Més informació sobre el CCUC

ESPINÓS, MONTSERRAT; MESTRE, MERCÈ; TORT, MARTA. **"El Catàleg Col·lectiu de les Universitats de Catalunya: control de qualitat i detecció de duplicats en la creació d'una base de dades bibliogràfica"**. En: *Jornades Catalanes de Documentació (5es: 1995: Barcelona). Biblioteques, centres de documentació i serveis d'informació*. [Barcelona : Societat Catalana de Documentació i Informació: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya. 1995. p. 507-517

La ponència fa una introducció sobre la teoria, els objectius i els problemes de la detecció i l'eliminació de registres duplicats en bases de dades bibliogràfiques. A continuació, s'explica el procés de creació del Catàleg Col.lectiu de les Universitats de Catalunya (CCUC) amb una breu introducció històrica. Finalment s'analitza i es descriu el procés d'elaboració i el funcionament del programa de detecció de duplicats del CCUC.

1. Introducció

Cada vegada més, els serveis bibliotecaris es basen en la cooperació per rendibilitzar recursos. Per això es creen les xarxes de biblioteques. Només l'augment progressiu de les relacions de cooperació interbibliotecària permetrà assolir l'objectiu final de qualsevol biblioteca: oferir tota la informació bibliogràfica disponible i proporcionar qualsevol document que se li demani.

La informàtica i les telecomunicacions s'usen per posar en comú recursos bibliotecaris locals que, transformats en eines col.lectives, milloren notablement els serveis bibliotecaris. Els països més avançats en l'àmbit bibliotecari han interrelacionat des de fa temps les seves biblioteques en xarxes cooperatives i consorcis que, per mitjà de serveis i instruments comuns, ofereixen més i millors prestacions a la comunitat que serveixen. Aquests projectes cooperatius es basen, sovint, en catàlegs col.lectius consultables des d'un ordinador central.

2. Teoria del control de duplicats

Els catàlegs col.lectius. La duplicació de registres bibliogràfics

Els catàlegs col.lectius, per assolir els seus objectius principals -consulta àgil de les col.leccions bibliogràfiques, millora en la gestió del préstec interbibliotecari, i facilitats en la catalogació per còpia- han de complir un objectiu de qualitat important: presentar un registre únic per a cada exemplar bibliogràfic referenciat.

Fins al moment de la irrupció dels catàlegs col.lectius automatitzats, no s'havia parat esment a definir exactament què es consideren registres bibliogràfics duplicats. El plantejament i la detecció de registres duplicats ha esdevingut un problema en el moment que les biblioteques han connectat els seus catàlegs en xarxes automatitzades1. I, sobretot, amb l'aparició dels catàlegs col×lectius, creats sovint a partir de la càrrega de diferents catàlegs locals, i on la catalogació posterior és compartida. Les pràctiques catalogràfiques de cada institució participant poden ser molt diferents, així com les normes utilitzades. Això fa que un mateix exemplar bibliogràfic pugui aparèixer referenciat en registres bibliogràfics diferents.

Una altra causa de duplicació és la dificultat de distingir reimpressions i edicions d'un mateix títol. Per això és molt important definir clarament què es considera duplicat. Per exemple, és freqüent tractar les reimpressions com a duplicats, i en canvi, les edicions paperback, que normalment són reimpressions de les hardback, com a registres bibliogràfics diferents.

Els duplicats incrementen innecessàriament la dimensió de la base de dades, dificulten i

fan perdre temps al personal bibliotecari en el procés d'adquisició i catalogació i, sobretot, confonen l'usuari a l'hora de decidir quin dels registres que li apareixen després d'una recerca és el que necessita2.

La detecció de registres duplicats. Planificació del programa de detecció de duplicats

Els registres duplicats idèntics és fàcil identificar-los. Un procediment que compari dues descripcions bibliogràfiques byte per byte i digui SÍ quan tots els bytes siguin iguals i NO quan hi hagi diferències, pot ser senzill de fer. Però això no soluciona el gran problema dels registres duplicats en una base de dades. Són els registres duplicats similars, i no els idèntics, els que són difícils de detectar. I el programa que s'elabori per detectar-los ha d'eliminar els duplicats sense eliminar registres únics.

Stephen R. Toney recomana que a l'hora de planificar un programa de detecció de duplicats es tinguin en compte els punts següents:

1. Tenir ben clar l'objectiu del projecte
2. Decidir qui determina quins registres són duplicats: el mateix programa informàtic que els detecti o bé el personal catalogador
3. Definir perfectament el procés de l'algoritme que ha d'utilitzar el programa per comparar els registres
4. Seleccionar acuradament els camps bibliogràfics que es faran servir per a la comparació de registres
5. Establir les claus de comparació
6. Determinar què es considera registre duplicat
7. Decidir què es fa una vegada identificats els registres duplicats
8. Fer proves del programa fins a assolir els resultats més satisfactoris

Per seleccionar els camps bibliogràfics a comparar en la detecció de duplicats, Toney proposa seguir els criteris següents:

1. Freqüència: No comparar mai camps que només apareguin en alguns registres
2. Consistència: Si les dades en el camp són inconsistents, seran menys fiables per a la comparació (per exemple, els camps de notes)
3. Simplicitat de programació i processament: Els camps curts i els camps numèrics són preferibles en la comparació

Sobre la decisió de què es considera un registre duplicat, sempre s'ha de tenir en compte des de quin punt de vista es defineix: des del punt de vista catalogràfic, a l'hora de crear un nou registre, o des del punt de vista informàtic, quan es tracta de confeccionar un programa que detecti duplicats. Per exemple, a l'OCLC Online Union Catalog, quan un catalogador vol introduir un registre nou, ha de comparar els següents elements per decidir si el registre ja existeix:

- Autor, títol, editor, data del copyright, data d'impressió, edició, format, col×lecció, paginació o nombre de volums, il.lustrador i traductor.

Si aquests elements no coincideixen, el catalogador ha de crear un registre bibliogràfic nou. En canvi, els programes informàtics de detecció de duplicats no poden comparar tants elements exhaustivament. És més normal que en comparin només una part (les primeres i últimes paraules del títol, uns caràcters de l'autor, etc.).

La qüestió és decidir entre utilitzar pocs elements de comparació i posteriorment realitzar una revisió manual, o utilitzar-ne més i evitar aquesta revisió. L'ús dels dos sistemes alhora és la metodologia que han seguit molts catàlegs col.lectius, perquè els resultats que s'obtenen assoleixen un nivell de qualitat elevat en relació amb els recursos emprats.

El següent quadre mostra els elements bibliogràfics que comparen alguns programes de detecció de duplicats existents:

	MELVYL	OCLC	RLIN	WLN	LINDA	CCUC
Tipus registre				*	*	
Nivell bibliogràfic		*		*	*	*
Codi reprod.	*	*				
Desig. mater.						*
LCCN	*	*	*	*		
ISBN					*	*
ISSN					*	*
País publicac.	*			*		
Autor	*	*		*	*	*
Títol	*	*	*	*	*	*
Núm. volum.					*	*
Edició	*		*			
Lloc publicac.		*	*			
Editor	*	*	*			
Data	*	*	*	*	*	*
Llengua					*	

- MLVYL (DLA): Catàleg Col.lectiu de la universitat de Califòrnia.
- OCLC: Online Computer Library Center.
- RLIN: Research Libraries Information Network.
- WLN: Washington Library Network.
- CCUC: Catàleg Col.lectiu de les Universitats de Catalunya.
- LINDA: Catàleg Col.lectiu de les biblioteques universitàries i científiques finlandeses.

Respecte a com s'ha d'actuar quan es detecten registres duplicats, existeixen tres possibilitats diferents:

1. Es tria un dels registres com a definitiu i els altres s'esborren

2. Es capturen tots els registres i s'agrupen en un registre únic (sense perdre informació bibliogràfica de cap d'ells)

3. Es tria un registre com a definitiu i s'hi afegeixen uns camps variants, prèviament definits, dels altres registres.

Les tres tècniques es consideren vàlides, segons la casuística de cada base de dades, però és molt important haver fet prèviament proves amb mostres de registres. El mètode s'ha d'anar modificant segons els resultats de cada prova fins que s'aconsegueixi un nivell d'èxit acceptable.

Factors que contribueixen a la duplicació de registres bibliogràfics

O'Neill, Rogers i Oskins[4], en un estudi sobre les característiques dels registres duplicats de l'OCLC, determinen els següents factors com a contribuents a la duplicació:

- Errors tipogràfics, camps i subcamps erronis, omissió d'informació, inconsistències entre camps fixos i variables.

Els errors tipogràfics, encara que no siguin molt freqüents, si apareixen en alguns elements crítics com en l'autor i el títol o el número de control, dificulten enormement la detecció de duplicats. I, tot i que aquests errors es poden donar en tots els camps, es troben més sovint en els camps textuals més llargs com són justament l'autor i el títol.

Els camps i subcamps erronis o intercanviats suposen un altre tipus d'error comú que dificulta la detecció de duplicats.

L'omissió d'informació és un altre punt important. Per exemple, si un registre té una menció de col.lecció i un altre que sembla igual no la té, es tendeix a considerar-los diferents, i, en realitat, la major part de les vegades es tracta de registres duplicats.

Finalment, la inconsistència que apareix sovint entre els camps fixos i els variables en els formats MARC és un altre factor important a tenir en compte en la detecció de duplicats. Per exemple, entre dos registres possibles duplicats pot diferir la informació donada en el camp fix comparat i en canvi, coincidir la informació corresponent donada en el camp variable.

Tots aquests factors, juntament amb les diferents pràctiques catalogràfiques utilitzades en els catàlegs locals i la convivència de diferents normatives catalogràfiques en aquests mateixos catàlegs, fan que la tasca de detecció de duplicats sigui molt complexa, i gairebé impossible elaborar un programa que solucioni el cent per cent dels casos de possible duplicació. Per causa de la situació bibliotecària catalana, aquests últims aspectes citats han tingut una especial incidència en el procés de creació del Catàleg Col.lectiu de les Universitats de Catalunya i en la redacció del seu programa de control de duplicats.

3. El Catàleg Col.lectiu de les Universitats de Catalunya

El projecte de creació del Catàleg Col.lectiu de les Universitats de Catalunya s'inscriu dins del projecte més ampli de creació d'una xarxa de biblioteques de les universitats de Catalunya. Les institucions participants són, ara per ara, les set universitats públiques de Catalunya: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat Rovira i Virgili i Universitat de Lleida, amb el patrocini del Comissionat per a Universitats i Recerca de la Generalitat de Catalunya. L'ordinador central s'ha ubicat a la Universitat Politècnica de Catalunya, que és la institució que lidera el projecte.

La primera fase del projecte, iniciada l'any 1993, consistia en la interconnexió dels

catàlegs de les biblioteques d'aquestes set universitats, que comparteixen el mateix sistema d'automatització5. La segona fase suposava la creació d'un catàleg col.lectiu de totes les universitats participants, com a instrument de consulta bibliogràfica, catalogació compartida i préstec interbibliotecari molt més potent que l'ús per separat dels catàlegs locals. El catàleg col.lectiu ha de ser l'instrument central i indispensable d'aquesta xarxa, però el projecte contempla anar més enllà incorporant altres serveis cooperatius, com poden ser l'accés compartit a bases de dades en CD-ROM o remotes, la instauració de nous serveis d'accés a fonts d'informació electrònica, etc.

La creació d'un catàleg col.lectiu, i la posterior consecució de tots els seus objectius, ha de comportar un doble benefici per a la comunitat universitària catalana: d'una banda, el benefici redunda a favor dels docents, discents i investigadors, ja que s'augmenten els recursos bibliotecaris disponibles de forma immediata i es disminueixen i rendibilitzen els esforços per localitzar i obtenir materials bibliogràfics. De l'altra banda, el benefici millora l'estructura bibliotecària, permetent-li estalviar costos econòmics i recursos humans, i disminuint la seva dependència d'organitzacions i de serveis estrangers.

Procés de creació del Catàleg Col.lectiu de les Universitats de Catalunya

La creació i la posada en marxa d'un catàleg col.lectiu requereixen tasques de planificació, coordinació i establiment de directrius generals, a més de la presa de decisions en diferents àmbits. En funció d'això, el mètode de treball que es va dissenyar va ser la constitució de diversos grups, i en alguns casos, subgrups, de treball, integrats per persones amb diferents nivells de responsabilitats i d'atribucions decisòries. D'acord amb els temes a tractar, es van formar tres grups de treball: el dels directors de les biblioteques participants, el dels responsables informàtics, i el dels tècnics bibliotecaris. Aquest darrer va rebre l'encàrrec d'elaborar les propostes tècniques. Un dels seus àmbits de treball es va centrar en l'elaboració d'un programa de detecció de registres bibliogràfics duplicats que assegurés la qualitat de la informació bibliogràfica en el futur catàleg col.lectiu6.

Els informes tècnics

La primera feina que va emprendre el grup de tècnics va ser l'elaboració de cinc informes. Aquests informes representaven el punt de partida imprescindible per a l'obtenció d'uns mínims acords tècnics respecte al funcionament del catàleg col.lectiu. Evidentment, amb posterioritat, molts altres acords tècnics han estat presos i molts altres temes consensuats. La redacció d'aquests informes, a més de la seva utilitat primera, va ajudar també que els tècnics participants coneguessin mutúament més a fons les particularitats pròpies dels catàlegs de totes les institucions.

Els informes, que tenen plantejaments i valor diferents -els uns eren estrictament descriptius mentre que altres eren propostes concretes d'actuació amb la inclusió de diferents alternatives-, són els següents:

1. Nivells de codificació bibliogràfica
   S'hi descrivien, i homogeneïtzaven, els valors en el nivell de codificació bibliogràfica que tindrien en el catàleg col.lectiu els diferents tipus de registres bibliogràfics aportats per cada institució, segons la qualitat i la completesa de la catalogació, tant en la càrrega inicial com en l'alimentació posterior del catàleg.

2. Informació sobre les col.leccions locals en el catàleg col.lectiu
   En aquest informe es plantejaven diverses opcions respecte a la inclusió de les dades locals (localització del document, disponibilitat, etc.) en el catàleg col.lectiu, tant pel que fa a les monografies com a les publicacions en sèrie.

3. Etiquetes a considerar en l'eliminació de registres bibliogràfics per detecció de duplicats en la càrrega inicial del catàleg col.lectiu
   En la primera part es feia una relació, institució per institució, de les etiquetes utilitzades per a la classificació i la indexació, amb indicació del seu contingut. Una

segona part contenia les etiquetes amb usos particulars que no havien de ser transferides al catàleg col.lectiu. Finalment, una tercera part feia una proposta sobre quines etiquetes conservar en casos d'eliminació de registres bibliogràfics per detecció de duplicats.

4. Tractament catalogràfic de les obres en volums
   Aquest informe contenia una exposició, institució per institució, de les diferents polítiques catalogràfiques utilitzades en el tractament tècnic de les obres en volums. En recollia l'estat de la qüestió, i constituïa una eina de treball per a futures decisions al respecte.

5. Registres i fons especials
   L'informe recollia el tractament, també institució per institució, dels registres bibliogràfics amb particularitats especials (registres de part component, registres CIP, etc.) o bé qüestions relatives a la propietat i disponibilitat dels fons associats a registres bibliogràfics. Finalment, es feia una proposta d'inclusió dels registres en el catàleg col.lectiu.

Aquests informes tècnics van servir de base per a la tasca del control de registres duplicats, mitjançant l'elaboració de les especificacions del programa.

Les bases de dades de prova

La confecció del programa va tenir una part teòrica, on s'establien quins elements es comparaven durant el procés i s'elaboraven les fórmules de comparació; i una part experimental, on es posava a prova. En aquest sentit, es va crear una base de dades de prova per veure el funcionament del programa i adaptar-lo a les necessitats de la base de dades real. Segons els resultats obtinguts, es modificava el programa i es comprovaven un altre cop els resultats que produïa sobre els registres, en un procés d'ajustament fins a trobar la fórmula més adequada. Aquesta base de dades de prova estava formada per l'aportació de 424 registres originals procedents dels catàlegs de quatre de les institucions participants, seleccionats expressament uns com a duplicats i altres com a conflictius, amb l'objecte de comprovar que els registres realment duplicats es fusionaven correctament, i en canvi, no es fusionaven els que no ho eren. Aquesta base de dades s'anava reprocessant a mesura que es modificaven les especificacions del programa i s'estudiaven els resultats.

Per tal d'assegurar el funcionament del programa, es va crear una segona base de dades amb 23.000 registres procedents dels catàlegs de les set universitats participants en el projecte i triats aleatòriament (entre els registres més antics, els del mig i els més nous de cada catàleg).

4. El control de duplicats en el Catàleg Col.lectiu de les Universitats de Catalunya

El programa per al control de duplicats del CCUC va ser elaborat amb l'objectiu de detectar i d'eliminar el major nombre possible de registres bibliogràfics duplicats, tenint en compte la gran diversitat de registres existents a les diferents bases de dades participants, per tal d'assolir el més alt nivell de qualitat.

L'algoritme per al control de duplicats

El control de duplicats és una part del procés de la càrrega inicial de registres en el CCUC. Un programa de càrrega ha de fer moltes tasques i tan ràpidament com sigui possible, i la comparança de registres és un d'aquests processos.

El programa de control de duplicats consta de dues fases diferenciades: una primera fase, anomenada Explode, on es barregen tots els registres de forma simultània per trobar-hi coincidències partint de la comparació de poques dades, i una segona fase, anomenada Control Complet de Duplicats, on s'examinen els registres amb més detall.

Aquest programa de detecció i eliminació de duplicats funciona mitjançant un algoritme, que és aquell procediment lògic que consisteix a complir un seguit ordenat i finit d'instruccions que condueix, un cop especificades les dades, a la solució del problema segons les dades considerades7.

Prèviament al control de duplicats, tots els registres es preparen convenientment per a la seva incorporació al catàleg col.lectiu: s'eliminen els camps que no es volen carregar, es codifiquen els registres amb els valors de nivell de codificació acordats en els informes tècnics, s'afegeix al registre la informació local en una etiqueta bibliogràfica, etc.

Els registres es carreguen dins el CCUC seguint els dos pasos ja citats: l'Explode i el Control Complet de Duplicats, i posteriorment es porta a terme la resolució dels possibles conflictes que hagin pogut sorgir. Durant la càrrega, si una comparança és positiva, els dos registres es fusionaran. El programa tria automàticament un dels dos registres com a definitiu, d'acord amb els valors de nivell de codificació assignats a cada registre, i hi afegeix les etiquetes de classificació i matèries del registre desestimat, prèviament determinades als informes tècnics. El registre no escollit s'esborra de la base de dades.

Complementàriament al programa de control de duplicats, va ser necessari elaborar una sèrie d'especificacions concretes per a diferents punts:

1. L'elaboració d'un algoritme per a la creació d'un número de control per a cada registre, on s'estableix l'ordre de preferència dels possibles elements usats en la seva creació (ISBN, ISSN, Dipòsit Legal o número identificador del registre bibliogràfic) i com s'ha de confeccionar a partir d'aquestes dades. Aquest número de control s'utilitzarà per a la comparació de registres en la càrrega inicial del CCUC, i serà indispensable, en el futur, per a la detecció de duplicats entre els catàlegs locals i el catàleg col.lectiu

2. La redacció d'unes especificacions per a l'establiment de la data de publicació a comparar, en el cas de les reimpressions

3. La redacció d'unes especificacions per a la càrrega de camps de classificació i matèries en el registre definitiu, en produir-se una fusió de registres.

L'Explode

La primera fase del control de duplicats consisteix a carregar al mateix temps tots els registres, procedents dels catàlegs locals, perquè es puguin comparar tots amb tots. L'objectiu de l'Explode és trobar coincidències entre registres comparant un nombre petit de dades.

L'establiment dels elements a comparar va comportar un procés d'estudi detallat per poder determinar quins són comuns a la majoria dels registres bibliogràfics i si la quantitat de caràcters utilitzada és suficient per detectar el màxim nombre de duplicats i prou fiable per evitar la fusió de registres diferents. Sempre és preferible que dos registres quedin separats quan haurien d'ajuntar-se que no que s'ajuntin quan haurien de quedar separats, ja que aquest segon cas suposaria la pèrdua d'un registre bibliogràfic diferent, i, en conseqüència, una pèrdua d'informació.

Les dades a comparar consisteixen en 48 caràcters procedents de diversos camps del registre bibliogràfic, en combinacions diferents per a monografies i per a publicacions en sèrie. Les dades procedents de camps textuals són normalitzades, és a dir, s'han ignorat els diacrítics, signes de puntuació, espais en blanc, majúscules/minúscules i indicadors de subcamp.

L'Explode actua agrupant, per a cada registres, totes aquestes dades en dues fòrmules de 24 caràcters cada una, que serviran per comparar els registres:

1. Si les dues fórmules són diferents, considera que els registres són únics i els carrega al catàleg

2. Si només coincideix una fórmula, marca els registres com a possibles duplicats per tornar-los a processar passant pel Control Complet

3. Quan les dues fórmules són idèntiques, considera que els registres són duplicats, els fusiona en un de sol i el carrega al catàleg. Els criteris per decidir quin dels dos registres ha de romandre i quin ha de desaparèixer venen determinats pels valors de nivell de codificació assignats prèviament als registres. A més, s'han de tenir en compte les especificacions de càrrega d'etiquetes de classificació i matèries.

El Control Complet de Duplicats (CCD)

El CCD compara els registres no solucionats a l'Explode (possibles duplicats) amb els ja carregats. És un procés més lent i més complex (l'Explode és cinc vegades més ràpid), ja que s'hi comparen més dades. La finalitat del CCD és detectar i eliminar el màxim nombre de registres duplicats, i deixar per a una posterior revisió manual aquells casos on és impossible determinar-ho.

Els elements per comparar en el CCD s'obtenen dels mateixos camps que els de l'Explode, però s'agafen molts més caràcters de cada un d'aquests camps i es comparen de forma diferent. El CCD està dissenyat per aplicar més vies de comparació entre els registres, abans de decidir si un registre és únic, duplicat, o possible duplicat. Quan tots els elements comparats coincideixen, el darrer element a comparar, i que farà decidir si dos registres són duplicats o possibles duplicats, és l'autor principal. Si els caràcters comparats de l'autor principal coincideixen exactament els registres es consideraran duplicats i es fusionaran en un de sol, tenint en compte les mateixes especificacions de valors de nivell de codificació i de càrrega d'etiquetes de classificació i matèries que a l'Explode.

Si els caràcters comparats de l'autor no coincideixen, es marca el registre per a una posterior revisió manual. En aquest cas, el programa permet visualitzar tres registres: el registre ja existent, el registre possible duplicat i el registre que resultaria de la fusió de tots dos. El catalogador haurà de decidir si els registres són duplicats. Si ho són, processarà el registre resultant de la fusió dels dos, i automàticament s'esborraran els dos registres individuals. Si no ho són, processarà el registre possible duplicat i automàticament s'esborrarà el registre mixt.

Resultats

Els resultats que es donen en aquesta ponència són els obtinguts en la base de dades de prova mencionada anteriorment, ja que en el moment de redacció del text (juny-juliol de 1995) encara no havien finalitzat els processos de càrrega de registres i control de duplicats en el CCUC.

Els resultats obtinguts van mostrar que l'algoritme definitiu resol un 83% dels duplicats existents, entenent per duplicats aquells registres iguals i tractats amb el mateix criteri catalogràfic. Aconseguir el 100% és impossible degut al marge d'error produït per omissions d'informació i errades tipogràfiques. D'altra banda, no s'ha detectat cap cas on dos registres s'haguessin fusionat erròniament.

5. Conclusions

- Una gran quantitat de registres duplicats dificulta la recerca i incrementa el nombre de registres que l'usuari ha de comparar i inspeccionar abans de trobar el que busca.

- La reducció del nombre de registres duplicats incrementa la productivitat de

catalogació i la satisfacció de l'usuari final, especialment en la consulta i en el préstec interbibliotecari.

- En una base de dades elaborada a partir de múltiples catàlegs acostumen a haver-hi inconsistències. Tota base de dades, per inconsistent que sigui, pot ser millorada amb un procediment de control de registres duplicats.

- El programa informàtic que s'elabori detectarà i/o corregirà un nombre determinat de duplicats. Cal ser conscient que, per causa de les inconsistències esmentades, el programa no solucionarà tots els casos de possibles duplicats. El que cal assegurar és que funcioni a ple rendiment, detectant el major nombre de duplicats sense córrer el risc de perdre informació.

- Per aquest motiu, és imprescindible la intervenció del catalogador per resoldre els possibles conflictes.

- Cal trobar l'equilibri entre el temps i els esforços dedicats a l'elaboració del programa i els resultats que s'esperen obtenir.

- Un cop acabada la fase de detecció de duplicats, cal garantir el manteniment de la qualitat del catàleg mitjançant l'establiment de pautes catalogràfiques comunes a totes les institucions participants.

Bibliografia

Ayres, F. H. "Duplicates and other manifestations: a new approach to the presentation of bibliographic information". Journal of Librarianship. Vol. 22 (1990), no. 4, p. 236-251.

Chapman, Ann. Quality of bibliographic records in a shared cataloguing database: a case study using the BLCMP database. Bath: Centre for Bibliographic Management, University of Bath, 1993. 34 p.

Coyle, Karen; Gallaher-Brown, Linda. "Record matching: an expert algorithm". ASIS Proceedings. Vol. 22 (1985), p. 77-80.
--------."Record matching: a discussion". Information Technology and Libraries. Vol. 4 (1985), no. 1, p. 57-59.

Hickey, Tomas B.; Rypka, David J. "Automatic detection of duplicate monograph Records". Library Automation. Vol. 12 (1979), no. 2, p. 125-141.

Lazinger, Susan S. "To merge and not to merge: Israel's Union List of Monographs in the context of merging algorithms". Information Technology and Libraries. Sept. 1994, p. 213-219.

O'Neill, Edward; Rogers, Sally A.; Oskins, W. Michael. "Characteristics of duplicate records in OCLC's Online Union Catalog". Library Resources & Technical Services. Vol. 37, (1993), no. 1, p. 59-71.

Ridley, M. J. "An expert system for quality control and duplicate detection in bibliographic databases". Program. Vol. 26 (1992), no. 1, p. 1-18.

Sonkkila, Tuija. "Forms of cooperation in the library information network of Finnish academic libraries". IATUL Proceedings. Vol. 3 (1994), p. 168-176.

Toney, Stephen R. "Cleanup and deduplication of an International Bibliographic Database". Information Technology and Libraries. Vol. 11 (Mar. 1992), p. 21-23.

Williams, Martha E.; MacLaury, Keith D. "Automatic merging of monographic data bases. Identification of duplicate records in multiple files: the IUCS scheme". Journal of Library Automation. Vol. 12 (1979), no. 2, p. 156-168.