades, integració en el programa de recuperació, creació de la interfície d'usuari, "premastering" i fabricació del disc. S'hi inclou una breu descripció dels costos implicats en el procés de la realització dels discos.

Introducció

En parlar de l'edició de discos CD-ROM, cal partir de la base que es tracta d'un procés editorial anàleg al d'altres productes sobre suports tradicionals, on hi ha la figura del productor de la informació, la figura de l'editor, la del fabricant, i la del distribuïdor. Parlem de procés editorial, perquè el CD-ROM és un suport de distribució de productes llegibles per ordinador, consistents en una gran quantitat de dades (textuals, numèriques, imatges...) i uns programes per a la recuperació i tractament d'aquestes dades.

Al llarg de l'article s'analitzaran les diferents fases d'aquest procés editorial des de la concepció del producte a editar, fins a la comercialització dels discos, incidint sobretot en les peculiaritats tècniques del desenvolupament de les aplicacions.

Generalment s'arriba al plantejament de la necessitat d'editar unes dades en CD-ROM quan s'adverteix que el volum del producte que fins ara es distribuïa en suport magnètic o paper fa incòmoda la seva distribució comercial. Arribat aquest punt, l'editor acostuma a trobar-se amb el problema que cal redefinir el producte per tal d'adaptar-lo a les peculiaritats del nou suport. Això és així perquè els programes de recuperació de la informació han d'estar preparats específicament per tal de garantir unes prestacions d'execució acceptables per a l'usuari final, ja que el CD-ROM és un dispositiu molt més lent que els discos durs magnètics.

* DOC6. Àrea d'Edició de CD-ROM.

1. Condicionaments del suport a la recuperació de les dades

Les peculiaritats del suport deriven del fet que es tracta d'un dispositiu que gira a Velocitat Lineal Constant, i que conté la informació emmagatzemada en una espiral que va des del centre del disc cap a fora.

Això implica que, per poder mantenir constant la velocitat amb què es llegeixen les dades, el disc ha de girar més a poc a poc a mesura que el capçal lector s'allunya del centre del disc.

L'accés a la informació és aleatori i es realitza mitjançant la tramesa de les adreces dels segments d'espiral que contenen les dades desitjades al controlador del lector del disc.

El capçal lector es desplaça únicament en una direcció, mogut per uns dispositius que el traslladen en ambdós sentits a petició del controlador. Quan es rep una adreça, el capçal es desplaça cap a la posició indicada. En arribar-hi, espera que la informació desitjada passi per sota i comença a llegir les dades.

Aquesta manera de procedir, implica que l'accés serà molt més ràpid als arxius que es trobin més a prop del centre del disc, ja que el capçal triga molt menys a arribar-hi i posicionar-se correctament des de l'adreça inicial del principi de l'espiral. Cal tenir en compte, a més a més, que com menys vegades hagi d'aixecar-se el capçal i posicionar-se en una altra adreça, menys pèrdua de treball útil d'aquest capçal per a la lectura de les dades.

Això implica que el programa de recuperació oferirà unes prestacions d'execució superiors si:

- Els arxius que conformen l'aplicació no s'emmagatzemen fragmentats en el disc per evitar que el capçal lector hagi de fer desplaçaments i posicionaments innecessaris.
- Els arxius que són utilitzats amb més freqüència pel programa de recuperació, són situats al principi de l'espiral per disminuir el temps amb què el capçal pot localitzar la informació requerida.
- Es redueix al màxim el nombre d'accessos que ha de fer el programa de recuperació per obtenir una informació del disc, mitjançant la creació d'uns índexs exhaustius i compactes, així com a través de l'aprofitament de memòries intermitges i arxius temporals en el disc dur per obtenir la major quantitat d'informació útil en el menor temps possible.

Si bé els dos requisits primers es poden aconseguir mitjançant un procediment d'optimització de les dades abans de gravar la informació en el CD-ROM, el tercer i més important només s'aconsegueix si el programa de recuperació ha estat dissenyat específicament per treballar en CD-ROM des d'un principi preveient les característiques especials d'aquest suport. Cal tenir en compte que l'adopció de programes per a la recuperació de dades dissenyats en principi per treballar sobre suport magnètic, o un procés d'optimització deficient o incorrecte,

pot fer que el temps emprat per a la recuperació de la informació sigui superior al que l'usuari final del producte està disposat a esperar.

A causa dels requeriments esmentats, en la majoria de casos es fa necessari redissenyar el producte que abans es distribuïa en suport magnètic per tal de poder editar-lo en CD-ROM. Generalment s'adopta un programa comercial per a la recuperació d'informació específic del CD-ROM i per a la interfície d'usuari, ja que els costos de desenvolupament d'un nou programa o l'adaptació de l'existent, serien massa elevats. Aquest costos, de fet, podrien impedir que el producte tingués una rendibilitat comercial, i no es tindria la seguretat d'aconseguir els objectius prefixats.

2. Selecció dels programes de recuperació

L'adopció d'un programa comercial ja creat presenta sense dubtes una gran quantitat d'avantatges enfront de nous desenvolupaments ad hoc, que es poden resumir bàsicament en:

- Garantia que el producte funcionarà correctament.
- Seguretat de revisions del programa –noves versions–, que permetran la remodelació del producte per anar adaptant-lo a la demanda del mercat.- Augment de la potencialitat comercial del producte mitjançant versions per a diferents sistemes operatius.
- Suport tècnic efectiu i continuat.

Lògicament existeixen els desavantatges que es deriven de les limitacions que aquest software pugui tenir en relació al desenvolupament del producte que es pretén editar; i la dependència en termes de propietat intel·lectual (que es tradueix indefectiblement en el pagament de llicències d'ús) que s'adquireix amb la casa creadora dels programes.

3. Fases de l'edició

Les fases per a l'edició de discos CD-ROM poden agrupar-se bàsicament en tres conjunts d'activitats:

- Les dirigides al desenvolupament del producte
 - Elecció del software de recuperació
 - Preparació de les dades (conversió de les dades a un format llegible pel software escollit)
- Integració de les dades
 - Definició dels camps de la base de dades
 - Definició dels camps a indexar i dels tipus d'indexació
 - Disseny de la interfície d'usuari
 - Disseny dels formats de visualització, impressió i exportació

- Creació del sistema d'ajuda
- Avaluació i depuració
- Les corresponents a la fabricació física del producte
 - "Premastering"
 - "Mastering"
 - Replicació
 - Serigrafiat dels discos
 - Impressió dels manuals
 - Fabricació de l'embalatge
 - Manipulació i encaixat
- Les destinades a la distribució dels discos
 - Selecció dels canals de distribució

Tot seguit farem una breu anàlisi de les fases més importants del desenvolupament de l'aplicació i de l'obtenció dels discos CD-ROM.

3.1. Preparació de les dades

A l'hora d'editar una base de dades en CD-ROM, així com en qualsevol altre suport, hi ha un procés previ de preparació de les dades.

Aquest procés serà més o menys complicat en funció de l'estat de les dades font i del producte que es vulgui obtenir. Així doncs, en primer lloc cal plantejar-se quines dades es tenen, com estan organitzades i quins objectius ha de complir el producte que es vol presentar. Segons aquests objectius es decidirà el contingut, els camps d'interrogació, els índexs, els formats de visualització, etc.

En aquest plantejament s'han de tenir en compte els factors econòmic i temporal. Per exemple, aconseguir que el contingut d'un camp com el de matèria sigui homogeni, pot implicar moltes vegades un procés de preparació llarg i costós.

Per altra banda, els problemes amb què ens podem trobar a l'hora de fer un nou disseny són de tipus estructural i de contingut.

a) **Estructura.** És indispensable que les dades tinguin una estructura homogènia. Quan provenen d'un gestor de bases de dades, ja sigui documental o relacional, el procés de preparació és mínim, ja que la base de dades està organitzada segons una estructura predefinida que es repeteix a cada registre.

Hi ha, però, una tipologia molt variada de possibles problemes que es poden plantejar, per exemple:

- Integrar el contingut de dues bases de dades amb diferents estructures en una sola.
- Integrar el contingut d'una bases de dades dins d'una altra segons un codi que relaciona el contingut d'un o més camps.

- Afegir un nou camp a una base de dades relacional amb camps de longitud fixa, el contingut del qual s'ha generat amb un processador de textos i és de longitud variable.
- Homogeneïtzar el contingut de la base de dades que s'ha generat amb un processador de textos.

En tots aquests casos hi ha un procés previ a la lectura de les dades, que segons el problema requereix una solució diferent. Per exemple, en el primer cas caldria dissenyar una nova estructura de la base de dades que inclogués tots els camps possibles, en el segon i en el tercer cas caldria crear un programa que integrés el contingut de les dues fonts i en el quart caldria homogeneïtzar el contingut dels registres (p. e. fer que cada registre estigués separat per dos retorns de carro i un signe "\$" i que cada camp estigués separat per un retorn de carro i un signe "#").

b) **Contingut.** Aquests problemes també tenen una tipologia variada:

- Diferents pàgines de codis: a Catalunya per tal de poder accentuar les majúscules s'utilitza la pàgina de codis 850, mentre que la majoria de programes autors per a la edició de CD-ROM són d'origen anglosaxó i treballen amb la pàgina de codis 437 que no les inclou. Aquest problema afecta els camps que es volen indexar, ja que el programa autor en comptes d'indexar una vocal majúscula accentuada indexa el símbol corresponent de la pàgina de codis 437.
- Diferents jocs de caràcters: molts programes utilitzen un joc de caràcters propi que representa les vocals accentuades amb uns codis diferents. Per exemple:

```
8101160050020000$G Cas.$C 20-JAN-81$C 24-FEB-81$C 27-MAR-81$C08
8101160050020001-FEB-84
8101160050090000$C 111$X La Hormiga de Oro$Y 675$Z 10
8101160050900000$I 84-350-0300-0$C 831.49U Bea
8101160051000000$A BEAUVOIR$B Simone de$D 1908-1986
8101160052450000$A La plenitud de la vida$C Simone de Beauvoir;
8101160052450001 [traducció de Silvína Bullrich]
8101160052600000$A Barcelona$B Edhasa$C 1980
8101160053000000$A 661 p.$C 20
8101160054400000$A Narrativas contemporànees
8101160055000000$A La force de l'age
8101160056000000$A BEAUVOIR, SIMONE DE (1908-1986)
8101160057000000$A Bullrich$B Silvína$E trad.
```

En aquests casos cal fer una substitució de caràcters per tal que aparegui la vocal corresponent.

- Falta de criteris dels productors de les bases de dades (p. e. en el camp AUTOR, dins d'una mateixa base de dades, noms entrats de forma directa i for-

ma inversa, o bé indicant el tipus de responsabilitat sense una puntuació homogènia).

- Integració de bases de dades de diferents productors (p. e. en el camp MATÈRIA, diferents encapçalaments per una sola matèria o fins i tot diferents idiomes, o bé, en el camp DATA, diferents formats dia-mes-any, mes-dia-any, etc.).

Moltes vegades la solució d'aquests problemes passa per un procés manual llarg i costós.

3.2. Integració de les dades

A partir del disseny establert i de l'estructura de la base de dades es procedeix a la lectura i integració de la informació en el software de recuperació. El productor normalment facilita les dades en fitxers llegibles per ordinador en suport magnètic (cintes, disquets, etc.)

Hi ha diferents tipus d'estructures segons el gestor de bases de dades utilitzat: camps etiquetats i sense etiquetes, que alhora poden ser de longitud fixa, variable o delimitada, o que segueixen la norma ISO 2709 (formats MARC). Per exemple :

Camps etiquetats de longitud variable

Autor= Mendoza, Eduardo Títol= Sin noticias de Gurb Lloc_public= Barcelona
Editorial= Círculo de Lectores Any_public= 1991 Autor= Tasis, Rafael Títol= La
Bíblia valenciana Lloc_public= València Editorial= Eliseu Climent Any_public=
1991

Camps sense etiquetes i de longitud fixa

Mendoza, Eduardo	Sin noticias de Gurb	Barcelona
Círculo de Lectores	1991Tasis, Rafael	La
Bíblia valenciana	València	Eliseu Climent
		1991

Camps sense etiquetes i de longitud variable delimitada

Mendoza, Eduardo#Sin noticias de Gurb#Barcelona#Círculo de Lectores#
1991\$\$Tasis, Rafael#La Bíblia valenciana#València#Eliseu Climent#1991\$\$

Registre ISO 2709

```
00816 XXXX2200217XXX450
00100130000000800410001308000430005408000430009711000260014024
00820016625500160024826000540026430000130031850301050033155400
10004365560031004466100036004776500043005139980009005569990033
00565^00^_a03000003^ m 19—196cat wi^_e 000 1 spap
^00^_a388.4:658.115(467.1 Ba Barcelona)(058)^00^_a658.115(467.1 Ba
Barcelona)(058):388.4^20^_aTranvías de Barcelona^10^_aDatos
```

estadísticos...^_dTranvías de Barcelona S.A., Sociedad Privada
 Municipal^^00^_a%%%%-[1968]^10^_a[Barcelona]^_bTranvías de Barcelona^_c
 [19—]-
 [1968]^00^_c21x30 cm^^00^_aContinuada per: Datos
 estadísticos - Sociedad Privada Municipal Transportes de Barcelona =
 0211-3236^^00^_aAnual^^00^_aDescripció basada en: 1966^^20^_aTrànviats
 de Barcelona^_xInformes^^00^_aTransport urbà^_zBarcelona^_xEstadística
 00^<->

Durant aquest procés es decideix quins camps es vol que apareguin a la base de dades, s'indica quins són repetibles, quins contenen text, quins dades numèriques, quins estan dividits en subcamps; també es pot dividir el contingut d'un camp en diversos camps o viceversa, es poden fer substitucions de caràcters o de cadenes de caràcters. Però sempre a partir d'una estructura homogènia ja existent.

3.3. Definició dels camps de la base de dades

Una vegada s'han llegit les dades cal decidir quina serà l'estructura de la base de dades consultable en CD-ROM. Aquesta estructura pot ser la mateixa que ja tenia o es pot haver modificat mitjançant el procés d'integració.

Així mateix, s'indica si es tracta d'un camp de text, numèric o de data, si aquest camp es podrà visualitzar, imprimir i/o exportar, i quina etiqueta ha de tenir cada camp.

3.4. Definició dels camps a indexar i dels tipus d'indexació

El pas següent seria decidir quins camps es volen indexar i quin tipus d'indexació s'hi aplica. Hi ha diferents tipus d'indexació:

- **Paraula:** indexa el contingut d'un camp textual paraula per paraula. Es poden definir fitxers de paraules buides i indicar si s'hi volen incloure o excloure els numerals.
- **Frase:** indexa el contingut d'un camp textual sencer.
- **Permutat:** indexa el contingut d'un camp textual sencer per cadascun dels termes que conté.
- **Númers:** indexa el contingut d'un camp numèric.
- **Dates:** indexa el contingut d'un camp data. Es pot indicar quin format de data es desitja.

Pel que fa a l'ordenació, pels camps textuals es pot triar entre ordenació lletra per lletra o paraula per paraula i pels camps numèrics si es volen ordenar alfabèticament (p. e. CDU) o numèricament.

Es poden definir diferents tipus d'indexació per un mateix camp.

3.5. Disseny de la interfície d'usuari

Alguns softwares per a l'edició de CD-ROM compten amb una interfície d'usuari estàndard que no es pot modificar, d'altres permeten definir la interfície d'acord amb les necessitats de l'editor de la base de dades.

En els casos de les interfícies estàndards només varien els camps pels quals es pot interrogar la base de dades (*Spirs* de SilverPlatter, *Online* utilitzat per Bowker, Chadwyck-Healey, etc.). En altres casos es poden fer modificacions de la interfície d'usuari, etc. (*CD-Author/CD-Answer*, de Dataware, *Re:SEARCH retrieval & authoring system*, de MicroRetrieval, ...etc.).

3.6. Disseny dels formats de visualització, impressió i/o exportació

El mateix procés que s'ha indicat per a la interfície d'usuari es pot aplicar al disseny dels formats de visualització, impressió i exportació. Alguns softwares tenen formats estàndards que només permeten modificar l'ordre de les dades (*Spirs* de SilverPlatter) i d'altres permeten la definició de formats diversos (*Online* utilitzat per Bowker, *CD-Answer* de Dataware, etc.), per exemple formats etiquetats, ISBD, MARC, etc.

Aquesta versatilitat depèn de la finalitat que tingui el software per a l'edició de CD-ROM. Per exemple, el software *Spirs* de SilverPlatter no està pensat per editar catàlegs de biblioteques.

3.7. Avaluació i depuració

Al llarg de tots aquests processos cal realitzar una tasca d'avaluació i depuració conjuntament amb el productor i l'editor de les dades, per tal de poder arribar al producte desitjat sense pèrdues de temps i d'esforços.

Cal que el sistema autor inclogui els mecanismes adients per tal de dur a terme l'avaluació de cadascuna de les fases que s'han descrit.

El resultat, superada la fase d'avaluació i depuració, és el producte final, de moment encara sobre suport magnètic, i que cal traslladar-lo al suport CD-ROM final.

3.8. "Premastering"

Entenem com a "premastering" la conversió dels fitxers que formen l'aplicació des del seu format inicial (DOS, MAC...) a un format ISO 9660 o a l'estàndard High Sierra (predecessor de l'anterior). En aquest procés cal tenir en compte tot el procediment d'optimització de la disposició dels fitxers que permetrà que l'aplicació s'executi correctament.

El resultat d'aquest procés és un volum d'arxius en el format escollit (ISO o HS) i que cal disposar en un suport que sigui acceptat per la fàbrica que ha de

realitzar l'estampació dels discos (cintes de 9 pistes, DAT, cartutxos de 8 mm, CD-ROM (WO), etc.).

Cal tenir en compte que durant el desenvolupament de l'aplicació, s'ha de fer el disseny de l'etiqueta del disc, dels manuals, i de l'embalatge, així com gestionar els números normalitzats que calgui (ISBN, ISSN...).

Un cop realitzat el "premastering", cal enviar el suport que conté el "premaster" a la factoria escollida per a l'estampació, juntament amb el disseny gràfic per al serigrafiat de les còpies.

3.9. "Mastering" i estampació

Aquest procés s'inicia a la fàbrica d'estampació amb la rebuda del "premaster", que és, de fet, una imatge del que serà el disc CD-ROM.

A partir del "premaster" es realitza un original del disc en vidre especial, que rep el nom de "master" i es comença l'obtenció de les còpies del disc pel procediment d'estampació.

4. Costos

El tema de l'anàlisi de costos en l'edició de CD-ROM, està en funció dels costos de desenvolupament de l'aplicació i depèn en cada cas de la complexitat del producte a desenvolupar. Els principals factors a tenir en compte són: la quantitat de manipulacions que requereixin les dades per tal de poder convertirles a un format integrable en el programa de recuperació; la quantitat de dades a tractar i la complexitat de la interfície d'usuari que necessiti l'aplicació.

Els costos per traslladar una aplicació sobre suport magnètic al suport CD-ROM es poden dividir entre les etapes següents:

"Premastering"

Té un cost que depèn del volum de dades a contenir en el disc. Aproximadament unes 500 PTA per cada Mb d'informació (moltes empreses apliquen una quota mínima que pot oscil·lar entre les 50.000 i les 100.000 pessetes).

"Mastering"

Cost fix que depèn de la factoria en concret. Al voltant del mig milió de pessetes.

Obtenció de les còpies

El preu per unitat varia segons el número de còpies que es vulguin. Per a una tirada mitjana (200 exemplars), cal comptar entre 500 i 800 pessetes/còpia (tam-

bé influeix en el preu el nombre de colors que es fan servir per al serigrafiat de l'etiqueta).

Embalatge i manipulació

Depèn en cada cas de l'embalatge escollit i de la complexitat de la manipulació del producte fins aconseguir la presentació desitjada.

Els discos surten de la casa d'estampació dins de les típiques caixes de discos compactes musicals. Afegir-hi etiquetes impreses i incloure el disc en una caixa més gran juntament amb els manuals, pot suposar entre 500 i 5.000 pesetes per disc.

Llicències d'ús del software de recuperació

Depèn de cada casa en concret. De si l'edició és o no comercial, i si es tracta d'una llicència per a un ús en xarxa o bé per a un ús en un sistema monousuari. Hi ha softwares que no requereixen cap tipus de llicència d'ús.

Les llicències monousuari més cares poden arribar quasi a les 20.000 PTA/disc i any.

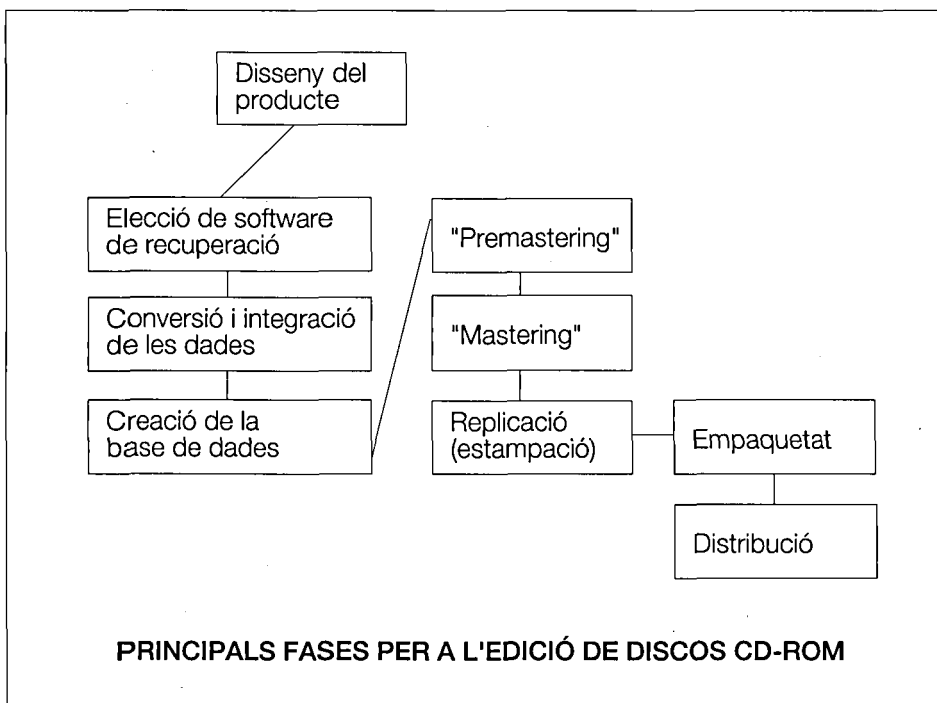
5. Tendències del mercat

El primer CD-ROM fet a Espanya, i un dels primers a Europa, va ser el *Diccionario médico Marín* desenvolupat per l'empresa ComCal de Barcelona l'any 1986. Aquesta primera temptativa no va tenir gaire ressò donat l'estat embrionari en què es trobava el mercat del CD-ROM al nostre país. Des d'aleshores s'han editat més de 15 títols comercials, nombre que creix lentament.

Per altra banda, la producció de discos no comercials s'incrementarà considerablement amb la introducció dels sistemes de gravació de sobretaula de discos CD-ROM WO (Write Once). Aquests dispositius estan pensats per obtenir una còpia (o una tirada curta de còpies) d'una aplicació informàtica, la qual cosa suposa un pas endavant en la normalització del CD-ROM com a suport de difusió de la informació. L'elaboració del disc final, fins ara patrimoni exclusiu de les factories d'estampació de Compact Discs, és ara a l'abast de qualsevol petita o mitjana empresa. Això suposa que el CD-ROM, a més de la distribució de dades, comença a ser utilitzat com a suport d'emmagatzematge i còpies de seguretat. En aquesta última accepció, presenta l'avantatge respecte al WORM de ser un dispositiu estàndard que pot ser llegit per qualsevol lector de CD-ROM.

Sens dubte, el fet de poder tenir una còpia del disc abans de fer-ne l'estampació en sèrie, permet fer un control de qualitat més exhaustiu a les empreses dedicades a l'edició de discos comercials, ja que no han de reïar-se de simulacions del comportament del CD-ROM mitjançant un software, o fer costoses in-

versions en hardware que permetin realitzar les emulacions necessàries per tal de comprovar el funcionament correcte de les aplicacions.



Bibliografia

- Cichocki, Edward M., i Susan M. Ziemer, "Design considerations for CD-ROM retrieval software." *Journal of the American Society for information science* 39(1) January 1988, p. 43-46.
- Davies, David H. "The CD-ROM medium." *Journal of the American Society for information science* 39(1) January 1988, p. 34-42.
- Pahwa, Ash, i Gordon Rudd, "Sending your CD-ROM data for mastering." *CD-ROM professional* September 1991, p.100-103.
- Soler i Carrascosa, Lluís, "El CD-ROM, emmagatzematge òptic d'informació." *Ciència* 64 abril 1990, p. 39-47.
- Thiel, Thomas J. "Costs of CD-ROM production : what they are and how to overcome them." *CD-ROM professional* March 1993, p.43-46.