

SISTEM SUPTOR PENTRU DECIZII DE BIBLIOTECĂ

Cornel Lepădatu (cornel_lepadatu@biblacad.ro)

Academia Română București, Biblioteca Academiei Române

Rezumat. Sistemele suport pentru decizii oferă cunoștințe și capacitatea de prelucrare a cunoștințelor esențiale în sesizarea situațiilor decizionale și în elaborarea deciziilor, îmbunătățind procesele decizionale și rezultatele luării deciziilor și relaxând limitele cognitive, temporale, spațiale și economice ale factorilor de decizie. Sprijinul bibliotecilor și bibliotecarilor în luarea deciziilor a variat în timp de la unul pasiv, colecțiile tradiționale de cărți și reviste, către unele extrem de active, asistenți decizionali. Bibliotecile digitale au oferit perspective noi pentru sistemele suport pentru decizii ale companiilor. Tehnicile data mining au devenit cruciale pentru gestionarea, organizarea informațiilor și diseminarea acestora către utilizatorii potriviți, la momentul potrivit. Bibliomining a oferit posibilitatea ca prin intermediul unui singur depozit de date să se prelucreze cunoștințe privind interconexiunile dintre rețele sociale diferite, comunitatea de autori și comunitatea formată din bibliotecă și utilizatorii săi. Sistemul suport pentru decizii al bibliotecii tinde să devină, în mod natural, un actor foarte important în alimentarea cu cunoștințe a sistemelor suport pentru decizii ale companiilor. Articolul prezintă unele rezultate obținute în urma abordării formalizate a construirii unui sistem suport pentru decizii de bibliotecă.

Cuvinte cheie: bibliometrie, biblioteconomie, descriere bibliografică, depozitare date, explorare date și descoperire de cunoștințe, sistem suport pentru decizii.

A LIBRARY DECISION SUPPORT SYSTEM

Abstract. Decision Support Systems provide knowledge and knowledge processing capacity essential for referral decision situations and decision-making, improving decision-making and decision-making results and relaxing the cognitive, temporal, spatial and economic limits of the decision-makers. Libraries and librarians support in decisions-making varied in time from a passive, traditional collections of books and journals, to some highly active, decision assistants. Digital libraries have provided new insights for corporate decision support systems. Data mining techniques have become crucial for the management, organization and dissemination of information to the right users at the right time. Bibliomining provides the opportunity through a single data warehouse to compile knowledge on the interconnections between different social networks, the community of authors and the community made up of the library and its users. Library decision support system tends to become a naturally important actor in the supply of knowledge to companies decision support systems. The paper presents some results obtained from a formalized approach to build a library decision support system.

Key words: Bibliometrics, Library and Information Science, Bibliographic Record, Data Warehousing, Data Mining and Knowledge Discovery, Decision Support System.

1. Introducere

Timp de secole, factorii de decizie au folosit conținutul cărților, periodicelor, scrisorilor și altor documente ca depozite textuale de cunoștințe. Cunoștințele încorporate într-un fragment de text pot fi descriptive, procedurale sau de raționament. Indiferent de tipul acestora, factorii de decizie caută și selectează piese de text pentru a dobândi mai multe cunoștințe, pentru a verifica impresii sau pentru a stimula idei. Bibliotecile digitale au oferit perspective noi pentru sistemele suport pentru decizii [1, 5, 9, 11, 14, 15, 19, 20, 22, 23].

Provocările cu care se confruntă un sistem suport pentru decizii de bibliotecă sunt: elaborarea politicilor de achiziție și de diseminare orientate către cerere, optimizarea fluxurilor și alocării resurselor, îmbunătățirea conservării colecțiilor, îmbunătățirea satisfacției utilizatorilor, comunicare mai bună cu partenerii, diversificarea și creșterea veniturilor culturale și comerciale.

Obiectivele sistemului sunt: furnizarea de indicatori, de stare și de performanță care să permită evaluarea în timp a conformității cu obiectivele bibliotecii, furnizarea unor instrumente de analiză a tendințelor, de sesizare a situațiilor decizionale și de sugerare a unor acțiuni corespunzătoare în vederea luării deciziilor, integrarea datelor și simplificarea accesului prin schimb transparent și diseminare accelerată a informațiilor, asigurarea unor funcționalități de alimentator de cunoștințe pentru sistemele suport pentru decizii ale unor companii.

Seviciile oferite de sistem constau în: furnizarea de indicatori de stare și de performanță și de rapoarte predefinite; interogări avansate și interogări personalizate la cerere; analize avansate implicând navigare multidimensională; funcții puternice de analiză, simulări și statistici avansate.

2. Arhitectura sistemului

Arhitectura generală a sistemelor suport pentru decizii permite evidențierea diferențierilor dintre categoriile distincte de astfel de sisteme, arhitecturile personalizate păstrând caracteristicile sugerate de modelul conceptual generic dar fiind orientate către una sau mai multe tehnologii de reprezentare și de prelucrare a cunoștințelor. În funcție de tehnologia dominantă, sistemele suport pentru decizii pot fi orientate către: texte, hipertext, baze de date, foi electronice de calcul, reguli sau rezolvatoare [7, 19].

Dacă factorul decizional are nevoie de capacitățile de prelucrare oferite de mai multe tehnologii de management al cunoștințelor există două opțiuni de bază. Prima opțiune constă în utilizarea mai multor sisteme suport pentru decizii fiecare orientat către o anumită tehnologie. A doua opțiune constă în utilizarea unui singur sistem suport pentru decizii care integrează mai multe tehnologii.

Un caz special de integrare [3], reprezentând și arhitectura aleasă pentru sistemul suport pentru decizii al bibliotecii (Figura 1), a constat din integrarea rezolvatoarelor analitice (*online analytical processing*) și de descoperire a cunoștințelor (*data mining and knowledge discovery*) cu depozitarea datelor (*data warehousing*).

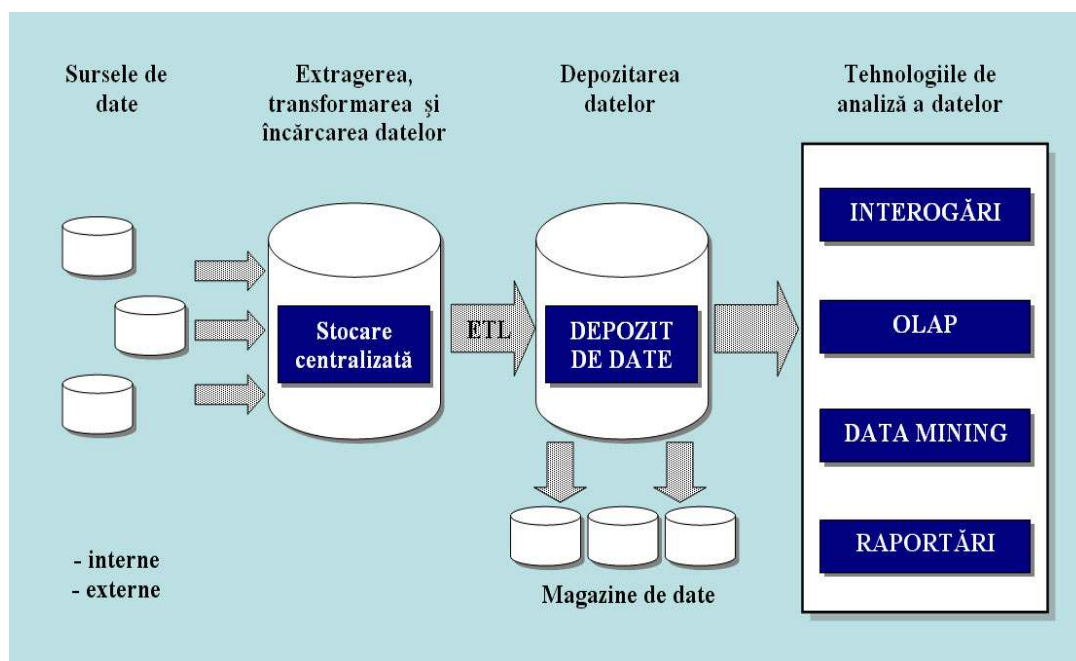


Figura 1. Arhitectura sistemului suport pentru decizii al bibliotecii

3. Depozitarea datelor

Proiectarea conceptuală a unui depozit de date este pasul cel mai important în reprezentarea corectă a unui domeniu de interes, fiind elementul esențial asupra căruia atât factorii de decizie cât și informaticienii sunt de acord. Este foarte important ca proiectanții unui depozit de date să urmeze o metodologie de proiectare conceptuală, consolidată și robustă dat fiind că dezvoltarea unui depozit de date este un proces foarte scump chiar și astăzi când există multe instrumente software oferind soluții prefabricate care acoperă toate etapele din ciclul de viață al depozitului de date [4, 8, 15, 20, 23, 25].

Există cel puțin două noțiuni specifice pe care orice model conceptual pentru baze de date multidimensionale trebuie să le includă într-o anumită formă și anume faptul și dimensiunea. Faptul este o entitate, a unei aplicații, care face obiectul unei analize orientată către decizie, reprezentabilă prin intermediul cubului de date, iar dimensiunea corespunde perspectivei din care faptele pot fi analizate în mod concludent. Aspectele specifice și măsurabile ale unui fapt, relevante pentru analiză, sunt numite măsuri.

Eficiența modelării în depozitarea datelor depinde strict de capacitatea de a descrie datele factuale în funcție de dimensiunile adecvate, adică în funcție de perspectivele din care datele pot fi analizate. Pentru a putea susține mai bine analiza datelor este util ca pentru fiecare dimensiune să se organizeze o ierarhie de niveluri obținabilă prin gruparea elementelor dimensiunii în funcție de nevoile analizei. Un nivel are asociate, de obicei, atribute descriptive (nume și descrieri). O dimensiune are deci trei componente principale: o mulțime de niveluri, o mulțime de descrieri de niveluri și o ierarhie între niveluri.

O colecție de măsuri ale aceluiași fapt este reprezentată prin metafora „cub de date”, având câte o dimensiune „fizică” pentru fiecare dimensiune „conceptuală” a măsurării: o coordonată a cubului de date specifică o combinație de membri ai nivelurilor mai multor dimensiuni iar celula corespunzătoare conține măsura (măsurile) asociată (asociate) unei astfel de combinații. În cazul în care membrii unui nivel pot fi agregați în membri ai unui alt nivel se spune că primul nivel se agregă la cel de al doilea nivel.

Proiectarea conceptuală a unui depozit de date poate fi realizată prin mai multe categorii de metode: orientate către date, orientate către cerințe și metode mixte sau hibride. Până în prezent rezultatele cele mai promițătoare au fost obținute prin metodele hibride secvențiale, metode care combină și integrează o etapă de abordare orientată către date cu o etapă de abordare orientată către cerințe și în care cele două etape sunt executate într-o ordine prefixată, ieșirea primei etape fiind utilizată ca intrare în a doua etapă [19, 26]. Pe scurt, etapele generale ale unei astfel de metode sunt: analiza cerințelor informaționale, reconcilierea cu sursele de date, modelarea multidimensională a datelor, generarea arborilor de atribute și remodelarea datelor.

4. Cerințe informaționale

Stabilirea cerințelor informaționale s-a bazat pe prevederile normative specifice domeniului bibliotecilor elaborate, recomandate și utilizate atât pe plan intern cât și pe plan internațional privind: descrierile bibliografice [2]; indicatorii operaționali [10]; indicatorii de performanță [12] și indicatorii bibliometrici [6, 24]. Practic, în urma unei abordări formalizate a construirii sistemului suport pentru decizii de bibliotecă [19, 20], a fost definit un sistem unitar, coerent și evolutiv de indicatori.

Cerințele bibliografice. Descrierea bibliografică a unui document este considerată ca o mulțime de informații privind patru aspecte diferite ale documentului descris: caracteristicile individuale ale unui exemplar al documentului, caracteristicile publicației sau manifestării de care aparține, caracteristicile expresiei sau conținutului intelectual/artistic și caracteristicile lucrării/creației abstracte la care se referă acest conținut. La fiecare din aceste patru niveluri de analiză, documentul descris este pus în relație cu o persoană sau cu o colectivitate/grup care a intervenit într-un mod specific la acest nivel. Fiecare dintre aceste noțiuni, la care se mai adaugă loc, eveniment, obiect și concept pot constitui subiecte ale unei lucrări.

Principalele entități utilizate în descrierile bibliografice ale documentelor/publicărilor ($d \in D$), sunt: autorii ($a \in A$), grupurile de autori sau centrele de creație artistică sau de cercetare la care sunt afiliați autorii ($g \in G$), publicațiile ($p \in P$), editorii/responsabilii de ediții ($r \in R$), editurile ($e \in E$), (sub)domeniile de creație artistică sau de cercetare ($c \in C$) și subiectele ($s \in S$) ca părți ale unui (sub)domeniu.

Cerințele biblioteconomice. Sistemul de indicatori operaționali ai bibliotecii permite descrierea stării curente și/sau dorite a sistemului instituției la un moment dat. Managementul performant al instituției se bazează pe un set de obiective fixate și un sistem de indicatori de performanță asociați, menite să faciliteze sesizarea situațiilor decizionale, identificarea acțiunilor posibile de urmat și luarea deciziilor.

Cerințele bibliometrice. Cercetătorii, ca autori, elaborează diverse tipuri de lucrări care conțin rezultate experimentale, teorii, recenzii etc. Cea mai simplă metodă pentru a măsura productivitatea unui cercetător sau a unui grup de cercetare este de a număra documentele publicate (publicările) de un anumit autor sau de un anumit grup dintr-un anumit interval de timp. Indicatorii destinați pentru a

măsura productivitatea cercetătorilor sau a grupurilor de cercetare sunt considerați indicatori cantitativi. Pentru a raporta constatările lor la rezultate anterioare autorii citează alte lucrări. Citările furnizează date care pot fi utilizate pentru a măsura statistic și matematic importanța relativă a unui articol sau a unei publicații precum și conectivitatea dintre domenii științifice, dintre departamente de cercetare sau dintre autori. Modalitatea actuală de a măsura impactul lucrărilor științifice se bazează pe numărul de citări. Indicatorii care ajută la identificarea nivelului de calitate al lucrărilor unui cercetător sau ale unui grup de cercetare și pot fi utilizați pentru a evalua impactul cercetărilor în comunitatea științifică sunt considerați indicatori de performanță.

5. Sursele de date

Pentru depozitul de date al unui sistem suport pentru decizii de bibliotecă principalele surse de date pot fi: bazele de date operaționale ale bibliotecii; bazele de date bibliografice respectiv cataloagele on-line de bibliotecă; biblioteci digitale precum *Europeana*, *WDL* ș.a.; publicații și baze de date internaționale, on-line, pentru documentare precum *Science Direct*, *SpringerLink*, *Wiley Blackwell*, *Taylor and Francis*, *DOAJ*, etc.; baze de date bibliometrice precum *ISI Web of Science* (Thomson Reuters), *Scopus* (Elsevier), *Google Scholar* (Google), ș.a.

Reconcilierea cerințelor informaționale privind descrierile bibliografice cu sursele de date se realizează prin maparea informațiilor existente în sursele de date pe informațiile considerate necesare în modelul *FRBR*. În raport cu cerințele respective sursele de date oferă mai puține informații obținându-se practic surrogate ale descrierilor bibliografice [19].

6. Modelarea multidimensională a datelor

Identificare fapte. Pentru mediul decizional al unei biblioteci subiectele majore de interes sunt:

- ↪ serviciile de bibliotecă,
- ↪ aparițiile editoriale,
- ↪ calitatea publicațiilor.

Definire dimensiuni. Perspectivele de analiză necesare pentru faptele identificate sunt:

- ↪ pentru serviciile de bibliotecă: timpul, operațiile și utilizatorii;
- ↪ pentru aparițiile editoriale: timpul, publicările, autorii, editorii, publicațiile, subiectele;
- ↪ pentru calitatea publicațiilor: timpul, publicările, autorii, referințele, publicațiile, subiectele.

Definire măsuri. Aspectele specifice și măsurabile ale faptelor, relevante pentru analiză, la nivelul minim de granularitate, sunt [19, 20]:

- ↪ la serviciile de bibliotecă: indicii de selecție (s), de realizare (r), duratele (τ) și costurile (c) unitare ale operațiilor;
- ↪ la aparițiile editoriale: indicii de publicare (π) și de cotare (ψ);
- ↪ la calitatea publicațiilor: indicii de notorietate (ε), de impact (ρ) și de relevanță a citărilor (σ).

Definire cuburi de date. Schemele cuburilor de date/arborilor de atribute sunt reprezentate prin diagrame specifice în care faptele sunt reprezentate prin dreptunghiuri, dimensiunile sunt reprezentate prin dreptunghiuri rotunjite, măsurile sunt reprezentate prin cercuri.

În continuare sunt reprezentate cuburile de date cu dimensiunile corespunzătoare faptelor identificate (servicii, publicări și citări).

Schema cubului de date „Servicii de bibliotecă”

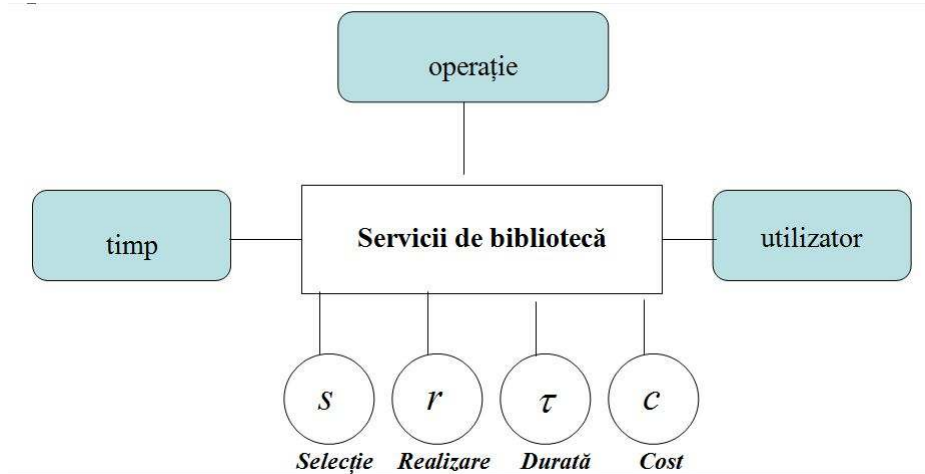


Figura 2. Cub de date privind serviciile bibliotecare

Schema cubului de date „Publicări”:

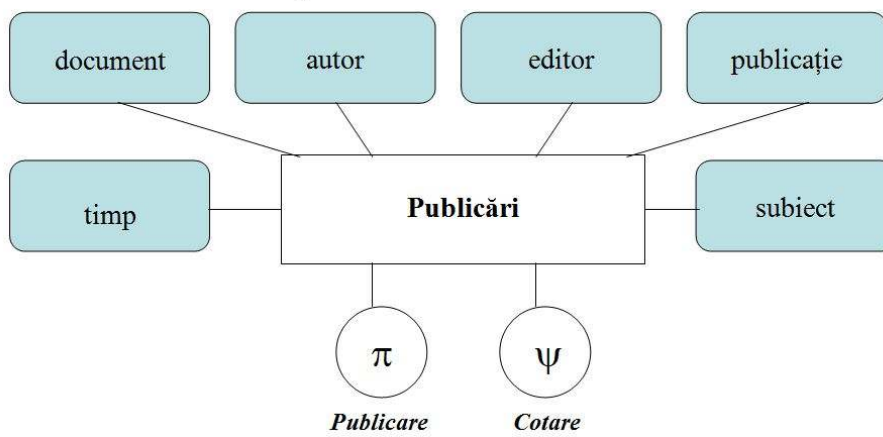


Figura 3. Cub de date privind aparițiile editoriale

Schema cubului de date „Citări”

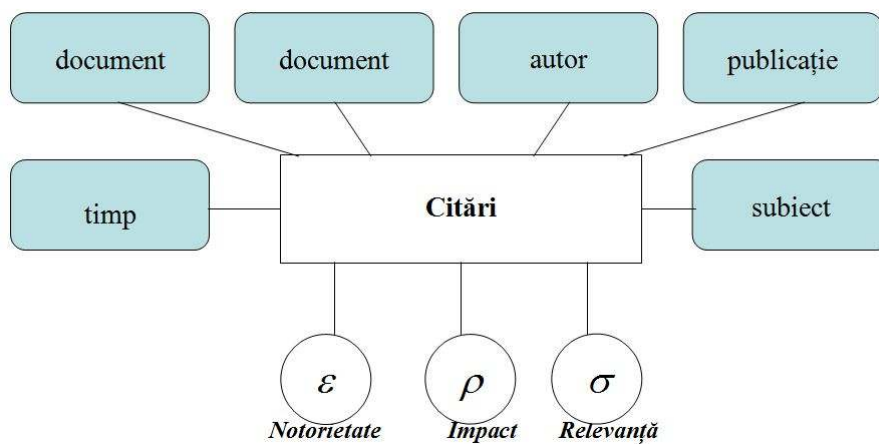


Figura 4. Cub de date privind calitatea publicărilor

Schema conceptuală a depozitului de date

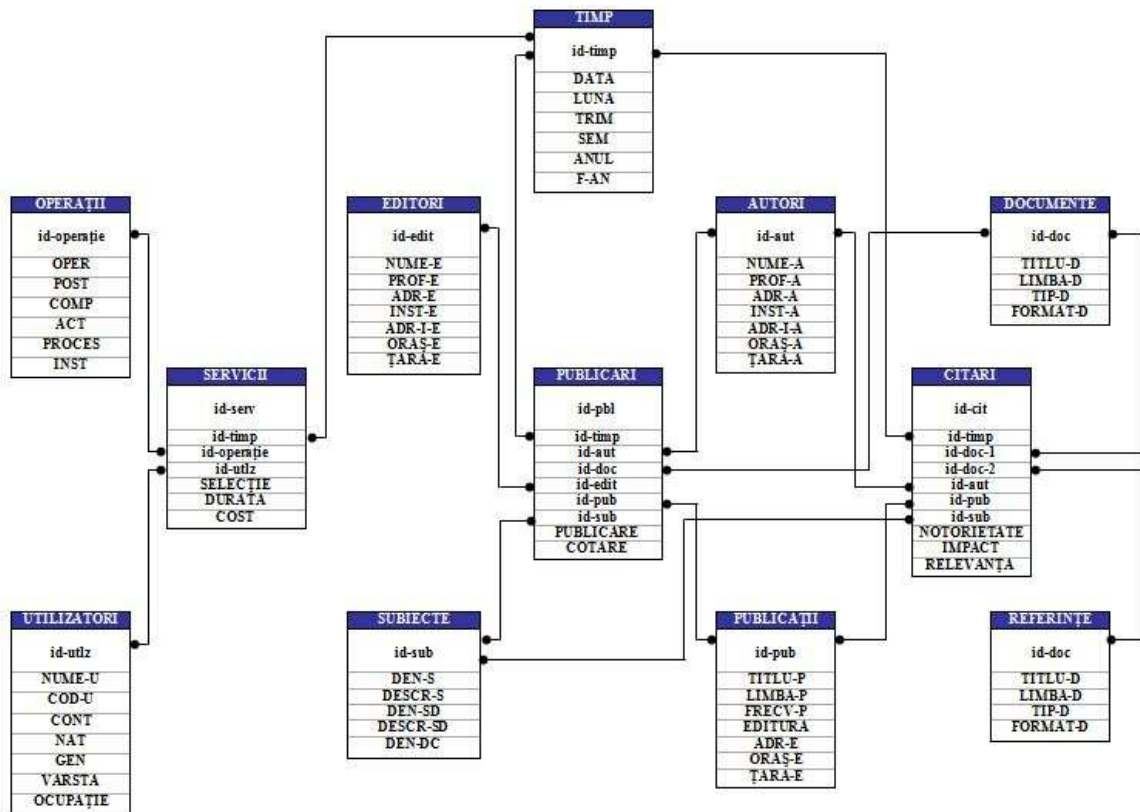


Figura 5. Schema „constelație” a depozitului de date

7. Explorarea datelor și descoperirea cunoștințelor

Tehnologia data mining permite descoperirea de pattern-uri structurale din date utilizând algoritmi suficient de robusți atât pentru a prelucra date imperfecte, corelate stohastic, cât și pentru a extrage corelații, uneori imprecise și reguli utilizabile ulterior în predicția, explicarea și înțelegerea evoluției structurii datelor analizate. Aportul data mining se rezumă la un număr limitat de acțiuni care, folosite în mod adecvat, se pot dovedi extrem de utile pentru numeroase probleme și situații din domeniul decizional. Între tipurile de probleme, rezolvabile cu data mining, cele mai frecvente sunt: analiza asocierilor, pattern-uri secvențiale, analiza grupurilor, clasificare, mulțimi rough, link mining.

Ceea ce se exploatează prin data mining sunt colecții de date disponibile, de volum mare sau foarte mare, provenite din surse interne ale organizației care au fost constituite, inclusiv ca structură, în perspectiva unor finalități și la care se adaugă date provenite din diverse alte surse externe organizației. Utilizarea data mining presupune: identificarea oportunității acestora și a datelor pe care se poate baza explorarea; extragerea informațiilor din depozitul/depozitele de date existente și prelucrarea acestora prin tehnici adecvate de data mining; adoptarea de decizii pe baza rezultatelor obținute și întreprinderea de acțiuni; măsurarea rezultatelor concrete pentru a identifica și alte modalități de exploatare a datelor disponibile [4, 9, 13, 14, 16, 17, 18, 19, 20, 21, 23].

Datele disponibile sunt privite ca reprezentând o serie de observații privind un set de caracteristici/variabile, care au fost măsurate pe un eșantion de obiecte/indivizi. Există două tipuri de variabile: explicative și de explicat. Mulțimea de variabile explicative/predictive, este constituită din variabile, fie toate cantitative, fie toate calitative, fie mixte. De asemenea, variabilele de explicat sau de predicție sau țintă, pot fi: cantitative și calitative cu două sau mai multe modalități.

Un prim demers constă în efectuarea unei explorări a acestor date: alura distribuțiilor, prezența datelor atipice, corelații și coerență, transformări eventuale ale datelor; clasificare. Demersul descriptiv și exploratoriu permite realizarea de rezumate și grafice mai mult sau mai puțin elaborate, descrierea mulțimilor de date și stabilirea de relații între variabile, fără a acorda un rol privilegiat vreunei variabile. Demersul exploratoriu se sprijină, în mod esențial, pe noțiuni elementare (medie și dispersie), pe reprezentări grafice și pe tehnici descriptive multidimensionale. Metodele exploratorii caută subspațiile de reprezentare (factoriale) de dimensiuni mici, care aproximează cel mai bine norii de puncte-indivizi sau de puncte-variabile, astfel încât vecinătățile măsurate în aceste spații să reflecte cât mai exact proximitățile reale. În demersul descriptiv și exploratoriu obiectivele principale urmărite sunt: explorarea multidimensională (bazată, cel mai frecvent, pe analiza în componente principale, analiza factorială discriminantă, analiza corespondențelor simple, analiza corespondențelor multiple și analiza canonică) și clasificarea (utilizând cel mai adesea metode precum: metoda ascendentă ierarhică și metoda norilor dinamici sau o metodă mixtă).

Un al doilea demers îl constituie modelarea, în scopul predicției unei/unor variabile țintă, prin variabilele explicative utilizând instrumente de modelare/învățare. Demersul inferențial și confirmatoriu permite validarea/infirmary, pornind de la teste statistice sau modele probabiliste, a ipotezelor formulate a priori (adică urmare a unui demers exploratoriu) și extrapolarea acestora de la nivelul eșantionului la cel al unei populații mai mari. Demersul confirmatoriu face apel, în special, la metodele numite explicative și previzionale destinate să explice apoi să prevadă, urmând anumite reguli de decizie, o variabilă privilegiată cu ajutorul uneia sau mai multor variabile explicative. În demersul inferențial și confirmatoriu obiectivul principal urmărit îl constituie modelarea/discriminarea respectiv deducerea unui model de previziune pentru variabila (variabilele) țintă. Metodele cele mai frecvent utilizate în atingerea acestui obiectiv sunt: modelele liniare, metodele conexioniste, analiza discriminantă, mașinile cu suport vectorial, metodele de segmentare, agregarea modelelor (*Bagging*, *Boosting*, *Random Forest*).

Demersurile sunt complementare, explorarea și descrierea trebuind, în general, să precedă etapele explicative și predictive. O explorare preliminară este adesea utilă pentru a avea o primă idee despre natura legăturilor între variabile și pentru a trata cu prudență variabilele corelate, și deci redundante, ce riscă să încarce inutil modelul. Succesiunea acestor două demersuri, explorare și apoi învățare, constituie fundamentul utilizării data mining.

Pentru a se oferi șanse mai favorabile de succes unui proces data mining este evident că preocupările legate de definirea obiectivelor și de analiză a datelor ar trebui să intervină cât mai devreme posibil. O practică bună de data mining necesită din partea asistenților decizionali să știe să articuleze toate metodele sarcină care nu poate fi îndeplinită decât cu condiția de a avea foarte bine clarificate obiectivele studiului. Spre deosebire de abordarea statistică tradițională în care observarea datelor este integrată în metodologie (planificarea experimentului) în data mining datele sunt prealabile analizei. Prezența unei expertize statistice rămâne inevitabilă pentru că necunoașterea limitelor și capcanelor metodelor utilizate poate conduce la aberații de natură să discrediteze demersul, făcând caduce investițiile consimțite.

8. Nucleu de autoritate

Definirea de nucleu de autoritate în diferite mulțimi de entități implică definirea, cât mai simplu posibil, a unor proceduri generale pentru a genera ierarhii. Ideea fundamentală a abordării [19] este de a considera entitățile analizate ca părți ale unui sistem, fiecare entitate putând fi caracterizată de către celelalte. Elementul de legătură este mulțimea de documente. Compararea a două entități înseamnă să se selecteze din mulțimea de documente o secțiune corespunzătoare și să se efectueze comparația în interiorul acelei secțiuni. În continuare sunt schițate două proceduri, prima pentru definirea unei ierarhizări a subiectelor în raport cu interesul utilizatorilor și a doua pentru definirea unei ierarhizări a autorilor care au tratat un anumit subiect, definirea de astfel de proceduri pentru alte entități precum publicațiile, editurile sau organizațiile științifice fiind asemănătoare.

Procedura 1. Fie $S(c)$ mulțimea subiectelor dintr-un anumit (sub)domeniu de interes c și $D(c, T)$ mulțimea documentelor care tratează subiectele $s \in S(c)$ consultate în perioada de timp analizată T . Se dorește o ierarhizare în interiorul mulțimii $S(c)$ pe perioada T .

Se presupune că accesarea unui document d de către un utilizator se realizează prin una din următoarele trei tipuri de operații: prin acces on-line la documentul d , $\theta^1_d \in \Theta^1$, împrumut $\theta^2_d \in \Theta^2$ sau prin împrumut interbibliotecar $\theta^3_d \in \Theta^3$. Indicele $r(\theta^i_d, \Theta^i)$ ($i \in \{1, 2, 3\}$) de realizare al operației θ^i_d este egal cu 1 dacă operația $\theta^i_d \in \Theta^i$ a fost realizată, în caz contrar fiind 0. Pentru punerea în aplicare a procedurii se presupune că mulțimea documentelor care abordează fiecare subiect s nu este vidă, $D(s) \neq \emptyset$ ($\forall s \in S(c)$) și că fiecare document a fost consultat cel puțin o dată, ($\forall s \in S(c)$) și ($\forall d \in D(s)$) ($\exists \theta_d \in \Theta = \bigcup_{i=1}^3 \Theta^i$: $r(\theta^i_d, \Theta^i) = 1$).

Pasul-1. Inițializarea procedurii

c := subdomeniul de interes

$S(c)$:= mulțimea de subiecte s din subdomeniul c

Pasul-2. Selectarea secțiunii de documente care tratează subiectele din c

$D := D(S(c)) = \bigcup_{s \in S(c)} D(s)$ unde: $D(s)$ = mulțimea documentelor cu subiectul $s \in S(c)$

Pasul-3. Selectarea mulțimilor operațiilor de consultare a documentelor $d \in D$ realizate în perioada de timp analizată T

$\Theta^i := \{O(i, s, T) \equiv O(\Theta^i_s, T) | s \in S(c)\}$ unde: $i \in \{1, 2, 3\}$; $\Theta^i_s = \{\theta_d | \theta_d \in \Theta^i, d \in D(s), t(\theta_d) \in T\}$

Pasul-4. Generarea valorilor funcțiilor de realizare a operațiilor de consultare pentru fiecare subiect și tip de consultare

$R^i := \{R(\Theta^i_s, T) | s \in S(c)\}$, unde: $i \in \{1, 2, 3\}$ și $R(\Theta^i_s, T) = \sum_{\theta \in O(i, s, T)} r(\theta, \Theta)$

Pasul-5. Generarea valorilor funcțiilor de realizare a operațiilor de consultare pentru fiecare subiect $s \in S(c)$

$R := \{R(\Theta_s, T) | s \in S(c)\}$, unde: $R(\Theta_s, T) = \sum_{i=1}^3 R(\Theta^i_s, T)$

Pasul-6. Furnizarea, în ordine descrescătoare, a valorilor din R

Sort descending $\{R(\Theta_s, T) | s \in S(c)\}$

Pasul-7. Finalizare

Delimitare nucleu de autoritate; Interpretări; Alte prelucrări.

Procedura 2. Fie s un anumit subiect de interes și fie \mathcal{S} o mulțime de subiecte care conține pe s . Mulțimea \mathcal{S} conține subiectul s dar include și alte subiecte apropiate din același (sub)domeniu pentru a evita căutarea într-un set prea îngust de subiecte. Se dorește o ierarhizare în interiorul mulțimii de autori în subiectul s . Pentru punerea în aplicare a procedurii se presupune că mulțimea documentelor care abordează subiectul s nu este vidă, $D(s) \neq \emptyset$ și, de asemenea, că există informațiile necesare, listele de clasificare notorietate-expert și mulțimile de ponderi.

Pasul-1. Inițializarea procedurii

s := subiectul de interes

\mathcal{S} := mulțime de subiecte ce conține pe s

\mathcal{A} := listele autorizate cu autorii de notorietate

\mathbf{w} := mulțimea de valori pentru ponderi

Pasul-2. Selecția documentelor care tratează subiectul s

$D(s) := \{d(s)\}$

Pasul-3. Selecția documentelor care citează documente din $D(s)$

$R(D(s)) := \{d' | \rho(d, d') = 1, d \in D(s)\}$

Pasul-4. Generarea secțiunii de documente privind subiectul s

$\mathcal{D} := D(s) \cup R(D(s))$

Pasul-5. Generarea mulțimii de autori

$\mathbf{A} := A(\mathcal{D}) \cup \mathcal{A}$, unde: $A(\mathcal{D}) = \{a | a \in A(d); d \in \mathcal{D}\}$; $\mathcal{A} = \{a | \alpha(a) > 0\}$

Pasul-6. Determinarea indicilor de notorietate \mathcal{E} pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathbf{A}

$N(\mathcal{S}, \mathbf{A}) := \{ \mathcal{E}(D(\mathcal{S}, a)) \mid a \in \mathbf{A} \}$, unde:

$D(\mathcal{S}, a) = D(\mathcal{S}) \cap D(a)$; $\mathcal{E}(D(\mathcal{S}, a)) = \sum_{d \in D(\mathcal{S}, a)} (w_E \mathcal{E}(E(d)) + w_P \mathcal{E}(P(d)))$

Pasul-7. Determinarea funcțiilor de impact I pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathbf{A}

$I(\mathcal{S}, \mathbf{A}) := \{ I(D(\mathcal{S}, a), t) \mid a \in \mathbf{A} \}$, unde: $D(\mathcal{S}, a) = D(\mathcal{S}) \cap D(a)$; $I(D(\mathcal{S}, a), t) = \sum_{d \in D(\mathcal{S}, a)} I(d, t)$

Pasul-8. Determinarea indicilor de notorietate-impact $\mathcal{E}I$ pentru submulțimile de documente cu subiecte din \mathcal{S} elaborate de autorii din \mathbf{A}

$\gamma(\mathcal{S}, \mathbf{A}) := \{ \mathcal{E}I(D(\mathcal{S}, a)) \mid a \in \mathbf{A} \}$, unde: $\mathcal{E}I(D(\mathcal{S}, a)) = w_1 \mathcal{E}(D(\mathcal{S}, a)) + w_2 I(D(\mathcal{S}, a), t)$

Pasul-9. Furnizarea, în ordine descrescătoare, a valorilor din $\gamma(\mathcal{S}, \mathbf{A})$

Sort descending ($\gamma(\mathcal{S}, \mathbf{A})$)

Pasul-10. Finalizare

Delimitare nucleu de autoritate; Interpretări; Alte prelucrări.

9. Grupare bazată pe densitate

Gruparea constă în identificarea de grupuri sau clustere într-o mulțime de date. Clusterizarea bazată pe densitate încearcă să identifice și să separe regiunile dense (foarte populate) ale unei mulțimi de puncte, p , dintr-un spațiu multidimensional [13].

Densitatea este definită ca fiind numărul de puncte existente într-o regiune de căutare, de dimensiune *eps* specificată, numită *eps-vecinătate*. Un punct este considerat *punct de bază* dacă *eps-vecinătatea* sa conține mai multe puncte decât un număr, *MinPts*, specificat. Punctele de bază sunt în interiorul unui cluster. Un punct este *punct de frontieră* dacă *eps-vecinătatea* sa conține un număr de puncte mai mic decât *MinPts* dar se află în *eps-vecinătatea* unui punct de bază. Un punct este considerat *punct de zgomot* dacă nu este nici punct de bază și nici punct de frontieră.

Definirea clusterelor se bazează pe noțiunea de accesibilitate în densitate. Un punct Q este *direct accesibil* în densitate dintr-un alt punct P dacă Q este conținut în *eps-vecinătatea* lui P și dacă P este punct de bază. P și Q fac parte din același cluster. Un punct Q este *accesibil* în densitate dintr-un alt punct P dacă există o secvență de puncte P_1, \dots, P_n cu $P_1 = P$ și $P_n = Q$ în care fiecare punct P_{i+1} , $i = 1 \div n - 1$, este direct accesibil în densitate din punctul P_i .

Relația de accesibilitate în densitate nu este simetrică. Dacă punctul Q ar fi situat la frontiera unui cluster, având deci insuficient de multe puncte vecine pentru a conta ca punct de bază, procesul de găsire a unui drum între Q și P s-ar întrerupe acesta oprindu-se la primul punct care nu este punct de bază. Începând însă procesul din punctul P drumul ar ajunge la Q și procesul s-ar opri aici, acesta fiind primul punct găsit care nu este punct de bază. Datorită acestei asimetrii, a fost necesară introducerea noțiunii de *conectare* în densitate.

Două puncte P și Q sunt *conectate* în densitate dacă există un punct O astfel încât ambele puncte P și Q sunt accesibile în densitate din O . Conectarea în densitate este simetrică. Un *cluster* este o submulțime de puncte a lui \mathcal{P} care satisface două proprietăți: toate punctele din cluster sunt reciproc conectate în densitate și dacă un punct este conectat în densitate cu orice alt punct din cluster atunci aceasta aparține clusterului.

În continuare este schițată o procedură de grupare bazată pe densitate derivată din [13]. Parametrii *eps* și *MinPts* reprezintă o estimare a densității punctelor din clustere și trebuie să fie specificați de către utilizator. Dacă *eps* este ales prea mic, o mare parte din puncte nu vor fi grupate în timp ce, pentru o valoare prea mare, grupurile vor fuziona și majoritatea punctelor vor fi în același cluster.

Pasul-1. Inițializarea procedurii

eps := distanța aleasă pentru *eps-vecinătate* (regiuneadeCăutare)

$MinPts$ = numărul minim de puncte ales pentru a forma un cluster

Pasul-2. Prelucrări preliminare

se selectează elementele mulțimii P

se determină coordonatele punctelor $P \in P$ în spațiul multidimensional

se determină distanțele dintre puncte

Pasul-3. Clusterizare($P, eps, MinPts$)

$C = 0$

pentru fiecare punct „nevizitat” P din mulțimea P

se marchează P ca „vizitat”

regiuneadeCăutare($P, eps, NrPts(P), Pts(P)$)

dacă $NrPts(P) < MinPts$

expandareZgomote($Pts(P), eps, MinPts, 0$)

altfel

$C = C+1$ (clusterul următor)

expandareCluster($P, eps, MinPts, C$)

↪ **expandareCluster**($P, eps, MinPts, C$)

se adaugă punctul P la clusterul C ($P \in C$)

pentru fiecare punct Q din mulțimea P

dacă $Q \in C$ și Q nu este „membru” al clusterului C

se marchează Q ca „membru” al clusterului C

regiuneadeCăutare($Q, eps, NrPts(Q), Pts(Q)$)

clusterul $C =$ clusterul C reunit cu $Pts(Q)$

se marchează punctele din $Pts(Q)$ ca „vizitate”

↪ **expandareZgomote**($Pts(P), eps, MinPts, 0$)

pentru fiecare punct Q din $Pts(P)$

regiuneadeCăutare($Q, eps, NrPts(Q), Pts(Q)$)

dacă (\forall) $Q \in Pts(P), NrPts(Q) < MinPts$

clusterul $0 =$ clusterul 0 reunit cu $Pts(Q)$

se marchează punctele din $Pts(P)$ ca „zgomote” (membri al „clusterului” 0)

se marchează punctele din $Pts(Q)$ ca „vizitate”

↪ **regiuneadeCăutare**($P, eps, NrPts(P), Pts(P)$)

se returnează numărul $NrPts(P)$ și punctele $Pts(P)$ din eps -vecinătatea punctului P , inclusiv P

Pasul-4. Finalizare

se returnează pentru fiecare cluster descoperit punctele din cluster

Exemplu. Fie $S = \{s_i \mid i = 1 \div 23\}$ mulțimea subiectelor accesate de utilizatorii bibliotecii într-un interval de timp analizat. Pentru fiecare subiect s_i sunt cunoscute: numărul de documente x_i care tratează subiectul $s_i \in S$ precum și numărul total de accesări y_i ale subiectului respectiv. Mulțimea $P = \{P_i(x_i, y_i) \mid i = 1 \div 23\}$ formează un nor de puncte în spațiul bidimensional. Aplicarea procedurii, cu $eps = 2,237$ și $MinPts = 3$, evidențiază zgomotele (☼) și clusteretele (🌸, ✨, ★).

$P = \{P_1(4, 29), P_2(5, 31), P_3(5, 34), P_4(6, 28), P_5(7, 24), P_6(7, 26), P_7(7, 37), P_8(8, 28), P_9(8, 34), P_{10}(9, 31), P_{11}(9, 36), P_{12}(9, 38), P_{13}(10, 29), P_{14}(10, 34), P_{15}(11, 37), P_{16}(13, 30), P_{17}(13, 32), P_{18}(14, 35), P_{19}(15, 37), P_{20}(16, 34), P_{21}(17, 36), P_{22}(18, 30), P_{23}(18, 32)\}$.

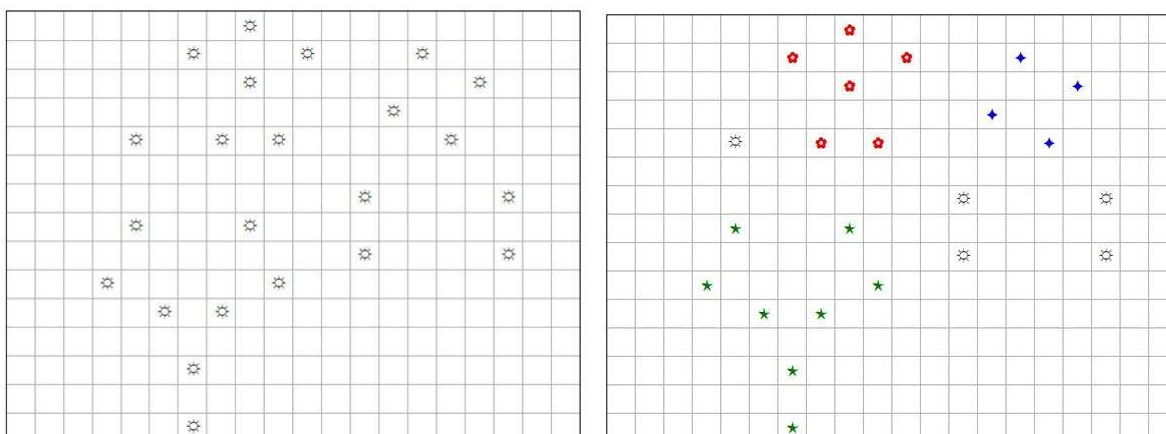


Figura 6. Norul de puncte $P \subset \mathbb{R}^2$ înainte și după procesul de grupare

10. Concluzii

Concepția și implementarea sistemului suport pentru decizii al unei biblioteci, ca de altfel ale oricărui sistem informatic, sunt influențate de către o serie de factori, între care pot fi menționați: obiectivele urmărite; personalul implicat; recomandările, normele și standardele utilizate; restricțiile impuse de către instituție; evoluția mediului; bugetul disponibil pentru realizare; termenele de finalizare.

Domeniile de activitate ale bibliotecii, posibile de îmbunătățit, sunt: dezvoltarea colecțiilor; accesibilitatea colecțiilor; accesul la publicații; utilizarea bibliotecii; digitalizarea colecțiilor; serviciile bibliografice, asistența și îndrumarea; potențialul de dezvoltare; conservarea colecțiilor; managementul. Pentru ca serviciile oferite de sistem să poată satisface cerințele fiecărei categorii de utilizatori principalele categorii de servicii sunt: căutările simple, care furnizează rapoarte predefinite și valori ale indicatorilor operaționali și/sau de performanță; interogările avansate și/sau personalizate; analizele avansate, care implică navigare multidimensională și funcții puternice de analiză; simulările și statisticile avansate.

Arhitectura sistemului suport pentru decizii al unei biblioteci este o combinație dintre o tehnologie de management a rezolvatoarelor flexibile și o tehnologie de management a bazelor de date menită să asigure integrarea depozitării datelor cu rezolvatoarele analitice și data mining. Pentru stabilirea cerințelor informaționale se impune aplicarea cu discernământ a prevederilor normative specifice domeniului bibliotecilor elaborate, recomandate și utilizate atât pe plan intern cât și pe plan internațional privind: descrierile bibliografice; indicatorii operaționali; indicatorii de performanță și indicatorii bibliometrici. Practic trebuie definit un sistem personalizat, unic, unitar, coerent și evolutiv de indicatori. În realizarea sistemului suport pentru decizii al bibliotecii este foarte important ca proiectanții să urmeze o metodologie de proiectare integrată, consolidată și robustă dat fiind că dezvoltarea acestuia este un proces foarte scump chiar în condițiile actuale când există instrumente software care oferă soluții prefabricate acoperind practic toate etapele din ciclul de viață al sistemului.

Pentru bibliotecă avantajele majore ale sistemului suport pentru decizii sunt: asigură informații de calitate și noi instrumente de management; rezolvă faze tehnice critice privind furnizarea, modelarea și stocarea datelor; satisface cerințe tehnice actuale și viitoare; satisface cerințele utilizatorilor; este adaptabil; susține trecerea la o cultură orientată către performanță și impune personalului dezvoltarea în consecință a abilităților; promite mari îmbunătățiri în modul de înțelegere a ceea ce se face în prezent și a ceea ce se preconizează pentru viitor. Pentru companii avantajele sistemului suport pentru decizii al bibliotecii constau în asigurarea unor puternice funcționalități de alimentator de cunoștințe pentru sistemele suport pentru decizii ale acestora prin diseminarea informațiilor/cunoștințelor către utilizatorii potriviți la momentul potrivit.

BIBLIOGRAFIE

- [1] BA, S.; HINKKANEN, A.; WHINSTON A. B. (1994) Digital Library as a Foundation for Decision Support Systems. In *Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries*, College Station, TX, Ed. J. Schnase, J. Leggett, R. Furuta, and T. Metcalfe, pp. 170 – 177.
- [2] BNF - Bibliothèque nationale de France (2012) *Fonctionnalités requises des notices bibliographiques : rapport final*. Trad. de : „Functional requirements for bibliographic records : Final Report” - 2e édition française, Paris, BNF, 100 p.
- [3] BURSTEIN, F.; HOLSAPPLE, C.W. (Eds) (2008), *Handbook on Decision Support Systems 1 : Basic Themes*, International Handbooks on Information Systems, Springer-Verlag, 854 p.
- [4] DEVA SARMA, P.K.; ROY, R. (2010) A Data Warehouse for Mining Usage Pattern in Library Transaction Data. In *Assam University Journal of Science & Technology : Physical Sciences and Technology*, Vol. 6, No. 2, pp. 125 – 129.
- [5] DUMITRESCU, G.; FILIP, F.-G.; IONIȚĂ, A.; LEPĂDATU, C. (2010). Open Source Eminescu’s Manuscripts: A Digitization Experiment. In *Studies in Informatics and Control*, vol 19 , no 1, pp 79 – 84.
- [6] DURIEUX, V.; GEVENOIS, P. A. (2010) Bibliometric indicators: quality measurements of scientific publication. In *RADIOLOGY*, 255 (2), pp. 342 – 351.
- [7] FILIP, F.-G. (2007) *Sisteme suport pentru decizii*, Ed. a 2-a, București: Editura Tehnică, 364 p.
- [8] GOLFARELLI, M.; RIZZI, S. (2009) *Data Warehouse Design: Modern Principles and Methodologies*, McGraw-Hill, 445 p.
- [9] HOMAYOUNVALA, E.; JALALIMANESH, A. (2012) Promoting research collaboration based on data mining techniques in library information systems. In *International Journal of Information Technology and Business Management*, Vol.8, No. 1, pp. 73 – 82.
- [10] INS - Institutul Național de Statistică (2012) Cult1 – Activitatea bibliotecilor. În *Chestionare statistice, Statistica culturii*. (<http://www.insse.ro/cms/files/chestionare/cult/CULT1%202012.pdf>)
- [11] IONIȚĂ, A.; LEPĂDATU, C.; DUMITRESCU, G. (2009) Digital Cultural Landscape Content. În: HERNIK, Jozef (edit.) *Cultural Landscape – Across Disciplines*, Oficyna Wydawnicza BRANTA, Kracow, pp. 255 – 277.
- [12] ISO (2009) *TR28118 Information and documentation – Performance indicators for national libraries*. (http://www.iso.org/iso/home/store/catalogue_ics/)
- [13] KRIEGEL, H.-P.; KRÖGER, P.; SANDER, J.; ZIMEK, A. (2011) Density-based clustering. In *WIREs Data Mining and Knowledge Discovery*, 1(3), pp. 231–240.
- [14] LEPĂDATU, C. (2007) Acquisition Policy of a Library and Data Mining Techniques. În: *Studies in Informatics and Control*, 2007, vol. 16, nr. 4, pp. 413 – 420.
- [15] LEPĂDATU, C. (2008) Sistem suport pentru decizii în cultura cunoașterii. In: *Revista română de biblioteconomie și științele informării*, Anul 4, nr. 2, pp. 41 – 45.
- [16] LEPĂDATU, C. (2011) *Soluții informatice pentru descoperirea cunoștințelor din date / mineritul datelor*. Referat doctoral nr. 1, Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu” al Academiei Române (http://www.racai.ro/media/Referatnr1SI-MDDCCornelLepadatu_sec.pdf)
- [17] LEPĂDATU, C. (2012) Explorarea datelor și descoperirea cunoștințelor - probleme, obiective și strategii. In *Revista Română de Informatică și Automatică*, vol.22, nr. 4, pp. 5 – 14.
- [18] LEPĂDATU, C. (2013) Metode exploratorii multidimensionale. In *Revista Română de Informatică și Automatică*, vol. 23, nr. 1, pp. 14 – 30.
- [19] LEPĂDATU, C. (2014) *Sistem pentru asistarea deciziilor bazat pe descoperirea cunoștințelor din date*. Referate doctorale nr. 2/3, Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu” al Academiei Române, aprilie 2012 / februarie 2014. (<http://www.racai.ro/media/Referat2-CornelLepadatu.pdf>) (<http://www.racai.ro/media/CLepadatu-Referat3.pdf>).
- [20] LEPĂDATU, C. (2014) Sisteme suport pentru decizii și bibliomining. In *Revista Română de Informatică și Automatică*, vol. 24, nr. 2, pp. 17 – 30.
- [21] MISHRA, R.-N.; MISHRA, A. (2013) Relevance of Data Mining in Digital Library. In *International Journal of Future Computer and Communication*, Vol. 2, No. 1, pp. 10 – 14.
- [22] MOORE, M.; LOPER, K. A. (2011) An Introduction to Clinical Decision Support Systems. În *J. Electron Resour Med Libr.*, University of Miami, 31p.

- [23] NICHOLSON, S. (2006) The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services. In *Information Processing and Management* 42(3), pp. 785 – 804.
- [24] PENDLEBURY, D.A. (2008) *Using Bibliometrics in Evaluating Research* (http://wokinfo.com/media/mtrp/UsingBibliometricsinEval_WP.pdf)
- [25] SREENIVASARAO, V.; PALLAMREDDY, V.-S. (2011) Advanced Data Warehousing Techniques for Analysis, Interpretation and Decision Support of Scientific Data. Springer-Verlag Berlin Heidelberg, D.C. Wyld et al. (Eds.): *ACITY 2011, CCIS* 198, pp. 162 – 174.
- [26] TRIA (DI), F.; LEFONS, E.; TANGORRA, F. (2012) Hybrid methodology for data warehouse conceptual design by UML schemas. In *Information and Software Technology*, 54, pp. 360 – 379.