

Reprezentarea cunoașterii

Reprezentarea cunoașterii de la schemele de clasificare la web-ul semantic (II)

Dr. Silvia-Adriana Tomescu
Biblioteca Centrală Universitară „Carol I”
Email: silvia.tomescu@bcub.ro

Dr. Silvia-Adriana Tomescu coordonează serviciul Cercetare. Metodologie din cadrul Bibliotecii Centrale Universitare „Carol I” din București. Principalele sfere de cercetare științifică sunt: e-learning, biblioteconomie, redactare academică, diplomație culturală și exil românesc. A publicat traduceri și studii în domeniul comunicării și al științei informării.

Rezumat

În acest eseu ne propunem să abordăm modalitățile de reprezentare a documentelor hibride în mediul online cu ajutorul standardelor specifice. Evoluția conținuturilor multimedia și diversificarea suporturilor și formatelor de stocare implică utilizarea unor instrumente elaborate de căutare, regăsire și filtrare. Aplicațiile dedicate au nevoie de un grad sporit de interoperabilitate ceea ce conduce spre nevoia acută de standardizare. Acest studiu încearcă să argumenteze teoretic nevoia de a încadra și reflecta logic proprietățile multimedia cu ajutorul standardelor de metadata și să ilustreze importanța taxonomiilor și a ontologiilor în maniera de reflectare online a acestora.

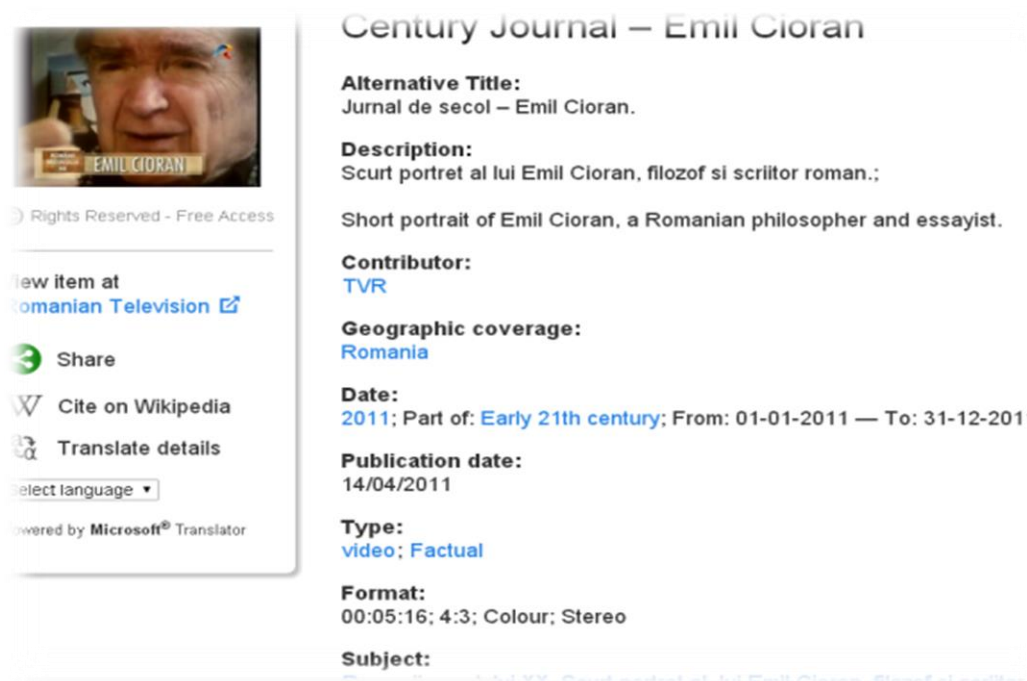
Cuvinte cheie: conținut media, metadata, taxonomii, ontologii, obiecte digitale

III. Reprezentarea cunoașterii prin conținuturi media

3.1. Instrumente de reprezentarea cunoașterii

Reprezentarea cunoașterii este procesul de ilustrare a conținuturilor cu ajutorul schemelor de clasificare, taxonomii, ontologii, structurate logic și semnatic pe web. Acest demers are la bază utilizarea limbajelor matematice și a formatelor de interschimb. Obiectele digitale sunt reprezentate prin arhitecturi de metadata, seturi de descriptori semantici grupați în jurul unui obiect pentru a-l explica, localiza, eficientiza în cataloage sau colecții. Orice tip de conținut media are nevoie de strategii de management al informației specifice pentru a optimiza procesul de creare, filtrare, și partajare, bazate pe metadata de conținut și context și standarde de interoperabilitate. Metadatale exprimă identitatea oricărui obiect digital, furnizând informații primare asupra conținutului.

Exemplul din figura alăturată descrie un film în catalogul Europeana. Dublin Core Metadata Element Set (1) este cadrul general de descriere al acestuia și cuprinde informații referitoare la: *titlu, creator, subiect, editor* etc. Nu este o schemă de metadata pentru materiale multimedia, însă are elemente comune cu acestea. Tipuri diferite de documente din baze de date sau cataloage cer standarde specifice de metadata.



Century Journal – Emil Cioran

Alternative Title:
Jurnal de secol – Emil Cioran.

Description:
Scurt portret al lui Emil Cioran, filozof și scriitor român.;
Short portrait of Emil Cioran, a Romanian philosopher and essayist.

Contributor:
TVR

Geographic coverage:
Romania

Date:
2011; Part of: [Early 21th century](#); From: 01-01-2011 — To: 31-12-2011

Publication date:
14/04/2011

Type:
video; Factual

Format:
00:05:16; 4:3; Colour; Stereo

Subject:
Filozofia; Jurnalul de secol XX; Scurt portret al lui Emil Cioran; Filozof și scriitor

Fig. 1. Metadata în Europeana (Material audio-video: Jurnal de secol: Emil Cioran)

Elementele de bază ale acestei scheme se referă la: (a) **Conținut:** titlu, subiect, descriere, limbă, relație; (b) **Drepturi intelectuale:** creator, editor, contributor, drepturi; (c) **Instanță:** dată, timp, format, identificator. *Titlul* este reprezentat de numele dat resursei; *Creatorul* este orice entitate responsabilă de producerea resursei; *Subiectul:* este dat de conținutul/topical resursei; *Editorul* este entitatea responsabilă de partajarea resursei. Noțiunile și conceptele sunt structurate pe principiul dicționarilor și al glosarelor și sunt organizate în acord cu regulile limbajului natural pentru a eficientiza procesul de căutare și regăsire. În plus, le atribuie categorii semantice pentru a rafina căutarea.

Managementul informației într-un mediu multimedia trebuie să ia în considerare:

- Utilizatorii, nevoile informaționale și abilitățile acestora;
- Fluxul de lucru pentru managementul informației și al metadatelor;
- Domeniile (multi)media;
- Graficul încărcării obiectelor digitale;
- Produsul final.

Instrumentele de reprezentare a cunoașterii în format digital sunt glosare – hărți conceptuale; tezaure – taxonomii; hărți de domeniu – ontologii; limbaje de reprezentare

specifice. Ontologiile sunt conceptualizări multifacetate ale unui domeniu al cunoașterii definit de asocierea *entitate – relație* în procesarea automatizată. Există ontologii specifice domeniilor: astrologie, biologie, clasificându-se astfel drept ontologii speciale, sau generale. Entitatea modelează clase de obiecte; attributele modelează proprietățile. Membrii unei clase reprezintă instanțele acestei categorii. Ontologiile sunt exprimate prin limbaje de programare ca KIF [Knowledge Interchange Format] sau prin intermediul limbajelor bazate pe logica descriptivă [OWL; OIL].

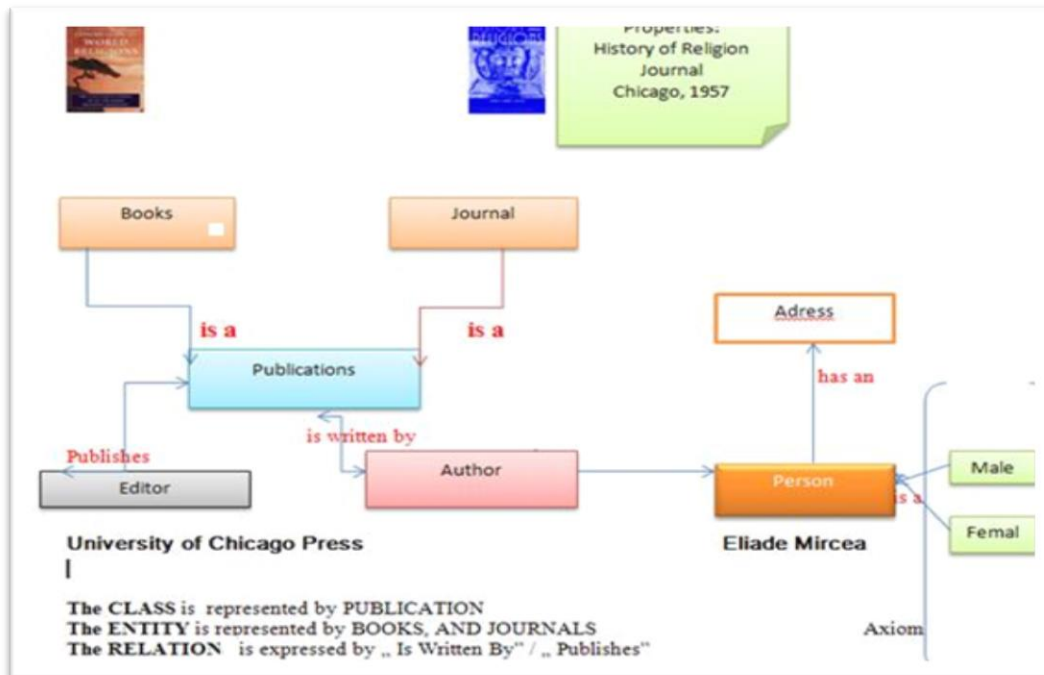


Fig. 2. Modelul Clasă-Entitate-Relație

Cu ajutorul ontologiilor realizăm asocieri semantice ale lucrurilor despre care vorbim; web ontologiile (2) sunt structurate în clase și instanțele care le definesc structura. Acestea oferă posibilitatea de a descrie relațiile cu ajutorul limbajului XML, pentru a fi procesate de computer. Ontologiile multimedia pot fi construite folosind OWL și modelate cu elemente din standardele MPEG 7 și MPEG 21.(3) Rolul ontologiei din punct de vedere pragmatic este acela de a furniza vocabulare controlate, a crea taxonomii pentru eficiență în navigare, prin ierarhii de termeni. Acest lucru poate fi util în analiza așteptărilor utilizatorilor în legătură cu conținutul. Un obiect digital este o reprezentare a resursei tradiționale sau digitale create. Metadatele sunt generate pentru a descrie resursa pentru a ajuta la identificarea, stocarea și distribuirea acesteia. Principalele tipuri de metadate pe care specialiștii folosesc trei clase sunt: (a) descriptive – ajută la identificarea și indexarea obiectului; (b) structurale – oferă informații despre structura internă a obiectului; (c) administrative – includ informații despre gestionarea drepturilor hard, soft utilizate. Cadrul de lucru pe care consoțul WWW îl agreează pentru procesarea datelor este RDF Resource Description Framework; acesta permite reprezentarea semantică și

sintactică a metdatelor, prin intermediul limbajelor XML, HTML și URI Universal Resource Locator [codare, schimb și procesare]. RDF este limbajul general pentru reprezentarea semantică a lumii reale în rețeaua WWW, permite schimbul metadatelor structurate, utilizează sintaxe XML și formatul namespace. Cu ajutorul RDF descriem evenimente, informații, conținutul pentru o bună reprezentare și eficiență în căutare.

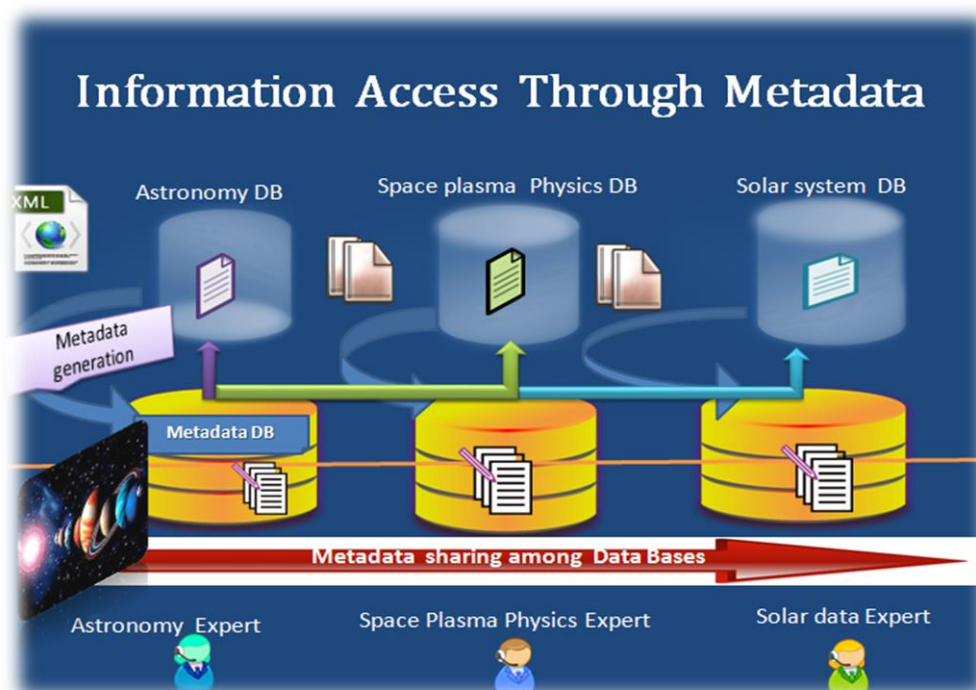


Fig. 3. Accesul la informații cu ajutorul metadatelor

3.2. Standarde de metadata pentru materiale multimedia

Deoarece un standard de metadata este un instrument de referințe cu atribute specifice pentru interoperabilitate, presupune o înțelegere semantică a conținuturilor media și a drepturilor legale care decurg. În acest mediu semantic, vocabularele controlate, schemele de clasificare și taxonomiile au un rol fundamental. Taxonomiile organizează informația descrisă de metadata. Informația este codificată sub conținutul media, care poartă o semantică implicită. Există numeroase standarde pentru sfera multimedia; o schema puternică de metadata specifică multimedia trebuie să țină cont de: *interoperabilitate*, *extensibilitate*, *modularitate* și *formatul media*. Odată structurate sursele cu ajutorul unei scheme solide, conținutul media poate fi căutat, identificat și furnizat către utilizatori. Designul arhitecturii multimedia implică *captură*, *adnotare*, *editare*, *publicare*, respective *partajare*. Materialele multimedia capătă valoare numai prin adăugarea unor metadata specifice. Cele mai importante standarde de metadata pe care experții în domeniu le utilizează sunt: (a) MPEG – Moving Picture Expert Group,

care a dezvoltat standardul pentru codarea conținuturilor digitale audio-video; (b) SMTPE – Society for Motion Picture and Television Engineering, care elaborează schema pentru industria de filme de animație; (c) DCMI – Dublin Core Metadata initiative, care elaborează standardul pentru conținuturi accesibile online; (d) IPTC – International Press Telecommunications Council, care întreține metadatele de interschimb pentru știri; (e) Forumul PRO-MPEG care dezvoltă standarde pentru echipamentele TV; (f) Forumul TV Anytime care dezvoltă standarde de date stocate pe platforme online. În modelul mai jos, am încercat să prezentăm modalități de reprezentare, utilizare, partajare a obiectelor media spre un public specializat [Artă și Industrie]. Strategia a implicat (a) subiectul; (b) obiectul media; (c) metadate și ontologii; (d) reprezentare și standardizare; MPEG 7 Moving Picture Expert Group este un standard de descriere semantică ce permite combinarea elementelor audio-video. Descrierea în manieră standardizată, conform ISO/IEC 15938-10, este prezentată mai jos:



Fig. 4. Structură ierarhică în MPEG 7

Această schemă este bazată pe descriptorii de reprezentare semantică și sintactică: (a) informații referitoare la crearea conținutului; (b) informații referitoare la utilizarea conținutului; (c) informații referitoare la stocare; (d) informații asupra coordonatelor temporale; (e) informații referitoare la eveniment; (f) informații referitoare la colecții; (g) informații referitoare la interacțiunea utilizator-conținuturi. Standardul include specificități de codare și peste 450 tipuri de metadate bazate pe XML ca limbaj de lucru. Cele două standarde MPEG 7 [descriere multimedia] și MPEG 21 [schimb și distribuție de obiecte digitale] au o mare importanță în televiziune, industriile media. CIDOC Conceptual reference Model (4) este un standard internațional de metadate cu

aplicabilitate în sfera culturală, care descrie semantica și regulile de bază ale documentelor. Acoperă 86 de clase și 137 proprietăți, oferind posibilitatea de a descrie sute de formate de metadata. Este folosit ca format de schimb și migrare în webul semantic. Interoperabilitatea semantică în media poate fi obținută cu ajutorul unei „ontologii de relaționări extinse”. CRM furnizează o descriere partajată și mai puțin prescrierea unei structuri comune de date. În cadrul CRM instanțele sunt codificate cu ajutorul limbajelor RDF și XML. Datele pot fi importate, exportate, schimbate fără a le altera semnificația.

Exemplificare: Conținutul „Mâna de aur a artistului”, autor Mihai Zgondoiu este stocată într-un server MySQL și constituie bază pentru crearea unui conținut 3D de metadata. Pentru mapare a fost utilizat modelul CIDOC-CRM. Prin serverul D2R se permite publicarea conținutului MYSQL DB în RDF pentru a fi operabil în webul semantic.



Fig. 5. Mihai Zgondoiu. The Golden Hand of the Artist (sculptură) – Reprezentarea conținutului 3D

3.3. Reprezentarea cunoașterii în mediul social creative media

Cunoașterea domeniului este structurată pe două tipuri de ontologii: *conținut de creative media* și *utilizatori de creative media*. În spațiul multimedia utilizatorii interacționează, comunică, colaborează și schimbă conținuturi ceea ce oferă posibilitatea de a analiza interacțiunile complexe dintre ele. În egală măsură, permite ghidarea utilizatorilor pentru a obține rezultatul cel mai adecvat în căutare. Reprezentarea

