

Mineria de dades en biblioteques: la bibliomineria

[[Versión castellana](#)]

JORGE CANDÁS ROMERO

Alumne del màster en Investigación en Documentación
de la Universidad Carlos III de Madrid.

jorgecandas@gmail.com

Opciones

 [Imprimir](#)  [Recomanar](#)  [Citació](#)  [Estadístiques](#)  [Metadades](#)

Resum [[Abstract](#)] [[Resumen](#)]

Aquest article presenta una introducció teòrica a l'aplicació de la mineria de dades en biblioteques, anomenada bibliomineria (proposta terminològica en català per a l'anglès *bibliomining*). Així mateix, s'hi presenten algunes de les possibles aplicacions pràctiques i com aquestes serveixen de suport a l'anomenada Biblioteca 2.0 i a la creació i gestió de serveis més orientats cap a l'usuari i millors, basats en noves tecnologies. Finalment s'hi analitza el problema de la privadesa en l'aplicació de la bibliomineria.

*I never waste memory on things that can easily
be stored and retrieved from elsewhere.*

Albert Einstein (1879-1955)¹

1 Minería de dades: conceptualització i utilitat

La mineria de dades és una activitat que està a mig camí entre la informàtica, l'estadística i la documentació, i que s'ha estat utilitzant en nombroses disciplines per a l'anàlisi de grans quantitats de dades. En biblioteques, la seva pràctica està poc estesa a pesar de les múltiples aplicacions que té.

L'objectiu d'aquest article és fer una introducció teòrica a la mineria de dades per aprofundir, tot seguit, en l'aplicació en biblioteques, i presentar aquestes com un camp de treball ideal. D'aquesta aplicació, denominada en anglès *bibliomining* (terme per a la traducció del qual al català es proposa emprar "bibliomineria"), se n'expliquen breument alguns dels usos pràctics, que van des de l'avaluació de biblioteques, la planificació d'adquisicions i l'organització de la col·lecció fins als estudis d'usuaris. Una altra de les aplicacions de major interès és l'elaboració de sistemes de recomanacions tal com s'estan aplicant en serveis web, i que entra en el camp de l'anomenada Biblioteca 2.0 (en anglès *Library 2.0*).

La bibliomineria, igual que qualsevol activitat de mineria de dades, comporta certs problemes respecte a la privadesa. Aquí es presenten algunes de les possibles mesures sobre això que es poden prendre.

1.1 Conceptualització

El gran creixement de les bases de dades i l'augment de les capacitats d'emmagatzemament d'informació han fet que tot tipus d'organitzacions puguin disposar d'una gran quantitat i varietat de dades relatives a la seva activitat diària. En moltes d'aquestes organitzacions s'han adonat del potencial que aquesta informació té per al suport a la gestió. El seu estudi permet adonar-se de l'evolució i el desenvolupament de les organitzacions i, per tant, traçar una línia de tendència que mostri per on poden anar els trets en el futur.

Així, l'estudi de les dades i la informació emmagatzemades en les bases de dades ofereix una visió perspectiva (què s'està fent i com s'està fent) i prospectiva (com pot evolucionar l'organització en el futur, a curt i a mitjà terminis) de l'organització, i és per això que té una funció de suport a la presa de decisions.

Aquest estudi de les bases de dades ha estat, i és, molt més habitual en organitzacions econòmiques i empresarials, des de supermercats fins a grans multinacionals, però també organismes científics que manegen grans quantitats d'informació han vist la utilitat d'aquest tipus d'estudis.

Pel fet que la informació de les bases de dades està més desestructurada (seguint el model DIKW)² que en altres fonts d'informació, per utilitzar-la és necessari un procés de tractament i anàlisi exhaustius. Aquestes tasques d'anàlisi de la informació de les bases de dades es denominen mineria de dades (en anglès *data mining*) o descobriment d'informació en bases de dades (en anglès *knowledge discovery in databases* o KDD)³. Sota aquesta perspectiva, Hand, Mannila i Smyth defineixen la mineria de dades com "la ciència d'extracció d'informació útil de grans conjunts de dades o de bases de dades" (Hand et al., 2001). Per alguns autors mineria de dades i KDD es refereixen al mateix, mentre que per altres KDD es refereix a un procés de diverses fases, entre les quals hi ha la mineria de dades (Hernández et al., 2004).

Tal com assenyalen Fayyad, Piatetsky-Shapiro i Smyth, la gran quantitat de dades que s'emmagatzemen en les organitzacions fa impossible la utilització de mètodes manuals per fer-ne l'anàlisi (Fayyad et al., 1996). Per això

són necessàries tècniques i eines informàtiques capaces d'ajudar l'ésser humà de manera "intel·ligent" i "automàtica" en l'anàlisi de grans quantitats de dades.

Aquests mateixos autors aporten la definició de mineria més esmentada en les obres sobre el tema: "mineria de dades és un procés no trivial d'identificació de patrons de dades vàlides, nous, potencialment usables i comprensibles". És a dir, es tracta d'un procés concret, específic, amb un objectiu que busca identificar repeticions i tendències en un conjunt de dades que resultin útils i siguin veraces. Per Hernández Orallo, Ramírez Quintana i Ferri Ramírez (Hernández et al., 2004), en aquesta definició es resumeixen les propietats del coneixement extret: vàlid, nou, potencialment útil i comprensible.

Aquest procés de mineria de dades, Fayyad, Piatetsky-Shapiro i Smyth (Fayyad et al., 1996) el descomponen en nou fases lineals i iteratives; de la mateixa manera, Kudyba i Hoptruff ho fan des del punt de vista empresarial (Kudyba; Hoptruff, 2001). Aquestes fases, encara que amb una denominació diferent, consisteixen en tres operacions: definició de la informació que es pretén obtenir, arplega i processament de dades, anàlisi dels resultats, i obtenció i consolidació de nova informació.

En el camp de les biblioteques, Guenther, basant-se en Brethenoux i Strange, simplifica aquestes fases i les resumeix en tres (Guenther, 2000):

- Selecció i adquisició de dades: definir quines dades es volen recopilar i quin mètode o procediment s'usarà per fer-ho.
- Preparació i processament de les dades.
- Interpretació i integració dels resultats.

En qualsevol de tots aquests passos, i en qualsevol de les propostes consultades, el procés fonamental és el de la selecció de les dades. Si el conjunt de dades (o bases de dades) que s'han d'utilitzar no és l'adequat per al tipus d'anàlisi que es pretén dur a terme, l'èxit del procés de mineria de dades resultarà frustrat.

El procés de la mineria de dades és similar al procés d'elaboració d'una estadística, que per a Montaña Lacambra, 2005) passa també per tres fases: la recopilació de les dades, el seu processament i la difusió del resultat final; si bé el producte acabat de les estadístiques és una de les matèries primeres de la mineria de dades.

El primer impuls important per al desenvolupament d'aquesta disciplina es va donar en *The First International Conference on Knowledge Discovery and Data Mining (KDD-95)* en la *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, coordinada per Usama Fayyad i Ramasamy Uthurusamy.

Llavors, el mateix Fayyad, juntament amb Simoudis, definien la mineria de dades (encara que predominava l'ús de l'expressió *knowledge discovery in databases*) com a "camp emergent que combina tècniques d'aprenentatge-màquina, reconeixement de patrons, estadística, bases de dades i visualització per extreure automàticament conceptes, conceptes interrelacionats i patrons d'interès des de grans bases de dades. La seva tasca principal és l'extracció de coneixement (o informació) de dades de baix nivell (bases de dades)" (Fayyad; Simoudis, 1995).

Així, doncs, la mineria de dades és un camp inicialment relacionat amb la intel·ligència artificial i l'estadística, en el qual la documentació i altres ciències de la informació han participat posteriorment, des del punt de vista de l'administració i gestió (*management*) de centres d'informació, i com a part de les tasques de gestió de la informació i del coneixement. D'aquesta manera, la mineria de dades representa una reivindicació del valor de les fonts de dades estadístiques per a la gestió.

1.2 Utilitat de la mineria de dades

Actualment, la mineria de dades és una activitat en expansió aplicada cada vegada en més disciplines, que han vist la utilitat de l'estudi de dades per donar suport a la presa de decisions. Està tenint un èmfasi especial en les àrees relacionades amb l'economia, mitjançant l'econometria, i en el camp d'intel·ligència dels negocis.

El cas més tractat (i pres com a referència per explicar diferents tipus d'algoritmes de mineria de dades) és el dels supermercats (Agrawal; Imielinski; Swami, 1993) (Brin; Motwani; Ullman; Tsur, 1997) (Chi-Wing Wong; Wai-Chee Fu; Wang, 2005) (Hernández et al., 2004), ja que conèixer el perfil de compra dels clients pot ajudar a organitzar l'estoc. D'una manera similar, encara que més exhaustiva, és la seva utilització en el comerç electrònic (Kohavi, 2001) (Kohavi; Provost, 2001), ja que en aquest cas és possible obtenir molta més informació del comportament global del client, i no solament conèixer les compres. Al seu torn, i en el pas anterior a la cadena de producció, els proveïdors i distribuïdors poden planificar la producció conforme a la demanda. Tanmateix, encara que és eficient, aquest model d'actuació mostra la seva debilitat en moments d'excepció, com ara els casos de supermercats desproveïts pel fet de no disposar d'estocs durant les vagues, o la incapacitat de fer front a grans demandes en moments de crisi (equipament i material després del pas de l'huracà Katrina o de l'enfonsament del petrolier "Prestige").

Relacionats també amb l'economia són els estudis per evitar pèrdues de clients i frau (mitjançant la detecció de canvis de comportament sospitosos) (Fawcett; Provost, 1997) (Cox; Eick; Wills; Brachman, 1997), així com la concessió de crèdits ràpids i les assegurances (Hernández et al., 2004). En tots aquests casos es poden realitzar previsions de risc mitjançant l'assignació dels individus a uns perfils dels quals es coneix un patró de comportament, després d'analitzar unes variables clau.

Així mateix, en la realització d'estudis de mercats pot ajudar a mostrar canvis en les tendències de consum a gran escala d'un país.

En medicina pot mostrar patrons entre malalts de la mateixa malaltia, i fer veure l'evolució d'epidèmies en grans poblacions. També en disciplines com ara l'astronomia, la bioinformàtica o la genètica ajuda a estudiar les grans quantitats de dades amb què es treballa.

Altres camps concrets en què la integració de bases de dades i el desenvolupament d'activitats de mineria de dades serien útils són la possibilitat de compartir dades d'investigació científica i de l'Administració pública per a la seguretat ciutadana (Clifton; Doan; Elmagarmid; Kantarcioglu; Schadow; Suci; Vaidya, 2004).

En un altre sector totalment diferent, Ma (Ma; Liu; Wong; Yu; Lee, 2000) presenta una aplicació per al desenvolupament de la mineria de dades en educació amb la finalitat de localitzar, de manera molt precisa, estudiants que necessitin classes de suport d'una matèria determinada.

En biblioteques, la mineria de dades pot resultar igualment molt útil per a la presa de decisions basades en l'evidència, tot i que es manegen quantitats de dades molt inferiors (Guenther, 2000).

2 Bibliomineria: aspectes teòrics

L'aplicació de la mineria de dades en biblioteques es denomina bibliomineria⁴ (*bibliomining*). El terme va ser encunyat per Nicholson i Stanton (2003)⁵ com una derivació dels termes bibliometria (*bibliometrics*) i mineria de dades (*data mining*), a fi d'afavorir la conceptualització i el treball dels investigadors d'aquest àmbit. Aquests autors defineixen la bibliomineria com "la combinació de mineria de dades, bibliometria, estadística i eines d'elaboració d'informes i extracció de patrons de comportament, basats en sistemes bibliotecaris".

Per Nicholson (2006), la utilització d'un nou terme està justificada perquè el terme *library* (o biblioteca) associat a mineria de dades es refereix principalment al conjunt d'algorismes que utilitza el programari, per la qual cosa pot donar lloc a errors i dificultats en la descripció i recerca d'informació sobre el tema.

Si bé la conceptualització és recent, la bibliomineria és una activitat que es realitza des del final de la dècada dels noranta (Banerjee, 1998). En són exemples els casos (Cullen, 2005) de les biblioteques de la University of Waterloo (Ontario, el Canadà) i de la University of Pennsylvania (Filadèlfia, EUA). Un altre cas recent (Geyer-Schultz et al., 2003) és el de la biblioteca de la Universität Karlsruhe (TH) (Karlsruhe, Alemanya).

Com s'ha assenyalat, la bibliomineria sol relacionar-se amb la bibliometria, ja que ambdues tasques s'ocupen de l'anàlisi i encreuament de dades mitjançant tècniques estadístiques per descobrir i establir patrons i tendències en les dades com a ajut a la presa de decisions (obtenir la dada per la dada, la mera presentació de resultats és una tasca vàcua i inútil). Si bé en bibliomineria es tracta de dades sobre l'actuació prèvia a la utilització de la informació, mentre que la bibliometria treballa amb les dades relatives a la informació que ha estat finalment usada.

Per Nicholson, el procés de bibliomineria és compon de sis fases (Nicholson, 2003), que a grans trets coincideixen amb les propostes sobre mineria de dades indicades anteriorment:

- Determinació de les àrees d'interès.
- Identificació de fonts de dades internes i externes.
- Recopilar, netejar i fer anònimes les dades en el *data warehouse* (magatzem de dades).
- Selecció de les eines d'anàlisi apropiades.
- Descobriments de patrons a través de la mineria de dades i creació d'informes amb eines tradicionals d'anàlisi.
- Anàlisi i implementació dels resultats.

Per fonts de dades internes s'entén les dades generades per la mateixa biblioteca en el transcurs de la seva activitat, allò que es denomina dades observacionals (Hand et al., 2001), és a dir, dades que han estat recopilades en l'activitat diària. Per fonts de dades externes s'entén les dades que provenen de fonts alienes a l'organització i que serveixen per contextualitzar les primeres; es tracta, essencialment, de dades demogràfiques.

A les biblioteques tradicionals, les dades internes procedeixen principalment de dues fonts: les consultes als OPAC (*Open Public Access Catalogue*), i les dades de circulació de materials i préstec (inclòs el préstec interbibliotecari). Aquestes dades, com s'indica més endavant, mostren informació sobre els materials que són més usats, el temps que és requerit (renovacions), materials relacionats o similars (retirats en un mateix préstec o per un mateix usuari) i les biblioteques amb les quals es té una major relació a través del préstec interbibliotecari.

Tanmateix, aquestes dades no recullen íntegrament l'actuació que l'usuari realitza a la biblioteca (en el cas de les biblioteques d'accés obert a la col·lecció), i en queden fora la consulta de fons a la prestatgeria, la consulta de fons que no es deixen (normalment obres de referència i publicacions periòdiques en paper) i que, al seu torn, solen ser els més cars, etc. Per això, els resultats de la bibliomineria han de ser considerats com a representatius d'una gran part de la utilització de la biblioteca, però no de la totalitat, per la qual cosa aquestes dades han de recolzar en altres informacions.

En el cas de les biblioteques digitals i de fons en línia, el ventall de dades pot arribar a ser molt més gran, ja que és possible crear i utilitzar *logs* (arxius de registre d'activitats) que marquin tot el recorregut i les accions que realitzen els usuaris de la col·lecció de la biblioteca (Nicholson, 2006). La identificació de l'usuari pot fer-se de dues maneres, depenent de la política de la biblioteca. Si es tracta d'un accés restringit, l'usuari es veu obligat a identificar-se, i en el cas de biblioteques digitals d'accés obert, poden estudiar-se les sessions que realitza cada usuari mitjançant la identificació de l'adreça IP (*Internet Protocol*) d'accés i l'ús de *cookies*.

3 Utilitat de la bibliomineria

De manera general, la utilització de les dades emmagatzemades (generades i recopilades) per la biblioteca pot realitzar-se de tres maneres: en primer lloc, mitjançant l'elaboració d'informes periòdics de variables determinades (el que es denomina *OLAP*, *Online Analytical Processing*, en segon lloc, mitjançant preguntes concretes a la base de dades; i en tercer, mitjançant l'exploració aleatòria de variables (el que es denomina *mineria de dades no dirigida*).

Amb això es poden realitzar tres tasques bàsiques:

- **Associacions:** veure quins elements estan relacionats, ja sigui per derivació, causa/efecte o per similitud.
- **Agrupacions (*clustering*):** crear grups de dades amb característiques similars.
- **Resum:** presentar de manera abreujada les dades sobre l'activitat diària per aconseguir-ne una comprensió millor.

Una qüestió de summa importància en l'estudi, la interpretació i la utilització dels resultats de bibliomineria és que, igual que en les dades obtingudes mitjançant estudis bibliomètrics, aquests no han de ser presos com a definitius i autoexplicatius, sinó que han de ser tinguts en compte en el seu context i sent comparats, contrastats i estudiats en funció d'altres variables i altres dades.

Kohavi i Provost presenten cinc desiderates per a l'èxit de la mineria de dades en el comerç electrònic, desiderates que pot fer-se extensibles a qualsevol altre camp i que en el cas de les biblioteques (per tant, bibliomineria) es compleixen a la perfecció (Kohavi; Provost, 2001):

- **Dades amb descripcions riques.** A les biblioteques, els ítems compten (o han de comptar) amb una descripció completa, rica i adequada, que, a més, s'ha elaborat d'acord amb unes normes, la qual cosa en facilita enormement l'ús.
- **Una gran quantitat de dades.** Els autors indiquen que a major quantitat de dades, es poden extreure models més fidedignes.
- **Col·lecció de dades controlada i correcta.** Com s'ha indicat, a les biblioteques la normalització de la informació és una assignatura coberta.
- **La capacitat d'avaluar resultats.** Aquesta és l'última finalitat del procés, estudiar els resultats i tenir en compte les conclusions per a actuacions futures. En cas contrari, és una tasca inútil.
- **Facilitat d'integració amb els processos existents.** L'automatització de les biblioteques a través de les bases de dades és una realitat des de fa anys, i aquestes formen part ineludible de l'activitat diària, i per això dur a terme tasques de bibliomineria només representaria afegir un pas més a la cadena de la base de dades.

Quant a aplicacions concretes en biblioteques, les possibilitats són múltiples i han estat tractades per diferents autors. Un cas és Papatheodorou: centrant-se en les biblioteques digitals, indica que la bibliomineria pot ajudar les biblioteques de diferents maneres (Papatheodorou; Kapidakis; Sfakakis; Vassilou, 2003):

- **Optimització de serveis:** ajuda els administradors a reorganitzar el contingut de la biblioteca, les autoritats i les interfícies.
- **Suport a la presa de decisions.**
- **Personalització:** ajuda els usuaris a identificar informació d'interès per a ells per la recomanació de matèries similars. I això comportaria una millora dels tan útils sistemes de DSI (difusió selectiva d'informació).

Quant a aplicacions més concretes, Nicholson i Stanton assenyalen les següents (Nicholson; Stanton, 2003):

- **Predicció de necessitats dels usuaris:** veure l'evolució de les temàtiques consultades pot ajudar a predir quines seran consultades posteriorment.
- **Identificació de materials no consultats:** per un mal procés de selecció o per una catalogació o classificació incorrectes.
- **Justificació del manteniment o supressió d'acords de préstec interbibliotecari:** despesa del préstec davant despesa d'adquisició.

En els apartats següents s'expliquen amb més detalls aquestes i altres possibles aplicacions de la bibliomineria.

3.1 Avaluació bibliotecària

En primer lloc, la bibliomineria pot oferir una nova perspectiva en l'avaluació de les biblioteques. Per una part, perquè proporciona dades directes i objectives dels usuaris i de l'ús que fan de la biblioteca, davant altres mètodes amb major o menor grau de parcialitat com ara les enquestes, les entrevistes, el mètode Delphi i l'incident crític (Sanz Casado, 1994). I per l'altra, perquè permet definir i calcular una sèrie d'indicadors de comparació (*benchmarking*), tal com recomana la International Federation of Library Associations and Institutions (IFLA) en les seves directrius per a biblioteques públiques, com són préstecs, ús de fons, usuaris reals i potencials, etc.

L'anàlisi estadística de dades quantitatives és un criteri tradicionalment molt usat en l'avaluació de biblioteques (com en el cas de la Xarxa de Biblioteques Universitàries Espanyoles <<http://www.crue.org/rebiun/>>, l'Agència per a la Qualitat del Sistema Univeristari de Catalunya <<http://www.aqucatalunya.org>>, i, en menor mesura, l'Agència Nacional d'Avaluació de la Qualitat i Acreditació <<http://www.aneca.es>> entre d'altres). Però la bibliomineria permet anar més enllà, ja que aquests indicadors simples seran posats en un context més ampli i

relacionats entre si per a l'obtenció de nous indicadors. És a dir, no es limita a un recompte, sinó que permet estudiar la relació entre indicadors per veure patrons de repetició en l'activitat en temps real.

Per a Lancaster, ja el 1988, l'estudi de patrons d'ús de la col·lecció i la identificació d'ítems poc usats és una de les formes d'avaluació de la col·lecció (Lancaster, 1993). Amb aquest enfocament, es podria seguir de primera mà, almenys en el que incumbeix l'ús de la col·lecció, una "qualitat basada en el client/usuari",⁶ és a dir, la idea que els usuaris tenen diferents necessitats i els productes que satisfan millor les necessitats esmentades presenten una major qualitat. Com més i millor es coneixin les necessitats i opinions dels usuaris, millor es podran adaptar els serveis a aquests criteris i, per tant, s'obté una major qualitat, centrant-se en el que també s'ha anomenat "qualitat percebuda pel client" (Domínguez Sanjurjo, 1996).

3.2 Gestió econòmica: selecció i adquisició de fons

En la gestió de les biblioteques, com en qualsevol organització, la presa de decisions des de la direcció és una activitat fonamental. És especialment delicada en l'assignació i la utilització dels recursos econòmics, i més tenint en compte que la major part de les biblioteques són finançades amb fons públics. En aquest sentit es pronuncien l'IFLA i la UNESCO en les directrius per a biblioteques públiques, que estableixen que tant la planificació com l'administració financeres resulten fonamentals perquè la biblioteca funcioni amb eficiència (obtenint resultats òptims), de manera econòmica (amb el menor cost) i amb eficàcia (donant uns beneficis màxims); és per això que la utilització d'aquests recursos econòmics s'ha de realitzar amb la major transparència i el millor criteri possible.

Així, la utilització de dades sobre el comportament dels usuaris en l'ús de la col·lecció resulta de gran ajut i importància per a la presa de decisions respecte a la selecció i l'adquisició de fons (el 2000, les biblioteques públiques van destinar el 14,5 % del seu pressupost a adquisicions) (Hernández, 2003). És a dir, conèixer quin tipus de material s'usa més i de quins dels ítems més sol·licitats es tenen pocs exemplars, resulta crucial i molt valuós per orientar adquisicions futures.

Un estudi de la utilització de la bibliomineria en l'assignació de pressuposts per a adquisicions és el model presentat per Kao, Chang i Lin per a biblioteques universitàries. El model esmentat, anomenat ABAMDM (*Acquisition Budget Allocation Model via Data Mining*), està basat principalment en les dades de circulació de materials i en el "pes" que tenen els diferents departaments universitaris en l'assignació del pressupost d'adquisició de materials (Kao; Chang; Lin, 2003).

3.3 Organització de la col·lecció

Encara que potser sigui de menor riquesa, l'estudi del catàleg bibliogràfic mitjançant tècniques de mineria de dades pot donar resultats interessants. Així, Lavoie, Dempsey i Silipigni indiquen que mitjançant aquestes anàlisis es poden descobrir noves maneres de presentació i organització de la col·lecció que la facin més interessant per als usuaris i suposen, per tant, un sistema més eficaç d'accés a la col·lecció (Lavoie et al., 2006). Així mateix, indiquen que la bibliomineria combinada amb l'ús del model *Functional Requirements for Bibliographic Records* (FRBR) també és d'utilitat a l'hora d'identificar les múltiples manifestacions i els nombrosos ítems de què es disposa en el centre. Un exemple d'això seria relacionar les diferents edicions de *La Tragèdia de Hamlet, Príncep de Dinamarca* amb les crítiques, amb les versions cinematogràfiques, amb obres que hi facin referència o amb les adaptacions de tot tipus de què es disposi en el centre, i així poder recomanar tot aquest conjunt d'ítems als usuaris que estiguin interessats en Shakespeare.

3.4 Llenguatges de classificació

També a través de la bibliomineria i analitzant l'ús de l'OPAC (més endavant s'aprofundirà en la bibliomineria aplicada als OPAC), pot estudiar-se la manera com els usuaris arriben als registres. D'importància especial seria veure quins termes s'utilitzen, tant si aquests pertanyen al llenguatge documental usat per la biblioteca com si no, per, així, enriquir la descripció temàtica i apropar-la a la realitat de l'usuari. En el benentès que no es proposa l'addició dels termes emprats pels usuaris al llenguatge usat per la biblioteca, sinó la seva utilització com a sinònims i per enriquir un sistema de recomanacions, com s'explica en l'apartat següent.

Ja que l'ús de qualsevol sistema de classificació respon a una manera de percepció de la realitat, el llenguatge documental serà més eficaç i comprensible com més s'adapti a la visió que té l'usuari de les coses; per tant, serà més útil i, per consegüent, millor. Això és buscar l'adopció de la "saviesa de les masses" (Surowiecki, 2004) i de les *folcsonomies*

Encara que certs estudis (Rius García, 1991) mostren que les necessitats normals dels usuaris queden satisfetes amb un registre que contingui només un mínim de descripció bibliogràfica —la qual cosa Bereijo denomina "plantejaments reduccionistes" (Bereijo Martínez, 1998)—, com més rica sigui la descripció dels fons, més fàcilment hi arribaran els usuaris i, per tant, major ús en faran. I és mitjançant la bibliomineria com més fàcilment es poden estudiar les tendències d'ús de termes d'indexació per a una major adequació d'aquests a l'usuari.

Així, amb tot això s'avançaria en la solució dels dos principals problemes a què s'enfronten els usuaris dels OPAC segons Borgman (1986), citada per Fernández Molina i Moya Anegón (1998): problemes mecànics (errors tipogràfics, no recuperacions per errors en els noms dels autors o errors en el sistema) i sobretot els problemes conceptuals (no-comprensió del procés de recerca o ús de termes incorrectes).

3.5 OPAC: presentació de resultats i sistema de recomanacions

Quant al catàleg, amb l'ús de la bibliomineria poden estudiar-se el comportament i els tipus de consulta dels usuaris en l'OPAC, per a la presentació de resultats. Així, conèixer quins ítems van ser considerats rellevants anteriorment en cerques iguals o similars permetria establir resultats recomanats per a les cerques, i també implementar sistemes d'ordenació de resultats de cerca per rellevància basats en qüestions d'adequació temàtica, i no únicament en la concurrència del terme de cerca en el registre. Igualment, es podrien recomanar termes similars als empleats en la cerca, que servirien als usuaris per enriquir i reorientar la seva estratègia de cerca tal com s'ha explicat anteriorment. A més, es podrien detectar errors ortogràfics comuns en termes de cerca i corregir-los de manera automàtica o bé proposar correccions (de l'estil de cercadors web com ara *Google* (<http://www.google.com>), ja que sempre és millor que després d'una cerca s'ofereixi un resultat erroni (corregit automàticament l'usuari) que no oferir-ne cap, o un simple missatge d'error.

Amb tot això, el que es pretén és desenvolupar per als OPAC un sistema similar a l'utilitzat en portals comercials com ara Amazon.com (<http://www.amazon.com>) (en la utilització de la informació sobre el comportament dels clients, Amazon.com és un exemple clar que cal que segueixin les biblioteques), basats en el que es denomina una anàlisi associativa, del tipus *Els compradors d'aquest llibre també varen comprar..., Cerques relacionades, T'agradaria... o També podria agradar-te*, del mateix Amazon.com. L'obtenció de relacions associatives a través de la mineria de dades ha estat i continua sent un dels aspectes més estudiats dins d'aquesta disciplina.

Si bé el més correcte i complet seria que, al costat d'aquestes recomanacions (de resultats i termes de cerca) obtingudes per mineria de dades, se n'incloguessin d'altres fetes pels professionals de les biblioteques mitjançant l'ús de tècniques com ara el *clustering*⁷ i el propi coneixement de la col·lecció, i, a més, permetre als usuaris definir recomanacions pròpies. Així es convertiria en un treball de col·laboració que, d'altra banda, permetria que la biblioteca s'adaptés millor a la comunitat en la qual s'ubica.

En resum, l'OPAC, com a principal mitjà d'accés a la col·lecció, és l'eina fonamental de la biblioteca. Seguint Lubeztky (Bereijo Martínez, 1998), el catàleg pot funcionar com a simple eina de recuperació (*finding list*), i amb això n'hi hauria prou amb una relació de les entrades del material disponible en el centre, o també pot funcionar com un instrument de referència (*reference tool*), explotant, així, tots els recursos documentals i informatius del centre. Amb una correcta utilització de la bibliomineria és aquesta segona utilització de l'OPAC la que prevaldria, i, així, n'augmentaria la utilitat.

3.6 Estudis d'usuaris: socialització i grups d'usuaris

En biblioteques especialitzades i universitàries, la bibliomineria ha de ser un mètode complementari a la realització d'estudis bibliomètrics sobre la seva comunitat. Com ja s'ha assenyalat, ambdós estudis són complementaris, ja que la bibliomineria oferirà la perspectiva d'utilització de la col·lecció com a usuaris, mentre que els estudis bibliomètrics oferiran la perspectiva d'aquests usuaris en el seu paper d'autors, marcant el nivell d'utilitat efectiva de la col·lecció i mostrant també els elements que s'utilitzen i que no formen part de la col·lecció (la qual cosa orientarà les adquisicions futures).

Una altra aplicació similar, també per a biblioteques d'investigació i universitàries, és el descobriment de tendències d'investigació dels seus usuaris (Bollen et al., 2003), que ajudaria també a orientar la política de selecció i adquisició de fons. Això és el que es denomina models de comportament: l'estudi i definició de perfils dels usuaris a través del comportament que tenen respecte a la col·lecció del centre.

La bibliomineria ofereix una altra possibilitat. Com a conseqüència de l'estudi de dades sobre cerques (i préstecs) i els estudis d'usuaris, poden configurar-se grups d'usuaris amb perfils i gustos similars. Si en biblioteques acadèmiques això pot ajudar a establir i consolidar grups d'investigació, en biblioteques públiques pot servir d'element socialitzador entre els usuaris. Aquesta funció podria desenvolupar-se fins i tot mitjançant la creació de perfils públics per part dels usuaris, que els permetessin intercanviar opinions i suggeriments. Es crearia, d'aquesta manera, una xarxa social amb la biblioteca com a punt d'unió, similar al que ofereixen serveis web com ara Last.fm - The Social Music Revolution (<http://www.last.fm>) en relació amb la música.

3.7 Biblioteca 2.0

Tota aquesta actitud de col·laboració dels usuaris en la construcció i millora de serveis descrita en els apartats anteriors és la base de l'anomenada Web 2.0. En el camp de les biblioteques es parla de Biblioteca 2.0 o B 2 (*Library 2.0* o *L2*) terme encunyat per Casey (2005) i la definició del qual més acceptada és la donada per Houghton, que diu que "Biblioteca 2.0 simplement significa fer el teu espai bibliotecari més interactiu, amb una major col·laboració, i dirigida cap a les necessitat de la comunitat [...]" (Houghton, 2005). És a dir, bàsicament el que qualsevol biblioteca ha de fer, però recolzada en les tecnologies i amb una major quantitat d'informació.

Per Crawford (que, a més, recopila 62 visions i set definicions de B 2), "Biblioteca 2.0 inclou un ventall de noves, i no tan noves, metodologies de programari (programari social, interactivitat, API,⁸ software modular...) programari modular...), que pot ser i serà útil que moltes biblioteques ofereixin nous serveis i facin accessibles els ja existents d'una manera nova i interessant". Si bé, com també diu aquest autor, no tot és nou en l'anomenada B2; així, els conceptes són els mateixos i les tecnologies són noves, i amb això apareixen nous horitzons, i molts dels existents en resulten ampliat (Crawford, 2006).

Que les biblioteques estan (o han d'estar) obertes als canvis i necessitats dels usuaris és lògic, però la relació amb l'usuari és sempre difícil, ja que quants usuaris que no troben el llibre que busquen realitzen una desiderata? Tanmateix, si a través de les estadístiques de l'OPAC, mitjançant la bibliomineria, es veuen reiterades recerques d'un llibre del qual no es disposa a la biblioteca, ja se sabrà aquesta demanda i es coneixerà el tipus d'usuari que buscava aquest llibre, i se'l podrà avisar una vegada estigui disponible. És a dir, la B 2 no és "només" tecnologia, però "també" és tecnologia.

Així, des del meu punt de vista, els OPAC, i més concretament les interfícies amb l'usuari, seran l'aspecte que més pot beneficiar-se i millorar si la B 2 es consolida. Especialment en el foment de la personalització i adequació de la informació a les necessitats i característiques de l'usuari en temps real i de manera intel·ligent.

A la B 2, la bibliomineria li ofereix la base d'estudi de dades necessària per dur a terme els processos de col·laboració i personalització que promou mitjançant els esmentats estudis dels usuaris i els seus grups, de la col·lecció, de les consultes a OPAC, de la utilització de llenguatges de classificació, i d'altres. Així, les propostes i millores de la B 2 són l'element cristal·litzador i conjuminador de la útil informació que es pot obtenir mitjançant la bibliomineria.

4 Bibliomineria i privadesa

Un dels elements problemàtics de la mineria de dades i, per tant, de la bibliomineria està en com es respecta la privadesa dels usuaris. Aquesta situació fa que molts usuaris puguin desconfiar del fet de cedir les seves dades (pròpies o generades) perquè siguin explotades, i més encara si la biblioteca és considerada un lloc lliure i segur on informar-se, i on es respecten (o han de respectar-se) tots els punts de vista i opinions de manera imparcial i objectiva. Per trencar aquesta reticència es pot intentar emfatitzar un sentiment d'empatia cap al conjunt d'usuaris: si tu ajudes la comunitat i li aportes alguna cosa, la comunitat també t'aportarà alguna cosa a tu. Novament la filosofia de Web 2.0 o Biblioteca 2.0 i el seu èmfasi en la col·laboració seran un factor que ajudi a superar aquesta situació.

La privadesa en biblioteques és un assumpte de gran amplitud i de preocupació, i això fa que reaparegui periòdicament en els debats professionals, i és un aspecte que ha de ser regulat no solament per la legislació vigent, com s'indica més endavant, sinó també en els codis ètics de la professió (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998).

Per Zwass (1996), la privadesa és l'aspecte ètic més important dels sistemes d'informació. I indica que aquesta qüestió ha estat present des d'abans de l'aparició de les noves tecnologies a les biblioteques, però són aquestes les que han augmentat el risc d'un possible mal ús. Ramos Simón (2004) també mostra preocupació per l'ús de la informació generada per l'activitat dels usuaris, en el seu cas en els sistemes de DRM *Digital Right Management* o gestió de drets digitals).

Domingo-Ferrer i Torra (que, a més, presenten una breu recopilació d'articles sobre el tema) assenyalen que la qüestió de la privadesa en la mineria de dades (*Privacy-Preserving Data Mining* o PPDM) ha estat estudiada des de diferents disciplines (Domingo-Ferrer; Torra, 2005):

- Estadística: la majoria de les lleis nacionals d'estadística inclouen referències a la salvaguarda de la privadesa.
- Filosofia: l'ètica en la societat de la informació recull la privadesa com un dels valors que cal defensar.
- Informàtica: la privadesa s'inclou com una part de la seguretat de les dades.

Com indiquen Geyer-Schultz, Neumann i Thede, per part de les biblioteques aquesta qüestió ha estat una de les que ha impedit l'aplicació de la bibliomineria, com també ho han estat, segons aquests autors, les limitacions pressupostàries i alguns aspectes tècnics (Geyer-Schultz et al., 2003). Encara que no és objectiu d'aquest document analitzar els elements tecnològics de la bibliomineria, els problemes relatius a la tecnologia són cada vegada menors a causa de la baixada de preus i l'augment de la potència del maquinari, així com el desenvolupament de programari específic i fàcil d'usar, per a aquestes tasques, sent moltes vegades utilitzable el propi sistema de gestió bibliotecària (Michail, 2000).

Com s'ha assenyalat, en algunes ocasions es fa necessari encreuar diverses fonts i bases de dades per obtenir resultats més precisos. En aquesta línia, per Clifton actualment resulta crític desenvolupar tècniques que permetin integrar i compartir dades sense violar la privadesa. Per a això, aquests autors dissenyen un esquema d'organització dels camps de les bases de dades que garanteix el manteniment de la privadesa tot i utilitzant dades personals (Clifton et al., 2004). Per la seva part, Yang, Zhong i Wright proposen un procediment d'encriptatge de les dades que impedeixi que el sistema sigui capaç d'establir correspondències entre les dades i el seu origen, ja que així també es protegeix la privadesa dels usuaris (Yang et al., 2005).

Si bé, per al cas que aquí es tracta, realment les dades necessàries per a la bibliomineria no tenen per què contenir informació personal ni permetre una identificació de l'usuari. Això es duria a terme utilitzant un codi aleatori per a la identificació de cada usuari, o bé utilitzant el que Nicholson (2006)⁹ denomina *substituts demogràfics* (*demographic surrogates*), és a dir, emmagatzemant simplement unes dades del perfil d'usuari que permetin classificar-lo sense identificar-lo. Nicholson i Smith descriuen una metodologia per a aquest procés; parlen de la *desidentificació* (*deidentification*) de l'usuari (Nicholson; Smith, 2005), el que també en Estadística es denomina *dades anonimitzades* (Montaña Lacambra, 2005). Si bé ambdues tècniques són perfectament compatibles.

Així, en el cas de dades sobre els préstecs en una biblioteca universitària, amb vista a la bibliomineria no es necessita saber que el curs 2004–05 María López, alumna de tercer curs de la llicenciatura en Dret, ha tingut en préstec quatre vegades el llibre *Derecho del trabajo* d'Ignacio Albiol Montesinos, sinó que aquest any el dit llibre ha estat deixat quatre vegades a l'usuari 1010011010, estudiant de la llicenciatura de Dret.

Una altra qüestió que mereix una reflexió és si resulta interessant per a la gestió de la biblioteca mantenir una relació dels préstecs que ha realitzat cada usuari. Aquestes dades poden ser útils fins i tot per al mateix usuari, i protegides correctament no han de generar problemes de respecte de la seva privadesa. És aquí on intervien qüestions tècniques de protecció d'informació i també qüestions ètiques per als professionals de les biblioteques en la utilització de les dades de què disposen.

Per una altra banda, un cas d'utilització de la mineria de dades com a mitjà de control de les activitats bibliotecàries dels ciutadans (com ara els registres de circulació, els perfils de l'usuari, les peticions de préstec, i arxius de registre d'ús d'Internet) és l'USA Patriot Act (Taylor; Black, 2004), cas al qual l'ALA (*American Library Association*) s'ha oposat rotundament, ja que xoca amb el seu codi ètic.

Actualment no hi ha una política generalitzada sobre el tractament de les dades bibliotecàries, i la problemàtica legal que implica el seu maneig difereix entre els estats. Per això no es pot generalitzar una mesura d'actuació sobre el tema i haurien d'estudiar-se detingudament cada cas i cada situació legal.

No obstant això, com bé sostenen Million i Fisher, encara que no hi hagi una llei que reguli la confidencialitat dels registres bibliotecaris, les biblioteques han de tenir una política sobre el tema, "apropiada, aprovada, ben comunicada al personal i inclosa en el manual de procediment" (Million; Fisher, 1998). En el cas que hi hagués una legislació sobre això, la política del centre hi haurà de ser congruent.

En el cas espanyol, la Constitució de 1978, en l'article 18¹⁰ ja recull la protecció de la informació i les dades personals en garantir el dret a la intimitat. Però és la Llei orgànica 15/1999, de 13 de desembre, de protecció de dades de caràcter personal (normalment abreujada *LOPD*) la que regula aquesta activitat, i és l'Agència Espanyola de Protecció de Dades l'autoritat administrativa independent encarregada de vetllar per l'aplicació de la legislació sobre això. A més, la *LOPD* "regula el deure de guardar el secret professional als qui intervenen en qualsevol procés automàtic de recollida de dades de caràcter personal, i imposa les sancions oportunes en el cas de manipular il·legalment dades sensibles estretament vinculades al concepte d'intimitat" (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998).

Conforme a l'article 2 de la *LOPD*,¹¹ que fixa que el seu àmbit d'aplicació són totes les dades de caràcter personal registrades en suport físic, i que s'aplica a totes les modalitats d'ús posterior d'aquestes dades per part dels sectors públic i privat, els fitxers de lectors de biblioteques estan regulats per aquesta llei. Per utilitzar-la amb vista a la bibliomineria, l'article 4¹² de la dita llei assenyalava que l'ús de les dades amb finalitats estadístiques no és incompatible amb l'activitat per a la qual van ser recollides, per la qual cosa aquest pot ser considerat un ús legítim

També és una qüestió que està protegida en diferents instruments del dret internacional i de protecció dels drets humans, com són la Declaració universal dels drets humans (art. 12), el Conveni internacional sobre drets civils i polítics (art. 17), la Convenció europea sobre drets humans (art. 8) i la Convenció internacional de telecomunicacions (art. 22) (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998).

Com indiquen Agrawal i Srinikant (Agrawal; Srinikant, 2000), són molts els aspectes que planteja la qüestió de la privadesa i que cal estudiar, alguns dels quals són assenyalats per Nicholson i Smith (2005):

- Quanta informació es perd en la desidentificació? Hi ha algorismes que puguin suplir o simular les dades perdudes?
- Quins estàndards es podrien crear per motivar les biblioteques a mantenir el mateix tipus de dades desidentificades?¹³ ¿Aquests estàndards permetrien desenvolupar *data warehouses* (dipòsits de dades) en xarxes o consorcis bibliotecaris?
- Quina opinió tenen els usuaris de les seves dades desidentificades?

5 Conclusions

Com hem estat explicant, la Biblioteconomia i la mineria de dades no són dues activitats incompatibles. És més, es podria dir que les biblioteques són un espai ideal per desenvolupar tasques de mineria de dades pel fet de tractar-se d'un espai controlat, amb una descripció correcta (això és el que en podem esperar) dels elements amb què treballa, acostumades a recopilar dades (préstecs, usuaris, fons, etc.) i amb professionals (això també és esperable) que saben manejar informació i treure'n partit. Es tenen les dades, les eines i els professionals, la utilitat de la bibliomineria és clara i és alta, què s'espera per aplicar-la? Ja s'han enriquit els OPAC en contingut, i permeten l'accés a múltiples fons i bases de dades; ara han d'enriquir-se i poden enriquir-se en funcionalitats.

Con el aumento de información disponible y la mejora de servicios que se pueden obtener de la bibliominería, en primer lugar se debe romper con la idea de que el usuario vaya a la biblioteca a buscar un libro concreto para después irse. Hay que potenciar la biblioteca como un lugar rico, abierto y completo donde descubrir y ampliar sus ideas y gustos.

Amb l'augment de la informació disponible i la millora de serveis que es poden obtenir de la bibliomineria, en primer lloc s'ha d'arraconar la idea que l'usuari vagi a la biblioteca a buscar un llibre concret per anar-se'n tot seguit. Cal potenciar la biblioteca com un lloc ric, obert i complet on descobrir i ampliar les seves idees i els seus gustos.

Com s'ha assenyalat, són molts els aspectes per continuar estudiant i moltes les vies d'investigació existents en l'aplicació de la bibliomineria. En aquest article simplement se n'han presentat algunes utilitzacions possibles, però les potencialitats són moltes, i les vies d'investigació, molt àmplies.

La bibliomineria és un camp en què la innovació i l'amplitud de mires a l'hora d'estudiar i encreuar dades són importants. Cada centre, coneixent la seva manera de treballar, els seus usuaris i les seves pròpies característiques, haurà d'identificar uns paràmetres d'interès i estudiar-ne uns patrons concrets. Idees més treball sempre equivalen a èxit. Si bé pot ser que de la bibliomineria o de l'anomenada *Biblioteca 2.0* falli l'aplicació real de totes aquestes idees. Només falten recursos o és, també, un problema de falta de motivació professional?

Bibliografía consultada

- Ackoff, Russell L. (1989). "From data to wisdom". *Journal of applied systems analysis*, no. 16, p. 3-9.
- Agrawal, Rakesh; Imielinski, Tomasz; Swami, Arun (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD Conference*. <<http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>>. [Consulta: 1-05-2006]
- Agrawal, Rakesh; Srinikant, Ramakrishnan (2000). "Privacy-preserving data mining". *Proceedings of the 2000 ACM SIGMOD conference on management of data*. p. 439-450. <<http://doi.acm.org/10.1145/342009.335438>>. [Consulta: 1-05-2006].
- American Library Association. *The USA patriot act in the library*. <<http://www.ala.org/ala/oif/ffissues/usapatriactlibrary.htm>>. [Consulta: 30-04-2006].
- Banerjee, K. (1998). "Is data mining right for your library?" *Computers in libraries*, vol. 18, no. 10, p. 28-31.
- Bellinger, Gene; Castro, Durval; Mills, Anthony (1994). *Data, information, knowledge, and wisdom*. <<http://www.systems-thinking.org/dikw/dikw.htm>>. [Consulta: 1-05-2006].
- Bereijo Martínez, Antonio (1998). "Caracterización del concepto de 'calidad' en la catalogación descriptiva: factores que atañen al diseño de objetivos". *Boletín Millares Carlo*, núm. 17, p. 319-355. <http://dialnet.unirioja.es/servlet/fichero_articulo?articulo=1700760&orden=37277>. [Consulta: 12/04/2006].
- Bollen, Johan; Luce, Rick; Vemulapalli, Soma Sekhara; Weining, Xu (2003). "Usage analysis for the identification of research trends in digital libraries". *D-lib magazine*, May, vol. 9, no. 5, <<http://www.dlib.org/dlib/may03/bollen/05bollen.html>>. [Consulta: 31-03-2006].
- Borgman, Christine L. (1986). "Why are online catalogs hard to use? Lessons learned from information-retrieval studies" *Journal of the American society for information science*, vol. 37, no. 6, p. 387-400. <<http://www3.interscience.wiley.com/cgi-bin/abstract/57783/>>. [Consulta: 7-05-2006].
- Brin, Sergey; Motwani, Rajeev; Ullman, Jeffrey D.; Tsur, Shalom (1997). "Dynamic itemset counting and implication rules for market basket data". *Proceedings of the 1997 ACM SIGMOD conference*. p. 255-264 <<http://doi.acm.org/10.1145/253260.253325>>. [Consulta: 1-05-2006].
- Casey, Michael (2005). "Working towards a definition of Library 2.0". *Library Crunch: bringing you a library 2.0 perspective*, October 21, 2005. <http://www.librarycrunch.com/2005/10/working_towards_a_definition_o.htm>. [Consulta: 16-05-2006].
- Chi-Wing Wong, Raymond; Wai-Chee Fu, Ada; Wang, Ke (2005). "Data mining for inventory item selection with cross-selling considerations". *Data mining and knowledge discovery*, July 2005, vol. 11, no. 1, p. 81-112.
- Cleveland, Harland (1982). "Information as resource". *The futurist*, December, p. 34-39.
- Clifton, Chris; Doan, Anhai; Elmagarmid, Ahmed; Kantarcioglu, Murat; Schadow, Gunther; Suci, Dan; Vaidya, Jaideep (2004). "Privacy preserving, data integration and sharing". *Data mining and knowledge discovery archive. Proceedings of the 9th ACM SIGMOD workshop on research issues in data mining and knowledge discovery*. p. 19-26. <<http://doi.acm.org/10.1145/1008694.1008698>>. [Consulta: 21-04-2006].
- Cox, Kenneth C.; Eick, Stephen G.; Wills, Graham J.; Brachman, Ronald J. (1997). "Brief application description; visual data mining: recognizing telephone calling fraud". *Data mining and knowledge discovery*, June, vol. 1, no. 2, p. 225-231.
- Crawford, Walt (2006). "Library 2.0 and 'Library 2.0'". *Cites & insights*, vol. 6, no. 2, p. 1-32. <<http://cites.boisestate.edu/civ6i2.pdf>> [Consulta: 16-05-2006].
- Cullen, Kelvin (2005). "Delving into data". *Library journal* August, vol. 130, no. 13, p. 30-32. <<http://www.libraryjournal.com/article/CA633325.html>>. [Consulta: 26-04-2006].
- Domingo-Ferrer, Josep; Torra, Vicenç (2005). "Privacy in data mining". *Data mining and knowledge discovery*, September, vol. 11, no. 2, p. 117-119.
- Domínguez Sanjurjo, M^a Ramona (1996). *Nuevas formas de organización y servicios en la biblioteca pública*. Gijón: Trea
- España. "Ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal". *Boletín Oficial del Estado*, 14 de diciembre de 1999, núm. 298, p. 43088-43099.
- Fawcett, Tom; Provost, Foster (1997). "Adaptive fraud detection". *Data mining and knowledge discovery*, September, vol. 1, no. 3, p. 291-316.
- Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From data mining to knowledge discovery: an overview". En: Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (ed.). *Advances in knowledge discovery and data mining*. California: AAAI Press, The MIT Press, p. 1-36.

Fayyad, Usama; Simoudis, Evangelos (1995). *Knowledge discovery and data mining*. <<http://www-aig.jpl.nasa.gov/public/kdd95/tutorials/IJCAI95-tutorial.html>> . [Consulta: 1-05-2006].

Fayyad, Usama; Uthurusamy, Ramasamy (1996). "Data mining and knowledge discovery in databases". *Communications of the ACM*, November, vol. 39, no. 11, p. 24-26. <<http://doi.acm.org/10.1145/240455.240463>> . [Consulta: 1-05-2006].

Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2001). *Directrices de la IFLA/UNESCO para el desarrollo del servicio de bibliotecas públicas*. 94 p. <<http://www.ifla.org/VII/s8/news/pg01-s.pdf>> . [Consulta: 4-04-2006].

Fernández Molina, Juan Carlos; Moya Anegón, Félix de (1998). *Los catálogos de acceso público en línea: el futuro de la recuperación de información bibliográfica*. Málaga: Asociación Andaluza de Bibliotecarios, 197 p.

Geyer-Schulz, Andreas; Neumann, Andreas; Thede, Anke (2003). "An architecture for behavior-based library recommender systems". *Information technology and libraries*, 2003, vol. 22, no. 4, p. 165-174. <<http://www.ala.org/ala/lita/litapublications/ital/2204geyer.htm> . [Consulta: 23-08-2006]

Gómez-Pantoja Fernández-Salguero, Aurora; Pérez Pulido, Margarita (1998). "El concepto de privacidad en servicios bibliotecarios actuales". *FESABID 98. VI Jornadas Españolas de Documentación: los sistemas de información al servicio de la sociedad*. <http://fesabid98.florida-uni.es/Comunicaciones/a_gomez.htm> . [Consulta: 15-06-2006].

Guenther, Kim (2000). "Applying data mining principles to a library data collection". *Computers in libraries*, vol. 20, no. 4, p. 60-63.

Hand, David; Mannila, Heikki; Smyth, Padric (2001). *Principles of data mining*. Cambridge: Massachusetts Institute of Technology.

Hernández, Hilario (dir.) (2003). *Las colecciones de las bibliotecas públicas en España: informe de situación*. Salamanca: Fundación Germán Sánchez Ruipérez.

Hernández Orallo, José; Ramírez Quintana, M^a José; Ferri Ramírez, César. (2004). *Introducción a la minería de datos*. Madrid: Pearson - Prentice Hall.

Houghton, Sarah (2005). "Library 2.0 Discussion: Michael Squared". *LibraryInBlack.net: resources and discussions for the 'tech-librarians-by-default' among us...* December 19. <http://librarianinblack.typepad.com/librarianinblack/2005/12/library_20_disc.html> . [Consulta: 16-05-2006].

Kao, S.-C.; Chang, H.-C.; Lin, C.-H. (2003). "Decision support for the academic library acquisition budget allocation via circulation database mining". *Information processing and management*, no. 39, p. 133-147. *Knowledge discovery and data mining conference (KDD-95)*. <<http://www-aig.jpl.nasa.gov/public/kdd95/>> . [Consulta: 1-05-2006].

Kohavi, Ron (2001). "Mining e-commerce data: the good, the bad, and the ugly". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 8-13. <<http://doi.acm.org/10.1145/502512.502518>> . [Consulta: 1-05-2006].

Kohavi, Ron; Provost, Foster (2001). "Applications of data mining to electronic commerce". *Data mining and knowledge discovery*, January, vol. 5, no. 1-2, p. 5-10.

Kudyba, Stephan; Hoptruff, Richard (2001). *Data mining and business intelligence: a guide to productivity*. Hershey: Idea

Lancaster, Frederick Wilfrid (1993). *If you want to evaluate your library...* London: Library Association Publishing.

Lavoie, Brian; Dempsey, Lorcan; Silipigni Connaway, Lynn (2006). "Making data work harder". *Library journal*, January, vol. 131. <<http://www.libraryjournal.com/article/CA6298444.html>> [Consulta: 26-04-2006].

Ma, Yiming; Liu, Bing; Wong, Ching Kian; Yu, Philip S.; Lee, Shuik Ming (2000). "Targeting the right students using data mining". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 457-464. <<http://doi.acm.org/10.1145/347090.347184>> . [Consulta: 1-05-2006].

Michail, Amir (2000). "Data mining library reuse patterns using generalized association rules". *International Conference on Software Engineering Archive: proceedings of the 22nd International Conference on Software Engineering*. p. 167-176. <<http://doi.acm.org/10.1145/337180.337200>> . [Consulta: 2-05-2006].

Million, Angela C.; Fisher, Kim N. (1998). "Library records: a review of confidentiality laws and policies". *The journal of academic librarianship*, vol. 11, no. 6, p. 346-349.

Montaña Lacambra, Rosa (2005). "Fuentes estadísticas y bibliotecas: estado de la cuestión". *BiD: Textos universitaris de biblioteconomia i documentació*, desembre, núm. 15. <http://www2.ub.edu/bid/consulta_articulos.php?fichero=15monta2.htm> [Consulta: 23-08-2006].

Nicholson, Scott (2003). "The bibliomining process: data warehousing and data mining for library decision-

making". *Information technology and libraries*, December, vol. 22, no. 4, p. 146-151. <<http://www.ala.org/ala/lita/litapublications/ital/2204nicholson.htm>>. [Consulta: 23-08-2006]

Nicholson, Scott (2006). "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services". *Information processing & management*, vol. 42. iss. 3, May, p. 785-804. <<http://dx.doi.org/10.1016/j.ipm.2005.05.008>>. [Consulta: 22-08-2006]. (ver nota 9)

Nicholson, Scott; Smith, Catherine Arnott (2005). "Using lessons from health care to protect the privacy of library users: guidelines for the de-identification of library data based on HIPAA". En: Andrew Grove (ed.). *Proceedings 68th annual meeting of the American Society for Information Science and Technology (ASIST)* vol. 42. <http://eprints.rclis.org/archive/00005255/01/Nicholson_Using.pdf>. [Consulta: 29-04-2006].

Nicholson, Scott; & Stanton, Jeffrey (2003). "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries". En: Nemati, Hamid R.; Barko, Christopher D. (eds.). *Organizational data mining: leveraging enterprise data resources for optimal performance*. Hershey, PA: Idea Group Publishing, p. 247-262. (ver nota 5)

Papatheodorou, Christos; Kapidakis, Sarantos; Sfakakis, Michalis; Vassilou, Alexandra (2003). "Mining user communities in digital libraries". *Information technology and libraries*, December, vol. 22, no. 4, p. 152-157.

Ramos Simón, Luis Fernando (2004). "DRM: protección versus accesibilidad de la información digital". *Hipertext.net*, mayo, núm. 2. <<http://www.hipertext.net/web/pag208.htm>>. [Consulta: 23-08-2006].

Ríos García, Yolanda (1991). "Catálogos en línea de acceso público". *Revista española de documentación científica*, vol. 14, núm. 2, p. 121-141.

Sanz Casado, Elías (1994). *Manual de estudios de usuarios*. Madrid: Fundación Germán Sánchez Ruipérez.

Sharma, Nikhil (2005). *The origin of the 'data information knowledge wisdom' hierarchy*. <http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm>. [Consulta: 1-05-2006].

Surowiecki, James (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Doubleday.

Taylor, Mayo; Black, William (2004). "In search of reason: libraries and the USA patriot act". *Journal of librarianship and information science*, vol. 36, no. 2, p. 51-54.

Yang, Zhiqiang; Zhong, Sheng; Wright, Rebecca N. (2005). "Anonymity-preserving data collection". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 334-343. <<http://doi.acm.org/10.1145/1081870.1081909>>. [Consulta: 21-04-2006].

Zeleny, Milan (1987). "Management support systems: towards integrated knowledge management". *Human systems management*, vol. 7, no. 1, p. 59-70.

Zwass, Vladimir (1996). "Ethical issues in information systems". En: Kent, Allen; Lancour, Harold. *Encyclopedia of library and information science*. New York: Marcel Dekker, p. 175-195.

Data de recepció: 28/07/2006. Data d'acceptació: 10/09/2006.

Notes

¹ Citació presa de Fayyad (Fayyad; Uthurusamy, 1996).

² *DIKW* (*data, information, knowledge, wisdom*: dades, informació, coneixement i saviesa) és un dels més coneguts i acceptats per representar l'estructura jeràrquica conceptual de la informació. La teoria més àmpliament acceptada (Sharma, 2005; Bellinger, Castro, Mills, 1994) sosté que aquest esquema està basat en un poema de T. S. Eliot titulat *The Rock* (1990), que és desenvolupat teòricament per Ackoff (1989), Cleveland (1982) i Zeleny (1987), entre d'altres.

³ En les obres consultades, l'ús del terme "minería de dades" per a l'anglès *data mining* és més comú que no pas "descobrimet d'informació en bases de dades" per a "knowledge discovery in databases", i per això s'ha optat per utilitzar d'ara endavant la primera en català i la segona en anglès.

⁴ Encara que es podria utilitzar l'expressió "minería de dades en biblioteques", per les mateixes raons que exposa Nicholson (2006) per a l'ús del terme *bibliomining*, s'ha optat per l'adaptació directa al català com a bibliomineria.

⁵ També està disponible en línia l'edició preliminar actualitzada el 2004: Nicholson, Scott; Stanton, Jeffrey (2003). "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries." <<http://www.bibliomining.com/nicholson/odmcom.html>>. [Consulta: 30-03-2006].

⁶ Smart, Seawright i Detiene (Bereijo Martínez, 1998) defineixen un esquema estructural per estudiar el concepte de

qualitat, basat en l'esquema bidimensional de Seawright i Young, sobre el qual apliquen les sis categories de Garvin: qualitat transcendent, qualitat basada en el procés de manufactura, qualitat basada en el producte, qualitat basada en el client, qualitat basada en el valor i qualitat estratègica.

⁷ Per Fernández Molina i Moya Anegón, les tècniques de *clustering*, o anàlisi clúster, "consisteixen a classificar de manera automàtica el contingut de les bases de dades. [...] En termes generals, aquestes [tècniques] s'ocupen de la manera com s'agrupen els termes d'indexació assignats als documents, per tal de posar de manifest la relació entre els documents [...] de matèries similars" (Fernández Molina; Moya Anegón, 1998).

⁸ *Application Programming Interface*, , interfície de programació d'aplicacions. Pot definir-se, en termes molt bàsics, com el conjunt d'operacions que indica com han d'interactuar els mòduls de programari entre si.

⁹ També està disponible en línia l'edició preliminar en: Nicholson, Scott (2005?). "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services." <<http://bibliomining.com/nicholson/nicholsonbibliointro.html>>. [Consulta: 30-03-2006].

¹⁰ Article 18. 1. Es garanteix el dret a l'honor, a la intimitat personal i familiar i a la pròpia imatge. [...].

¹¹ Article 2. 1. La present Llei orgànica serà aplicable a les dades de caràcter personal registrades en suport físic, que les faci susceptibles de tractament, i a totes les modalitats d'ús posterior d'aquestes dades pels sectors públic i privat. [...].

¹² Article 4. 1. Les dades de caràcter personal només es podran recollir per al seu tractament, i sotmetre-les al dit tractament, quan siguin adequades, pertinents i no excessives en relació amb l'àmbit i les finalitats determinades, explícites i legítimes per a les quals s'hagin obtingut.

Article 4. 2. Les dades de caràcter personal objecte de tractament no podran usar-se per a finalitats incompatibles amb aquelles per a les quals hagin estat recollides. No es considerarà incompatible el tractament posterior d'aquestes dades amb finalitats històriques, estadístiques o científiques. [...].

¹³ En l'actualitat, l'establiment d'estàndards per compartir la informació obtinguda per mineria de dades és una de les principals tasques de l'àrea.

