

## Aplicatia SIMBNR, Sistem integrat de management al Bibliografiei nationale romanesti

1 SIMBNR este Proiectul de cercetare coordonat și realizat de Institutul Național de Cercetare-Dezvoltare în Informatica ICI București, în colaborare cu Biblioteca Academiei Române, proiect desfășurat în perioada 2009-2011.

2 foto

### 3 Sistemul SIMBNR

- este un instrument bibliografic de interes național, care oferă posibilitatea de acces on-line la informațiile cuprinse în bibliografia națională retrospectivă a cărții românești, utilizând sisteme de baze de date și tehnici multimedia
- este un sistem integrat destinat administrării, regăsirii și valorificării fondului de informații culturale oferit de BRV și BRM
- este un serviciu Web gratuit, de asigurare a accesului integrat la copiile digitale ale documentelor conținute în cele două bibliografii, făcute disponibile atât pe serverul Bibliotecii Academiei Române, instituție deținătoare a bibliografiei naționale retrospective, cât și pe serverul Institutului Național de Cercetare-Dezvoltare în Informatica ICI București, proiectantul și realizatorul sistemului

### 4 Sistemul dispune de două componente principale:

- componenta de creare, încărcare și gestionare a bazei de date bibliografice, și
- componenta on-line de regăsire după diferite criterii și de consultare a informațiilor bibliografice.

5 SIMBNR și-a propus să creeze o imagine clară, reală și accesibilă on-line a Bibliografiei naționale retrospective a cărții românești, un instrument de lucru în mâna specialiștilor domeniului, dar și la îndemâna altor cercetători preocupați de istoria științei și culturii românești.

Inițiative luate la nivel european încurajază crearea de conținut digital și constituirea de baze de date pentru o mai bună diseminare a informației, pentru a permite utilizatorilor accesul on-line la resurse, pentru încurajarea procesului de creație în format electronic, pentru promovarea și prezervarea digitală a patrimoniului cultural național și european în format electronic.

România este în măsura să îmbogățească memoria culturală colectivă cu un patrimoniu cultural și științific de valoare, oferit utilizatorilor de pretutindeni, într-un mod coerent, în concordanță cu normele europene.

Trecerea de la editarea documentelor pe hârtie la documente digitale a pus structurile de informare și documentare în fața unor noi abordări a activităților proprii precum: crearea unei infrastructuri electronice care să integreze eficient colecțiile tradiționale în formă tipărită; cooperarea și colaborarea cu alte instituții similare și chiar cu sectorul privat; negocieri cu producătorii și furnizorii de informație electronică în scopul asigurării accesului la resursele electronice; obligația de prezervare și arhivare a documentelor în formă electronică și obligația de a asigura transferul în formă electronică a unor documente solicitate de diverși utilizatori; obligativitatea de prelucrare a informației electronice în scopul facilitării accesului (realizarea de indexuri, rezumate, descriptori etc.).

6 SIMBNR, a parcurs etapele necesare, rezultatele obținute materializându-se în: identificarea criteriilor de structurare a conținutului bibliografiei de carte românească veche și modernă; definirea arhitecturii funcționale a sistemului bazat pe instrumente multimedia pentru valorificarea bibliografiei naționale retrospective; analiza normelor, formatelor și a tipurilor de suporti pentru arhivarea conținutului digital al bibliografiei naționale retrospective; identificarea normelor de publicare pe Web a conținutului digital; elaborarea criteriilor de regasire a publicațiilor românești cuprinse în bibliografia națională retrospectivă; definirea procedurilor de prelucrare a informațiilor digitizate; proiectarea și realizarea interfeței Web pentru SIMBNR; testarea funcționalității platformei tehnologice pentru SIMBNR; implementarea sistemului SIMBNR la Biblioteca Academiei Române.

7 Încărcarea în baza de date a informațiilor provenite din diverse surse, conform structurii bazei de date, necesită utilizarea unui proces de tip ETL (Extract, Transform, Load - extragere, transformare, încărcare). Acest proces implică: extragerea de date din surse externe, transformarea acestora pentru a se potrivi structurii funcționale a bazei de date în care urmează să fie încărcate, și încărcarea în baza de date țintă.

Ținând cont de structura diferită a surselor de date corespunzătoare celor două bibliografii - BRV și BRM - pentru maparea datelor existente în sursele de date pe structura bazei de date SIMBNR, au fost definite proceduri ETL specifice fiecărei structuri de date.

8 Sursa de date pentru tabela „brv” a bazei de date SIMBNR o constituie fișierele imagine, de tip „tif”, obținute prin scanarea celor patru volume ale Bibliografiei Românești Vechi, operație realizată în cadrul Bibliotecii Academiei Române.

9 Platforma tehnologică utilizată pentru scanarea celor patru volume BRV a fost compusă din:

- calculator Intel(R) Core (TM) 2 duo CPU E 7300 x 2.66 GHz 2.67 GHz și 200 MB of RAM
- sistem de operare Microsoft Windows XP professional version 2002;
- scanner Copy Book 12 S KRBGS.

Structura datelor obținute în urma scanării, corespunzătoare celor patru volume, este de aproximativ 2500 de fișiere

10 Pentru a putea fi vizualizate aceste fișiere pe Web și a încărca descrierea lor în baza de date SIMBNR, în vederea consultării on-line, datele și informațiile au fost supuse unor serii de transformări.

- Convertirea fișierelor din format „tif” în format „jpg” s-a realizat cu ajutorul opțiunii Image Processor din programul Adobe Photoshop CS4.
- Aplicarea procedurii de recunoaștere optică a caracterelor asupra fișierelor imagine ce conțin cuprinsul cărților descrise în fiecare volum.
- Proceduri ETL pentru încărcarea și indexarea datelor referitoare la Bibliografia Românească Veche.

Toate procedurile au fost realizate în cadrul unei baze de date temporare creată cu Microsoft Access 2003, datele finale fiind exportate prin intermediul unei conexiuni de tip ODBC (Open Database Connectivity) în baza de date MySQL a sistemului SIMBNR.

11 Sursa de date pentru tabela „brm” a bazei de date SIMBNR o constituie fişierele text, de tip „doc”, obţinute prin ocerizarea celor patru volume ale Bibliografiei Româneşti Moderne, operaţie realizată în cadrul Bibliotecii Academiei Române.

12 Platforma tehnologică utilizată pentru ocerizarea celor patru volume a fost compusă din:

- calculator Intel(R) Celeron (R) CPU 240 GHz, 504 MB of RAM;
- sistem de operare Microsoft Windows XP professional version 2002;
- program de ocerizare Inteligent OCR 9.0 Professional Edition Abbyy Fine Reader Essentials

Textele obţinute în urma ocerizării sunt grupate pe autori, în ordine alfabetică, materialul obţinut necesitând o corectură amănunţită.

13 Procedurile pentru prelucrarea informaţiilor ocerizate pentru BRM sunt

- Transformarea informaţiilor din format „text” în format „tabel”. Prin ocerizare, a fost obţinut un fişier „doc”. Pentru încărcarea acestor informaţii în baza de date „brm”, informaţiile au fost transformate în format „tabel”

- Proceduri ETL pentru încărcarea şi indexarea datelor referitoare la Bibliografia Românească Modernă.

Similar cu procedurile pentru Bibliografia Românească Veche, şi în acest caz toate procedurile au fost realizate în cadrul unei baze de date temporare creată cu Microsoft Access 2003, datele finale fiind exportate prin intermediul unei conexiuni de tip ODBC (Open Database Connectivity) în baza de date MySQL a sistemului SIMBNR.

14 Interfaţa Web cu utilizatorul este componenta care permite dialogul dintre utilizator şi aplicaţie prin intermediul unui limbaj cvasinatural, comunicând mecanismului de inferenţă cererile utilizatorului, iar utilizatorului, rezultatele prelucrărilor.

La proiectarea interfeţei cu utilizatorul pentru sistemul SIMBNR s-a avut în vedere ca aceasta să fie unitară şi ergonomică, uşor de înţeles şi de utilizat în mod intuitiv, pentru a permite utilizatorilor un acces facil la informaţii, fiind necesare doar cunoştinţe minime de lucru cu calculatorul.

15 Pentru afişarea informaţiilor conţinute în baza de date „brv” au fost proiectate trei tipuri de pagini Web.

Prima pagină BRV conţine o descriere sumară a celor patru volume componente.

În al doilea tip de pagină sunt afişate, în ordinea anilor de apariţie, o serie de informaţii referitoare la cărţile tipărite în perioada corespunzătoare fiecărui volum

Informaţiile sunt afişate în pagini succesive, cu posibilitate de navigare între pagini.

Al treilea tip de pagină a fost proiectat pentru a permite utilizatorilor să vizualizeze şi să navigheze printre paginile corespunzătoare unei cărţi, selectate din lista de mai sus.

Astfel, zona principală a acestui tip de pagină este împărţită în două coloane: prima coloana conţine pictogramele paginilor corespunzătoare cărţii selectate, iar în a doua coloană se afişează imaginea completă a conţinutului unei pictograme selectate.

16 foto

17 foto

18 Structura interfeței Web pentru afișarea informațiilor conținute în Bibliografia Românească Modernă (BRM) este similară cu structura interfeței Web pentru BRV. Astfel, ea cuprinde patru tipuri de pagini proiectate pentru:

- descrierea sumară a celor patru volume componente;
- afișarea indexului alfabetic al autorilor, prin activarea link-ului corespunzător unei litere
- afișarea descrierii bibliografice corespunzătoare unui autor selectat
- vizualizarea unor imagini atașate descrierilor bibliografice ale autorului selectat.

19 foto

20 foto

21

În concluzie, prin efortul conjugat al celor două instituții, s-a putut crea o alternativă modernă, prin folosirea tehnologiilor multimedia, pentru accesul publicului utilizator de Internet la conținutul digital al Bibliografiei naționale retrospective a cărții românești din perioada 1508-1918.