

La gestió del document electrònic: l'experiència de la Biblioteca de la UOC

Marta Enrech Larrea. menrech@uoc.edu

David Maniega Legarda. dmaniega@uoc.edu

Jordi Serrano Muñoz. jserrano@uoc.edu

Núria Comellas Gaya. ncomellas@uoc.edu

- Resum
 - Introducció
 - Els documents electrònics: dins o fora del catàleg?
 - Document electrònic: criteris de selecció
 - Document electrònic: adquisició
 - Document electrònic: descripció i/o organització dins la web de la biblioteca
 - Dins el catàleg de la BUOC
 - A la web de la BUOC
 - Document electrònic: accés i control
 - Documents de bibliografia recomanada: digitalització dels sumaris i resums a la Biblioteca de la UOC
 - Sumaris de revistes
 - Procés de creació de les pàgines de sumaris
 - Futurs projectes en la digitalització de sumaris
 - Bibliografia
-

RESUM

En aquest treball es fa una aproximació al tractament dels documents electrònics (sobretot recursos d'Internet) en el context de la biblioteca d'una universitat a distància. També es relata l'experiència de la Biblioteca de la Universitat Oberta de Catalunya en el procés de creació de pàgines

electròniques de sumaris de revistes i monografies que formen part de la col·lecció de la biblioteca. Tots aquests serveis de valor afegit estan directament orientats vers uns usuaris que es posen en contacte amb la biblioteca des del Campus Virtual de la UOC.

INTRODUCCIÓ

La gestió del document electrònic genera una sèrie d'accions per part de qualsevol biblioteca, més encara si es tracta de la biblioteca d'una universitat a distància i que treballa en base a un campus virtual, mitjançant el qual es té accés a tots els serveis universitaris. El model de biblioteca que necessita i exigeix l'usuari de la UOC forçosament ha d'incloure i potenciar l'accés al document primari des del seu ordinador. La ponència presentarà els següents aspectes de la gestió del document electrònic: els criteris de selecció, l'adquisició, el tractament tècnic, l'accés, la difusió i el control dels documents.

ELS DOCUMENTS ELECTRÒNICS: DINS O FORA DEL CATÀLEG?

Entre tota la brossa que circula per Internet molt sovint fem troballes interessants per als usuaris d'una biblioteca universitària. Es tracta d'informació generada per universitats d'arreu del món, per instituts de recerca o per altres organitzacions públiques i privades que posen a l'abast de tothom documents i dades de tota mena (estadístiques, textos literaris, diccionaris, informes, etc). Si aquests documents ens arribessin a la biblioteca en qualsevol dels suports tradicionals, i si la informació que contenen fos interessant per al desenvolupament de la nostra col·lecció, no dubtaríem en incorporar-los al catàleg de la biblioteca i posar-los a disposició dels nostres usuaris.

Ara bé, la qüestió que se'ns planteja últimament és: podem fer el mateix amb els documents electrònics? Com podem gestionar documents que no tenim "físicament" a la nostra biblioteca? Quina és la millor manera de facilitar-ne l'accés a l'usuari?

A la BUOC ens és molt necessari incorporar els documents electrònics a la biblioteca virtual, i fer-los accessibles de manera ràpida i còmode als ordinadors dels nostres usuaris. Això vol dir dissenyar una interfície i un mode d'accés tan fàcils de consultar per un estudiant que disposa d'un Pentium a Barcelona com per aquell que viu a Figueres i treballa amb un 486.

Hi ha dues maneres bàsiques d'integrar aquests documents a la col·lecció d'una biblioteca: incorporar-los al catàleg o bé crear una pàgina web. A la BUOC actualment estem treballant amb ambdues possibilitats, i cadascuna d'elles té les seves propietats. Integrar un document electrònic al catàleg significa que se li aplicarà el mateix tipus de tractament tècnic (descripció, indexació, classificació) que a qualsevol altre document i que, per tant, les eines de recuperació per part de l'usuari seran les que ja ha après i que usa quotidianament quan consulta el catàleg de la biblioteca.

D'aquesta manera trobarà la versió electrònica de la Constitució espanyola de 1978 "al costat" de la seva versió impresa, només que per a accedir-hi el seu ordinador haurà de connectar-se amb un servidor mitjançant una adreça http. A la BUOC, tal com expliquem més endavant, hem aprofitat les possibilitats de l'etiqueta 856 de l'USMARC per tal d'agilitzar la consulta d'aquests recursos electrònics des del mateix catàleg.

L'altra opció és fer una pàgina web on s'organitzin d'una manera sistemàtica i clara els recursos d'Internet que la biblioteca ha seleccionat per als seus usuaris. Fins ara la pàgina web de la BUOC ha estat oferint en un dels seus apartats la visualització de diversos recursos d'Internet organitzats sota el títol "Biblioteca electrònica" (per al curs 97/98 aquesta pàgina haurà sofert canvis importants).



Figura 1. Pàgina web de la Biblioteca Electrònica de la UOC

Un aspecte a destacar i que la BUOC vol aportar en el tema de la gestió del document electrònic és el procés de convertir en electrònica una informació que abans ha estat analògica, com es el cas dels sumaris de llibres de la bibliografia recomanada i de revistes. Aquest tema el desenvolupem a partir del punt 7.

DOCUMENT ELECTRÒNIC: CRITERIS DE SELECCIÓ

Els criteris de selecció que s'apliquen als documents electrònics són els mateixos que s'apliquen als documents en qualsevol altre suport: la seva temàtica ha d'estar relacionada amb els estudis que s'imparteixen a la UOC, amb l'educació a distància i amb Internet. Per descomptat, també es valoren el rigor acadèmic de l'entitat que l'ha generat així com l'interès i la utilitat del recurs en qüestió per a la major part de la comunitat, però un dels factors més valorats és la disponibilitat del text complet del document. No oblidem que els alumnes de la UOC estudien a distància i els interessa poder accedir des del seu ordinador a la major quantitat d'informació possible sense intermediaris.

Com ja sabem, degut a la lògica protecció de la propietat intel·lectual i a que les editorials encara són un negoci, els documents accessibles a text complet per Internet no són tan nombrosos com l'usuari de la BUOC desitjaria. No obstant, la tipologia dels documents electrònics seleccionats és diversa: monografies sense problemes de drets de reproducció (obres clàssiques recopilades per eines com el Project Gutenberg), textos legals de domini públic, comunicacions i ponències de congressos, bibliografies, guies electròniques, revistes, bases de dades i tot tipus d'obres de referència.

Val la pena destacar que en la selecció hi participa tant el personal de la biblioteca com el professorat, així com els estudiants i el personal de gestió. Pel que fa a la selecció de recursos per part del professorat, s'han creat les vies necessàries per tal d'agilitar al màxim l'intercanvi d'informació amb la biblioteca sobre les novetats aparegudes a Internet en les respectives àrees d'interès.



Figura 2. Pàgina web dels estudis de Dret dins el Campus Virtual de la UOC

Dins el Campus Virtual de la UOC els recursos d'Internet seleccionats pels professors són accessibles tant des de la pàgina web de cada estudi (vegeu la figura 2) com des de la biblioteca. D'aquesta manera i a través del professorat s'encoratja els alumnes a consultar la biblioteca, a considerar-la una eina útil per al seu aprenentatge i a treure'n el màxim profit des del seu propi ordinador.

DOCUMENT ELECTRÒNIC: ADQUISICIÓ

Tot i que el concepte "adquirir" sembli una mica estrany dins el context d'Internet, cal recordar que moltes bases de dades (i sovint les millors) malgrat tenir interfícies web no són de consulta gratuïta. El Campus Virtual de la UOC és una intranet. Aquest fet ens permet gaudir d'una situació "privilegiada" a l'hora de negociar llicències per la consulta de bases de dades. La BUOC negocia amb el proveïdor un únic identificador i un password. Quan un membre de la comunitat UOC s'hi connecta, el proveïdor el reconeix com a usuari autoritzat i el sistema executa l'identificador i el password de manera "amagada".

Pel que fa als recursos i bases de dades accessibles a Internet que són d'accés gratuït, des de la BUOC informem per correu electrònic al gestor de la web de cada recurs seleccionat que hem enllaçat amb la seva pàgina i els demanem la seva autorització. En el cas que no hi estiguéssin d'acord (encara no s'ha donat aquesta situació) anul·lariem l'enllaç. En el mateix missatge els demanem que ens avisin si canvien d'URL o adreça Internet.

DOCUMENT ELECTRÒNIC: DESCRIPCIÓ I/O ORGANITZACIÓ DINS LA WEB DE LA BIBLIOTECA

DINS EL CATÀLEG DE LA BUOC

La inclusió de documents electrònics accessibles per Internet al catàleg de la Biblioteca de la UOC ha estat possible mitjançant una adaptació de l'etiqueta 856 de l'USMARC. L'etiqueta 856 conté la informació necessària per a identificar i localitzar un recurs a Internet, possibilitant a més la seva recuperació a través d'una connexió remota al servidor del recurs. L'etiqueta 856 té molts subcamps, però els que fem servir de moment a la UOC per tal de descriure i donar accés als documents electrònics seleccionats són bàsicament els tres següents: \$u, \$2 i \$n. El \$u és l'espai dedicat a l'adreça URL del recurs i el \$2 ens indica quin és el seu mètode d'accés (normalment: http). Ara bé, el \$n té a la BUOC una aplicació diferent de la que li dona l'USMARC ("Localització del host del \$a", en el cas que n'hi hagi). Nosaltres fem servir aquest subcamp de l'etiqueta 856 per a

que el sistema generi diverses icones a la part inferior de la fitxa bibliogràfica, segons el terme que el catalogador escrigui després del \$n:

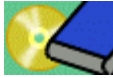




\$n	Icona
Document electrònic	
Sumaris	
WWW	
Fitxer informàtic	
Resum	

Figura 3: Icones generades per el \$n

Cal tenir en compte que aquesta etiqueta 856 no es visualitza a la fitxa bibliogràfica; el que veu l'usuari és una o més icones al peu de la descripció que li donen accés directe als diversos tipus de document electrònic (document a text complet, sumari d'una revista o d'una monografia, web creat per l'equip editorial de la revista descrita, etc.).



Figura 4. Fitxa bibliogràfica de la revista Internet World i accessos a recursos electrònics relacionats (sumari i web de la revista)

A LA WEB DE LA BUOC

Els documents electrònics s'organitzen dins la "Biblioteca Electrònica" de la UOC en grans blocs temàtics i, dins de cada bloc, per la forma bibliogràfica. Aquests grans blocs temàtics coincideixen amb els estudis impartits actualment per la UOC (Ciències Empresarials, Dret, Humanitats i Psicopedagogia), encara que hi ha també una important secció d'obres de referència accessibles per Internet. Quan l'usuari consulta, per exemple, el llistat de recursos de Ciències Empresarials (vegeu figura 5), troba una pàgina hipertext que a més dels enllaços corresponents a cada recurs li ofereix un resum indicatiu del tipus d'informació que trobarà a cadascun d'ells. Hem de tenir en compte

que aquesta primera selecció evita a l'usuari haver de carregar moltes pàgines abans de trobar la que realment li interessa.

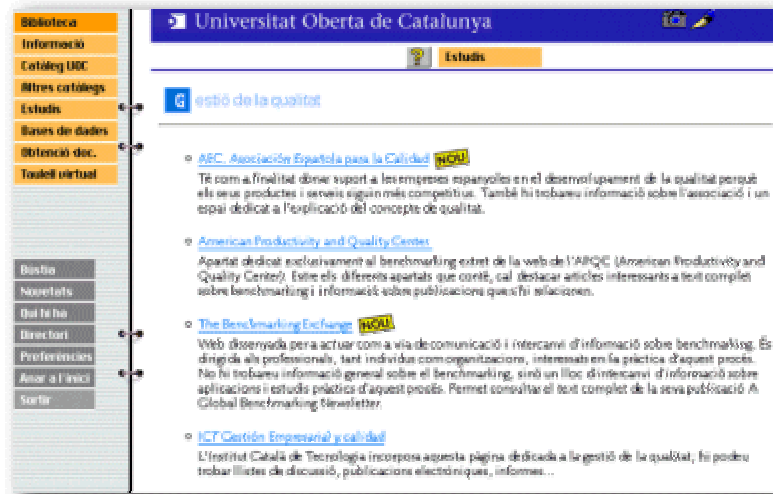


Figura 5. Recursos de Ciències Empresarials a la Biblioteca Electrònica de la UOC

DOCUMENT ELECTRÒNIC: ACCÉS I CONTROL

L'usuari intern de la UOC té accés a tots els documents electrònics incorporats a la BUOC a través del Campus Virtual simplement clicant a sobre de l'enllaç o de la icona corresponents al document que vol visualitzar. A més a cada pàgina de la web de la biblioteca es troba recolzat per missatges i/o formularis d'ajuda que el guien en la navegació. Solen ser consells sobre com configurar el telnet quan cal, quin és l'horari més idoni per a consultar certs recursos, etc. L'usuari extern només té accés a la selecció de recursos gratuïts que pot consultar mitjançant la pastilla "Biblioteca Virtual" de la web externa de la UOC (<http://www.uoc.es>) o accedint directament a la web de la biblioteca (<http://biblio.uoc.es>).

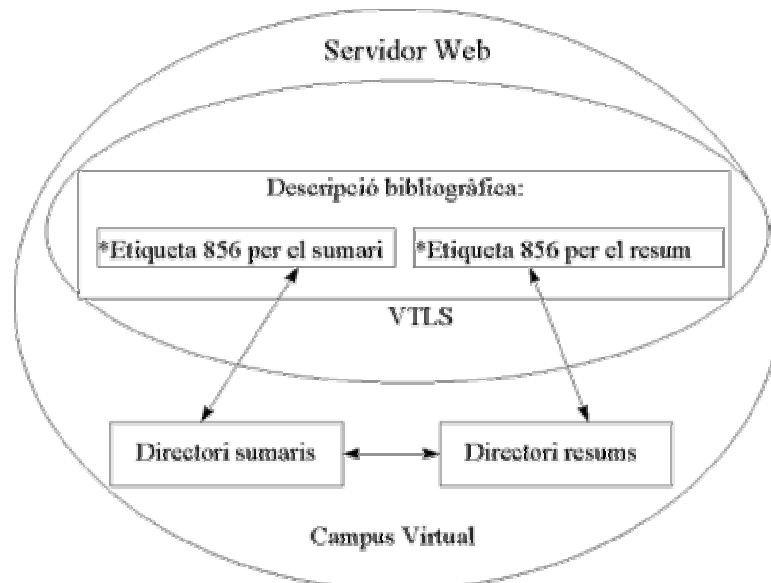
Els documents electrònics enriqueixen la col·lecció "física" de les biblioteques; però requereixen atencions especials. Tot el que circula per Internet és viu i canviant, i si un document electrònic canvia de format o (el que és pitjor) d'adreça URL la biblioteca ha de ser prou àgil per tal de mantenir al dia la seva selecció de recursos Internet. Si no, correm el perill de donar "pistes falses" a l'usuari, adreçant-lo a llocs d'Internet que ja no existeixen i, en conseqüència, desanimant-lo en l'ús de les eines de la biblioteca. Per tal d'evitar aquesta situació a la BUOC estem avaluant diversos programaris (com NetCarta) dissenyats per a comprovar periòdicament els enllaços establerts amb Internet i verificar la seva permanència a la xarxa. També estem fent proves amb serveis de web que ofereixen prestacions semblants (Netmechanics és un d'ells).

DOCUMENTS DE BIBLIOGRAFIA RECOMANADA: DIGITALITZACIÓ DELS SUMARIS I RESUMS A LA BIBLIOTECA DE LA UOC

En una altra comunicació d'aquestes Jornades, s'exposa el procediment de préstec als nostres usuaris per mitjà de la interfície de la web del Campus Virtual de la UOC. Com a conseqüència de la implementació d'aquesta aplicació, vam detectar la manca d'informació sobre el contingut de l'obra al demanar-la en préstec i la necessitat de subsanar aquesta mancança.

La impossibilitat de fullejar el contingut dels documents físics des de la interfície, ens va dur a cercar alternatives sobre com guiar a l'usuari en la selecció dels possibles documents a demanar en préstec. La solució ha estat el reconeixement òptic (OCR) i conversió en HTML, del sumaris i del resums o comentari de les obres de la bibliografia recomana per el professorat dels diferents estudis.

Aprofitant l'aplicació per web del catàleg i les possibilitats de l'etiqueta 856 l'usuari té la possibilitat, un cop feta la cerca i abans de demanar el document en préstec, de fullejar el seu contingut en el sumari i en el resum de la mateixa forma que en una biblioteca presencial. L'estructura del registre bibliogràfic segueix el següent esquema:



Esquema 1. Estructura del registre bibliogràfic

A la figura 6, podem comprovar el resultat de la selecció d'un document i les seves possibilitats.

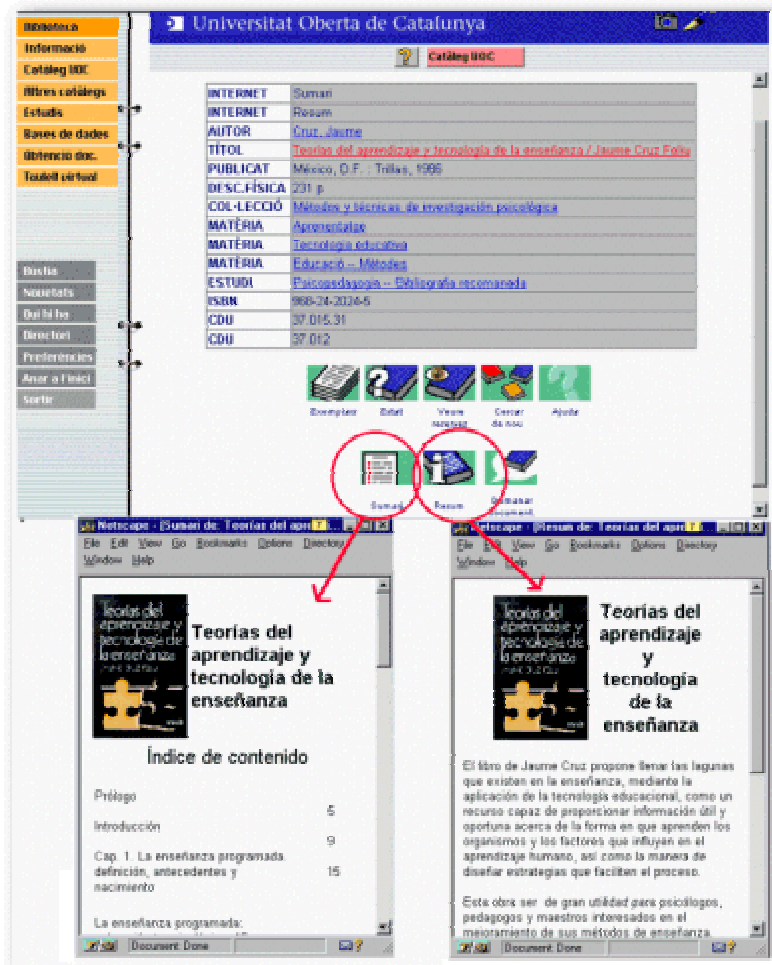


Figura 6. Document de bibliografia recomanada

SUMARIS DE REVISTES

Una situació similar es produïa en les publicacions periòdiques: la dispersió geogràfica dels usuaris, i de les revistes entre les diferents biblioteques dels Centres de Suport, no permetia la facilitat d'accés a les publicacions, a part de no ser econòmicament viable la duplicitat de col·leccions arreu de Catalunya.



Figura 7. Accés al sumari des del registre bibliogràfic



Figura 8. Índex de les publicacions periòdiques

A partir d'aquestes dues possibilitats l'usuari accedeix a la pàgina índex de la publicació que conté la relació dels números de la revista disponibles en ordre invers: del número més recent al més antic (figura 9).

A més d'accedir al contingut dels sumaris dels números rebuts, l'usuari té altres possibilitats:

- Subscriure's a la llista de distribució de la publicació per a rebre els sumaris a la seva bústia de correu electrònic.
- Accedir al catàleg per saber a quines biblioteques de la UOC pot trobar la revista.
- Consultar el servidor de Web de la publicació (si en té).

- Accedir als índexs acumulatius i/o fóruns de discussió de la publicació si aquests es troben disponibles a la web de la publicació.



Figura 9. Diferents opcions a partir de l'índex dels sumaris

Un cop el sumari està elaborat, s'actualitzen els índexs, s'envia als usuaris subscrits a la llista de la publicació i en el mateix sumari s'ofereix la possibilitat de demanar còpia dels articles en els que es pugui estar interessat per mitjà d'un formulari del Servei d'Obtenció de Documents (figura10),

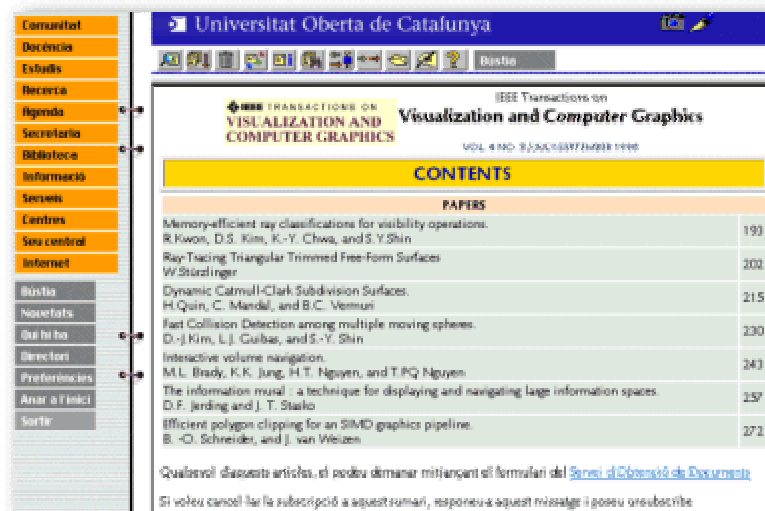


Figura 10. Exemple de sumari

PROCÉS DE CREACIÓ DE LES PÀGINES DE SUMARIS

El procés de creació d'aquestes pàgines de sumaris, en una primera etapa tant sols per als sumaris, constava de dos formats: HTML per els sumaris recuperats des del servidor de web de la publicació, reconeixement òptic, i una imatge escanejada en format PDF per aquells sumaris que per la seva presentació i/o disposició del text dificultaven el reconeixement òptic.

Un cop posat en marxa el procés de creació i traspàs dels sumaris al servidor, vam observar quatre inconvenients respecte al format PDF que afectava als nostres usuaris i a un desenvolupament futur d'aquest servei:

- La necessitat de cridar una segona aplicació per visualitzar-los (ACROBAT) i en molts casos desorientació o sorpresa per part de l'usuari.
- La diversitat d'equipaments informàtics (486 o pentiums a diferents velocitats) i especialment la diversitat de formes de connexió al Campus Virtual: des de la Xarxa de telefonia bàsica amb mòdems de 14.400 bauds a 33.600 bauds, passant per XDSI o Infovia. Per tant la velocitat i facilitat de recuperar els sumaris en format PDF afectava a una part dels nostres usuaris.
- El tamany dels fitxers va ser un altre aspecte: mentre els fitxers PDF tenien un tamany mig de 379,4 Kb., la mitjana dels sumaris en HTML és de 4,88 Kb.
- Finalment, aquesta duplicitat de formats, afectava a posteriors desenvolupaments d'aquest servei: per una banda el tamany dels PDF afectava a la distribució de sumaris per correu electrònic i també afectaria al nou projecte que tenim previst posar en funcionament el proper mes de gener: la creació d'un motor de cerca que faciliti l'accés al contingut d'aquests sumaris. Els fitxers gràfics en format PDF impedirien la cerca en el seus continguts i per tant obtindriem uns resultats parcials en les cerques ens els continguts de les nostres revistes .

Per tant, vam procedir a reconvertir els fitxers PDF a HTML de forma manual donada la seva complexitat.

Actualment, els procés de creació d'aquest sumaris en HTML, té diferents orígens:

- Reconeixement òptic i conversió a HTML dels sumaris (revistes i bibliografia recomanada)
- Sumaris recuperats des de Internet a partir dels servidors de Web de les publicacions.
- Sumaris transcrits manualment.

En la majoria dels casos l'origen del sumari d'una publicació periòdica, és per reconeixement òptic en un 55,2% dels sumaris (no dels títols), els recuperats del propis servidors de la publicació el 31,5%, i els sumaris de les publicacions periòdiques més conflictives i transcrits manualment representen un 13,3%.

Els costos humans d'aquest procés, en un principi semblarien elevats, però hem desenvolupat un conjunt de procediments que simplifiquen la tasca. El valor temps, en una mostra dels sumaris processats durant la darrera setmana de juny i ajustant-nos a la distribució percentual del seu origen ha estat:

- 55,2% dels sumaris processats per reconeixement òptic: 2,5 minuts per sumari.
- 31,5% dels sumaris recuperats del servidor (connectar, cercar i gravar): 1,5 minuts per sumari.
- 13,3% dels sumaris processats manualment amb l'editor d'HTML: 8 minuts per sumari.

Per a la seva presentació definitiva i donar uniformitat a tots els sumaris, hem creat una sèrie de plantilles a l'editor d'HTML que utilitzem i de macros per a Word amb l'Internet Assistant, on de forma molt senzilla l'elaboració de la pàgina HTML definitiva ens dona aquests temps:

- Per a sumaris processats per reconeixement òptic: 3,5 minuts
- Per a sumaris recuperats del servidor: 2 minut

- Processats manualment: ja inclòs en l'apartat anterior.

Finalment, l'actualització de la pàgina índex de la publicació representaria 0,5 minuts, i la distribució per mitjà de la llista als usuaris subscrits, el traspàs de les dades i comprovació final de que tot és correcte (enllaços o "links") 1,5 minuts. Per tant, aquesta mostra ens donaria els següents valors de temps:

- Sumari per reconeixement òptic: 8 minuts.
- Sumari recuperats del propi servidor: 5,5 minuts.
- Sumari processats manualment: 10 minuts.

Aquest valors ponderats amb el percentatge que representen en el nostre servidor donaria una mitjana aproximada de procés d'uns 7,5 minuts per sumari.

FUTURS PROJECTES EN LA DIGITALITZACIÓ DE SUMARIS

Durant aquest curs 1997-98, tenim previst tres grans camps d'actuació:

- Instal·lació d'un motor de cerca en el contingut dels sumaris que posteriorment s'aplicaria als sumaris de la bibliografia recomanada, a la biblioteca electrònica i a la web de la biblioteca en general que permeti cerques globals o parcials per a cadascuna d'aquestes opcions.
- Distribució electrònica dels articles que ens demanin per mitjà del formulari del Servei d'Obtenció de Documents (ARIEL o aplicació similar)
- Escurçar els temps de processament dels sumaris de revistes cercant alternatives del tipus distribució de sumaris com EBSCO o altres serveis com l'UNCOVER. Malgrat tot les publicacions nacionals requeriran en la majoria dels casos el mateix tractament que fins ara.

BIBLIOGRAFIA

Cataloging Internet resources: a manual and practical guide, Nancy B. Olson, ed. ([Dublin, Ohio: OCLC], 1995).

Estivill, A. "Organització dels recursos Internet", *Item*, núm. 18 (1996), p. 42-74.