



Cercetări cu privire la securitatea informației în contextul colaborativ al bibliotecilor virtuale europene

Autor: **Laurențiu Mihai F. DOINEA**

Lucrare realizată în cadrul proiectului "Cultura română și modele culturale europene - cercetare, sincronizare, durabilitate", cofinanțat din FONDUL SOCIAL EUROPEAN prin Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007 – 2013, Contract nr. POSDRU/159/1.5/S/136077.

Titlurile și drepturile de proprietate intelectuală și industrială asupra rezultatelor obținute în cadrul stagiului de cercetare postdoctorală aparțin Academiei Române.

* * *

*Punctele de vedere exprimate în lucrare aparțin autorului și nu angajează
Comisia Europeană și Academia Română, beneficiara proiectului.*

DTP, complexul editorial/redacțional, traducerea și corectura aparțin autorului.

Descărcare gratuită pentru uz personal, în scopuri didactice sau științifice.

Reproducerea publică, fie și parțială și pe orice suport,
este posibilă numai cu acordul prealabil al Academiei Române.



ISBN 978-973-167-308-0

CUPRINS

REZUMAT	5
SUMMARY	7
1. INTRODUCERE	9
2. SOLUȚIE PRIVIND MANAGEMENTUL FLUXULUI DE DATE ÎN SISTEMELE DE BIBLIOTECI VIRTUALE	12
2.1. Biblioteci virtuale în contextul globalizării	12
2.2. Partajare și colaborare în bibliotecile virtuale	17
2.3. Digitizare, proces generativ de conținut digital	21
2.4. Metodologie pentru securitatea conținutului digital	25
3. SOLUȚIE PRIVIND CLASIFICAREA INFORMAȚIEI STOCATĂ ÎN VOLUME MARI DE DATE	31
3.1. Big Data în era datelor cu acces liber	31
3.2. Taxonomii de securitate în sistemele integrate de bibliotecă pe bază de ontologii	41
3.3. Analiza și preprocesarea datelor	45
3.4. Clasificarea și ierarhizarea conținutului digital	52
4. SOLUȚIE PRIVIND PROTEJAREA CONȚINUTUL DIGITAL LA NIVELUL SISTEMELOR INTEGRATE DE BIBLIOTECĂ	59
4.1. Managementul drepturilor digitale în sistemele integrate de bibliotecă	59
4.2. Protejarea conținutului digital pe bază de tehnici DRM	65
4.3. Verificarea autenticității conținutului digital folosind coduri cu răspuns rapid	73
5. DISEMINARE	83
5.1. Soluții originale avansate în lucrarea de cercetare	83
5.2. Diseminarea în reviste de specialitate și conferințe internaționale	85
5.3. Stagiul de mobilitate – Universitatea din Catania	86
5.4. Direcții viitoare de cercetare	87
6. CONCLUZII	89
BIBLIOGRAFIE	91
Anexa 1 – Lista de acronime	100
Anexa 2 – Lista de figuri	101
Anexa 3 – Lista de tabele	103

CONTENTS

REZUMAT	5
SUMMARY	7
1. INTRODUCTION	9
2. SOLUTION FOR DATA FLOW MANAGEMENT IN VIRTUAL LIBRARY SYSTEMS	12
2.1. Virtual libraries in the globalization era	12
2.2. Sharing and collaboration in virtual libraries	17
2.3. Digitization, generative process of digital content	21
2.4. Methodology for digital content security	25
3. SOLUTION FOR CLASSIFICATION OF INFORMATION STORED IN LARGE VOLUMES OF DATA	31
3.1. Big Data in era of open access data	31
3.2. Ontology based security taxonomies for integrated library systems	41
3.3. Data analysis and preprocessing	45
3.4. The classification and ranking of digital content	52
4. SOLUTION FOR DIGITAL CONTENT PROTECTION IN INTEGRATED LIBRARY SYSTEMS	59
4.1. Digital rights management in integrated library systems	59
4.2. Digital content security based on DRM techniques	65
4.3. Genuineness verification of digital content using quick response codes	73
5. DISSEMINATION	83
5.1. Original solutions advanced in the research	83
5.2. Dissemination in scientific journals and international conferences	85
5.3. Research Mobility – University of Catania	86
5.4. Future research directions	87
6. CONCLUSIONS	89
BIBLIOGRAPHY	91
Annex 1 – List of Acronyms	100
Annex 2 – List of Figures	101
Annex 3 – List of Tables	103

REZUMAT

Proiectul de cercetare postdoctorală își propune realizarea unui sistem de evaluare pe bază de ontologii a conținutului digital din cadrul bibliotecilor virtuale, cu scopul de a implementa măsuri de securitate care să contribuie la promovarea valorilor culturale de o manieră sigură și corectă din perspectiva drepturilor digitale.

Obiectivul general al acestei cercetări îl reprezintă identificarea, analiza și implementarea de soluții de securitate pentru protecția conținutului digital în contextul colaborativ al sistemelor online de biblioteci virtuale din spațiul european.

Pe baza rezultatelor evaluării obiectelor digitale sunt realizate clasificări și ierarhizări ale acestora. Un sistem de protecție a conținutului bibliotecilor virtuale pe bază de drepturi digitale este propus astfel încât obiecte de o valoare culturală importantă să fie protejate de posibile utilizări necorespunzătoare. Sunt folosite ca bază de date informațiile din cadrul unor biblioteci virtuale. Sunt identificate taxonomii reprezentative iar obiectele sunt ierarhizate pe baza unui set de criterii folosit ulterior la atribuirea unor niveluri de securitate.

Știința informației reprezintă disciplina care se ocupă de studiul comunicării informației în societate prin metode științifice. Știința informației este prin natura acesteia o știință interdisciplinară care vizează analiza, colectarea, clasificarea, manipularea, stocarea, extragerea și diseminarea informației.

Bibliotecile virtuale, datorită resurselor de care dispun, fac cunoscute resursele digitale într-o formă eficientă iar servicii precum achiziția, organizarea, stocarea, regăsirea și prelucrarea informației sunt oferite publicului la un nivel înalt de calitate. Procesarea și partajarea informației reprezintă procese complexe care vizează atât aspectele calitative cât și pe cele cantitative ale acesteia. Accesul online la documente unicat de valoare inestimabilă, regăsirea rapidă de informații prin intermediul sistemelor de căutare precum Online Public Access Catalog, procesarea conținutului digital utilizând tehnici de data mining cu scopul de a crea cunoștințe, toate acestea constituie avantaje nete care fac din sistemele online de biblioteci virtuale unelte vitale în realizarea progresului și a completării patrimoniului cultural european.

În cadrul sistemelor informatice care stau la baza activităților ce se desfășoară într-o bibliotecă virtuală, datorită caracterului deschis pe care trebuie să îl promoveze un astfel de sistem, intervin o serie de vulnerabilități exploatate de amenințări prin intermediul unor atacuri de natură malițioasă.

Automatizarea proceselor de prelucrare generează riscuri legate de integritatea, confidențialitatea, disponibilitatea, non-repudierea și autenticitatea informației. Securitatea are rolul de a asigura caracteristicile de calitate ale informației. Fiecărei caracteristici de calitate a informației îi revin mecanisme de securitate care au rolul de a preveni și combate riscurile la care se expun în momentul în care sunt accesibile prin intermediul rețelei internet.

Datorită complexității și diversității informațiilor procesate în cadrul bibliotecilor virtuale sunt fundamentate premisele utilizării unor instrumente complexe și diverse, precum ontologiile. Ontologiile sunt capabile să identifice relații complexe între conceptele vehiculate în cadrul înregistrărilor publicate online. Ontologia ajută la determinarea sensurilor exacte ale termenilor care fac parte din descrierea obiectelor digitale iar acest lucru ajută mai departe algoritmi de clasificare utilizați să stabilească clasele de apartenență funcție de un set de criterii evaluate.

Clasificarea corectă a obiectelor conduce la stabilirea unui nivel de securitate corespunzător pentru administrarea drepturilor digitale, concept cunoscut sub numele de Digital Rights Management.

Prin intermediul drepturilor digitale, bibliotecile virtuale au șansa de a-și proteja conținutul digital, patrimoniu cultural, de utilizările și preluările neautorizate. Efortul susținut al bibliotecarilor de a digitaliza materialele achiziționate prin licitații, donații sau achiziție directă de la colecționari a unor elemente de o valoare culturală ridicată este astfel protejat de însușirea ilegală a acestor bunuri digitale.

Protejarea conținutului digital este un aspect vital în momentul în care informațiile sunt partajate într-un sistem colaborativ la nivelul mai multor biblioteci virtuale interconectate prin diverse protocoale de comunicare fie ele gratuite sau proprietare. Drepturile digitale asupra informațiilor dintr-o bibliotecă virtuală reglementează ce se poate face cu conținutul digital și cum se poate accesa acesta în scopul prelucrării.

Rezultatele au fost publicate în cadrul unor conferințe naționale și internaționale cu scopul validării soluțiilor propuse de către comunitatea științifică precum și în cadrul unor jurnale de specialitate.

Această lucrare a fost realizată în cadrul proiectului “Cultura română și modele culturale europene: cercetare, sincronizare, durabilitate”, cofinanțat de Uniunea Europeană și Guvernul României din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013, contractul de finanțare nr. POSDRU/159/1.5/S/136077.

SUMMARY

This postdoctoral research project aims to achieve the development of an ontology based evaluation system of digital content within the virtual libraries in order to implement security measures that will help promote the cultural values in a safe and correct manner from a digital rights perspective.

The main objective of this research is to identify, analyze and implement security solutions for digital content protection in the context of online, collaborative virtual library systems in European countries.

Based on the assessment results of digital objects, classifications and rankings are achieved. A content protection system in virtual libraries is proposed based on digital rights management techniques. Its role is to assure that objects of high cultural value are protected against possible misuse from malicious users. Representative taxonomies are identified and objects are ranked based on a set of criteria used later to establish subsequent levels of security.

Information science is a discipline that deals with the study of communication in the information society through scientific methods. Information science is by its nature an interdisciplinary science that aims to analyze, collect, classify, handle, store, retrieve and disseminate information.

Huge amounts of electronic resources that are stored in a virtual library are efficiently publish on the web and services such as acquisition, organization, storage, retrieval and processing of information are offered to the public at a high level of quality. Processing and sharing information are both complex processes that aim qualitative and quantitative aspects alike. Online access to unique documents of inestimable value, quick retrieval of information through online search engines such as Online Public Access Catalog and as well processing digital content using data mining techniques in order to create knowledge, all these are great advantages that make the online systems for virtual libraries vital tools in achieving progress and completion of European cultural heritage.

In online systems that support the activities undergone within a virtual library, due to the openness characteristic of such systems that must be promoted, a huge number of vulnerabilities exploited by threats through different attacks of malicious nature occur.

Process automation entails risks relating to the integrity, confidentiality, availability, non-repudiation and authenticity of information. Security must ensure that information quality characteristics are preserved. A security mechanism or techniques correspond to each information's characteristic quality. These security measures are

designed to prevent and combat the risks the information faces when is freely accessible on the Internet.

Due to the complexity and diversity of information processed within a virtual library, premises for using tools for semantic analysis like ontologies are very well grounded. Ontology helps in identifying complex relationships between concepts extracted from the description of digital objects that are published online. Ontology helps also determine the precise meaning of concepts that are part of the description of digital objects, meaning which is used further in classification algorithms to establish membership classes according to a set of criteria assessed.

The correct classification of objects helps to determine a correct level of security represented by adequate security measures in the field of digital rights management.

By using digital rights measures, digital libraries have the chance to protect their digital content, cultural heritage, of misuse and unauthorized access. Librarian effort to digitize materials of high cultural value, purchased through auctions, donations or direct purchases from collectors is protected by the misappropriation of these digital assets.

Protecting digital content is a vital aspect when this is freely accessible over the Internet or shared in a collaborative architecture across multiple virtual libraries interconnected by various communication protocols either free or proprietary. Digital rights over electronic resources found in a virtual library governs what can be done with the digital content and how it can be accessed in order to process it with a high level of protection.

The results were published in national and international conferences in order to validate the solutions proposed by the scientific community and as well in specialized journals.

This paper is supported by the Sectorial Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number SOP HRD/159/1.5/S/136077.

CUPRINS

REZUMAT	5
SUMMARY	7
1. INTRODUCERE	9
2. SOLUȚIE PRIVIND MANAGEMENTUL FLUXULUI DE DATE ÎN SISTEMELE DE BIBLIOTECI VIRTUALE	12
2.1. Biblioteci virtuale în contextul globalizării	12
2.2. Partajare și colaborare în bibliotecile virtuale	17
2.3. Digitizare, proces generativ de conținut digital	21
2.4. Metodologie pentru securitatea conținutului digital	25
3. SOLUȚIE PRIVIND CLASIFICAREA INFORMAȚIEI STOCATĂ ÎN VOLUME MARI DE DATE	31
3.1. Big Data în era datelor cu acces liber	31
3.2. Taxonomii de securitate în sistemele integrate de bibliotecă pe bază de ontologii	41
3.3. Analiza și preprocesarea datelor	45
3.4. Clasificarea și ierarhizarea conținutului digital	52
4. SOLUȚIE PRIVIND PROTEJAREA CONȚINUTUL DIGITAL LA NIVELUL SISTEMELOR INTEGRATE DE BIBLIOTECĂ	59
4.1. Managementul drepturilor digitale în sistemele integrate de bibliotecă	59
4.2. Protejarea conținutului digital pe bază de tehnici DRM	65
4.3. Verificarea autenticității conținutului digital folosind coduri cu răspuns rapid	73
5. DISEMINARE	83
5.1. Soluții originale avansate în lucrarea de cercetare	83
5.2. Diseminarea în reviste de specialitate și conferințe internaționale	85
5.3. Stagiul de mobilitate – Universitatea din Catania	86
5.4. Direcții viitoare de cercetare	87
6. CONCLUZII	89
BIBLIOGRAFIE	91
Anexa 1 – Lista de acronime	100
Anexa 2 – Lista de figuri	101
Anexa 3 – Lista de tabele	103

CONTENTS

REZUMAT	5
SUMMARY	7
1. INTRODUCTION	9
2. SOLUTION FOR DATA FLOW MANAGEMENT IN VIRTUAL LIBRARY SYSTEMS	12
2.1. Virtual libraries in the globalization era	12
2.2. Sharing and collaboration in virtual libraries	17
2.3. Digitization, generative process of digital content	21
2.4. Methodology for digital content security	25
3. SOLUTION FOR CLASSIFICATION OF INFORMATION STORED IN LARGE VOLUMES OF DATA	31
3.1. Big Data in era of open access data	31
3.2. Ontology based security taxonomies for integrated library systems	41
3.3. Data analysis and preprocessing	45
3.4. The classification and ranking of digital content	52
4. SOLUTION FOR DIGITAL CONTENT PROTECTION IN INTEGRATED LIBRARY SYSTEMS	59
4.1. Digital rights management in integrated library systems	59
4.2. Digital content security based on DRM techniques	65
4.3. Genuineness verification of digital content using quick response codes	73
5. DISSEMINATION	83
5.1. Original solutions advanced in the research	83
5.2. Dissemination in scientific journals and international conferences	85
5.3. Research Mobility – University of Catania	86
5.4. Future research directions	87
6. CONCLUSIONS	89
BIBLIOGRAPHY	91
Annex 1 – List of Acronyms	100
Annex 2 – List of Figures	101
Annex 3 – List of Tables	103

REZUMAT

Proiectul de cercetare postdoctorală își propune realizarea unui sistem de evaluare pe bază de ontologii a conținutului digital din cadrul bibliotecilor virtuale, cu scopul de a implementa măsuri de securitate care să contribuie la promovarea valorilor culturale de o manieră sigură și corectă din perspectiva drepturilor digitale.

Obiectivul general al acestei cercetări îl reprezintă identificarea, analiza și implementarea de soluții de securitate pentru protecția conținutului digital în contextul colaborativ al sistemelor online de biblioteci virtuale din spațiul european.

Pe baza rezultatelor evaluării obiectelor digitale sunt realizate clasificări și ierarhizări ale acestora. Un sistem de protecție a conținutului bibliotecilor virtuale pe bază de drepturi digitale este propus astfel încât obiecte de o valoare culturală importantă să fie protejate de posibile utilizări necorespunzătoare. Sunt folosite ca bază de date informațiile din cadrul unor biblioteci virtuale. Sunt identificate taxonomii reprezentative iar obiectele sunt ierarhizate pe baza unui set de criterii folosit ulterior la atribuirea unor niveluri de securitate.

Știința informației reprezintă disciplina care se ocupă de studiul comunicării informației în societate prin metode științifice. Știința informației este prin natura acesteia o știință interdisciplinară care vizează analiza, colectarea, clasificarea, manipularea, stocarea, extragerea și diseminarea informației.

Bibliotecile virtuale, datorită resurselor de care dispun, fac cunoscute resursele digitale într-o formă eficientă iar servicii precum achiziția, organizarea, stocarea, regăsirea și prelucrarea informației sunt oferite publicului la un nivel înalt de calitate. Procesarea și partajarea informației reprezintă procese complexe care vizează atât aspectele calitative cât și pe cele cantitative ale acesteia. Accesul online la documente unicat de valoare inestimabilă, regăsirea rapidă de informații prin intermediul sistemelor de căutare precum Online Public Access Catalog, procesarea conținutului digital utilizând tehnici de data mining cu scopul de a crea cunoștințe, toate acestea constituie avantaje nete care fac din sistemele online de biblioteci virtuale unelte vitale în realizarea progresului și a completării patrimoniului cultural european.

În cadrul sistemelor informatice care stau la baza activităților ce se desfășoară într-o bibliotecă virtuală, datorită caracterului deschis pe care trebuie să îl promoveze un astfel de sistem, intervin o serie de vulnerabilități exploatate de amenințări prin intermediul unor atacuri de natură malițioasă.

Automatizarea proceselor de prelucrare generează riscuri legate de integritatea, confidențialitatea, disponibilitatea, non-repudierea și autenticitatea informației. Securitatea are rolul de a asigura caracteristicile de calitate ale informației. Fiecărei caracteristici de calitate a informației îi revin mecanisme de securitate care au rolul de a preveni și combate riscurile la care se expun în momentul în care sunt accesibile prin intermediul rețelei internet.

Datorită complexității și diversității informațiilor procesate în cadrul bibliotecilor virtuale sunt fundamentate premisele utilizării unor instrumente complexe și diverse, precum ontologiile. Ontologiile sunt capabile să identifice relații complexe între conceptele vehiculate în cadrul înregistrărilor publicate online. Ontologia ajută la determinarea sensurilor exacte ale termenilor care fac parte din descrierea obiectelor digitale iar acest lucru ajută mai departe algoritmi de clasificare utilizați să stabilească clasele de apartenență funcție de un set de criterii evaluate.

Clasificarea corectă a obiectelor conduce la stabilirea unui nivel de securitate corespunzător pentru administrarea drepturilor digitale, concept cunoscut sub numele de Digital Rights Management.

Prin intermediul drepturilor digitale, bibliotecile virtuale au șansa de a-și proteja conținutul digital, patrimoniu cultural, de utilizările și preluările neautorizate. Efortul susținut al bibliotecarilor de a digitaliza materialele achiziționate prin licitații, donații sau achiziție directă de la colecționari a unor elemente de o valoare culturală ridicată este astfel protejat de însușirea ilegală a acestor bunuri digitale.

Protejarea conținutului digital este un aspect vital în momentul în care informațiile sunt partajate într-un sistem colaborativ la nivelul mai multor biblioteci virtuale interconectate prin diverse protocoale de comunicare fie ele gratuite sau proprietare. Drepturile digitale asupra informațiilor dintr-o bibliotecă virtuală reglementează ce se poate face cu conținutul digital și cum se poate accesa acesta în scopul prelucrării.

Rezultatele au fost publicate în cadrul unor conferințe naționale și internaționale cu scopul validării soluțiilor propuse de către comunitatea științifică precum și în cadrul unor jurnale de specialitate.

Această lucrare a fost realizată în cadrul proiectului “Cultura română și modele culturale europene: cercetare, sincronizare, durabilitate”, cofinanțat de Uniunea Europeană și Guvernul României din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013, contractul de finanțare nr. POSDRU/159/1.5/S/136077.

SUMMARY

This postdoctoral research project aims to achieve the development of an ontology based evaluation system of digital content within the virtual libraries in order to implement security measures that will help promote the cultural values in a safe and correct manner from a digital rights perspective.

The main objective of this research is to identify, analyze and implement security solutions for digital content protection in the context of online, collaborative virtual library systems in European countries.

Based on the assessment results of digital objects, classifications and rankings are achieved. A content protection system in virtual libraries is proposed based on digital rights management techniques. Its role is to assure that objects of high cultural value are protected against possible misuse from malicious users. Representative taxonomies are identified and objects are ranked based on a set of criteria used later to establish subsequent levels of security.

Information science is a discipline that deals with the study of communication in the information society through scientific methods. Information science is by its nature an interdisciplinary science that aims to analyze, collect, classify, handle, store, retrieve and disseminate information.

Huge amounts of electronic resources that are stored in a virtual library are efficiently publish on the web and services such as acquisition, organization, storage, retrieval and processing of information are offered to the public at a high level of quality. Processing and sharing information are both complex processes that aim qualitative and quantitative aspects alike. Online access to unique documents of inestimable value, quick retrieval of information through online search engines such as Online Public Access Catalog and as well processing digital content using data mining techniques in order to create knowledge, all these are great advantages that make the online systems for virtual libraries vital tools in achieving progress and completion of European cultural heritage.

In online systems that support the activities undergone within a virtual library, due to the openness characteristic of such systems that must be promoted, a huge number of vulnerabilities exploited by threats through different attacks of malicious nature occur.

Process automation entails risks relating to the integrity, confidentiality, availability, non-repudiation and authenticity of information. Security must ensure that information quality characteristics are preserved. A security mechanism or techniques correspond to each information's characteristic quality. These security measures are

designed to prevent and combat the risks the information faces when is freely accessible on the Internet.

Due to the complexity and diversity of information processed within a virtual library, premises for using tools for semantic analysis like ontologies are very well grounded. Ontology helps in identifying complex relationships between concepts extracted from the description of digital objects that are published online. Ontology helps also determine the precise meaning of concepts that are part of the description of digital objects, meaning which is used further in classification algorithms to establish membership classes according to a set of criteria assessed.

The correct classification of objects helps to determine a correct level of security represented by adequate security measures in the field of digital rights management.

By using digital rights measures, digital libraries have the chance to protect their digital content, cultural heritage, of misuse and unauthorized access. Librarian effort to digitize materials of high cultural value, purchased through auctions, donations or direct purchases from collectors is protected by the misappropriation of these digital assets.

Protecting digital content is a vital aspect when this is freely accessible over the Internet or shared in a collaborative architecture across multiple virtual libraries interconnected by various communication protocols either free or proprietary. Digital rights over electronic resources found in a virtual library governs what can be done with the digital content and how it can be accessed in order to process it with a high level of protection.

The results were published in national and international conferences in order to validate the solutions proposed by the scientific community and as well in specialized journals.

This paper is supported by the Sectorial Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number SOP HRD/159/1.5/S/136077.

1. INTRODUCERE

Obiectivul general al acestei cercetări îl reprezintă identificarea, analiza și implementarea de soluții de securitate pentru protecția conținutului digital în contextul colaborativ al sistemelor online de biblioteci virtuale din spațiul european.

Necesitatea este reprezentată de nivelul ridicat de digitizare la care sistemele informaționale au ajuns. Trăim într-o eră cu totul digitizată în care fiecare proces poate fi înregistrat, măsurat iar informația vehiculată poate fi transmisă aproape instantaneu către toate părțile interesate de felul în care s-a desfășurat procesul respectiv. Mai mult, costul tehnologiei din domeniul IoT (Internet of Things) s-a redus foarte mult deoarece perioadele de timp la care apar noi tipuri de senzori și instrumente electronice automate este foarte mic.

Metodologia de lucru presupune utilizarea de metode și instrumente de analiză specifice unui domeniu de cercetare aflat la intersecția dintre știința informației și securitatea informatică.

Proiectul de cercetare postdoctorală își propune realizarea unei evaluări a conținutului digital al bibliotecilor virtuale folosind ontologii lexicale. Analiza lexicală furnizată de ontologie este folosită în cadrul algoritmilor de clasificare utilizați cu scopul de a stabili niveluri de importanță diferitelor obiecte digitale analizate în cadrul bibliotecilor virtuale. Pe baza rezultatelor evaluării sunt realizate clasificări și ierarhizări ale informației. Taxonomiile identificate, reprezentative pentru domeniul analizat, au la bază un set de criterii folosit ulterior la atribuirea de niveluri de securitate. Sunt propuse tehnici de protejare a conținutului bibliotecilor virtuale pe bază de drepturi digitale și steganografie, astfel încât obiecte de o valoare culturală ridicată să fie protejate împotriva utilizării necorespunzătoare.

Lucrarea este structurată pe șase capitole care au rolul să prezinte nivelul actual de cercetare în domeniile abordate și să propună o abordare originală în ceea ce privește securitatea conținutului digital în sistemele de biblioteci virtuale din spațiul european.

Capitolul intitulat *Soluție privind managementul fluxului de date în sistemele de biblioteci virtuale* își propune prezentarea următoarelor problematici:

- descrierea conceptului de biblioteci virtuale, instrumentele utilizate, actanții implicați în managementul acestora, conținutului digital vehiculat precum și cerințele necesare contextului european de partajare a conținutului digital în scopul globalizării;

- evidențierea aspectelor importante în procesul de partajare de resurse precum și particularitățile procesului colaborativ în astfel de sisteme integrate de bibliotecă;
- prezentarea procesului de digitizare în scopul creării de conținut digital; care sunt etapele de lucru, ce instrumente sunt utilizate pentru realizarea eficientă a acestui proces precum și formele de publicare a conținutului digital după finalizarea procesului de digitizare;
- abordarea aspectelor de securitate ce intervin în cadrul procesului de partajare de resurse între mai mulți actanți de tip bibliotecă virtuală în contextul european precum și propunerea unei metodologii pentru protejarea conținutului digital din cadrul bibliotecilor virtuale pe baza analizei lexicale a obiectelor digitale și a mecanismelor de securitate specifice.

În capitolul *Soluție privind clasificarea informației stocată în volume mari de date* sunt prezentate următoarele aspecte:

- descrierea conceptului de informație stocată în volume mari de date cu acces liber la resurse; ce particularități au aceste volume mari de date și cum este influențată procesarea având în vedere că accesul la acest tip de resurse este liber;
- descrierea unor taxonomii specifice sistemelor integrate de bibliotecă prin care conținutul digital să poată fi manipulat în scopul protejării; sunt utilizate ontologii lexicale cu scopul de a îmbunătăți clasificarea supervizată a obiectelor din cadrul unei biblioteci;
- operațiile efectuate în cadrul etapei de preprocesare a datelor, după ce acestea au fost extrase din volumele mari de date cu scopul de a le transforma în informații; preprocesarea datelor este o etapă importantă fără de care rezultatele procesărilor ulterioare ar fi puternic influențate de redundanță, factorii indirecti de tip cauză-efect care influențează rezultatul final;
- clasificarea și ierarhizarea conținutului digital folosind algoritmi de clasificare supervizată al cărei rezultat este îmbunătățit de utilizarea unei analize semantice folosind ontologiile lexicale.

Soluție privind protejarea conținutului digital la nivelul sistemelor integrate de bibliotecă este capitolul în care sunt descrise:

- aspecte care țin de legislația în domeniu precum și aspecte tehnice ale managementului drepturilor digitale cu referințe la sistemele integrate de bibliotecă;
- soluție pentru protecția conținutului digital utilizând mecanisme de protecție a integrității și confidențialității pe bază de înveliș digital;

- soluție pentru verificarea autenticității conținutului digital existent într-o bibliotecă virtuală pe bază de coduri cu răspuns rapid.

Lucrarea își propune să dezbată probleme de securitate identificate în sistemele integrate de biblioteci virtuale. Sunt prezentate caracteristicile unor astfel de sisteme, ce instrumente se folosesc în interiorul acestora, ce conotație are conceptul de partajare și cum este implementată colaborarea, care sunt mecanismele de creare de conținut digital și cum se poate proteja acesta de manipulări și întrebuințări ilegale. Este descrisă o metodă prin care conținutul digital existent într-o bibliotecă virtuală este analizat semantic, ierarhizat și clasificat, cu scopul de a identifica domenii sensibil expuse preluărilor și prelucrărilor digitale neautorizate. În scopul protejării conținutului digital sunt prezentate metode de securitate bazate pe tehnici de management al drepturilor digitale și steganografie care să urmărească documentele, să identifice sursa de proveniență și să restricționeze accesul la acestea pentru utilizatorii neautorizați.

Această lucrare a fost realizată în cadrul proiectului “Cultura română și modele culturale europene: cercetare, sincronizare, durabilitate”, cofinanțat de Uniunea Europeană și Guvernul României din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013, contractul de finanțare nr. POSDRU/159/1.5/S/136077.

2. SOLUȚIE PRIVIND MANAGEMENTUL FLUXULUI DE DATE ÎN SISTEMELE DE BIBLIOTECI VIRTUALE

2.1. Biblioteci virtuale în contextul globalizării

Sistemele integrate de biblioteci virtuale, denumite în continuare sisteme ILS (Integrated Library System), reprezintă un ansamblu de instrumente care ajută la achiziția, organizarea, stocarea, regăsirea și prelucrarea informației.

Cu ajutorul acestor sisteme, utilizatorii, atât cei care le întrețin cât și cei care beneficiază de resursele puse la dispoziție de acestea, au posibilitatea de a accesa volume mari de informații corelate, funcție de criterii multiple astfel încât, prin procese complexe de *data mining*, (Witten, Frank, și Hall, 2011), să poată fi generate cunoștințe utile în luarea deciziilor. Știința informației, (Vickery, 2009), este cea care are ca obiectiv analiza, colectarea, clasificarea, manipularea, stocarea, extragerea și diseminarea informațiilor.

Procesul de data mining derulat în cadrul sistemelor integrate de bibliotecă reprezintă un mecanism specific prelucrării unui volum mare de date stocat în format digital astfel încât să se poată extrage informații și genera cunoștințe pe baza resurselor existente într-o bibliotecă virtuală, figura 2.1.

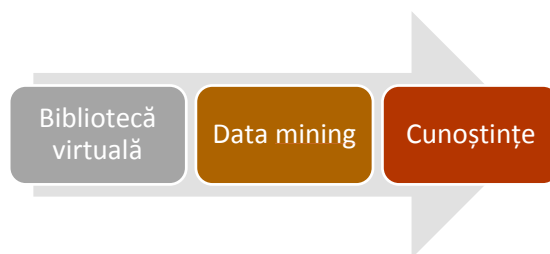


Figura 2.1 – Contribuția ILS în formarea de cunoștințe

Bibliotecile virtuale pun la dispoziția utilizatorilor o modalitate simplă de acces la conținut, pe bază de interfață web, denumit OPAC (Online Public Access Catalog). OPAC este un instrument care asigură accesul la documentele stocate în cadrul bibliotecilor virtuale. Interfața web oferă utilizatorilor instrumente ușor de utilizat, de căutare avansată pe bază de obiecte de tip *facets*. Acestea reprezintă criterii definite prin intermediul unui triplet <A,B,C> format din:

- A. interfața web – aplicația online, cu conținut dinamic, care permite utilizatorului să formuleze, construiască, cererile de regăsire;

- B. cerere de regăsire – sintaxă care face apel la algoritmi de căutare avansați ce primesc ca input un set de concepte, iar pe baza unor indecși construiți și întreținuți dinamic sunt furnizate rezultate care să satisfacă criteriile specifice;
- C. indecși definiți la nivelul documentelor – o structură de date optimizată asupra căreia sunt aplicate cereri de regăsire, (SRU, 2015), construită cu scopul de a îmbunătăți timpul de căutare.

Scopul principal al sistemelor integrate de biblioteci alături de alte soluții pentru bibliotecile digitale este de a gestiona conținutul digital la un nivel de calitate superior (English, 2009), oferind instrumente de management pentru utilizatorii externe și pentru personalul bibliotecii.

Patrimoniul cultural stocat în cadrul unor astfel de sisteme este un atu important care necesită să fie protejat împotriva manipulărilor malițioase. Cultura în cadrul unei societăți informaționale bazată pe cunoștințe influențează în mod direct economia și nivelul de dezvoltare al acelei țări (Filip și Cojocaru, 2010).

Aceste sisteme trebuie să ofere acces facil la conținutul digital specific bibliotecilor, muzeelor sau arhivelor, pe baza suportului oferit de diverse instrumente software care sunt interconectate între ele.

În cele ce urmează sunt prezentate o serie de instrumente software dezvoltate de către furnizori specializați de soluții automate pentru biblioteci care pot fi interconectate cu scopul de a oferi informații pentru un acces facil al utilizatorilor finali:

- soluția Primo (Primo, 2015), utilizată pentru a oferi soluții pentru identificarea și obținerea unui spectru complet al materialelor disponibile în biblioteci: imprimate, electronice și digitale, indiferent de formatul sau locația în care se regăsește;
- MetaLib (Metalib, 2015), folosit pentru a permite căutarea în cadrul unor multiple baze de date;
- Rosetta (Rosetta, 2015), instrument folosit în scopul conservării datorită faptului că bibliotecile virtuale se ocupă de cantități mari de materiale;
- Aleph (Aleph, 2015), un instrument pentru managementul colecțiilor tipărite, este un sistem integrat de bibliotecă care servește utilizatorilor în efortul acestora de a cataloga colecțiile fizice;
- DigiTool (DigiTool, 2015), o soluție performantă folosită pentru afișarea materialelor digitale într-o manieră elegantă.

Un sistem clasic ILS implică următorii actanți și operații care au ca obiectiv principal managementul resurselor digitale stocate de acesta după cum este descris în figura 2.2.

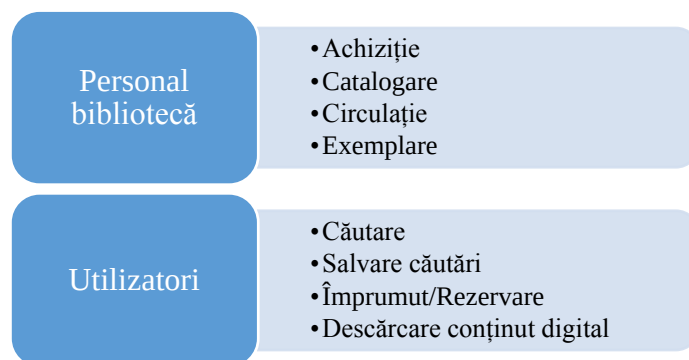


Figura 2.2 – Actanții și operațiile la nivelul unui ILS

Acest tip de sistem ILS rulează în cadrul unei arhitecturi distribuite care este foarte eficientă. Dar o altă generație de sisteme integrate de biblioteci se află în dezvoltare, o generație care oferă mai multe avantaje decât cea actuală. Un exemplu de ILS aflat în cloud este ALMA (Exlibris ALMA, 2015), dezvoltat de către Exlibris care este succesorul prezentului produs de tip client-server denumit Aleph ce poate gestiona o dimensiune mare de documente. ALMA este un sistem integrat de bibliotecă care operează în cloud și care oferă suport pentru implementarea tuturor proceselor specifice bibliotecilor virtuale. Prelucrările în cadrul unui astfel de sistem sunt realizate în cloud astfel încât instituțiile care îl implementează să beneficieze de avantaje precum resurse de procesare pe măsura gradului de încărcare, modalități de back-up și restaurare eficiente precum și acces permanent, fără întreruperi datorate perioadelor de mentenanță.

ALMA este un sistem integrat de bibliotecă care asigură colaborarea și cu alte sisteme destinate prelucrării documentelor digitale. Datorită centralizării proceselor, astfel de sisteme pun la dispoziția administratorilor componente statistice importante ce asigură un nivel constant de suport în ceea ce privește resursele digitale partajate. O caracteristică importantă este cea legată de achiziția de noi materiale.

Sistemul, pe baza analizei efectuate asupra cererilor de regăsire furnizate de utilizatori, identifică și propune achiziția de noi materiale astfel încât să crească gradul de interes asupra patrimoniului cultural pus la dispoziție de bibliotecă și în același timp să ofere informații cu privire la cele mai puțin relevante documente care, eventual, necesită să fie arhivate în scopul optimizării costului de stocare. În urma implementării unui astfel de sistem, capabil să gestioneze la un nivel avansat conținutul digital, apare necesitatea protejării documentelor, (Jones, 2007), datorită costurilor ridicate care stau la baza achiziției, stocării, prelucrării și arhivării acestora.

Un sistem ILS trebuie să furnizeze soluții pentru servicii multiple pentru diferite tipuri de activități care se desfășoară în cadrul bibliotecilor, muzeelor sau arhivelor, pe baza suportului oferit de diverse instrumente software care sunt interconectate.

Astfel de instrumente puternice pentru sisteme ILS sunt dezvoltate de Exlibris, cu o experiență de mai bine de 20 de ani, pornind de la software cadru dedicat serviciilor desfășurate în cadrul bibliotecilor, continuând cu sisteme distribuite implementate într-o arhitectură de tip client-server și ajungând, în dezvoltare acum, la infrastructuri de tip cloud pentru managementul bibliotecilor. Spre exemplu, LIBISnet (LIBISnet, 2015) este o rețea de biblioteci din regiunea flamandă a Belgiei, ce cuprinde cel puțin 30 de membrii precum instituții guvernamentale, organizații publice și private. Datorită faptului că oferă servicii care centralizează resursele digitale ale mai multor biblioteci, astfel de nuclee sunt implicate în multe proiecte care fac posibilă identificarea, proiectarea și implementarea de soluții fiabile și complete pentru sistemele de biblioteci. Proiecte precum VEP, Europeana inside, Flandrica sau Open Vlacc, care este una dintre cele mai mari rețele LIBIS, contribuie în cadrul comunității cu idei și cunoștințe care au scopul de a identifica multiple soluții eficiente pentru ILS. Open Vlacc reprezintă o bază bibliografică organizată într-un catalog centralizat a bibliotecilor din Anvers, Bruges, Bruxelles, Gent, Hasselt și Leuven. Combinat cu date locale, acesta creează un sistem PBS (Provincial Library System), folosit în cadrul bibliotecilor publice. Open Vlacc este un catalog central pentru biblioteci publice în cadrul regiunii flamande, fiind automat încărcat cu înregistrări CDR și Boekenbank de pe întreg teritoriul țării.

În rețeaua PBS există în jur de 30 de membrii din cadrul provinciei belgiene “Vlaams-Brabant”. Rețeaua PBS suportă acces în cadrul colecțiilor online, utilizatorii pot vizualiza locația materialelor digitale având posibilitatea de a lansa o cerere de împrumut pentru un material disponibil. Rețeaua PBS are la bază un ILS cu arhitectură client-server, Aleph, versiunea 21, care gestionează întreaga rețea, structurând conținutul în diferite baze de date în funcție de particularitatea fiecărui element. Înregistrările bibliografice sunt stocate într-o bază de date PBS01 cu înregistrări structurate în diferite seturi logice în funcție de locația fizică.

Sistemul integrat de bibliotecă ALMA, cu suport în cloud, își propune să fie succesorul actualului produs Aleph, care funcționează pe o arhitectură de tip client-server. În prezent, sistemul Aleph gestionează un număr foarte mare de documente, în jur de 7.3 milioane, și un număr de aproximativ 2.3 milioane de copii. Rețeaua PBS oferă sprijin pentru integrarea componentei ALMA cu scopul de a oferi suport, în viitor, prin intermediul platformei SaaS, System as a Service.

LIBIS, în calitate de partener de dezvoltare al sistemului ALMA, a început integrarea sistemului URM (Unified Resource Management), la începutul anului 2014,

având la nivelul superior un sistem de timp URDD (Unified Resource Discovery and Delivery), bazat pe instrumentul Primo, denumit Limo, precum este prezentat în figura 2.3.



Figura 2.3 – Arhitectura unui sistem de bibliotecă

Sistemul ALMA oferă suport pentru toate procesele desfășurate în cadrul unei biblioteci precum operațiile de selecție, achiziție, managementul elementelor de tip metadata, digitizarea, partajarea resurselor, capacitatea de căutare bazată pe motoare Apache Lucene precum și posibilitatea de integrare externă în cadrul altor tipuri de sisteme de biblioteci.

Avantajele principale pe care le oferă implementarea sistemului ALMA sunt reprezentate de trecerea de la arhitectura client-server la paradigma bazată pe sistemul cloud, în care accentul se pune pe resursele electronice, tinzând la noi așteptări în ceea ce privește caracteristica de disponibilitate, precum și pe un motor statistic mai puternic care poate identifica ariile care necesită o abordare strategică în scopul de a obține eficiență și calitate maximă.

Prin utilizarea unei arhitecturi în cloud, ALMA oferă capacități de procesare îmbunătățite. Dacă resursele nu mai creează o problemă, atunci întreaga atenție se concentrează spre modalitatea în care conținutul digital este gestionat, iar acest aspect este foarte important în contextul bibliotecilor virtuale. ALMA nu doar procesează conținut digital stocat în cadrul sistemului pentru prezentarea sa publicului larg, ci și preia date de la utilizatori pentru a analiza comportamentul acestora și a determina anumiți parametri utilizați în procesul de achiziție de noi materiale sau optimizarea surselor existente.

Acesta este un proces pe baza căruia bibliotecile își îmbunătățesc patrimoniul lor cultural prin achiziții pe baza titlurilor cerute de către utilizatori. Pe baza procesului de achiziție orientat spre utilizatori, bibliotecile vor tinde să cumpere sau să includă în cataloagele lor titluri care sunt cerute frecvent de către utilizatori și nu sunt disponibile.

Cealaltă categorie fiind treptat optimizată prin procese de arhivare, astfel reducând cantitatea de materiale care nu este relevantă grupului țintă.

Conținutul digital care nu este frecvent accesat este redirecționat spre o componentă specializată de conservare sau arhivare. Astfel, sistemul este menținut curat iar motorul de căutare rulează mai eficient.

O altă modalitate de implementare a acestei particularități orientate spre utilizatori, precum este prezentat în (Doinea și Pocatilu, 2015), se bazează pe o caracteristică importantă, motoarele analitice. Datorită capacității de a rula sub o cupolă multiplă, caracteristică denumită *multi tenancy*, sistemul este capabil să ofere statistici relevante despre materialele care sunt cele mai vizualizate, precum și teste de performanță realizate la diferite niveluri de stres.

Un aspect important în contextul sistemelor integrate de biblioteci care gestionează cantități foarte mari de date este acela de a fi pe cât se poate de rapid atunci când utilizatorii trimit cereri multiple de căutare în același timp, caracteristică denumită acces concurrent. Nu numai că trebuie să răspundă într-o manieră rapidă, dar și rezultatele trebuie să fie cât se poate de relevante. Pentru acest lucru, conceptul de căutări pe bază de chei este prezentat în (Gollub et. al., 2013), conform căruia o căutare bazată pe chei reprezintă un context dinamic format din descriptori pentru documente, aceștia fiind definiți implicit de către un index și de un model de regăsire pe bază de căutare de referințe: căutările pe bază de chei pentru un document sunt căutările minimale care returnează documentul în rezultatele de pe primele locuri.

În (Gollub et. al., 2013) fiecare document este gândit astfel încât să conțină căutări cheie denumite și *keyqueries* în locul clasicelelor cuvinte cheie cunoscute sub denumirea de *keywords*. Aceste noi tipuri de sintagme vor ajuta motoarele de căutare integrate în bibliotecile virtuale pentru a găsi cât de rapid posibil documentele care conțin acele sintagme. Pentru a determina căutările cheie pentru un document în (Gollub et. al., 2013) se prezintă un algoritm de căutare exhaustiv alături de strategii eficiente de tăiere, o tehnică bazată de arbori de decizie.

2.2. Partajare și colaborare în bibliotecile virtuale

Partajarea reprezintă metoda prin care resursele unui sistem integrat de bibliotecă sunt utilizate de mai multe entități, fie acestea procese sau utilizatori, pe baza unui set de restricții, norme, ce reflectă felul în care acestea urmează să fie întrebuințate. Partajarea trebuie să țină cont de o serie de elemente ce definesc atât resursele cât și entitățile care le folosesc, precum:

- dimensiunea mulțimii partajate; după dimensiunea resurselor se poate discuta de următoarele tipuri de partajări:
 - partajare de tip 1 la n: aceeași resursă la mai mulți utilizatori; partajarea unei resurse la mai mulți utilizatori implică mecanisme specifice care să gestioneze accesul concurent la aceasta precum și tehnici de înregistrare a operațiilor și crearea unui jurnal al tuturor acțiunilor care au avut loc la nivelul acesteia;
 - partajare de tip n la n: mai multe resurse la mai mulți utilizatori presupune existența unui sistem pe baza căruia să fie alocate aceste resurse fără a crea dependențe între procesele de alocare.
- natura resurselor partajate; natura acestora determină modalități diferite de partajare, după cum partajarea unui obiect digital, stocat în cadrul unei biblioteci virtuale, este diferită de partajarea zonei de memorie din cadrul unui sistem informatic sau de cea a procesorului utilizat pentru efectuarea calculelor a două procese distincte.

Colaborarea este descrisă în (Mican, Tomai și Cocoș, 2009) ca fiind procesul prin care un sistem are capacitatea de a gestiona un număr important de utilizatori care lucrează împreună spre atingerea aceluiași obiectiv, folosind resurse proprii; o astfel de abordare se bazează pe un sistem complex de management al documentelor și al proceselor distribuite. Colaborarea reprezintă un proces de partajare de resurse cu un nivel suplimentar de restricții care să specifice felul în care aceste resurse sunt partajate.

În cadrul unui proces colaborativ entitățile care utilizează setul de resurse lucrează pentru atingerea aceluiași obiectiv, fiecare procesare a resursei fiind direct dependentă de o alta, apriorică, cu excepția primei procesări.

În cadrul sistemelor integrate de bibliotecă mecanismele de partajare și colaborare sunt bine definite, întrucât intersecția utilizatorilor și a proceselor care folosesc resursele generează o matrice a drepturilor de acces foarte complexă, (Sanchez et. al., 2009). Această complexitate poate fi tratată doar dacă la nivelul sistemului integrat de bibliotecă sunt prevăzute măsuri prin care resursele acestuia sunt utilizate corespunzător.

Un exemplu de partajare de resurse, în cadrul unui sistem integrat de bibliotecă, îl reprezintă felul în care fișele bibliografice sunt utilizate de către utilizatorii de sistem cu roluri diferite, funcție de specificul activității acestora.

Începând de la momentul creării unei fișe bibliografice, la nivelul serviciului de achiziție, continuând cu completarea acesteia de către serviciul de catalogare și până la momentul în care fișa intră în gestiunea serviciului de preservare digitală, o fișă

bibliografică suferă o serie de transformări menite să aducă un plus valoare patrimoniului cultural.

Aceste modificări aduse de fiecare grup de utilizatori, respectiv de procesele care rulează în mod automat în plan secund, în cadrul sistemului integrat de bibliotecă cu rolul de a elimina redundanțele sau de a omogeniza conținutul, trebuie să păstreze caracteristicile de bază ale informației stocate la nivelul fișei bibliografice, fără a denatura în vreun fel calitatea expusă de aceasta.

Partajarea resurselor în sistemele integrate de bibliotecă este tratată diferențiat la următoarele niveluri logice, toate componentele fiind perfect sincronizate pentru a nu periclita integritatea resurselor gestionate:

- baza de date în care sunt stocate obiectele bibliotecii virtuale; înregistrările stocate la nivelul bazei de date sunt protejate de mecanisme specifice de blocare a accesului în momentul în care sunt în curs de editare; astfel se evită suprascrierile nedorite în momentul în care mai mulți utilizatori doresc să acceseze aceeași resursă cu scopul de a o modifica;
- aplicația distribuită în arhitectură client-server care permite utilizatorilor să utilizeze resursele; înregistrările sunt marcate la acest nivel prin permisiuni de editare ierarhizate pe niveluri de prioritate astfel încât odată ce nivelul de prioritate crește, grupurile de utilizatori cu valori sub nivelul curent să nu mai poată realiza modificări asupra conținutului decât prin solicitări justificate și aprobate de către un utilizator autorizat;
- serverul web care accesează resursele web gestionate de sistemul integrat de bibliotecă; la acest nivel înregistrările sunt mai puțin vulnerabile deoarece accesul la resurse se realizează mai mult în modul consultare; la acest nivel sunt utilizate mecanisme de actualizare corespunzătoare astfel încât conținutul afișat utilizatorilor să fie în permanență corelat cu ceea ce se află stocat în baza de date a bibliotecii virtuale.

La nivel național, geopolitic, cultural sau teritorial se impune o sincronizare a tuturor sistemelor de biblioteci virtuale pentru eliminarea redundanței și pentru a eficientiza procesul de catalogare și digitizare în scopul creării de conținut digital unitar, ca element valoric al patrimoniului cultural, (Filip, 2006). Această armonizare între mai multe instituții culturale care gestionează cataloagele unor biblioteci virtuale presupune utilizarea unor protocoale standardizate de partajare a conținutului digital precum Z39.50 (Information Retrieval, 2015) sau OAI-PMH (Lagoze et. al, 2015).

Dezvoltarea instrumentelor specifice automatizării proceselor în bibliotecile virtuale a cuprins chiar și zona programelor software cu acces liber (Breeding, 2009). Standardul internațional ISO 23950, (ISO, 2014), care referă protocolul Z39.50 are rolul de a putea lansa cereri de regăsire către sistemele integrate de bibliotecă fără a cunoaște sintaxa de căutare specifică acelui sistem.

Protocolul returnează o listă cu rezultate care au legătură cu termenii cheie utilizați în cererea de regăsire trimisă spre procesare. Protocolul OAI-PMH, (Lagoze et. al, 2015), folosit pentru partajarea resursele în mediul online, este foarte eficient întrucât utilizează un set restrâns de expresii, denumite verbe, pentru a lansa cereri de regăsire în cadrul depozitelor de date care publică la rândul lor meta-date tot prin intermediul acestui protocol. Partajarea resurselor prin OAI-PMH implică doi actanți:

- furnizori de date – sistemele care au implementat protocolul OAI-PMH cu scopul de a publica meta-date și de a răspunde cererilor de regăsire primite de la clienți;
- furnizori de servicii – reprezintă clienții care lansează cererile de regăsire către furnizorii de date cu scopul de a colecta meta-date din cadrul depozitelor de date.

În cadrul protocolului OAI-PMH se disting trei entități cu roluri diferite în managementul meta-datelor stocate în aceste depozite, precum se observă în figura 2.4.

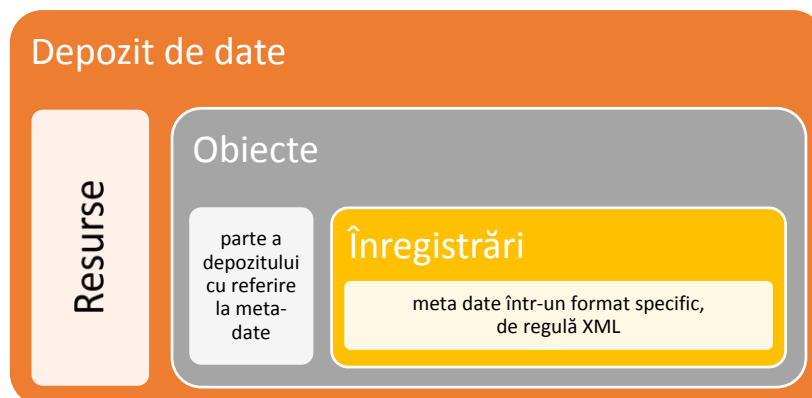


Figura 2.4 – Structura depozitului de date din perspectiva OAI-PMH

Protocolul OAI-PMH, (Goebert, Harriehausen-Mühlbauer și Furnell, 2014), utilizat pentru partajarea resurselor între diverse instituții culturale dispune de mecanisme de transfer de meta-date de tip unidirecțional, dinspre furnizorii de date spre clienții care lansează cererile de regăsire. Acest tip de protocol nu permite implementarea unor mecanisme colaborative în mod implicit.

Pentru implementarea unor procese colaborative între mai mulți actanți culturali este necesară asigurarea unei relații de transfer de meta-date bidirecționale astfel încât fiecare să poată prelua, modifica, apoi publica versiuni actualizate ale meta-datelor inițiale, model de partajare în medii eterogene prezentat în (Houssos et. al., 2014).

Pe lângă această relație bidirecțională este utilă și implementarea unei funcții care să anunțe toți furnizorii de servicii, dezvoltatori de colecții, de îndată ce un nou furnizor de date este disponibil, precum descrie (Houssos et. al., 2014).

Colaborarea între sistemele integrate de bibliotecă, la nivelul instituțiilor culturale europene, are rolul de a construi progresiv patrimoniul cultural digital al spațiului european, asigurând durabilitatea formelor digitale precum și conservarea formelor tipărite ale căror expunere este cu mult diminuată.

Modelul colaborativ propus este format din următorii actanți care participă activ la schimbul de resurse electronice pentru a construi un depozit central la care să aibă acces toți utilizatorii spațiului virtual. Componentele modelului sunt prezentate în figura 2.5.

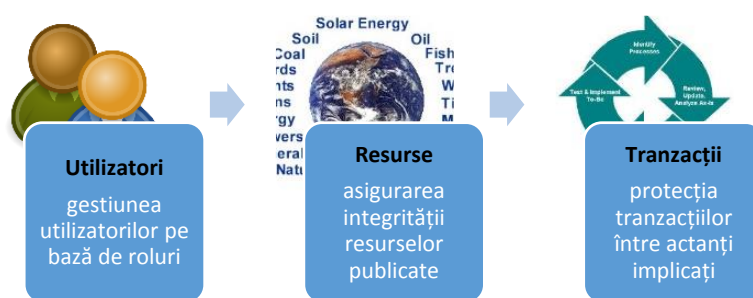


Figura 2.5 – Componentele modelului colaborativ și caracteristicile fiecărui grup

La nivelul acestor componente este necesară derularea unor procese de sincronizare și management al versiunilor conținutului electronic partajat astfel încât utilizatorilor să li se ofere în permanență cele mai bune rezultate pentru cererile de regăsire trimise cu ajutorul protocolului Z39.50.

2.3. Digitizare, proces generativ de conținut digital

O modalitate de îmbogățire a unei biblioteci virtuale cu conținut interesant și atrăgător este prin digitizarea celor mai căutate titluri de pe piață. Digitizarea presupune un know-how specific bibliotecilor precum și costuri cu resursele implicate în proces.

Din acest motiv, rezultatele procesului de digitizare a materialelor, conținut digital cu valoare adăugată, trebuie protejate folosind controale speciale de securitate. Digitizarea este un proces complex ce cuprinde multiple stagii, precum este descris în figura 2.6, conducând la un produs finit care va fi încărcat și afișat în cadrul portalurilor de bibliotecii virtuale.

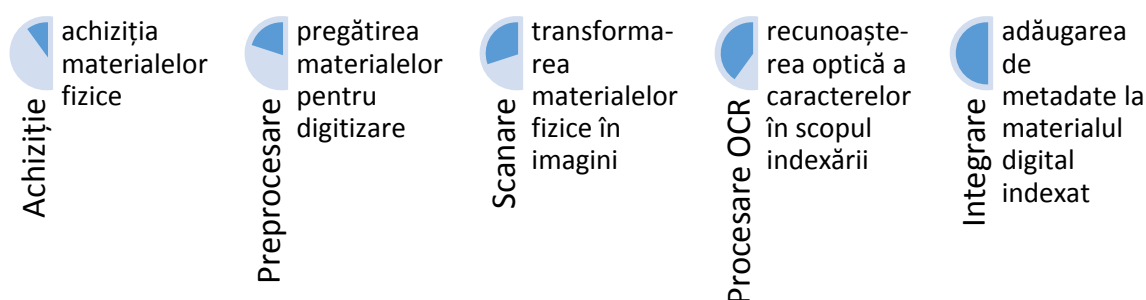


Figura 2.6 – Etapele procesului de digitizare a conținutului digital

Procesul de digitizare pornește de la forma tipărită a materialului care este pregătită pentru a fi transformată într-o resursă electronică. Pregătirea materialului fizic constă în numerotarea fiecărui element cheie din cadrul documentului actual precum descrieri speciale, tabele, figuri și alte tipuri de elemente care nu pot trece de faza de recunoaștere optică a caracterelor.

Odată ce documentul este pregătit, procesul de scanare poate începe, transformând materialul tipărit într-o imagine digitală detaliată ce reprezintă date de intrare pentru următoarea etapă.

Procesul de recunoaștere optică a caracterelor este aplicat la nivelul fiecărei imagini digitale, astfel încât să fie identificate toate fragmentele de text cu o rată cât mai mică de eroare, în scopul unei indexări cât mai corecte.

Procesul de adăugare a elementelor de tip metadată, pentru obținerea conținutului digital final, utilizează tehnici complexe multimedia care adaugă date în documentul digital actual sau separat, folosind descriptori speciali care sunt mai ușor de citit de către sistemele de management al drepturilor digitale.

Pentru crearea elementelor de tip metadate în ALMA cu scopul de a conecta conținutul digital de înregistrările asociate, specialiștii folosesc aplicația Filemaker prin intermediul modului Digicorder. Aplicația este folosită pentru a descrie conținutul unei cărți, precum detalii despre figuri și tabele, pentru a numerota automat paginile unei cărți, pentru a adăuga notații specifice care apar în cartea originală și nu sunt altfel reflectate facil în materialul digital.

Aplicația, de asemenea, descrie structura capitolelor prin crearea automată a unui cuprins care servește ca un ghid pentru asocierea conținutului digital regăsit în imaginile scanate.

Procesul de digitizare pornește prin utilizarea de instrumente precum OCR (Optical Character Recognition) și NER (Name Entity Recognition), iar materialul digitizat este

stocat în arhive sau prezentat utilizatorilor finali. Datele de ieșire pot fi introduse în LIAS, un instrument pentru arhivare care are la bază sistemul Rosetta, succesor Digtol pentru vizualizarea online a resurselor digitale. Termenul de LIAS se referă la aplicațiile și serviciile LIBIS pentru arhivare.

Depozitul central pentru stocarea conținutului digital și pentru arhivare este integrat cu sistemul de management al metadatelor, orientat pe domenii pentru a oferi capacitatea de arhivare pentru biblioteci precum și pentru muzee. LIAS implementează o structură ierarhică, oferă mecanisme de livrare orientate pe conținut și impune politici de drepturi de acces în scopul prevenirii accesului neautorizat la conținutul digital.

LIRIAS este un instrument universitar de arhivare a publicațiilor pentru cercetătorii afiliați mediului academic sau unei instituții de cercetare. LIRIAS este un instrument cu acces gratuit care permite înregistrarea, stocarea, indexarea, arhivarea și distribuția cercetării în format digital din cadrul asociației KU Leuven. Acest instrument este folosit pentru a arhiva toate tipurile de rezultate ale cercetătorilor pentru care se atribuie un identificator unic, vizibil de oriunde. De asemenea, în scopul conservării digitale, Rosetta este un instrument utilizat pentru a permite universităților să obțină o preservare pe termen lung a conținutului digital din cadrul bibliotecilor, arhivelor și muzeelor.

În scopul facilitării procesului de livrare a conținutului către alte sisteme de biblioteci, un cadru MIF (Metadata Interoperability Framework) poate fi folosit pentru a partaja conținutul digital la nivelul mai multor platforme. Procesul de încărcare implică definirea elementelor de tip metadata, caracteristicile de previzualizare și validare, servicii de încărcare a datelor pe baza componentei Sword, extragerea datelor folosind protocoale specializate precum și suport pentru transformare și asociere.

Pe lângă conținutul digital care poate fi rezultat al unui proces de digitizare, sistemul ILS necesită deținerea elementelor de tip metadata cu scopul de a fi capabil să descrie materialul electronic atașat. OCR și NER sunt instrumente create în interiorul proiectului SUCCEED (Support Action Centre of Competence in Digitization) pentru realizarea procesului de digitizare. Scopul este de a îmbunătăți instrumentele OCR pentru documente text istorice într-o manieră cât mai automată posibil. Instrumentele NER au fost dezvoltate pentru identificarea persoanelor, organizațiilor și locațiilor din cadrul elementelor text care sunt procesate folosind instrumentele OCR. Fluxul de lucru implementat pentru a atinge rezultatele propuse conține următoarele etape:

- digitizarea, etapa în care materialele fizice sunt digitizate, transformate din forma lor fizică în materiale electronice stoca prin intermediul imaginilor;

- atestare, crearea unui etalon pentru evaluarea OCR; etalonul reprezintă modalitatea în care trebuie să arate o pagină;
- setarea evaluării, crearea de seturi pentru antrenare și testare folosind datele de ieșire de la etapa anterioară;
- îmbunătățirea calității, etapa în care se antrenează sistemul OCR pentru recunoașterea caracterelor speciale prin utilizarea unui dicționar specializat furnizat de Institutul de Lexicologie Olandeză; datele de ieșire ale etapei curente reprezintă modelul utilizat în procesul OCR;
- executarea OCR, procesul efectiv de extragere a caracterelor din cadrul documentului imagine.

În cursul implementării etapelor definite în fluxul de lucru, corectitudinea procesului de recunoaștere a fost într-un procent de 80% precum și 80% dintre numele entităților au fost identificate, lucru ce demonstrează succesul acestui proiect. Există, de asemenea, câteva dezavantaje ale sistemului de care trebuie să se țină cont, cu privire la generalizarea mecanismului. Anumite etape trebuie reevaluate în momentul schimbării materialului digital cu scopul recalibrării sistemului cu noile formate de cărți.

LIBIS a dezvoltat, de asemenea, LibisCoDe care suportă servicii de transformare din MARC în EDM și din LIDO în EDM pentru a facilita procesul de încadrare cu succes a metadatelor exportate din cadrul bibliotecii virtuale în Europeana. Aceste caracteristici sunt furnizate ca servicii pentru integrarea CMS sau pot fi folosite de un client REST.

LIMO este un alt exemplu de implementare cu succes a produsului Exlibris Primo. Acesta ajută utilizatorii să caute publicații imprimate sau electronice din cadrul mai multor surse. Astfel, implementarea Limo unește căutările din cadrul tuturor resurselor bibliotecilor, chiar și între resurse externe din cadrul altor sisteme de management al conținutului digital. Folosind Limo se poate căuta prin cataloage LIBISnet, depozitul academic Lirias precum și prin Primo Central care conține date extrase de la diferiți editori. Fiind un sistem de descoperire, Limo folosește o interfață utilizator simplă care nu afectează interogările de căutare ale utilizatorilor. Motorul de căutare utilizat este Apache Lucene Core, un motor de căutare de înaltă performanță prin definirea unui format XML personalizat pentru reguli de normalizare denumit PNx, Primo XML file. Limo definește fațete obiectelor prin care conținutul poate fi filtrat folosind diferite criterii care descriu cuprinzător setul de date.

2.4. Metodologie pentru securitatea conținutului digital

În scopul implementării partajării resurselor pe baza protocolului OAI-PMH, între diferite sisteme integrate de bibliotecă ale instituțiilor culturale deținătoare de biblioteci virtuale, următoarele premise de lucru trebuie asigurate:

- fiecare actant trebuie să fie în același timp atât furnizor de date cât și consumator de servicii de meta-date;
- existența unui mecanism de notificare cu privire la actualizările efectuate la nivelul furnizorilor de meta-date;
- implementarea unui model de gestiune a versiunilor pentru înregistrările care suferă modificări la nivelul oricărui actant.

Securitatea unor procese colaborative prezintă o serie de caracteristici particulare datorită modului în care se desfășoară acestea în raport cu resursele pe care le antrenează în sistem. Colaborarea într-un sistem, după cum este menționat în (Doinea și Van Osch, 2010), poate servi unor scopuri distincte, după cum securitatea în astfel de sisteme pune accentul pe elemente diferite funcție de contextul colaborativ, precum:

- informare – permit publicarea de conținut digital în același spațiu informatic de mai mulți utilizatori cu scop informativ; ex. rețeaua Facebook;
- negociere – permit desfășurarea unor procese de licitație online, având la bază obiecte digitale; ex. rețeaua Bidson sau rețele de tranzacționare Forex;
- conlucrare – permit lucrul în echipă pentru atingerea aceluiași obiectiv, fiecare utilizator folosind însă resurse individuale; ex. Microsoft Project;
- cooperare – permit utilizarea aceleiași resurse de către mai mulți utilizatori pentru îndeplinirea unor obiective comune; ex. Google Documents; rețeaua Dropbox.

Modelul colaborativ asigură un nivel de partajare de meta-date de tip bidirecțional, care presupune dezvoltarea incrementală a unui patrimoniu cultural european. În figura 2.7 este prezentată o diagramă SWOT cu principalii factori pentru un astfel de sistem.

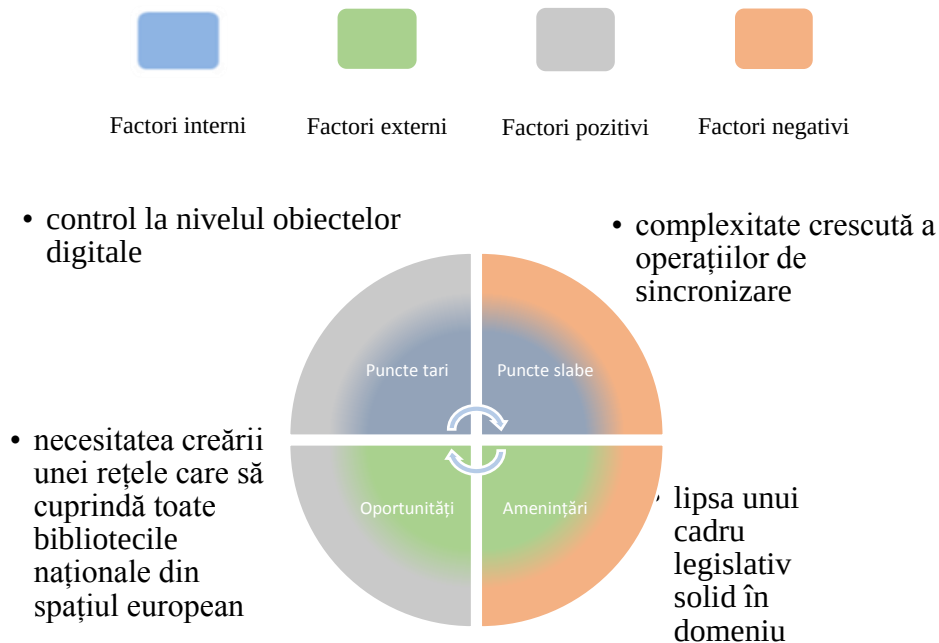


Figura 2.7 – Diagrama SWOT a modelului colaborativ la nivelul bibliotecilor virtuale

Funcție de aceste aspecte pe care le vizează un proces colaborativ, securitatea ridică diferite probleme care trebuie avute în vedere. Natura proceselor colaborative este decisivă atunci când se analizează securitatea acestora, (Carminati și Ferrari, 2009), după cum urmează:

- securitatea sistemelor colaborative care pun accentul pe publicare de conținut digital cu rol de informare trebuie să urmărească disponibilitatea acestora; astfel de sisteme nu permit existența unor timpi de inoperabilitate, fiind necesară prezența unor sisteme cu copii de rezervă sau a unor sisteme care să preia sarcinile celor care nu mai sunt funcționale;
- în cazul sistemelor colaborative care au ca principală activitate negocierea securitatea trebuie să vizeze aspectele temporale ale tranzacțiilor precum și confidențialitatea acestora; în astfel de sisteme este foarte important să se cunoască cu o precizie la nivel de milisecunde când a fost efectuată o tranzacție iar detaliile acesteia să fie cunoscute doar de persoanele autorizate;
- sistemele colaborative al căror principal obiectiv este partajarea de către mai mulți utilizatori a unor resurse divizibile și independente au ca principală caracteristică a securității, integritatea conținutului sub toate aspectele acestuia: al transferului pe canalele de comunicație; al stocării în baze sau depozite de date; al publicării pe paginile de internet;
- în sistemele în care se utilizează tehnici de partajare a aceleiași resurse de către mai mulți utilizatori, pentru obținerea aceluiași obiectiv, caracteristica de bază a

securității, vitală în acest context, este aceea de non-repudiere; fiecărui utilizator trebuie să-i fie foarte bine atribuite operațiile care au fost efectuate în sistem pentru o cât mai bună urmărire a evoluției procesului colaborativ.

În cadrul unui model colaborativ implementat între mai multe sisteme de biblioteci virtuale toate aceste aspecte ale colaborării sunt prezente iar securitatea trebuie asigurată la următoarele niveluri de lucru:

- utilizatori – din perspectiva acestora, securitatea trebuie să asigure principiul funcționării pe bază de roluri, implementarea unei politici pe bază de roluri de acces; un astfel de model este descris în (Demurjian et. al., 2009); rolurile trebuie definite după o analiză minuțioasă a operațiilor din sistem și a necesarului de resurse; trebuie cunoscut exact cine are acces și la ce resurse are acces, pentru a păstra o imagine clară asupra modificărilor aduse acestora în urma proceselor colaborative care le accesează;
- resurse – în vederea unei manipulări corecte a resurselor, măsurile de securitate trebuie să asigure integritatea acestora la nivelurile unde acestea sunt utilizate; măsurile de integritate la nivelul stocării sunt asigurate de sistemele de gestiune a bazelor de date; transmiterea datelor prin rețea are implementate mecanisme la nivelul stivei de protocoale TCP/IP pe bază de sume de control, (Stone și Partridge, 2000), care asigură transmiterea corectă a pachetelor; tot la nivelul resurselor se impune implementarea de măsuri pentru crearea copiilor de siguranță în cazul avariilor de orice natură ce pot apărea;
- tranzacții inter/intra biblioteci – acestea sunt cele mai sensibile elemente ale sistemului pentru că pot altera conținutul de o manieră ireversibilă; din perspectiva securității, tranzacțiile efectuate între diferite instituții trebuie să aibă un caracter confidențial, eliminând orice posibilitate de a capta informația vehiculată între acestea; în acest scop pot fi implementate sisteme de criptare pe bază de chei publice precum Diffie–Hellman sau RSA, (Tao, et. al., 2014), bazat pe standardul PKCS#1 sau sisteme simetrice cu chei private precum AES sau DES, (Damjanovic și Simic, 2013).

Obiectele online ce se regăsesc la nivelul bibliotecilor digitale reprezintă principala resursă a patrimoniului cultural, moștenire a erei digitale, în scopul căruia au fost studiate modele cibernetice pentru a se realiza o formalizare din punct de vedere informatic (Filip, 2001). În cadrul unei biblioteci virtuale gestionată de un sistem integrat care manipulează resursele este propusă o metodologie, (Don, 2004; Kumar, 2005), pentru managementul

conținutului digital în scopul protejării acestuia de utilizările necorespunzătoare. Etapele acesteia sunt prezentate în figura 2.8.

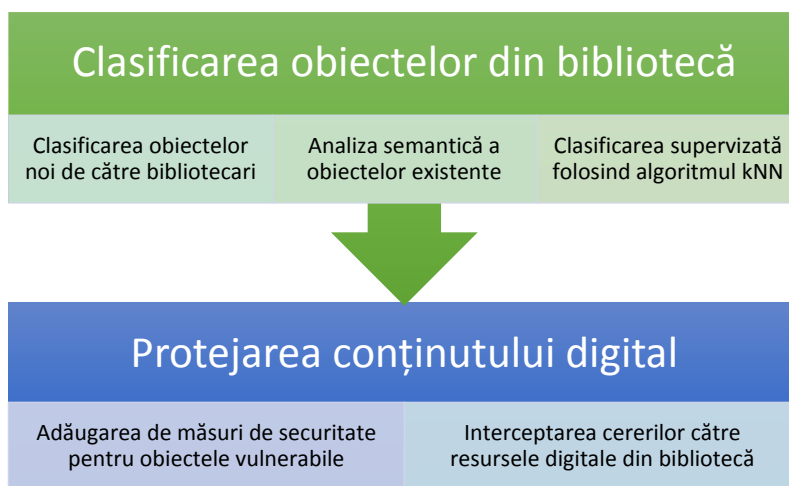


Figura 2.8 – Etapele metodologie de protecție a conținutului digital

În cadrul unei biblioteci virtuale există resurse care necesită digitizare, denumite resurse noi, iar cele care sunt digitizate, denumite resurse existente. În cazul resurselor noi care urmează să fie introduse în sistemul integrat de bibliotecă, catalogatorii, experți ai domeniilor în care activează au capacitatea de a le clasifica atunci când realizează munca de catalogare și digitizare. Această clasificare se va realiza funcție de o serie de caracteristici pe care catalogatorii le urmăresc în momentul în care realizează documentarea resursei digitale. Expertiza acumulată de aceștia, în domenii precum manuscrise - carte rară, numismatică, stampe, microfilme, hărți sau muzică îi ajută să stabilească o clasificare și o ierarhizare pe trei niveluri de importanță: scăzut, mediu, ridicat.

Catalogatorii realizează o descriere detaliată a fiecărui obiect în parte atunci când îl introduc în sistemul integrat de bibliotecă.

În schimb, pentru obiectele deja existente în sistem se impune o abordare diferită. Numărul mare de obiecte existente în baza de date, obiecte care au fost deja catalogate și digitizate nu permite supraîncărcarea catalogatorilor cu această operație nouă de clasificare și ierarhizare. În acest sens se va folosi un algoritm de clasificare supervizată care va realiza încadrarea fiecărui obiect existent într-una din cele trei clase, pe baza unui set de clasificări manual realizate aprioric de către catalogatori.

Însă, în vederea îmbunătățirii clasificării, se va realiza și o analiză semantică a obiectelor care urmează să fie clasificate utilizând o ontologie lexicală. Algoritmul de clasificare supervizată calculează o distanță între obiectul ce trebuie clasificat și fiecare din obiectele aprioric clasificate manual. Această distanță urmează să fie ajustată cu o măsură a similarității semantice între fiecare pereche de obiecte, evaluată prin intermediul

ontologiei lexicale. Ca exemplu, dacă în cadrul descrierii unui obiect este utilizat cuvântul inimă, analiza semantică va evidenția dacă sensul acestui cuvânt este cel medical sau dacă este folosit doar într-o expresie precum ”în inima teritoriului”. Determinarea sensului semantic corect are o importanță semnificativă în clasificarea ulterioară folosind algoritmul supervizat, îmbunătățind precizia cu care obiectul este încadrat în unul din cele trei niveluri. În felul acesta, toate obiectele din biblioteca virtuală vor avea o etichetă care le va încadra corespunzător, astfel încât să poată fi aplicate măsurile de securitate necesare protejării conținutului digital, diminuând munca catalogatorilor.

Fiecare obiect din biblioteca virtuală, care va fi clasificat automat, va avea două niveluri de evaluare, clasificarea finală realizându-se pe baza a două componente, după cum este prezentat în figura 2.9.

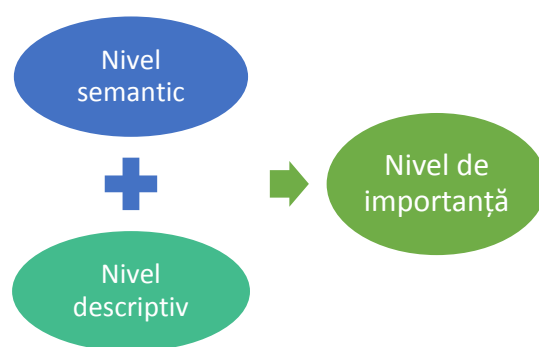


Figura 2.9 – Componentele procesului de clasificare

Nivelul semantic are la bază analiza realizată cu ajutorul ontologiei lexicale, aceasta furnizând o distanță semantică. Nivelul descriptiv, bazat pe caracteristicile cantitative ale unui obiect digital, este caracterizat de rezultatul algoritmului de clasificare supervizată, bazat pe o distanță euclidiană. Setul de caracteristici utilizat este format din date existente în descrierea obiectului digital precum: anul apariției documentului fizic pe care îl reprezintă, numărul de obiecte fizice existente, valoarea de achiziție, numărul de accesări ale obiectului digital, precum și alte caracteristici vitale unei bune clasificări.

Eficiența clasificării ține de cât de bine sunt alese aceste caracteristici, iar acestea sunt decise de către catalogatori.

Decizia de a încadra un obiect digital într-o anumită categorie este dată de rezultatul unei funcții care agregă valorile furnizate de analiza semantică și distanța calculată pe baza clasificării supervizate.

După ce toate obiectele digitale sunt clasificate, mecanismele de securitate care asigură integritatea, confidențialitatea și autenticitatea vor acționa pe baza etichetelor pe care obiectelor le vor avea atașate, ca răspuns la cererile lansate de utilizatorii bibliotecii virtuale.

Scopul unei biblioteci virtuale este să stocheze variantele digitale, copii fidele ale exemplarelor fizice, și să le prezeve în scopul constituirii unui patrimoniu cultural la nivel digital. Date fiind aceste restricții, variante modificate, funcție de nivelul de importanță, vor fi expuse utilizatorilor cu scopul de a le proteja de utilizările necorespunzătoare. Aceste versiuni sunt furnizate utilizatorilor la cerere, procesarea realizându-se în timp real asupra obiectelor solicitate.

3. SOLUȚIE PRIVIND CLASIFICAREA INFORMAȚIEI STOCATĂ ÎN VOLUME MARI DE DATE

3.1. Big Data în era datelor cu acces liber

Cloud computing, împreună cu tehnologiile mobile, reprezintă direcția de dezvoltare prioritară la nivelul aplicațiilor informatice curente, (Synergy Research Group, 2014; AWS, 2015; Google, 2015). Din acest considerent, cloud computing este un subiect important, larg dezbătut la nivelul instituțiilor academice și de cercetare în domeniul informatic, și cu precădere al științelor informaționale, cu exemple elocvente prezentate în (Pocatilu, Alecu și Vetrici, 2010; Morar, Muntean și Silaghi, 2011; Garrison, Kim și Wakefield, 2012).

În domeniul cloud computing se întâlnesc trei tipuri de modele reprezentative, arhitecturi prezentate în (Pocatilu, Alecu și Vetrici, 2010; Mell și Grace, 2011; Garrison, Kim și Wakefield, 2012):

- Infrastructure as a service (IaaS) – utilizează modele de virtualizare furnizând servicii pentru acces la resursele hardware: memorie, spațiu de stocare; putere de procesare;
- Platform as a Service (PaaS) – se orientează pe dezvoltarea de aplicații, distribuirea automată a unor soluții software; furnizorii de astfel de soluții pun la dispoziția utilizatorilor un mediu specific pentru procesele de dezvoltare software desfășurate;
- Software as a Service (SaaS) – permite utilizatorilor să folosească aplicațiile informatice direct în cloud, opțiune care nu necesită niciun fel de resursă din partea utilizatorilor ci doar o conexiune la rețeaua Internet.

O altă caracteristică a sistemelor de cloud computing o reprezintă varianta în care acestea sunt furnizate, așa cum se prezintă în (Yeluri and Castro-Leon, 2014):

- cloud public – într-o astfel de configurație serviciile și resursele sunt partajate între toți utilizatorii;
- cloud privat – oferă avantajul de a avea o limită de resurse prestabilită care este folosită doar de un anumit utilizator, chiar dacă acele resurse nu sunt întotdeauna utilizate la maxim;
- cloud hibrid – reprezintă o formă combinată de arhitectură publică și privată, funcție de contextul problemei de rezolvat.

Din perspectiva tehnică un sistem cloud operează cu câteva concepte cheie care trebuie menționate:

- sistemul cloud se bazează pe noduri care sunt reprezentate de mașini virtuale sau VMs;
- nodurile sunt grupate la nivel logic în componente denumite clustere;
- virtualizarea este gestionată de un supervizor care crează și monitorizează mașinile virtuale ce rulează pe unul sau mai multe servere fizice;
- clienți unui sistem cloud rulează aplicațiile pe dispozitive multiple precum tablete, calculatoare personale, telefoane mobile inteligente, servere.

Domenii emergente precum mobile cloud computing reprezintă o zonă importantă de dezvoltare a acestor tipuri de servicii, precum este prezentat în (Popa, Avornicului și Besfelean, 2012).

Sistemele integrate de biblioteci virtuale actuale își însușesc din plin prerogativele implementării în cloud datorită unui cumul de factori care a condus la o explozie a conținutului digital. Datorită acestui trend apar noi direcții ce necesită o abordare temeinică din perspectiva utilizatorului care se găsește bombardat de un volum foarte mare de date ce poate fi acum procesat în cadrul sistemelor cu implementare în cloud.

Dreptul de a fi informat și a avea acces la surse de date exacte, cu privire la bunurile și serviciile destinate utilizatorilor este unul din cele șapte nevoi legitime de bază ale consumatorului. Acest lucru este stipulat de la adoptarea Ghidului de Protecție a Consumatorului, de către Națiunile Unite, în 1985, precum și a revizuirilor ulterioare din 1999 și 2013.

Datorită globalizării și a creșterii capacității de procesare și stocare, o cantitate enormă de date este generată în fiecare zi și distribuită în rețeaua Internet, fiind accesibilă unui număr foarte mare de utilizatori, aproape instant. Acest lucru contribuie din plin la imposibilitatea de a ține evidența surselor de la care aceste date provin, pierzându-și originea și devenind date care nu pot fi certificate ca având o sursă de încredere.

Volumul mare de date care se adaugă constant la ceea ce există deja în Internet, presupune o nouă abordare în ceea ce privește analiza acestora cu scopul de a păstra intacte caracteristicile de bază ale datelor și informațiilor generate pe baza acestora. Mecanisme de etichetare la nivelul bunurilor puse la dispoziția utilizatorilor în rețeaua Internet au fost create pentru a contribui la o cât mai bună informare a consumatorului final, precum cele folosite în cadrul produselor ecologice (Dinu, Schileru și Atanase, 2012).

Din acest motiv, toate părțile implicate în mod direct sau indirect la dezvoltarea, livrarea, stocarea precum și în fazele de vânzare a produselor, necesită un cadru de

interoperabilitate foarte bine pus la punct între serviciile informatice guvernamentale și sectorul privat (Constantinescu, 2013), astfel încât utilizatorului final să i se ofere toate informațiile cu privire la condițiile de utilizare, garanție și returnare.

Mediul online este plin de informații contradictorii care bulversează utilizatorul atunci când acesta este în căutare de indicii pentru a se documenta. Un exemplu de domeniu în care există un nivel scăzut al adoptării tehnologiei informaționale îl reprezintă fermele agricole care, în multe țări, cu precădere cele emergente și subdezvoltate (Moga, Constantin și Antohi, 2012), nu reușesc să promoveze produsele pe care acestea le furnizează. Însă, în ciuda acestui lucru, aceste bunuri, din perspectiva economică, reprezintă bunuri de primă necesitate sau bunuri normale, iar lipsa implementării aspectelor tehnologice nu influențează foarte mult consumul sau strategiile de dezvoltare (COM, 2009). Dar, dacă vorbim despre informații și despre datele care stau la baza acestora, este foarte important să fie menționată sursa, actualitatea precum și părerea altor utilizatori care au referit aceste informații în scopul fundamentării unor decizii, (Filip, 2012).

Preocuparea asupra datelor cu acces liber, la nivel mondial, începe în Ianuarie 2004, când ministerele științelor și tehnologiilor din grupul celor mai dezvoltate țări care fac parte din OECD (Organization for Economic Cooperation and Development) s-au întâlnit la Paris și au discutat despre nevoia elaborării unui ghid internațional de acces liber la date pentru domeniile de cercetare. Obiectivele specifice descrise în acest ghid precum și principiile dezbătute, prezentate în (OECD, 2007), sunt:

- să informeze comunitățile publice de cercetare din cadrul țărilor membre OECD și nu numai, despre însemnătatea datelor cu acces liber și a partajării acestora în scopul unei cercetări cât mai eficiente;
- să promoveze bunele practici în privința accesului la date și partajarea la nivel public;
- să informeze publicul țintă despre potențialele costuri și beneficii ce decurg din utilizarea datelor cu acces liber;
- să sublinieze reglementările în materie de utilizare și partajare a datelor cu acces liber;
- să stabilească un cadru de principii operaționale pentru realizarea de acorduri de acces la date din domeniul cercetării la nivelul țărilor membre;
- să informeze statele membre asupra mecanismelor de îmbunătățire a cercetării la nivel internațional.

În 2007 a fost semnat un protocol la nivelul țărilor membre OECD care presupune că toate arhivele de date care au fost finanțate din bani publici să fie date spre accesul liber al publicului. Înainte de acest acord, în 2006, fundația OKF (Open Knowledge Foundation), (OKF, 2006), a propus o definiție la ceea ce numim în acest moment conținut deschis, și anume: „Anumite date sau un conținut digital este denumit ca fiind cu acces deschis dacă oricine are libertatea de a-l folosi, reutiliza și redistribui – drepturi supuse condiției de a-l atribui și/sau distribui în condiții identice”. În ceea ce privește datele, după cum este menționat în (Davies, Perini și Alonso, 2013), accesul liber presupune ca un set de date să fie accesibil, de obicei fiind publicat pe Internet, fără nici un cost și fără nici un fel de restricție tehnologică care să permită reutilizarea.

Uniunea Europeană a fost interesată mulți ani de problema datelor cu acces liber, privindu-le ca pe o resursă destinată creării de produse și servicii inovative și ca un mijloc de a aborda provocări sociale și de a promova transparența guvernamentală (EU, 2013). S-a observat, după cum este evidențiat într-un raport al Uniunii, că utilizarea corectă a datelor, incluzându-le și pe cele guvernamentale, poate îmbunătăți economia, servind ca premisă pentru o întreagă gamă de produse și servicii informaționale și îmbunătățind eficiența sectorului public și a unor sectoare din industrie (EU; G8UK, 2013). Deoarece Uniunea Europeană dorește promovarea datelor cu acces liber, în cadrul unui raport cu privire la acestea, s-au consimțit următoarele (EU, 2013):

- identificarea și publicarea a unor seturi de date de interes major disponibile la nivelul UE;
- publicarea datelor pe portalul Uniunii, EU Open Data Portal;
- promovarea aplicării principiilor descrise în raportul despre datele cu acces liber la nivelul tuturor statelor membre;
- realizarea de activități pentru a sprijini, mobiliza, consulta și implica promovarea datelor cu acces liber;
- împărtășirea experiențelor de lucru în domeniul datelor cu acces liber.

Încă de la început, cea mai mare întrebuințare a datelor cu acces liber a fost în sectorul guvernamental. Având la bază date cu acces liber provenite de la nivel guvernamental, organizațiile le pot reutiliza fără nicio restricție în scopuri inovative.

Vivek Kundra, din cadrul colegiului Harvard a arătat în cadrul lucrării (2011, pp. 16), faptul că Weather Channel, o importantă rețea americană de televiziune, și Garmin, o firmă care dezvoltă produse și tehnologii pentru navigație marină și aeriană (cu o valoare pe piață la nivelul a 7 miliarde de dolari la sfârșitul lunii Ianuarie, 2013) au fost construite exclusiv pe bază de date guvernamentale. În această situație, putem afirma că datele cu

acces liber prezintă o importanță majoră atât pentru sectorul public cât și pentru cel privat, tabelul 3.1.

Tabel 3.1 – Beneficiile economice ale datelor cu acces liber

Adaptare sursă: The Open Data Economy Unlocking Economic Value (Tinholt, 2013)

	Obținerea de venituri din mai multe domenii	Reducerea costurilor și îmbunătățirea eficienței	Generarea de locuri de muncă și dezvoltarea de noi abilități	Construirea unei societăți transparente
Sectorul public	Creșterea veniturilor fiscale prin intensificarea activității economice Obținerea de venituri prin vânzarea informații care produc valoare adăugată	Reducerea costurilor tranzacționale Creșterea eficienței serviciului prin intermediul datelor partajate	Crearea de locuri de muncă în vremuri economice dificile Încurajarea antreprenoriatului	Transparența este esențială pentru sectorul public pentru a îmbunătăți performanța acestuia
Sectorul privat	Crearea de noi oportunități de afaceri	Costuri reduse datorită utilizării gratuite a datelor guvernamentale brute Luare unor decizii mai bune bazate pe informații corecte	Atragerea de forță de muncă calificată	Pentru sectorul privat transparența este un material important pentru noi proiecte inovatoare de afaceri.

Disponibilitatea datelor cu acces liber a crescut semnificativ în ultimii ani (Bătăgan, 2014). Cele mai importante aspecte care determină utilizarea datelor cu acces liber sunt date de faptul că acestea generează venituri fără ca organizațiile să fie nevoite să suporte costuri pentru achiziția lor, îmbunătățesc și eficiența proceselor organizaționale iar volumul acestora devine din ce în ce mai mare (Janssen, Charalabidis și Zuiderwijk, 2012).

Datele cu acces liber sunt aproape indispensabile în domeniul dezvoltării serviciilor publice dar sunt foarte des utilizate și la nivelul altor domenii din societate, precum în cazul transportului public, serviciilor de sănătate și educație (Janssen, Charalabidis și Zuiderwijk, 2012). Rolul acestora este unul esențial în piață pentru că oferă utilizatorilor:

- o diversitatea a opiniilor – fiecare utilizator își poate exprima opinia;
- independența opiniilor – opiniile utilizatorilor sunt independente;

- descentralizare – fiecare utilizator poate ajunge la cu totul altă decizie pe baza acelorași date studiate;
- agregare în scopul măririi eficienței – soluțiile similare pot fi materializate într-o singură soluție eficientă.

Institutul pentru analiza datelor cu acces liber, ODI (Open Data Institute), evidențiază în cadrul unui raport din 2013, impactul generat la nivelul societății al acestui tip de date. Se menționează în (Davies et. al, 2013) faptul că, în orașele în care au fost executate proiecte pe bază de date cu acces liber, pot fi identificate modificări importante în domenii precum: transparența și eficiența guvernamentală, sustenabilitatea mediului, incluziunea socială, creșterea economică și activitățile antreprenoriale, figura 3.1.

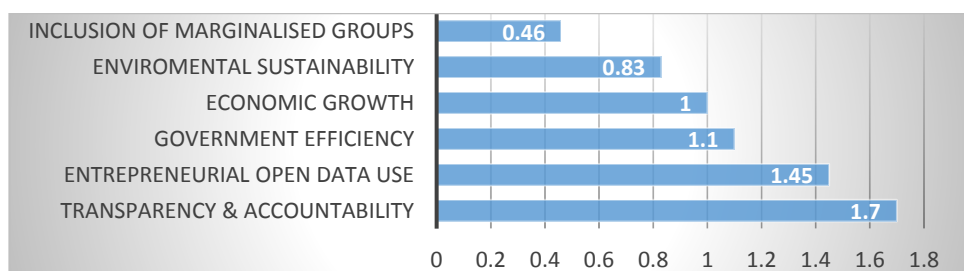


Figura 3.1 – Impactul datelor cu acces liber

Sursă: Open Data Barometer, 2013 Global Report– ODI, (Davies et. al, 2013)

Un număr foarte mare de țări, printre care și Statele Unite ale Americii, Franța, Regatul Unit al Mării Britanii, Danemarca, Spania și Finlanda au observat că datele cu acces liber au un efect cuantificabil și tangibil la nivelul volumului de afaceri desfășurat. Acest lucru se datorează faptului că organizațiile care utilizează astfel de date generează profituri foarte mari prin dezvoltarea de noi produse și servicii care au la bază o resursă fără costuri.

În fiecare ecosistem de date cu acces liber care este de succes sunt identificate trei componente majore, figura 3.2: partea guvernamentală (date cu acces liber produse sau colectate de către sectorul public), partea comercială și industria (date cu acces liber generate sau colectate de către sectorul privat) și clienții (date personale sau non-personale ale utilizatorilor publicate în mediul online).

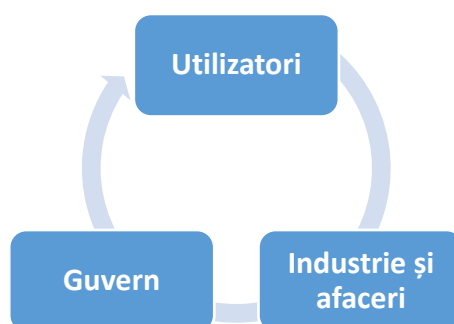


Figura 3.2 – Ciclul datelor cu acces liber

Cele mai populare domenii în care se regăsește utilizarea datelor cu acces liber sunt prezentate în cadrul unei analize efectuate de (Tinholt, 2013), figura 3.3.

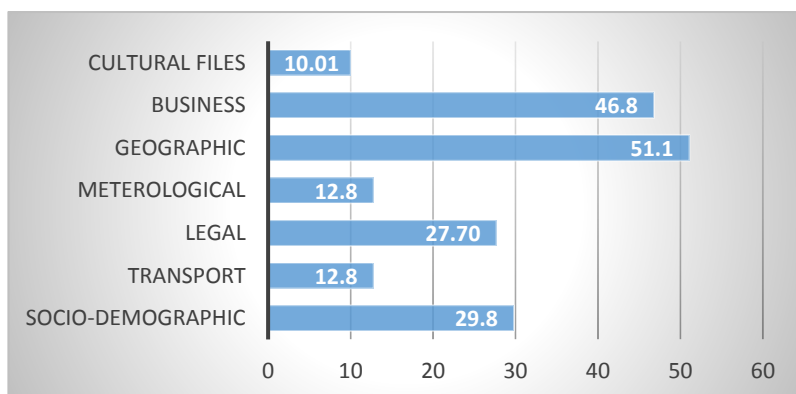


Figura 3.3 – Domenii importante ale datelor cu acces liber

Sursă: The Open Data Economy, 2013 Capgemini Consulting, (Tinholt, 2013)

Conform unei analize aprofundate prezentată în (Filip și Herrera-Viedma, 2014), este evidențiat faptul că prin utilizarea datelor cu acces liber națiunile vor realiza trecerea spre o economie dirijată de acest tip de date, la această tranziție contribuind și caracteristicile stabilite de către Comisia Europeană în materie de cum se folosesc datele deschise. Națiunile care utilizează masiv datele cu acces liber și implementează politici pe baza acestora sunt împărțite în trei categorii. Rezultate privind această clasificare, publicate în (Tinholt, 2013), indică faptul că doar câteva țări, 22% sunt denumite țări care dau trendul în acest domeniu. Aproape 78% din țări nu folosesc datele cu acces liber conform scopului pentru care au fost create. În această situație, țările se împart în țări începătoare și țări care urmează trendul stabilit de alții, tabelul 3.2.

Tabel 3.2 – Clasificare în funcție de utilizarea datelor cu acces liber

Adaptare sursă: The Open Data Economy, 2013 Capgemini Consulting, European Public Sector Information Platform, 2013 and The Global Competitiveness Report 2013–2014, World Economic Forum, 2013, (Tinholt, 2013)

Începători	Adepti	Lideri
Portugalia	Italia	USA
Irlanda	Danemarca	UK
Belgia	Noua Zeelandă	Canada
Arabia Saudită	Spania	Australia
Grecia	Finlanda	Franța
Turcia	Norvegia	Germania
Romania	Hong Kong	Suedia
Ucraina	Estonia	

Volumul mare al datelor cu acces liber precum și complexitatea acestora a determinat dezvoltarea unor instrumente de lucru specifice pentru a le gestiona și a le prelucra. De asemenea, în materie de securitate, au fost dezvoltate tehnici și metode specifice protecției datelor cu acces liber stocate în volume mari. Din momentul în care datele înregistrate sunt publicate în depozite mari de date cu acces liber, este doar o chestiune de câteva secunde și de cultură și responsabilitate informațională până când acestea sunt accesate și răspândite la nivelul întregii rețele Internet. Acest lucru facilitează procesul decizional atât la nivelul utilizatorului normal cât și la nivel organizațional sau guvernamental. Volumele mari de date cu acces liber prezintă următoarele caracteristici:

- conștientizează utilizatorii de importanța de a contribui la dezvoltarea acestora;
- necesită deprinderea unor abilități specifice pentru extragerea informațiilor din cadrul acestor depozite mari de date;
- necesită utilizarea unor instrumente de urmărire a sursei de proveniență pentru identificarea autenticității și originalității;
- creșterea productivității pe baza utilizării datelor extrase din cadrul sistemelor informaționale;
- necesită implementarea de controale de securitate pentru protecția datelor cu acces liber împotriva utilizării necorespunzătoare;
- sunt utilizate instrumente speciale pentru stocarea și arhivarea datelor cu acces liber.

Instrumentele dezvoltate din domeniul depozitelor mari de date au rolul de a extrage informații și de a crea cunoștințe care să ajute la fundamentarea deciziilor. De-a lungul vremii, în cadrul sistemelor informaționale (Vickery, 2009), s-au dezvoltat programe specializate care interacționează cu alte tipuri de sisteme precum sisteme suport de decizie, sisteme de inteligență artificială, sisteme de data mining dar și cu sisteme de gestiune a bazelor de date spațiale.

Cercetătorii au abordat acest domeniu de foarte mult timp iar acum, instrumente valoroase dezvoltate de-a lungul timpului oferă importante indicii cu privire la ce înseamnă cu adevărat conceptul de volum mare de date și la ce pot fi utilizate:

- procesul de optimizare – optimizarea privită ca un proces de management al resurselor, (Stefănoiu, et. al., 2014), este extrem de importantă pentru evoluție, pentru avansul tehnologic; pentru a finaliza un proces de optimizare sunt necesare cunoștințe cu privire la două stări: starea inițială și starea finală, cea la care se ajunge după ce criteriul de optimizare a fost aplicat; dacă starea finală este superioară calitativ sau cantitativ stării inițiale atunci se poate afirma că a fost

realizat un proces de optimizare; pentru acest lucru sunt necesare informații detaliate care să reflecte cât mai precis cele două stări; aceste informații sunt extrase cu ajutorul analizei de tip data mining;

- fundamentarea deciziilor pe baza a cât mai multe informații cu privire la aceasta – conceptul de decizie bazată pe informații reprezintă un proces condus de prelucrarea automată a datelor existente în depozitele mari de date; dacă deciziile au la bază informații obținute ca rezultat al unor procese de analiză a datelor de mari dimensiuni în timp real atunci acestea pot influența pozitiv procesul decizional;
- înțelegerea comportamentului – comportamentul a fost și va fi mereu o variabilă importantă în ecuațiile care simulează procesele economice; nu este suficient să se cunoască intrările unui sistem pentru a se determina rezultatul acestuia; este necesară și o cunoaștere și înțelegere cât mai profundă a conceptului de cutie neagră, pentru a extrapola cât mai bine rezultatele; în acest scop analiza datelor de mari dimensiuni prin procesarea a cât mai multe seturi de date, perechi intrare/ieșire, poate determina anumite tipare ale cutiei negre din fiecare sistem;
- antrenarea cât mai eficientă a resurselor în cadrul proceselor economice – pentru a realiza acest lucru este absolut necesar să fie cunoscute cât mai multe detalii cu privire la cum sunt folosite aceste resurse și care este impactul generat de acestea în rezultatul final; analize predictive pot fi folosite pentru a calcula soluții alternative bazate pe diferite configurații de factori de intrare.

Volumele mari de date sunt acum aprovizionate de către o multitudine de senzori ai domeniului cunoscut sub denumirea de Internetul Lucrurilor sau IoT, rețea de senzori interconectată în timp real care furnizează constant date ce intră în analiza instrumentelor de tip data mining.

Așa cum se prezintă în (Tene și Polonetsky, 2012), economia are de câștigat între 5 și 6 procente datorită deciziilor bazate pe analizele de tip data mining. O analiză de acest tip are avantajul faptului că nimic nu se poate pierde dacă este gestionat corespunzător. În foarte multe domenii de activitate sunt necesare procese secundare care să înregistreze, urmărească, verifice starea curentă proceselor primare, toate acestea generând date suplimentare care stocate, pot fi folosite în coroborare cu procesul principal pentru analize complexe cu scopul de a crește calitatea produsului final.

Pentru a beneficia de acest volum imens de date, sistemele actuale trebuie să facă față unei alte probleme cu privire la capabilitățile de natură hardware. Pentru a oferi un

răspuns la întrebarea referitoare la care este necesarul de resurse de procesare pentru o analiză de tip big data, trebuie mai întâi să răspundem următoarelor aspecte din domeniu:

- cât de repede trebuie furnizate rezultatele analizei big data – pentru a avea rezultate rapide un sistem necesită o putere de procesare mare;
- cât de în amănunțită se dorește analiza big data – pe măsură ce analiză avansează în profunzimea datelor, și din ce în ce mai multe date sunt procesate, rezultatele oferite sunt mai relevante.

Această balanță, figura 3.4, între obținerea rezultatelor unei analize big data într-un timp rapid și nivelul ridicat de relevanță dat de o analiză în profunzime a volumului mare de date este foarte greu de atins.

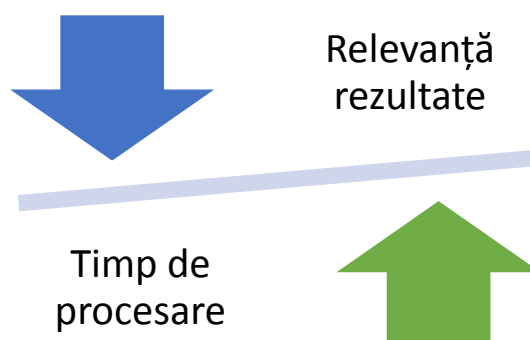


Figura 3.4 – Compromisul între timp și calitatea rezultatelor

Pentru a se adapta la nevoile de procesare constant mărite din domeniul depozitelor mari de date, componentele hardware trebuie dimensionate corespunzător după cum se prezintă în (Dilpreet și Chandan, 2014). În acest sens există două abordări care conduc la sporirea necesarului de procesare, după cum urmează:

- dimensionare pe orizontală – acest lucru presupune distribuția nevoilor de procesare pe orizontală, diminuând gradul de încărcare la nivel de mașină fizică; rețelele de tip peer-to-peer precum și arhitectura Apache Hadoop (Apache Hadoop, 2015) sau paradigma Spark reprezintă câteva exemple de astfel de implementări;
- dimensionare pe verticală – pentru acest tip de dimensionare, mașinile fizice necesită o reconfigurare ce se bazează pe adăugarea de noi unități de procesare, CPU și memorie suplimentară; un exemplu de astfel de implementare este un cluster de unități multiple de procesare de tip CPU (Central Processing Unit) sau GPU (Graphical Processing Unit).

Analiza big data ridică câteva probleme legate de redundanța datelor și uneori de irelevanța acestora. Aceste aspecte pot influența drastic rezultatul final deoarece:

- în cazul în care datele redundante nu sunt eliminate prin intermediul algoritmilor de reducere a dimensiunii, acele aspecte multiple vor influența rezultatul final într-

un mod negativ deoarece ponderea lor în cadrul întregului set de date analizat este mult mai mare;

- direcțiile de analiza care au prea puține date sau datele dintr-un anumit domeniu sunt irelevante conduc și ele la denaturarea rezultatului final.

3.2. Taxonomii de securitate în sistemele integrate de bibliotecă pe bază de ontologii

În contextul tehnologiei informației, ontologia este definită ca fiind un set de concepte împreună cu legăturile dintre ele destinată modelării unui domeniu al cunoașterii (Jrme și Pavel, 2010). De cele mai multe ori, ontologiile sunt echivalente cu ierarhizarea claselor și a definițiilor lor în taxonomii, dar ontologiile nu trebuie să fie limitate la aceste forme. Definițiile din cadrul ontologiilor trebuie să adauge cunoștințe, nu doar o terminologie (Enderton, 2001), figura 3.5.

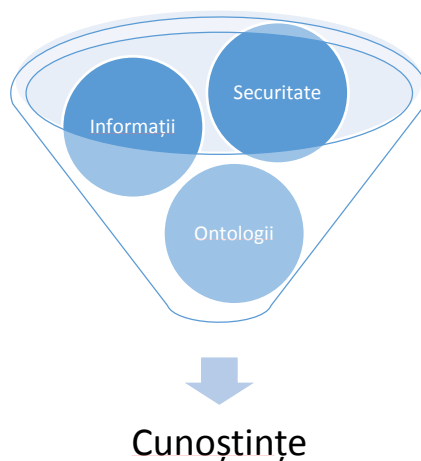


Figura 3.5 – Utilizarea ontologiilor în sistemele ILS

Ontologia, în cadrul tehnologiei informației, constă în reprezentarea formală a cunoștințelor printr-un set de concepte și legături, fiind utilizată într-un domeniu limitat în aceeași măsură cu aplicarea acesteia la un nivel mai general. Obiectivul unei ontologii constă în reprezentarea unei concepte, partajabile și reutilizabile, în care sunt ignorate detaliile specifice aplicațiilor. Componentele care fac parte din structura unei ontologii sunt:

- indivizi, reprezentați prin instanțe sau obiecte;
- clase, colecții, seturi sau tipuri de obiecte;
- attribute, proprietăți, caracteristici sau parametri ai unui individ sau clase;
- relații, felul în care clasele și indivizii sunt interconectați;

- restricții, descrieri formale a ceea ce trebuie să fie adevărat în scopul acceptării anumitor aserțiuni;
- reguli, declarații ce descriu inferențe logice ce pot fi deduse dintr-o aserțiune;
- axiome, aserțiuni logice cuprinzând teoria globală unanim acceptată pentru ontologia ce descrie un domeniu de aplicare;
- evenimente, schimbarea atributelor sau a relațiilor.

Legăturile dintre concepte au o reprezentare arborescentă la nivelul acestora calculându-se distanțe prin care se determină intensitatea acestei conexiuni. Arborele este constituit din cuvinte denumite și *synsets*, care reprezintă nodurile structurii precum și relații semantice descrise de arcele arborelui.

Ontologia este utilizată după ce o clasificare supervizată (Chen et. al, 2007; Kolbe, Zhu și Pramanik, 2010) a unui set inițial de documente digitale a fost aplicată cu scopul de a determina care sunt domeniile în care se regăsesc conceptele utilizate. Pe baza clasificării inițiale dată de algoritmi specializați, (Chen et. al, 2007), se va procesa întregul set de documente digitale utilizând următorii pași prezentați în figura 3.6.

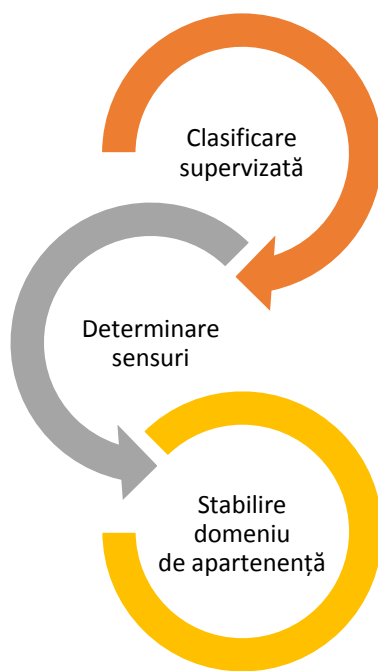


Figura 3.6 – Clasificarea obiectelor digitale

Ontologiile au rolul de a determina sensul corect al unui concept și de a stabili în care din domeniile identificate anterior prin procesul de clasificare supervizată se încadrează. Scopul determinării domeniilor folosind ontologiile la nivelul întregului set de documente digitale este acela de a calcula un set de indicatori care să pună în evidență ce domenii sunt importante și care au un grad ridicat de vulnerabilitate în ceea ce privește drepturile digitale asupra acestora.

Ontologiile sunt instrumente complexe ce servesc unor obiective specifice care includ o abordare lexicală. Ontologiile, în această abordare a securității sistemelor integrate de bibliotecă, sunt utilizate cu următoarele scopuri:

- identificarea sensului corect al unui concept într-un anumit context, analiză semantică realizată pe baza conceptelor stocate și a relațiilor dintre ele;
- determinarea unor clase de apartenență în care se pot încadra obiectele digitale din cadrul bibliotecilor virtuale.

Folosirea ontologiilor conduce la optimizarea unui aspect al securității care este foarte sensibil la percepția externă a posesorului informației extrasă din cadrul unui sistem ILS. Acest lucru implică construirea unui instrument care poate urmări materialele digitale din cadrul surselor lor la orice locație în care au fost folosite. Astfel, celelalte persoane care se folosesc de informația extrasă sunt obligați să introducă în materialele lor referințe către posesorul informațiilor extrase în cazul în care doresc să le introducă în cercetările lor personale.

Metodologia ce descrie implementarea ontologiilor cu scopul de a îmbunătăți un aspect vital al securității unui sistem ILS conține următoarele elemente:

- etapele corespunzătoare pentru care ontologia este folosită pentru a optimiza securitatea;
- sistemul informațiilor stocate în biblioteci alături de mulțimea vehiculată de date;
- controale de securitate pentru managementul drepturilor digitale.

Securitatea este un aspect foarte important în astfel de tipuri de sisteme deoarece orice eroare care apare între datele stocate în biblioteca virtuală și realitatea actuală conduce la o propagare care poate altera procese importante care au la bază acele date.

Astfel, următoarele caracteristici ale securității trebuie menținute în contextul sistemelor analizate:

- integritatea, datele nu pot fi modificate de către o entitate externă care nu are permisiunea de a realiza această operație;
- disponibilitatea, caracteristica ce permite utilizatorilor să dispună de informații în orice moment de timp sau spațiu, în cazul în care sistemul permite acest lucru;
- confidențialitatea, datele care nu sunt accesibile utilizatorilor care nu au drepturi de a le vedea sunt protejate prin parole sau sunt stocate într-o formă criptată;
- nerepudiare, abilitatea de a crea o relație indisolubilă între posesor și datele care sunt partajate;
- posesie, caracteristica ce permite unui deținător să aibă control asupra datelor sale în orice moment de timp;

- utilitate, asigură faptul că datele pot fi folosite doar dacă deținătorul lor poate să le acceseze, indiferent de alte restricții care sunt aplicate asupra lui.

Accesul prin intermediul managementului drepturilor digitale servește la protecția materialelor digitale atunci când sunt executate operații de partajare sau copiere asupra conținutului. Un set de reguli de acces sunt construite cu scopul de a determina cu ușurință între adevăratul deținător și persoana ce folosește datele respective. Tehnicile DRM (Digital Rights Management) sunt aplicate asupra materialelor digitale, având următoarele scopuri:

- protejarea conținutului digital în cazul accesului sau procesării neautorizate;
- partajarea în siguranță a materialelor digitale la nivelul întregii biblioteci virtuale.

Astfel, fiecărui nivel de importanță, determinat pe baza clasificării supervizate și analizei semantice a obiectului, îi corespunde un nivel de securitate care va influența felul în care vor fi accesate resursele digitale, precum este prezentat în figura 3.7.



Figura 3.7 – Nivelurile de securitate la nivelul conținutului digital

Mecanismele DRM vor acționa pe baza unui set de drepturi digitale care vor permite utilizatorilor acces asupra conținutului în funcție de nivelul de vulnerabilitate stabilit pentru fiecare domeniu identificat pe baza folosirii ontologiilor lexicale. Următoarele drepturi digitale sunt menite să restricționeze accesul utilizatorilor în funcție de următoarele niveluri:

- drepturi de procesare, se referă la dimensiunea culturală a materialelor digitale care au fost extrase din surse externe precum și la posibilitatea de a oferi sensuri noi dar cu obligația de a cita sursa originală;
- drepturi de folosire, drepturile ce permit utilizatorilor să includă materialul în lucrările lor în forma sa originală fără a altera în orice mod conținutul și sunt obligați să citeze forma originală;
- drepturi de consultare, se referă la drepturile de a accesa un material doar în scopuri de consultare.

Drepturile de procesare sunt drepturi aferente materialelor cu acces nerestricționat la resurse, cu nivel scăzut de importanță. Sunt drepturi acordate asupra materialelor cu o valoare de referință scăzută în domeniul de specialitatea pe care îl vizează.

Drepturile de utilizare se aplică acelor materiale asupra cărora utilizatorii au dreptul de utilizare însă sunt obligați să refere sursa de proveniență. Materialele sunt marcate prin coduri de tip răspuns rapid, care includ elementele bibliografice ale obiectului digital.

Drepturile de consultare sunt specifice unui nivel ridicat de importanță și se aplică acelor materiale asupra cărora nu se pot aduce modificări sub nicio formă și care sunt protejate la vizualizare pe bază unui înveliș digital, folosind instrumente software dedicate pentru vizualizare.

Împărțirea pe niveluri de importanță a fost realizată cu scopul de a filtra conținutul digital care urmează a fi procesat ca urmare a unei cereri lansate de un utilizator și a nivelurilor de securitate stabilite.

Toate aceste niveluri de securitate, definite prin diferite drepturi pe care utilizatorii le au asupra conținutului digital original aflat în baza de date, se traduc prin creare de noi forme procesate a obiectelor având diferite mecanisme de securitate atașate. Acestea vor fi livrate în timp real de fiecare dată când un utilizator accesează forma originală. Din acest motiv, documentelor care au o importanță scăzută sau redusă nu li se vor aplica elemente de securitate. În schimb, înregistrările clasificate în secțiunea de importanță medie sau crescută vor fi procesate diferit.

3.3. Analiza și preprocesarea datelor

Preprocesarea datelor reprezintă componenta de filtrare, omogenizare și reducere a dimensiunii din cadrul procesului general de prelucrare a datelor, (Borda, 2011). Pornind de la conceptul de date primare, acestea necesită diferite prelucrări asupra setului inițial de documente (bază de cunoștințe) în scopul determinării structurii sale omogene. Astfel, etapele parcurse în acest proces de transformare sunt:

1. omogenizarea structurii documentelor digitale cu scopul de a evidenția caracteristicile importante;
2. separarea documentelor în cuvinte sau fraze pentru o prelucrare la nivel atomic;
3. eliminarea redundanței prin eliminarea cuvintelor de legătură, etapă denumită Tokenizare;
4. reducerea dimensiunii reprezentării prin extragerea rădăcinii cuvintelor, Stemming;
5. reprezentarea conceptelor folosind reprezentarea vectorială sau arborescentă;

6. reducerea dimensiunii documentelor;
7. analize preliminare asupra datelor în scopul creării de structuri și modele analitice.

În (Burileanu et. al, 1999) este propus un algoritm de preprocesare format din stemming de tip corpus-based stemming. Schema generală a etapei de procesare lingvistică este prezentată în figura 3.8.

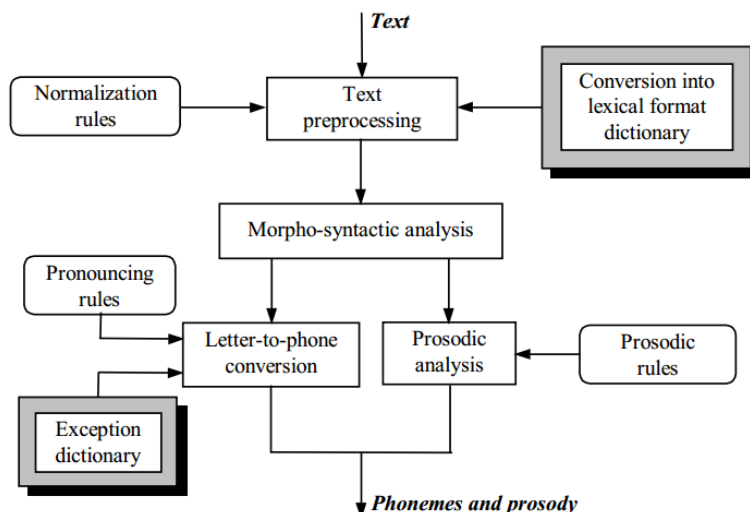


Figura 3.8 – Schema generală a etapei de procesare lingvistică

Sursă: (Burileanu et. al, 1999)

Preprocesarea ca etapă apriorică aplicării algoritmilor de clasificare supervizată și a ontologiilor lexicale are rolul de a scoate în evidență caracteristicile importante și de a elimina redundanța dată de elementele care nu joacă un rol esențial în clasificare. Datorită acestei etape rezultatele obținute ulterior reflectă mult mai bine realitatea și asigură un nivel înalt de corectitudine a prelucrărilor.

Astfel, pentru a determina metodele specifice aplicate în procesul de preprocesare a elementelor definitorii obiectelor analizate, trebuie realizată o separare a tipurilor de date analizate. Din punct de vedere a etapelor de preprocesare, se definesc date:

- numerice, în care informația este stocată la nivel cantitativ de dimensiunea naturală sau reală, precum: număr de pagini, anul apariției, valoarea de achiziție al obiectului, număr accesări online,
- text, în care informația este stocată la nivelul unei descrieri text sau al unor variabile care au datele stocate ca text, situație în care analiza este realizată la un alt nivel față de obiectele numerice, o analiză lexicală folosind ontologii lexicale precum ontologia WordNet, precum: subiectul descris, lista de autori, subiectele secundare tratate în document;

- text neasociat cu concepte WordNet, situație în care nu se pot procesa datele respective întrucât nu pot fi transpuse în informație stocată generatoare de cunoștințe suplimentare.

De asemenea, analizând caracteristicile de tip text care descriu obiectele procesate, se impun două direcții de cercetare:

- analiză lexicală, în cazul caracteristicilor text ce conțin informații stocate în cuvinte care pot fi transpuse, folosind algoritmi specializați precum cei de stemming și lematizare, în concepte disponibile în ontologii lexicale precum ontologia WordNet; asupra unor astfel de caracteristici se poate aplica algoritmul de calcul a distanței lexicale, distanță care se transpune în gradul de similitudine dintre oricare două concepte identificate, lând valori numerice aflate în intervalul [0;1];
- analiză comparativă, în cazul caracteristicilor text ce conțin concepte nestocate în ontologii; în astfel de situații, se aplică distanța folosind formula:

$$d_c(t_1, t_2) = \begin{cases} 0, & \text{dacă } t_1 = t_2 \\ 1, & \text{dacă } t_1 \neq t_2 \end{cases}$$

Există două tipuri de ontologii, descrise în (Trausan-Matu, 2004), în funcție de domeniul în care acestea sunt folosite:

- ontologii destinate sistemelor bazate pe cunoștințe, sunt caracterizate printr-un număr relativ redus de concepte, dar legate printr-un număr mare și variat de relații; conceptele sunt grupate în scheme conceptuale complexe sau scenarii, iar pentru fiecare concept pot exista una sau mai multe particularizări;
- ontologii lexicale, care cuprind un număr foarte mare de concepte legate printr-un număr redus de relații; în cazul ontologiei WordNet, conceptele sunt reprezentate prin mulțimi de cuvinte sinonime; aceste ontologii sunt folosite în sistemele de prelucrare a limbajului uman.

WordNet este o bază de cunoștințe care conține informații despre lexicul limbii engleze. Concepută original ca un model de scară largă a organizării semantice, a fost în scurt timp acceptată în domeniul de procesare a limbajului natural NLP (Natural Language Processing). Ontologia WordNet a devenit baza de date aleasă în NLP, Kilgariff spunând că nefolosirea acestei resurse necesită explicații și justificări, (Boyd, et. al., 2005). Popularitatea ontologiei este dată de accesul liber și de aria vastă de cuprindere (Lin și Sandkuhl, 2008).

Ontologia WordNet este creată și menținută de Universitatea Princeton, baza de date putând fi descărcată de la adresa (WordNet, 2015). Aceasta conține substantive, verbe, adjective și adverbe. Sensurile lexicale au relații între ele. Cuvintele cu sensuri similare

sunt structurate în seturi numite *synsets*. Ultima versiune WordNet 3.0 conține aproximativ 155.000 de cuvinte organizate în 117.000 *synsets*, (Hessami, Mahmoudi și Jadidinejad, 2011). Un *synset* este format din cuvintele asemănătoare ca sens, alături de o descriere, precum și exemple de folosire a acestor cuvinte.

Ontologia WordNet este integrată în cadrul reprezentării și procesării documentelor ca o componentă care răspunde principalelor probleme apărute în procesul de prelucrare a documentelor text, (Elberrichi, Rahmoun și Bentaalah, 2008), legate de inexistența unor relații explicite între cuvintele din text și de dimensiunea mare a spațiului de reprezentare.

În (Passos și Wainer, 2009), structura WordNet este văzută ca fiind intuitivă, formată din cuvinte care au mai multe sensuri, fiecare sens formând un *synset*, structura atomică a ontologiei WordNet, și relații dintre cuvinte precum sinonime, antonime, reprezentate prin legături într-un graf.

Reprezentarea arborescentă, (Blanchard, et. al., 2005; Boyd-Graber, et. al., 2005), a legăturilor dintre conceptele ontologiei WordNet are la bază crearea unui arbore format din cuvinte/*synsets* reprezentate prin noduri și relații semantice dintre conceptele WordNet reprezentate de arcele grafului. Reprezentarea top-bottom este formată dintr-o rădăcină, punctul de la care se bifurcă toate legăturile existente dintre concepte, rădăcină care poartă numele de entity, (Gonzalez, Rigau și Castillo, 2012). Astfel, pornind de la nivelul rădăcină și păstrând legăturile existente în ontologia WordNet între concepte, se ajunge la nivelul de frunză determinat de elementul analizat. Figura 3.9 conține o reprezentare arborescentă a unor concepte care derivă din rădăcina *artefact*, cu subdomeniu *motor vehicle*.

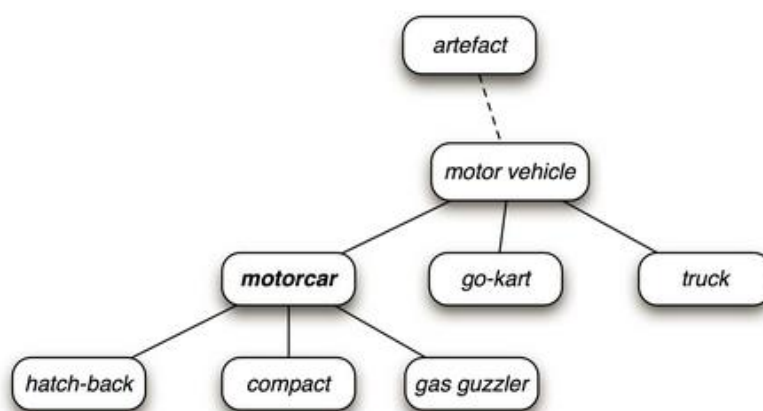


Figura 3.9 – Reprezentarea arborescentă din ontologia WordNet spre nodul compact

Sursa: <http://www.nltk.org/book/ch02.html>

Obiectivul principal al preprocesării și analizei datelor stocate în fragmente text este de a extrage informații generatoare de cunoștințe care sunt stocate în cadrul acestor caracteristici text analizate. Această reprezentare arborescentă a conceptelor din care sunt

compuse fragmentele text are la bază nodul care reprezintă un synset din cadrul ontologiei WordNet. Fiecare synset este format din subcomponente definite în cadrul figurii 3.10.

Property	Domain	Range	Prolog clause
containsWordSense word	Synset WordSense	WordSense Word	s s
lexicalForm	Word	xsd:string	s
synsetId	Synset	xsd:string	s
tagCount	Synset	xsd:integer	s
gloss	Synset	xsd:string	g
frame	VerbWordSense	xsd:string	fr
hyponymOf	Synset	Synset	hyp
entails	Synset	Synset	ent
similarTo	Synset	Synset	sim
memberMeronymOf	Synset	Synset	mm
substanceMeronymOf	Synset	Synset	ms
partMeronymOf	Synset	Synset	mp
classifiedByTopic	Synset	Synset	cls
classifiedByUsage	Synset	Synset	cls
classifiedByRegion	Synset	Synset	cls
causes	Synset	Synset	cs
sameVerbGroupAs	Synset	Synset	vgp
attribute	Synset	Synset	at
adjectivePertainsTo	Synset	Synset	per
adverbPertainsTo	Synset	Synset	per
derivationallyRelated	WordSense	WordSense	der
antonymOf	WordSense	WordSense	ant
seeAlso	WordSense	WordSense	sa
participleOf	WordSense	WordSense	ppl
classifiedBy meronymOf	Synset Synset	Synset Synset	cls mm,ms,mp

Figura 3.10 – Elementele component ale unui concept din cadrul ontologiei WordNet

Sursa: <https://wordnet.princeton.edu/wordnet/documentation/>

Fiecare concept din cadrul ontologiei lexicale poate conține unul sau mai multe sensuri, în funcție de contextul semantic în care acesta apare. Pentru a alege ce sens se încadrează pentru fiecare concept, figura 3.11 descrie componentele stocate în ontologie în cazul conceptului country. Pentru fiecare sens în parte, se atașează o cheie unică de înregistrare, numărul de apariții al aceluși sens în cadrul unei baze de antrenare disponibilă în WordNet cu care s-au antrenat și testat toate valorile inserate, partea de propoziție, o scurtă definiție și exemple de utilizare a respectivului sens în propoziții.

WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (frequency) {offset} <lexical filename > [lexical file number]
(gloss) "an example sentence"
Display options for word: word#sense number (sense key)

Noun

- (68){08185877} <noun.group>[14] [S: \(n\) state#4 \(state%1:14:00::\), nation#1 \(nation%1:14:00::\), country#1 \(country%1:14:00::\), land#9 \(land%1:14:00::\), commonwealth#2 \(commonwealth%1:14:00::\), res publica#1 \(res_publica%1:14:00::\), body politic#1 \(body_politic%1:14:00::\)](#) (a politically organized body of people under a single government) *"the state has elected a new president"; "African nations"; "students who had come to the nation's capitol"; "the country's largest manufacturer"; "an industrialized land"*
- (29){08562388} <noun.location>[15] [S: \(n\) country#2 \(country%1:15:00::\), state#7 \(state%1:15:00::\), land#5 \(land%1:15:02::\)](#) (the territory occupied by a nation) *"he returned to the land of his birth"; "he visited several European countries"*
- (12){08183451} <noun.group>[14] [S: \(n\) nation#2 \(nation%1:14:01::\), land#8 \(land%1:14:01::\), country#3 \(country%1:14:01::\)](#) (the people who live in a nation or country) *"a statement that sums up the nation's mood"; "the news was announced to the nation"; "the whole country worshipped him"*
- (11){08662297} <noun.location>[15] [S: \(n\) country#4 \(country%1:15:01::\), rural area#1 \(rural_area%1:15:00::\)](#) (an area outside of cities and towns) *"his poetry celebrated the slower pace of life in the country"*
- (3){08514304} <noun.location>[15] [S: \(n\) area#1 \(area%1:15:01::\), country#5 \(country%1:15:02::\)](#) (a particular geographical region of indefinite boundary (usually serving some special purpose or distinguished by its people or culture or geography)) *"it was a mountainous area"; "Bible country"*

Figura 3.11 – Sensuri concept „country” în ontologia WordNet

Sursa:<http://wordnetweb.princeton.edu/perl/webwn?c=2&sub=Change&o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&i=-1&h=00000&s=country>

Având context creat pentru transformarea fiecărui cuvânt regăsit în cadrul unui fragment text în concept WordNet, preprocesare în contextul caracteristicilor text implică primirea ca date de intrare fragmentul text și extragerea, pe baza de algoritmi specializați, a conceptelor alături de sensurile lor contextuale, lista de concepte devenind date de ieșire din cadrul etapei premergătoare analizei efective a datelor.

Etape de eliminare folosind semnele de punctuație este aferentă segmentării fragmentelor text în cuvinte, urmând procesul de eliminare a literelor mari prin transformarea lor în litere mici, case folding. La nivel conceptual, se aplică algoritmi de stemming și lemmatization pentru extragerea rădăcinii cuvintelor, rezultând o reducere a cardinalității numărului de cuvinte regăsite în textul analizat. De asemenea, folosind o listă de cuvinte aprioric cunoscută sau pe baza apariției lor în documente, se elimină cuvintele de legătură.

În morfologia lingvistică și în cadrul teoriei regăsirii informației, stemming reprezintă procesul de reducere a multiplelor forme și derivate ale unui cuvânt la nivelul

bazei sale dată de rădăcină. Algoritmul Porter, descris în (Willett, 2006), este format din cinci pași executați consecutiv, fiecare pas fiind format dintr-un set de reguli de forma <condiție> <sufix> -> <sufix_nou>. Marea majoritate a algoritmilor de reducere a unui cuvânt la rădăcina sa a fost dezvoltată pentru vocabularul limbii engleze, însă cercetări precum (Burileanu, 1999) au abordat problema preprocesării fragmentelor text și pentru particularitățile limbii române.

Lematizarea în teoria lingvistică este procesul de grupare a diferitelor forme ale unui cuvânt astfel încât să fie analizate ca un singur concept. Obiectivul general al proceselor de tokenizare și lematizare este de a reduce diferitele forme ale cuvintelor la o bază comună redusă în dimensiune. Diferența majoră dintre cele două procese este dată de procesul euristic brut de eliminare a prefixelor unui cuvânt în cazul tokenizării față de lematizare, în care se face o analiză morfologică și a vocabularului.

Figura 3.12 conține un exemplu de rulare a algoritmului de preprocesare folosind principiul de lematizare prin integrarea ontologiei lexicale WordNet ca bază generatoare de grupuri de cuvinte comune (caz A), precum și algoritmul Porter (caz B).

Figura 3.12 – Exemplu rulare algoritm Porter și WordNet Lemmatizer

Sursa: <http://text-processing.com/demo/stem/>

Folosind o analiză preliminară unei testări automate a rezultatelor obținute, se observă că algoritmul WordNet Lemmatizer generează rezultate mai bune față de algoritmul Porter, în sensul generării de concepte similare cu cele din ontologia WordNet. Pentru a compara cele două metode propuse, se alege un set de test format din fragmente de text și se rulează cele două metode urmând, ca la final, să se compare procentul de cuvinte care se regăsesc în concepte WordNet față de numărul total de cuvinte existente în text.

3.4. Clasificarea și ierarhizarea conținutului digital

În scopul încadrării obiectelor digitale în clase de apartenență specifice, se utilizează un algoritm de clasificare supervizată denumit kNN (k Nearest Neighbour) care asigură procesarea automată a tuturor obiectelor digitale existente în biblioteca virtuală. Având ca obiectiv principal clasificarea obiectelor digitale stocate în cadrul unei biblioteci virtuale în clase de securitate, securitate coborâtă, medie și înaltă, algoritmul care rezolvă problematica enunțată este dat de combinația dintre kNN, algoritm de clasificare supervizată împreună cu analiza lexicală folosind ontologia WordNet în cazul acelor caracteristici de tip text exemplificate în capitolul precedent. Astfel, analiza se împarte în două componente, o clasificare supervizată folosind doar caracteristicile numerice ale obiectelor digitale și o clasificare supervizată a obiectelor determinate de caracteristicile stocate în format text prin integrarea distanței semantice a conceptelor disponibile în ontologia WordNet. În urma analizei comparate a rezultatelor celor două clasificări folosind o bază inițială de testare, formată din obiecte clasificate manual de către specialiști, se agregă cele două metode în scopul îmbunătățirii clasificatorului propus.

Elementele componente în clasificarea supervizată folosind algoritmul kNN sunt descrise în tabelul 3.3 alături de notațiile aferente.

Tabel 3.3 – Variabilele folosite în cadrul clasificării supervizate kNN asupra caracteristicilor numerice

NOTAȚIE	DESCRIERE
n	Cardinalitatea obiectelor clasificate aprioric prin intermediul specialiștilor
O	Mulțimea obiectelor reprezentate de înregistrările din cadrul bazei de date aferente bibliotecii virtuale
o_i	Obiectul de pe poziția i din cadrul mulțimii totale de obiecte O
o_{ij}	Valoarea caracteristicii j a obiectului aflat pe poziția i din cadrul mulțimii O
c_i	Clasa asignată pentru obiectul i ; asignarea este făcută manual în cazul celor n obiecte care sunt clasificate aprioric și automat, pe baza algoritmului kNN în cazul obiectelor neclasificate inițial

NOTAȚIE	DESCRIERE
α, β, γ	Cele trei clase de asignare a obiectelor, clase ce determină nivelul de importanță acordat fiecărui obiect în parte din cadrul bazei de date
k	Cardinalitatea mulțimii formată din cele mai apropiate obiecte de obiectul neclasificat
x	Obiect neclasificat din cadrul mulțimii de obiecte O
C	Numărul de caracteristici numerice ale obiectelor din cadrul mulțimii O
N	Cardinalitatea mulțimii totale de obiecte O
OA	Mulțimea de obiecte folosite pentru procesul de antrenare a algoritmului de clasificare kNN
OT	Mulțimea de obiecte folosite pentru procesul de testare a algoritmului de clasificare kNN
OC	Mulțimea de obiecte clasificate prin intermediul specialiștilor
$d(o_i, o_j)$	Funcția distanță dintre două obiecte aflate pe pozițiile i și j în cadrul mulțimii de obiecte O , $d: O \times O \rightarrow [0; 1]$
D	Matricea distanțelor, unde d_{ij} reprezintă distanța dintre obiectele aflate pe pozițiile i și j
$p_\alpha, p_\beta, p_\gamma$	Probabilitățile de asignare a unui obiect analizat și neclasificat x la nivelul fiecărei categorii de clasificare
OD_x	Mulțimea celor k obiecte aflate cel mai aproape de obiectul neclasificat x

Algoritmul de clasificare kNN este format din trei componente: antrenare, clasificare efectivă și testare. Etapa de antrenare este formată din trei pași și cuprinde următoarele activități:

P₁₁. Crearea bazei de antrenare și testare prin intermediul unor specialiști în scopul clasificării manuale a n obiecte din mulțimea inițială de obiecte, astfel formându-se mulțimile OA și OT , $n = Card(OA) + Card(OT)$.

P₁₂. Separarea mulțimii de obiecte clasificate manual în mulțimea de obiecte destinată antrenării precum și mulțimea de obiecte destinată testării, două mulțimi disjuncte care, prin completare, formează mulțimea totală OC , ținând cont de restricțiile:

$$\begin{cases} OC = OA \cup OT \\ \emptyset = OA \cap OT \end{cases}$$

P₁₃. Încărcarea matricei distanțelor în scopul optimizării rulării algoritmului de clasificare.

$$D = \{d_{ij} | i = \overline{1, n}, j = \overline{1, n}, d_{ij} = d(o_i, o_j)\}$$

Clasificarea propriu-zisă este formată din patru etape și cuprinde activitățile:

P₂₁. Fiind dat obiectul neclasificat x , $x \in O \setminus OC$, aflat pe poziția l din cadrul mulțimii O , se calculează distanța dintre acest obiect și toate celelalte obiecte din cadrul mulțimii OA , sau se extrage din cadrul matricei distanțelor valorile anterior calculate.

P₂₂. Se sortează crescător distanțele obținute și se extrag primele k obiecte care au distanțele minime între obiectul neclasificat x și cele k obiecte selectate, astfel generându-se mulțimea de obiecte OD_x .

P₂₃. Se încarcă valorile $p_\alpha, p_\beta, p_\gamma$ folosind relația:

$$p_i = \frac{Card\{o_j | o_j \in OD_x, c(o_j) = i\}}{k}, i \in \{\alpha, \beta, \gamma\}$$

P₂₄. Se alege clasa de asignare pentru obiectul x folosind metoda votului majoritar neponderat ca fiind cea clasa în care sunt asignate majoritatea celor mai apropiate k obiecte de obiectul x .

$$c(x) = \max_{i \in \{\alpha, \beta, \gamma\}} p_i$$

Componenta de testare implică evaluarea gradului de corectitudine privind algoritmul de clasificare ales pentru asignarea în clase de apartenență a obiectelor neclasificate în etapa de antrenare. Această componentă folosește mulțimea de obiecte despre care se cunoaște clasa de asignare, clasificare făcută aprioric etapei de antrenare, asupra căreia se aplică algoritmul implementat. Această componentă cuprinde activitățile:

P₃₁. Pentru fiecare obiect din cadrul mulțimii OT se aplică algoritmul de clasificare pe baza celor patru etape descrise anterior, astfel rezultând vectorul de valori:

$$ca = (ca_1, ca_2, \dots, ca_i, \dots, ca_{Card(OT)})$$

unde $ca_i = c(o_i)$, $o_i \in OT$, $i = \overline{1, Card(OT)}$ reprezintă clasa de asignare rezultată în urma aplicării algoritmului de clasificare asupra obiectului o_i .

P₃₂. Confruntarea rezultatelor obținute cu valorile asignate inițial de către specialiști, folosind o metrică de evaluare definită astfel:

$$CT = \frac{Card\{o_i | o_i \in OT, c(o_i) = ca_i\}}{Card(OT)}$$

unde CT reprezintă procentul de obiecte clasificate corect folosind ca algoritm de clasificare algoritmul kNN aplicat asupra setului de obiecte OT aprioric clasificate manual.

P₃₃. Analiza privind gradului de corectitudine se va face folosind un prag minimal de validare a metodei implementate, fie acesta $\theta, \theta \in [0; 1]$.

Componenta de calcul a distanței implică alegerea celei mai bune funcții pentru evaluarea distanței dintre obiectele analizate, distanță care se transpune într-un spațiu C dimensional, spațiu ce are cardinalitatea egală cu cardinalitatea numărului de caracteristici descriptive ale obiectelor. Cum tipul caracteristicilor este un număr real, tabelul 3.4 conține un set de distanțe aplicabile acestei context (Shavitt, și Tankel, 2004).

Tabel 3.4 – Calculul distanței dintre două puncte reprezentate într-un spațiu ortogonal n -dimensional

Distanță	Formulă distanță	Codomeniul valorilor
Euclidiană	$d_E(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$	$[0, \infty)$
Manhattan	$d_{MA}(x, y) = \sum_{i=1}^n x_i - y_i $	$[0, \infty)$
Minkowski	$d_{MI}(x, y) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{1/p}$	$[0, \infty)$
Cosinus generalizat	$d_{COS}(x, y) = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}}$	$[0, 1]$
Canberra	$d_{CA}(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{ x_i + y_i }$	$[0, \infty)$
Bray-Curtis	$d_{BC}(x, y) = \frac{\sum_{i=1}^n x_i - y_i }{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}$	$[0, \infty)$
Standardizată	$d_{ST}(x, y) = \sum_{i=1}^n \left(\frac{x_i - y_i}{s_i} \right)^2$	$[0, \infty)$
Mahalanobis	$d_{MH}(x, y) = (x - y)^t \cdot S^{-1}(x - y)$	$[0, \infty)$

Făcând analiza la nivel lexical și semantic, variabilele reprezentate prin text sunt separate de cele numerice, urmând o nouă clasificare folosind algoritmul kNN, clasificare ce urmează pașii anterior menționați, singura modificare fiind realizată la nivelul de evaluare a distanței dintre două obiecte. Această distanță este extrasă din teoria dezvoltării

ontologiei lexicale WordNet, distanța semantică. Clasificarea prin căutarea automată a celor mai apropiate k obiecte de obiectul clasificat, folosind kNN și respectiv agregarea rezultatelor prin metoda votului majoritar, condensează informația semantică la nivel de structură text.

Evaluarea distanței semantice dintre două caracteristici reprezentate prin informație text implică soluționarea următoarelor etape:

- preprocesarea informației text prin separarea textului în cuvinte, extragerea din analiză a cuvintelor de legătură și a celor comune și aplicarea algoritmilor de reducere a cuvintelor la rădăcina lor;
- evaluarea mulțimii de cuvinte rădăcini obținute și extragerea acelor cuvinte care au un corespondent într-un concept din cadrul ontologiei WordNet, (Buhanitzky și Hirst, 2006);
- pentru fiecare concept extras, se analizează și se identifică sensul cuvântului în funcție de contextul în care acesta apare; un astfel de context poate fi extras prin intermediul unui număr de cuvinte apropiate sau folosind ca bază de analiză întreaga frază în care apare conceptul respectiv;
- aplicarea funcției de calcul a distanței semantice dintre sensurile conceptelor identificate în cele două fragmente text diferite.

Pentru a calcula distanța dintre două fragmente text se impune rularea unui algoritm de extragere a sensurilor contextuale pentru fiecare concept identificat. Figura 3.13 conține un exemplu de rulare pentru două propoziții.

1.	Input mode	☐ Word ☒ Sentence
2.	Sentence 1	Eventually, a huge cyclone hit the entrance of my house.
3.	Sentence 2	Finally, a massive hurricane attacked my home.
4.	Submit	Calculate Semantic Similarity

Figura 3.13 – Calculul gradului de similitudine, distanță, dintre două fragmente text

Sursa: <http://ws4jdemo.appspot.com/?mode=s&s1=Eventually%2C+a+huge+cyclone+hit+the+entrance+of+my+house.&s2=Finally%2C+a+massive+hurricane+attacked+my+home.>

Fiecare cuvânt identificat în cele două fraze și se asociază procente cele mai mari de intersecție dintre conceptele extrase dintr-o frază cu toate conceptele extrase din cea de-a doua frază, (McCarty, et. al., 2004; Kim și Baldwin, 2005). Figura 3.14 conține matricea

distanțelor dintre conceptele generate, folosind funcția distanță de evaluare a două concepte WordNet reprezentate arborescent, precum este descris în capitolul 3.3.

	Eventually /RB	,	a /DT	huge /JJ	cyclone /NN	hit /VBD	the /DT	entrance /NN	of /IN	my /PRP\$	house /NN	.
Finally/RB	-	-	-	-	-	-	-	-	-	-	-	-
,/	-	-	-	-	-	-	-	-	-	-	-	-
a/DT	-	-	-	-	-	-	-	-	-	-	-	-
massive/JJ	-	-	-	-	-	-	-	-	-	-	-	-
hurricane/NN	-	-	-	-	<u>0.5000</u>	-	-	<u>0.0625</u>	-	-	<u>0.0714</u>	-
attacked/VBD	-	-	-	-	-	<u>0.5000</u>	-	-	-	-	-	-
my/PRP\$	-	-	-	-	-	-	-	-	-	-	-	-
home/NN	-	-	-	-	<u>0.0833</u>	-	-	<u>0.1429</u>	-	-	<u>1.0000</u>	-
./.	-	-	-	-	-	-	-	-	-	-	-	-

Figura 3.14 – Matricea distanțelor dintre conceptele rezultate în urma preprocesării

Pentru a condensa informațiile descrise în figura 3.14 se extrage, pentru fiecare concept prezent în cadrul unei fraze, concept regăsit în ontologia WordNet și care nu este cuvânt de legătură, cea mai mare probabilitate de intersecție cu acele concepte din care este formată cealaltă frază analizată. Formula care definește distanța dintre două fraze conduce la obținerea disimilitudinii dintre conceptele existente, astfel:

$$ds(o_i, o_j) = \frac{\sum_{i=1}^{n_{c1}} \max_{j=1, n_{c2}} d_{PATH}(c1_i, c2_j)}{n_{c1}}$$

unde:

- $d_{PATH}(c1_i, c2_j) = \frac{1}{lg(c1_i, c2_j)}$ reprezintă distanța semantică dintre conceptele $c1_i$ și $c2_j$ folosind formula Path Length disponibilă în ontologia WordNet, (Pedersen, Patwardhan și Michelizzi, 2004);
- n_{c1} și n_{c2} reprezintă numărul de concepte din care sunt formate cele două fraze.

Variabilele suplimentare folosite în cadrul analizei lexicale și semantice față de cele prezentate în aplicarea algoritmului kNN asupra caracteristicilor numerice sunt prezentate în tabelul 3.5.

Tabel 3.5 – Variabilele clasificării supervizate kNN asupra caracteristicilor text

NOTAȚIE	DESCRIERE
CT	Numărul de caracteristici reprezentate prin text ale obiectelor din cadrul mulțimii O
$ds(o_i, o_j)$	Funcția distanță care evaluează gradul de similitudine din punct de vedere semantic dintre două obiecte aflate pe

NOTAȚIE	DESCRIERE
	pozițiile i și j în cadrul mulțimii de obiecte O , $ds: O \times O \rightarrow [0; 1]$
DS	Matricea distanțelor semantice, unde d_{ij} reprezintă distanța dintre obiectele aflate pe pozițiile i și j
$ps_{\alpha}, ps_{\beta}, ps_{\gamma}$	Probabilitățile de asignare a unui obiect analizat și neclasificat x la nivelul fiecărei categorii de clasificare
ODS_x	Mulțimea celor k obiecte aflate cel mai aproape de obiectul neclasificat x

Rezultatele obținute în cadrul aplicării algoritmului kNN asupra obiectelor digitale reprezentate doar din caracteristicile text extrase din cadrul bazei de date sunt transpuse în vectorul format din probabilitățile de asignare: $ps_{\alpha}, ps_{\beta}, ps_{\gamma}$.

Prin agregarea celor două aplicări a clasificării supervizate, folosind caracteristicile numerice precum și caracteristicile text, se obține un nou set de probabilități, folosind media aritmetică, de asignare în cadrul clasei de securitate majoritară, astfel:

$$pa_i = \frac{p_i + ps_i}{2}, \forall i \in \{\alpha, \beta, \gamma\}$$

Agregarea celor două tehnici de căutare automată folosite pentru clasificarea supervizată a obiectelor digitale existente în cadrul unei baze de date a bibliotecilor virtuale, obiecte descrise prin caracteristici text și numerice, înglobează avantajele aduse de fiecare analiză în parte conducând la o optimizare a clasificării în clase de securitate pornind de la baza de antrenare dezvoltată de specialiști. Multiple studii au integrat componentele text descriptive ale obiectelor în cadrul clasificării conducând la rezultate îmbunătățite verificate pe seturi de antrenare și testare (Elberrichi, Rahmoun și Bentaalah, 2008).

4. SOLUȚIE PRIVIND PROTEJAREA CONȚINUTUL DIGITAL LA NIVELUL SISTEMELOR INTEGRATE DE BIBLIOTECĂ

4.1. Managementul drepturilor digitale în sistemele integrate de bibliotecă

Sistemele integrate de bibliotecă au devenit omniprezente în domeniul bibliotecilor digitale datorită capacității lor de a gestiona conținut digital în aproape fiecare etapă a acestuia, de la achiziție până la arhivare. Prin utilizarea unui ILS angajații unei biblioteci au eficiență în îndeplinirea activităților de bază ce includ conținutul digital iar procesele desfășurate de aceștia sunt mult mai riguroase, având la dispoziție capabilitățile unui sistem care permite efectuarea de corecții la nivelul datelor stocate în sistem.

După cum este menționat în (Mehmood, et. al., 2015), ca urmare a faptului că în prezent mulți utilizatori au acces la o conexiune rapidă la Internet, volume foarte mari de conținut digital sunt ușor accesibile prin intermediul diferitelor tipuri de portaluri. Acest lucru se datorează și faptului că procesul de digitizare, cel prin intermediul căruia ia naștere conținutul digital, este unul foarte accesibil în contextul actual.

Acest aspect reprezintă o problemă de securitate, amplificând domeniul pirateriei, cel al utilizării ilegale de conținut digital. Legislația a ținut pasul cu noile amenințări și industria de dezvoltare software a contracara cu instrumente specializate, bazate pe tehnici de management al drepturilor digitale.

Din punctul de vedere al unui sistem integrat de bibliotecă problemele sunt mult mai ușor de prezentat spre deosebire de soluții, care nu sunt întotdeauna atât de ușor de pus în aplicare. Un ILS gestionează conținut digital, patrimoniu cultural al umanității, care trebuie să fie protejat nu doar de amenințări care îi pot afecta integritatea, dar și din perspectiva legală a drepturilor digitale.

Patrimoniul cultural existent în format digital este, de asemenea, produsul bibliotecii, al instituției care l-a creat. În afară de valoarea dată de proprietatea intelectuală a autorului, un material digitizat prezintă și o valoare adăugată, dată de procesele și resursele care au stat la baza creării acestuia. O mulțime de instituții culturale și biblioteci dețin echipamente specializate gestionate de personal foarte bine calificat. Activitatea se desfășoară în laboratoare de digitizare, unele instituții având chiar și un departament separat care preia obiecte fizice, cărți, în special cele vechi, și le transformă în adevărate opere de artă digitale, tot acest proces consumând foarte multe ore de muncă și inovare.

În scopul de a îmbogăți o versiune electronică a unei cărți sau a unei reviste, cei care digitizează trebuie să utilizeze mai multe instrumente care să le permită să scaneze, să extragă text automat, să identifice și să marcheze tabelele și figurile precum și să adauge fiecărei secțiuni din materialul digital, metadate care să descrie în detaliu obiectul pentru a fi cât mai ușor de accesat și studiat.

După cum este prezentat în (Agosti, et. al., 2014) toate metadatele sunt adăugate la formatul digital cu scopul de a maximiza șansele ca materialul să fie găsit și utilizat în cadrul altor sisteme. Adnotările reprezintă un nivel semantic adițional descrierii obiectului digital, îmbogățind astfel valoarea acestuia cu activitatea asiduă a celor care realizează digitizarea. În cele din urmă, la patrimoniului cultural se va adăuga o altă capodoperă digitală care va îmbunătăți experiența utilizatorilor finali.

Din acest motiv este necesar un nou strat de securitate care trebuie să fie prezent într-un sistem integrat de bibliotecă. Obiectivul său principal trebuie să fie protecția materialului digital pentru care bibliotecile au consumat resurse pentru a-l crea și distribui. Un strat de securitate bazat pe metode de management al drepturilor digitale ar împiedica utilizarea neautorizată de materiale digitale astfel încât să nu fie încălcate drepturilor de autor.

Datorită faptului că majoritatea biblioteci virtuale oferă acces liber la resursele digitale expuse și nu prin intermediul unui abonament, pe bază de subscripție, accesul la cantități mari de materiale digitale este facilitat utilizatorilor din întreaga lume. Din acest motiv este util ca fiecare bibliotecă virtuală să își protejeze resursele electronice, punând în practică sisteme de securitate care să restricționeze utilizările necorespunzătoare, să țină evidența resurselor accesate și să monitorizeze acțiunile utilizatorilor în cadrul sistemului.

Normele legislative trebuie să delimiteze foarte bine granița dintre piraterie și utilizarea corectă a conținutului digital. Din această cerință apare nevoia de punere în funcțiune a unor instrumente de securitate în cadrul sistemelor integrate de bibliotecă care gestionează resursele bibliotecii virtuale cu scopul de a:

- detecta și consolida vulnerabilități existente în sistem, punctele slabe ale acestuia;
- detectează atacurile care se desfășoară asupra sistemului;
- combate atacurile utilizând metode și tehnici specifice;
- diminua impactul unor atacuri asupra sistemului care nu au putut fi combătute.

Operațiile de bază dintr-o bibliotecă virtuală vizează în principal resursele digitale care reprezintă materialul cultural al acesteia. La nivelul unui sistem integrat de bibliotecă, conținutul digital este protejat doar de mecanismele de protecție aflate la nivelul

bazei de date, care asigură integritatea datelor sale precum și accesul neautorizat. Scopul principal al conținutului digital dintr-o bibliotecă virtuală este de a fi prezentat la nivelul interfețelor web pentru ca utilizatorii să-l poată accesa, descărca sau să modifice materialul original, dacă aceștia au drepturile necesare. Datorită acestui scop conținutul digital trebuie protejat de diverse forme de manipulare care nu sunt dorite.

Aceste tipuri de manipulări trebuie să fie menționate în mod explicit de legislația în vigoare și instrumente software specializate trebuie să fie dezvoltate pentru a pune în aplicare normele legislative în consecință. Întrucât tehnicile DRM se ocupă în principal cu protejarea materialelor digitale de manipulare ilegală de orice natură, o altă cale în DRM a fost abordată cu rezultate interesante.

Pe lângă abordarea normală în care managementul drepturilor digitale trebuie să împiedice utilizatorii rău voitori să vizualizeze conținut la care nu au acces, și să prevină încălcarea drepturilor de autor, un modalitate nouă și poate mult mai eficientă este de a schimba așteptările utilizatorilor cu privire la care sunt drepturile pe care aceștia le au asupra materialelor digitale, precum se descrie în (Samuelson, 2003). Acesta este un punct de vedere diferit asupra conținutului digital din perspectiva utilizatorului final.

Schimbarea așteptărilor consumatorilor cu privire la ceea ce au sau nu dreptul cu privire la materialele digitale, ar putea declanșa apariția unor tipuri noi și diferite de conținut digital care să nu pot fi manipulate cu ușurință de către persoanele rău intenționate. Pe lângă rolul de a proteja conținutul digital, sistemele DRM controlează și întregul lanț de distribuție specific materialelor multimedia. Din acest motiv, implicațiile comerciale care rezidă în dreptul de proprietate asupra unui material digital și asupra veniturilor generate de acesta, determină mai multe tipuri de sisteme de DRM, din care sunt două prezintă un interes sporit:

- sisteme DRM care oferă protecție contra manipulării ilegale de conținut digital;
- sisteme DRM care gestionează un întreg model de afaceri în care conținutul digital este accesat pe bază de subscripție sau de plată-per-vizualizare.

Datorită implicațiilor majore pe care sistemele DRM le presupun, dispoziții legislative detaliate trebuie să fie incluse în lege cu scopul de a proteja atât consumatorii cât și producătorii. Multe țări au adoptat în propriile lor legi și regulamente aspecte referitoare la sistemele de management al drepturilor digitale, la calitatea de autor al unui conținut digital sau la conceptul de proprietate asupra unei resurse digitale.

Din punct de vedere tehnologic, un sistem DRM, după cum este menționat și în (Bechtold, 2004), trebuie să aibă următoarele două roluri distincte care trebuie să se completează reciproc, după cum urmează:

- controlul accesului la conținutul digital – restricționează accesul utilizatorilor neautorizați la conținutul digital;
- controlul asupra utilizării conținutului digital – protejează drepturile de autor și proprietate limitând ceea ce un utilizator autorizat este îndreptățit să vizualizeze.

Un aspect de ordin moral la nivelul sistemelor DRM este discutat în (Cohen, 2003), reflectând necesitatea de confidențialitate pentru utilizatorii care folosesc conținut digital distribuit prin intermediul sistemelor DRM. Această lucrare ridică problema conform căreia cei care construiesc regulile de utilizare a resurselor digitale prin intermediul sistemelor DRM cât și factorii de decizie politică trebuie să ofere garanția confidențialității pentru utilizatorii de astfel de produse. Sistemele DRM contribuie la invadarea spațiului privat deoarece conținutul vizualizat de utilizatorii autorizați este în permanență monitorizat, în felul acesta cunoscându-se toate detaliile legate de utilizatorii de conținut. Din acest considerent trebuie introduse mecanisme de securitate la nivelul sistemelor DRM care să ofere garanția confidențialității pentru utilizatori bine intenționați.

Un aspect important în ceea ce privește omogenitatea conținutului în sistemele DRM este dezbătută în (Jonker și Linnartz, 2004). Problema de a integra conținut digital extern care a circulat prin diferite canale de comunicație nesigure, în cadrul unui sistem DRM securizat este foarte solicitantă. Această problemă reprezintă un alt aspect important care necesită reglementări din partea tuturor părților interesate, astfel încât consumatorii să fie protejați de legislație.

Datorită faptului că sistemele DRM nu sunt încă complet sigure iar legea eludează o serie de aspecte care sunt încă importante pentru protecția conținutului digital, dezvoltatorii de astfel de sisteme au recurs la adoptarea unei căi contractuale între aceștia și utilizatorii finali care obligă consumatorii să utilizeze sistemele în condiții specifice, figura 4.1.

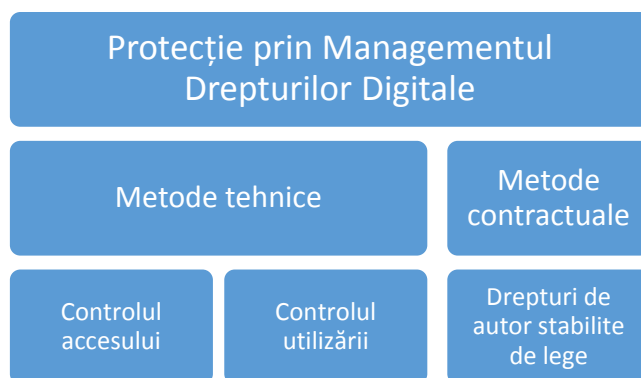


Figura 4.1 – Metode de protecție a conținutului digital utilizând DRM

În (Kubesch și Wicker, 2015) sunt prezentate moduri prin care tehnologiile specifice sistemelor DRM sunt folosite pentru a limita drepturile consumatorilor, de

asemenea, și prin utilizarea de acorduri contractuale pentru a restrânge și mai mult acțiunile utilizatorilor asupra conținutului digital. Aceste acorduri îl obligă pe consumator să suporte sancțiunile stipulate prin contract în cazul în care încearcă să folosească cu rea intenție sistemul de management al drepturilor digitale. Aceasta este o practică comună pentru furnizorii de conținut digital, protejând în acest fel resursa cât și sistemul DRM.

În concluzie, se poate afirma că sistemele DRM folosesc, în afară de soluții tehnologice, și acorduri contractuale pentru protejarea conținutului digital distribuit. Pentru a realiza acest lucru, conținutul digital este legat de un fișier de metadate, denumit fișier de expresie a drepturilor, RELs (**R**ights **E**xpression **L**anguages), care permite sistemelor DRM să verifice o întreagă listă de reguli de utilizare care intersectează conținutul digital cu utilizatorul căruia i se adresează. O implementare frecventă a regulilor stabilite prin metadate RELs este utilizarea de fișiere XML specifice drepturilor digitale, XrML (**E**xtensible **r**ights **M**arkup **L**anguage) care descrie drepturile asupra unui material digital precum: copierea, ștergerea, modificarea, execuția, descărcarea, adnotarea, marcarea, instalarea, distribuirea, împrumutul, imprimarea, afișarea, transferul, închirierea, vânzarea, restabilirea, verificarea, salvarea și alte drepturi adiacente.

Un control de securitate (Whitman și Mattord, 2011), precum controalele de tip DRM, urmărește asigurarea următoarelor caracteristici ale securității la nivelul conținutului digital:

- Integritate – certifică faptul că documentul digital nu a suferit modificări în urma transferului de la o entitate la alta;
- Confidențialitate – se referă la asigurarea unui acces restricționat la nivelul conținutului digital pe baza unor criterii precum nume utilizator și parolă;
- Disponibilitate – face posibilă accesarea materialului în orice moment, atâta timp cât sunt îndeplinite toate condițiile care să permită acest acces;
- Non-repudiare – creează o legătură indisolubilă între entitatea care pune la dispoziție conținutul digital și materialul respectiv;
- Autenticitate – asigură utilizatorul că materialul digital provine de la o sursă autentică.

Controalele de securitate au următorul rol în cadrul oricărui sistem la nivelul căruia sunt implementate, figura 4.2.

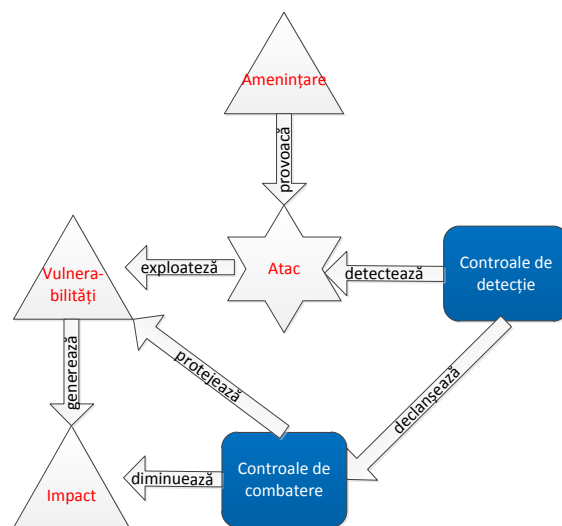


Figura 4.2 – Rolul controalelor de securitate din cadrul unui sistem

Tehnicile bazate pe managementul drepturilor digitale (Tessel, 2006) au rol un bine conturat în procesele de partajare a resurselor digitale. Datorită fluxului mare de operații de acces la conținutul digital, mecanisme rapide și robuste de urmărire, identificare și restricție asupra elementelor partajate, constituie principala metodă de prevenție și oprire a încălcării drepturilor digitale de către utilizatori rău intenționați. Etapele prin care se dorește protejarea conținutului digital la nivelul unui sistem integrat de bibliotecă sunt prezentate în figura 4.3.

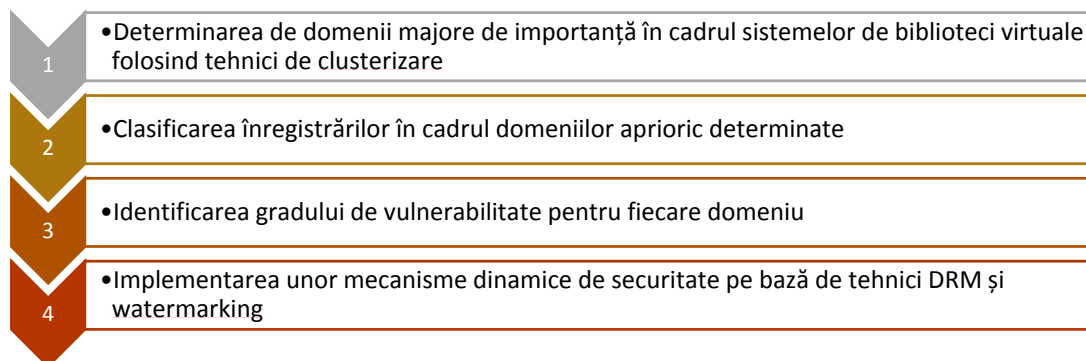


Figura 4.3 – Etapele procesului de protecție a conținutului digital

Scopul unui sistem DRM este acela de a se interpune între conținutul digital expus prin intermediul sistemelor de biblioteci virtuale și utilizatorii externi care îl accesează fără a ține cont de drepturile digitale pe care instituția ce-l pune la dispoziția lor le are. În această situație sunt puse în evidență următoarele aspecte:

- protejarea conținutului digital de utilizările și preluările neautorizate;
- gestiunea proceselor de partajare a conținutului digital între diverse biblioteci virtuale (OAIPMH, 2014; IR, 2014);
- definirea unui set de drepturi digitale specifice bibliotecilor virtuale:

- drepturi de consultare asupra conținutului digital – utilizatorii au dreptul de a viziona documentele și de a face mențiuni la acestea, specificând în mod explicit sursa la care se face trimitere;
- drepturi de preluare și utilizare a materialelor publicate – utilizatorii pot prelua conținutul și îl pot integra ca atare în propriile materiale, cu condiția să fie menționată sursa acestora;
- drepturi de prelucrare a elementelor partajate – se referă la dimensiunea culturală a conținutului digital preluat din surse externe și la posibilitatea modificării acestuia, oferindu-i noi valențe în scopul republicării cu menționarea sursei originale.

Toate aceste aspecte asigură o mai bună protecție a patrimoniului digital publicat la nivelul bibliotecilor virtuale, fiind astfel protejată munca persoanelor care s-au îngrijit de achiziția, crearea prin digitizare, prelucrarea metadatelor și elaborarea produsului final și totodată costurile indirecte care au condus la publicarea patrimoniului.

4.2. Protejarea conținutului digital pe bază de tehnici DRM

Implementările existente ale sistemelor bibliotecilor virtuale sunt bazate pe soluții software care utilizează sisteme informatice specializate ce permit manipularea materialului digital. Aceste tipuri de sisteme se concentrează mai mult pe manipularea materialelor, modul în care se fac căutările, modalitatea de afișare a lor cu scopul de a fi atrăgătoare utilizatorilor finali.

Protecția conținutului digital este realizată prin intermediul mecanismelor de securitate care se regăsesc la nivelul de stocare, al bazei de date. Din acest motiv, o configurație de protecție a datelor este analizată spre a fi propusă pentru implementare pe baza particularităților bibliotecilor virtuale. Instrumentul de securitate își propune să fie o interfață între interacțiunile dintre utilizatorii anonimi și catalogul cu acces public, Online Public Access Catalog, al unei biblioteci virtuale. O soluție pentru protejarea conținutului digital din cadrul unei biblioteci este de a utiliza tehnici specializate de management al drepturilor digitale care să restricționeze sau să limiteze accesul la materiale pe baza unor drepturi de acces.

Un mod prin care acționează un sistem DRM pentru protecția conținutului digital este de a transmite informația într-o formă criptată către consumator. Dispozitivele consumatorilor primesc materialele criptate și au capacitatea de a le decodifica la cerere atunci când utilizatorul încearcă să acceseze folosind conturi autorizate pentru care i-a fost

permis accesul. Acest tip de protecție este denumită soluție DRM prin intermediul unui suport digital de stocare.

Tehnicile de criptare au avansat constant deoarece cele mai vechi erau mereu în vizorul atacatorilor care reușeau să le corupă. Un nou model DRM pe bază de criptare este propus în (Mehmood, et. al., 2015), care utilizează sistemul de criptare integrat pe bază de curbe eliptice (ECIES) precum și o funcție *hash* unidirecțională pentru generarea cheilor de criptare/decriptare de tip singular. Precum se menționează în (Mehmood, et. al., 2015), o parte din cheie va fi stocată în licență iar cheia nu va fi niciodată refolosită sau stocată în cadrul echipamentelor furnizate utilizatorilor. Acest lucru asigură faptul că dacă în cazul în care o încercare de aflare a cheii a avut succes, iar aceasta a fost publicată, accesul la materialul digital protejat nu se va finaliza cu succes datorită faptului că dispozitivele nu au cheia stocată. În această situație nu trebuie decât înlocuită cheia compromisă cu una nouă și sistemul este din nou protejat.

În (Bechtold, 2004), sunt prezentate mai multe modele de sisteme de management al drepturilor digitale. Rolul lor este de a proteja sau limita accesul la conținutul digital pentru utilizatorii neautorizați, utilizând:

- containere digitale care folosesc criptarea pentru limitarea accesului;
- arhitecturi pentru blocarea drepturilor folosite pentru a permite utilizatorilor să acceseze propriile materiale digitale din cadrul diferitelor tipuri de dispozitive;
- sisteme de management al generării de copii, sau CGMS, limitează numărul de copii pe care un utilizator le poate face la un conținut digital; după ce această limită este atinsă, conținutul devine neutilizabil.

Sistemele DRM se confruntă cu două tipuri majore de atac care pot permite utilizatorilor malițioși să acceseze conținut digital fără permisiunile aferente:

- atac pentru a găsi cheile de criptare; fiecare utilizator care are acces la cheile de criptare are permisiunea de a vedea sau manipula conținut digital;
- captarea conținutului necriptat; pentru aceste atacuri, utilizatorii trebuie să exploateze breșele din cadrul procedurilor standardizate ale sistemelor DRM.

Pentru a realiza o protecție a materialului digital, sistemele DRM trebuie să folosească date în scopul validării, date atașate la conținutul multimedia original. Această informație este cunoscută sub numele de metadata, date ce descriu un obiect specific, ce pot fi atașate ca un antet special sau chiar încorporate în cadrul materialului prin tehnici speciale de filigranare digitală sau steganografie.

Metadatale permit citirea unor secțiuni speciale ale conținutului digital, zone denumite descriptori care conțin informații despre:

- cine a furnizat materialul în cele mai multe cazuri fiind vorba de calitatea de autor;
- cui îi este permis să acceseze, entitate denumită și utilizator sau deținător/posesor;
- ce tipuri de operații sunt permise pe baza privilegiilor asigurate posesorilor.

Sistemele DRM nu acționează doar ca un nivel de protecție la nivelul conținutului digital. Acestea sunt folosite și în scopuri comerciale în arhitecturi de tip plată-per-vizualizare, cunoscute sub denumirea de pay-per-view. În acest mod, sistemele DRM gestionează cum conținutul digital este transmis către utilizatorii finali și modalitatea în care acesta este accesat. Au fost dezvoltate chiar și sisteme DRM avansate care folosesc arhitecturi cloud, precum este prezentat în (Lee, et. al., 2015). Unul din motive este faptul că sistemele DRM trebuie să furnizeze servicii către utilizatori în orice moment de timp sau spațiu. De exemplu, sistemul UltraViolet oferit de DECE, (Kalker, Samtani și Wang, 2012; UltraViolet, 2015; UvuWiki, 2015), suportă partajarea conținutului digital între dispozitivele utilizatorilor prin intermediul tehnicilor de autentificare digitală în cloud. În (Lee, Seo și Shin, 2013) a fost propusă o arhitectură de tipul DRM-ca-serviciu ce oferă funcționalități diverse specifice DRM ca servicii în mediu cloud, fiind cunoscute sub numele de DRM Cloud.

Sistemele DRM sunt asociate, de asemenea, cu sistemele bazate pe încredere, (Cooper și Martin, 2006), într-o combinație ce descrie o arhitectură deschisă pentru managementul drepturilor digitale pus în aplicare pe platforme protejate care împuternicește consumatorii să selecteze sistemul lor de operare și aplicațiile, incluzând operații cu acces liber, fără a slăbi puterea funcțiilor de securitate. Tehnicile DRM sunt folosite pentru un anumit nivel a obiectelor media care sunt sensibile și protejate.

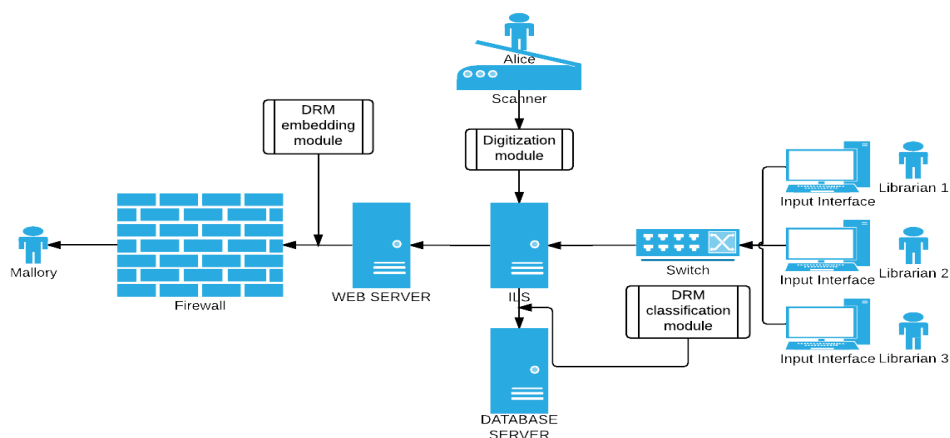


Figura 4.4 – Model de protecție a resurselor bazat pe DRM în bibliotecă virtuală

Modelul propus de protecție a datelor în cadrul figurii 4.4 are scopul de a fi o interfață între interacțiunile lui Mallory, utilizatorul rău intenționat, cu conținutul digital care a fost creat de Alice, utilizatorul care deține drepturile digitale asupra conținutului. Modelul dorește să capteze cererile lui Mallory pentru conținutul digital și să-i ofere

acestui conținut digital protejat funcție de un set de drepturi digitale adăugate fiecărui obiect.

Modulul de integrare a protecției din prezentul model se activează automat de fiecare dată când serverul web primește o solicitare de conținut digital. Modulul de clasificare DRM este declanșat, de asemenea, automat de fiecare dată când un conținut nou este stocat în cadrul sistemului.

Modulul pentru scrierea identificatorilor utilizați în verificarea integrității se bazează pe mecanisme de ascundere a informațiilor în conținutul digital ce se dorește a fi protejat, prin tehnici de steganografie sau de tip marcaje invizibile, fragile. Tehnicile de securitate care vizează caracteristicile de integritate și confidențialitate au la bază principii care au devenit standarde de securitate precum:

- principiul lui Kerckhoffs: “Dacă se presupune că sunt cunoscute date cu privire la sistemul de steganografie implementat, nimeni nu poate găsi indicii ale existenței mesajului, decât cu ajutorul cheii secrete”;
- principiul “Securitate prin obscuritate” ce presupune alterarea metodelor de design și păstrarea acestora secretă pentru a nu putea identifica mesajul;
- principiul susținut de NIST, „keep it simple”: „Securitatea sistemului nu trebuie să depindă pe ascunderea metodelor de implementare și a componentelor acestuia”.

Steganografia este știința de a ascunde existența unor mesaje transmise de la expeditor către destinatar prin intermediul unui purtător. Etimologia cuvântului steganografie are la bază două concepte ce provin din limba greacă, steganos care înseamnă *acoperiș* și grapho cu semnificația de *scriere*, de unde și denumirea de *scriere ascunsă*.

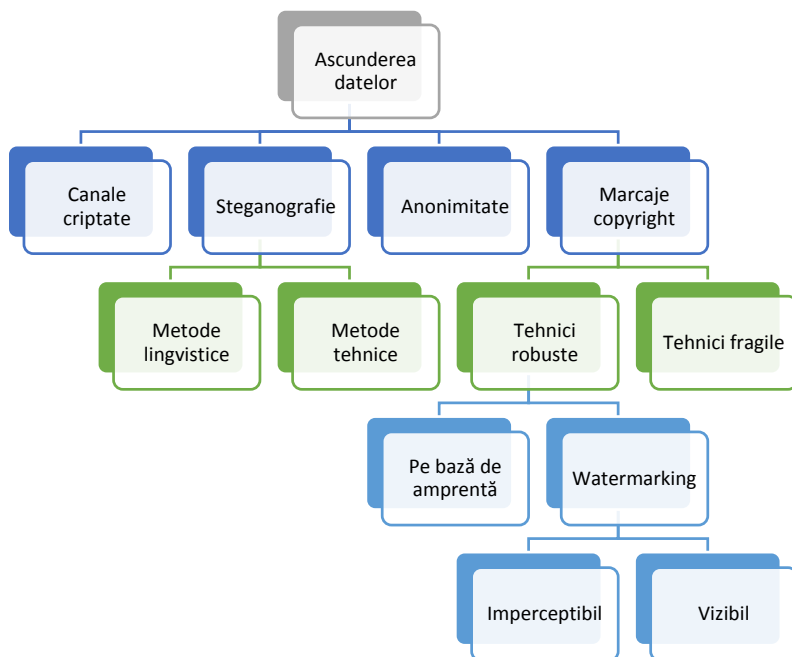


Figura 4.5 – Mecanisme de ascunde a datelor

Documentul în care se ascunde mesajul poartă denumirea de fișier gazdă iar cel care este ascuns se numește fișier sursă. În cadrul sistemului implementat a fost folosit pe post de fișier gazdă, un document digital de tip imagine bitmap. Formatul bitmap este un format standard cu următoarele caracteristici:

- este un tip de organizare a memoriei folosit pentru stocarea imaginilor digitale;
- este o matrice informațională simplă formată din pixeli reprezentați prin puncte de anumite culori;
- imaginea este voluminoasă și dependentă de scala de vizualizare;
- nu se poate adapta unei scări variabile de vizualizare;
- imaginea stocată este prezentată ca o hartă de biți.

Pentru fiecare punct din matricea informațională, denumit pixel, se folosește un sistem de reprezentare a culorilor denumit RGB. Acesta este un sistem de culori aditiv, care permite crearea unei culori prin suprapunerea a două sau trei culori de bază, roșu, verde și albastru, figura 4.6.

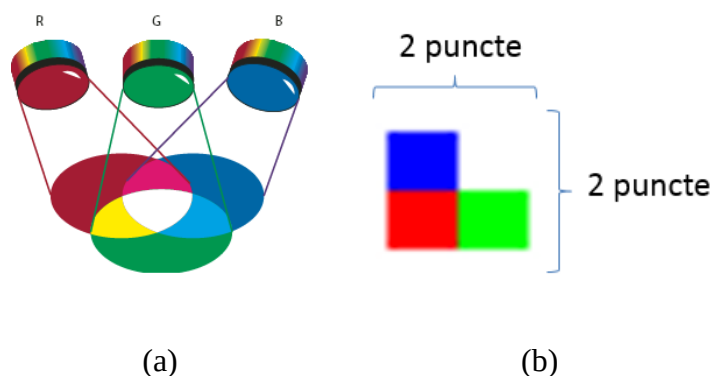


Figura 4.6 – (a) Modelul RGB de reprezentare (b) imagine a 4 puncte

Lipsa oricărei culori din reprezentarea RGB conduce la obținerea unui punct negru, acesta reprezentând absența culorilor, a luminii. Fișierele BMP stochează imaginile în format bitmap cu următoarea structură, precum este prezentat în figura:

- **BITMAPHEADER**, 14 octeți, conține informații variate despre antetul unui astfel de fișier, precum:
 - **2 octeți**, semnătura fișierului care este BM, valoarea *4D42h*;
 - **4 octeți**, dimensiunea fișierului;
 - **4 octeți**, o zonă rezervată;
 - **4 octeți**, deplasamentul la care reprezentarea imaginii începe;
- **BITMAPINFOHEADER**, lungime fixă, 40 octeți:

- **4 octeți**, prezintă dimensiunea zonei despre antetul informației, info header, valoare de 28h;
- dimensiunile imaginii date de înălțime(**4 octeți**) și lățime(**4 octeți**);
- **2 octeți**, numărul de plane;
- **2 octeți**, profunzimea culorii dată de numărul de biți per pixel;
- **4 octeți**, compresia, în cazul în care există;
- **4 octeți**, dimensiunea totală a imaginii dacă compresia este specificată;
- **4 octeți**, rezoluția orizontală și **4 octeți**, rezoluția verticală;
- **4 octeți**, numărul de culori utilizate si **4 octeți** numărul de culori importante.
- **OPTIONAL PALLETE** reprezentată de tabela de culori, este specificată dacă numărul de biți per pixel este mai mic sau egal cu 8;
- **IMAGE DATA** este zona în care informația utilă este stocată.

Structură		Legendă
Bitmap Header		zonă de 14 octeți
	1	4B , semnătura fișierului, <i>4D42h</i>
	2	4B , dimensiunea fișierului, <i>00000046h</i>
	3	4B , zonă rezervată, <i>00000000h</i>
	4	4B , deplasamentul la care se regăsesc datele, <i>00000036h</i>
Bitmap InfoHeader		zonă de 40 octeți
Bitmap Data (pixels)		puncte stocate în format RGB / 1B per culoare, un total de 3B
	5	3B pixel – culoarea roșie, <i>FF0000h</i>
	6	3B pixel – culoarea verde, <i>00FF00h</i>
	7	2B adăugați pentru deplasament
	8	3B pixel – culoarea albastră, <i>0000FFh</i>
	9	3B pixel – pixel alb, <i>FFFFFFh</i>
	10	2B adăugați pentru deplasament

Figura 4.7 – Descrierea structurii unui fișier BMP

Zonele de date notate cu numere de la 1 la 10 sunt prezentate în versiunea hexazecimală în figura 4.8.

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00000000	42	1D	46	00	20	00	00	00	00	36	00	00	00	28	00	
00000010	00	00	02	00	00	00	02	00	00	00	01	00	18	00	00	00
00000020	00	00	10	00	00	00	00	00	00	00	00	00	00	00	00	00
00000030	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000040	00	FF	FF	FF	FF	00	10	00								

Figura 4.8 – Structura unui fișier BMP în format hexazecimal

Tehnica de scriere a datelor de control în cadrul fiecărui document digital se bazează pe folosirea biților cel mai puțin semnificativi, astfel încât imaginea inițială să nu sufere alterări vizibile. Metoda presupune următoarele procese:

1. utilizarea unei chei în scopul modificării datelor de control pe baza acesteia;
2. utilizarea unei funcții de dispersie în scopul răspândirii datelor în interiorul fișierului gazdă;
3. combinarea cheii cu funcția de dispersie pentru minimizarea gradului de detecție.

Ca răspuns la o cerere de descărcare a unui obiect digital stocat în biblioteca virtuală, sistemul va furniza utilizatorului o versiune securizată care este capabilă să evidențieze orice modificare adusă obiectului prin utilizarea de coduri de control, figura 4.9.

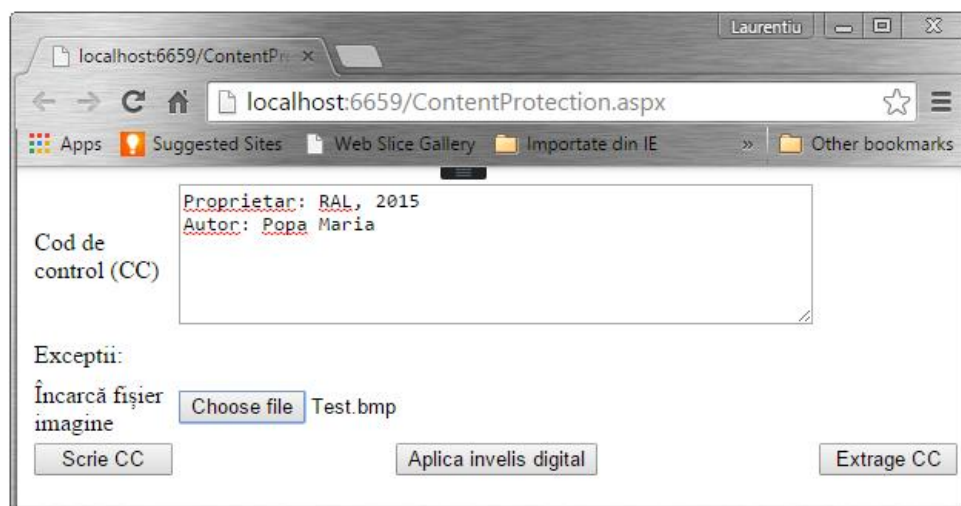


Figura 4.9 – Scrierea codului de control la nivelul unui obiect digital de tip imagine

Obiectul furnizat utilizatorilor rezultă ca urmare a procesului de marcare digitală, acesta conținând un cod de control ce permite verificarea dacă conținutul digital a fost sau nu modificat. În cazul în care utilizatorul nu modifică imaginea, codul extras din cadrul fișierului gazdă, va fi același cu cel scris pentru control.

Dacă însă, după ce imaginea a fost descărcată, utilizatorul modifică conținutul acesteia sau îi schimbă dimensiunile, codul de control stocat în imagine prin intermediul algoritmului fragil de scriere este denaturat, lucru ce se poate observa la o testare ulterioară modificării în încercarea de a-l extrage, figura 4.10.

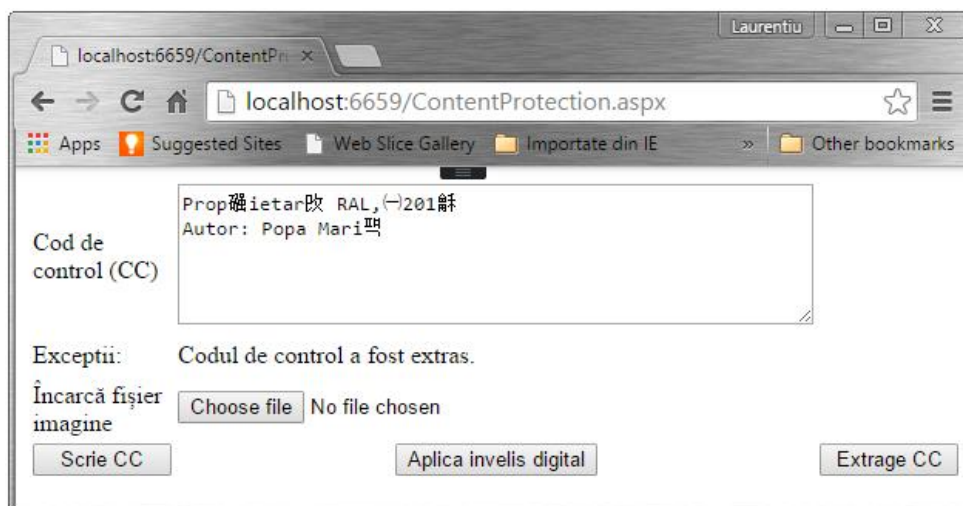


Figura 4.10 – Extragerea codului de control dintr-o imagine modificată

Verificarea integrității se bazează pe faptul că algoritmul pentru scrierea codului de control este unul fragil, astfel orice modificare adusă imaginii inițiale are ca efect alterarea codului inițial. În figura 4.10, opțiunea *Aplică înveliș digital* are rolul de a securiza imaginea din punct de vedere al caracteristicii de confidențialitate. Aceasta aplică asupra unei resurse digitale de tip imagine, un înveliș digital care o protejează de vizualizările neautorizate. Accesul la fișierul rezultat, tot de tip imagine, însă protejat, se realizează doar prin intermediul unei aplicații special dezvoltate, care permite citirea structurii noi imaginii și afișarea conținutului, figura 4.11.



Figura 4.11 – Decodificarea unei imaginii cu un înveliș digital

Metoda de realizare a învelișului digital se bazează pe utilizarea unei parole de acces la conținut, a cărei amprentă este adăugată imaginii inițiale, astfel rezultând o nouă imagine care pentru a fi vizualizată necesită decodificarea pe baza parolei corecte. Pașii algoritmului de decodificare sunt următorii:

- A. introduce parolă pentru vizualizare, p_1 ;
- B. creare valoare hash a parolei introduse, ph_1 ;
- C. extragere din șirul de octeți aferent imaginii codate, un număr de k octeți aferenți valorii hash calculate anterior, ph_1 ;
- D. transformarea celor k octeți în valoarea hash a parolei existente în imaginea codată, ph_2 ;
- E. compararea celor două valori hash, ph_1/ph_2 ;
- F. dacă cele două valori sunt identice se extrage imaginea și se afișează utilizatorului.

Asigurarea confidențialității unei resurse pe bază de DRM cu înveliș digital este o metodă eficientă care intermediază accesul la obiectele existente într-o bibliotecă virtuală pe bază de parole, certificate digitale, amprente sau alte mecanisme de identificare viabile unui sistem integrat de bibliotecă.

4.3. Verificarea autenticității conținutului digital folosind coduri cu răspuns rapid

Aplicațiile mobile sunt definite ca fiind modele noi care oferă un suport informațional în condiții de variabilitate pozițională. Această categorie de produse software extinde beneficiile puterii computaționale generată de calculatoare la nivelul dispozitivelor mobile care sunt mai ușor de transportat și accesat de către utilizatori, indiferent de locația în care aceștia se află.

În aceste circumstanțe, o aplicație mobilă poate fi descrisă ca un concept optimizat care poate rula pe calculatoare performante. Datorită limitării lor, cele mai importante aspecte ale unei aplicații pot fi rezumate în:

- consumul energiei care, în cazul aplicațiilor mobile, trebuie să fie redus, în caz contrar bateria se va epuiza rapid iar cea mai importantă caracteristică a unor astfel de dispozitive, mobilitatea, se va pierde;
- puterea de procesare trebuie de asemenea să fie limitată în termeni de operații pe ciclu, date fiind restricțiile hardware;
- capacitatea memoriei; acest subiect tinde să fie depășit datorită existenței tranzistoarelor cu capacitate mare de stocare;
- ecrane de dimensiune mai mică forțează o nouă abordare în managementul conținutului, astfel noi instrumente au apărut pentru a face față acestui dezavantaj.

Multe instrumente software pentru dispozitive mobile au fost dezvoltate pentru a acoperi nevoia din această arie dar, în același timp, ca un rezultat al caracteristicilor dispozitivelor mobile, au generat multe probleme de securitate. Codurile pe bază de

răspuns rapid , QR (**Q**uick **R**esponse), pe cât de simple par, sunt foarte sensibile la atacurile malițioase prin intermediul conținutului ascuns pe care acestea îl integrează.

Utilitatea codurilor cu răspuns rapid este vizibilă în foarte multe domenii de cercetare și economice precum și în mediile de afaceri. Pornind de la managementul produsului, situație în care codurile QR sunt folosite pentru identificare prin etichetare, continuând cu zona de publicitate, în care cel mai important rol al codurilor QR este de a încorpora diferite date precum adresa URL pentru un acces rapid, și încheind cu realitatea extinsă, situație în care codurile QR pot fi folosite ca o bază pentru extragerea și afișarea informațiilor adiționale la un nivel virtual care se suprapune fluxului video real, codurile QR și-au adus pe deplin contribuția indiferent de domeniu.

Realitatea extinsă sau îmbunătățită, (Azuma, 1997; Zhou, Duh și Billingham, 2008; Krevelen și Poelman, 2010), este un instrument puternic care aduce un plus de percepție realității înconjurătoare pe baza unor informații adiționale extrase în timp real din cadrul unui sistem informatic. Tehnicile de realitate îmbunătățită sunt folosite în diferite arii care implică prezentarea vizuală în scopuri informative. Codurile QR sunt folosite pentru a controla într-o anumită măsură procesul de transformare a mediului real cu ajutorul informației virtuale extrase din cadrul unui sistem. Codurile QR adaugă date suplimentare care pot altera calea pe care un sistem de realitate extinsă o urmează atunci când un conținut este afișat.

Un cod de bare constă într-o reprezentare optică a datelor care codifică informații referitoare la un obiect sau produs asupra căruia îi este atașat, fiind folosit în multe arii din industrie așa cum se prezintă în (Turcu și Doinea, 2015).

O descriere a evoluției și particularităților specifice codurilor cu răspuns rapid, precum este descris în (Turcu și Doinea, 2015), este prezentată în cele ce urmează. Un cod de bare clasic, precum se observă în figura 4.12, constă într-o secvență liniară de bare albe sau negre care variază în dimensiune, depinzând de tipul de informație care este codificată în cadrul codului de bare; setul de simboluri utilizat este, de regulă, limitat la câteva litere, numere sau caractere speciale.



Figura 4.12 – Cod de bare universal

Pentru a face față la creșterea cantității de date care trebuie împachetată într-un cod de bare simplu, s-au dezvoltat codurile bidimensionale: acestea reprezintă imagini bidimensionale care stochează informația atât orizontal, asemănător situației codurilor

clasice de bare, precum și vertical. Acestea au o formă dreptunghiulară și pot stoca până la 6000 de caractere, mult mai multe decât capacitățile codului de bare tradițional. Printre cele mai des întâlnite coduri de bare matriciale se amintesc codurile QR, MaxiCode, DataMatrix și Aztec Code.

Codul QR, precum se observă în figura 4.13, este o marcă înregistrată pentru codurile de bare, folosit pentru prima dată în industria automobilelor din Japonia.

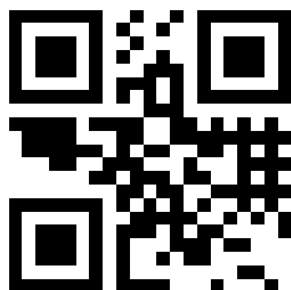


Figura 4.13 – Exemplu de QR Code

Odată cu apariția lor, codurile QR au atras atenția în cazul multor arii de interes, (Rouillard și Laroussi, 2008; Lorenzi, et. al., 2012). În marketingul de promovare, sistemele de e-learning, e-guvernamentale și aproape orice este alimentat de sistem informatice, se poate integra componenta ce cuprinde coduri QR.

Cantitatea de informație care poate fi codificată în cadrul acestor tipuri de coduri de bare se află în strânsă relație cu tipul datelor, versiunea (dimensiunea) codului de bare precum și de nivelul de corecție a erorii.

Codurile QR au fost dezvoltate într-o asemenea manieră încât chiar dacă aparent cea mai mare parte din el lipsește, datele pot fi încă citite: există patru niveluri de corecție a erorilor care pot fi integrate într-un cod Quick Response, în funcție de mediul în care acesta se dorește a se aplica:

- L (scăzut) – până la 7% din cod pot fi restaurate;
- M (mediu) – până la 15% din cod pot fi restaurate;
- Q (quartilă) – până la 25% din cod pot fi restaurate;
- H (ridicat) – până la 30% din cod pot fi restaurate.

Precum se observă în tabelul 4.1, capacitatea maximă a unui cod QR constă în 4296 de caractere alfanumerice (având dimensiunea maximă și nivelul de corecție a erorii minimă), oferind suport pentru date de tipul Kanji/kana.

Tabel 4.1 – Capacitatea maximă a unui cod QR

<i>Tipul de date</i>	<i>Numărul maxim de caractere</i>	<i>Intervalul de caractere</i>
Numeric	7.089	0 - 9

<i>Tipul de date</i>	<i>Numărul maxim de caractere</i>	<i>Intervalul de caractere</i>
Alphanumeric	4.296	0 - 9, A – Z, space, \$, %, *, +, _, ., /, :
Byte	2.953	ISO 8859-1
Kanji/kana	1.817	Shift JIS X 0208

Precum se observă în figura 4.14, codul QR conține diferite zone cu informație specifică despre: versiunea (dimensiunea) codului, formatul (nivelul de corectare a erorii și un model folosit pentru crearea QR), key pentru corecția erorilor (informație folosită pentru decriptarea codului), o zonă liniștită (pentru rezultate valide trebuie să existe un spațiu între două sau mai multe coduri alăturate) și modele necesare pentru detectarea poziției codului QR.

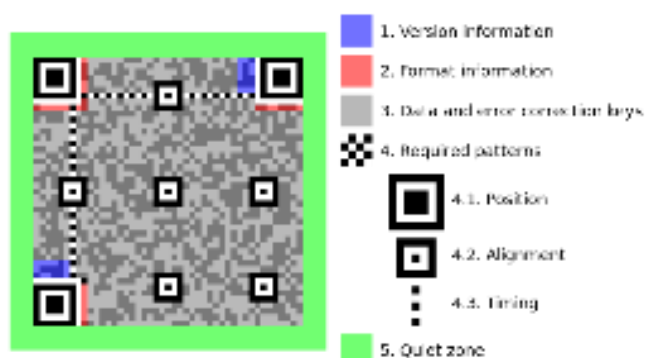


Figura 4.14 – Structura unui cod QR

Chiar dacă acest cod cu răspuns rapid a fost creat pentru utilitate industrială, a devenit o parte importantă din viața de zi cu zi, mai ales în zona de publicitate. Este foarte comun pentru o campanie de marketing să introducă în strategia sa coduri QR plasate la nivelul oricărui element de publicitate, precum se observă în figura 4.15, fișe, clipuri publicitare la TV sau internet, tricouri sau chiar tatuaje.



Figura 4.15 – Reclame cu coduri QR

Acest tip de plasare a codurilor QR se adresează clienților care dețin un telefon mobil inteligent sau smartphone: codul poate fi citit folosind aplicații diverse disponibile gratuit la nivelul fiecărui furnizor: Android, iOS, Blackberry OS sau WindowsPhone. În

cazul Android și iOS, nu există un cititor de coduri QR nativ, dar există mai mult de 50 de aplicații gratuite sau contra cost disponibile pe AppStore sau Google Play, care sunt capabile în același timp să decodifice mesajul și să redirecționeze utilizatorul în cazul în care se recunoaște un format: adresă de internet, adresă de email, ID Facebook, etc. Astfel, de îndată ce o persoană sesizează un panou cu un cod QR aceasta folosește telefonul pentru a scana codul și navighează imediat către pagina de internet dorită.

Aceste coduri bidimensionale pot fi, de asemenea, folosite pentru a stoca informații despre conturi bancare sau carduri sau pot fi adaptate pentru a opera în cazul unor aplicații de plăți speciale. De asemenea, codurile QR sunt foarte folosite în campaniile de marketing, la nivelul paginilor de internet sau a rețelelor online de socializare, sunt folosite și în cazul situațiilor din viața reală: un cod plasat pe un plic, o cutie de pizza sau un semn de stradă, situații în care utilizatorilor li se oferă posibilitatea de acces rapid la informații despre acel obiect; magazinele cu amănuntul pot pune coduri QR imediat în apropierea produselor, redirecționând spre pagini de internet care conțin aspecte detaliate sau comentarii despre ele; de asemenea, agenții imobiliari pot folosi coduri QR la nivelul panourilor publicitare de închiriere sau de vânzare care redirecționează spre o prezentare video a imobilului, precum este prezentat în figura 4.16.



Figura 4.16 – Anunț de vânzare cu coduri QR

Scopul aplicațiilor de citire QR este de a furniza o soluție rapidă la nevoile constante ale persoanelor care doresc prelucrarea rapidă de informații care sunt transmise prin intermediul codurilor QR, aplicații compatibile cu orice dispozitiv Android care deține o cameră de fotografiat și rulează pe versiunea minimă de 2.2 a acestui sistem de operare. O problemă comună pentru toate cititoarele de coduri QR disponibile pentru dispozitive inteligente este dată de incapacitatea de a decodifica coduri multiple amplasate în aceeași zonă de vizibilitate: aplicațiile vor recunoaște primul cod QR care devine disponibil pentru citire, precum se observă în figura 4.17. Dacă mai multe astfel de coduri sunt amplasate apropiat, aplicația va decodifica unul la întâmplare, ieșind din imaginea care le conține și

deschizând o fereastră care să indice ceea ce codul a scanat precum și opțiunile suplimentare disponibile.



Figura 4.17 – Modul de scanare al unei aplicații QR

În schimb o aplicație care oferă utilizatorilor posibilitatea de a vedea în timp real conținutul tuturor codurilor de bare existente în aria de vizualizare a camerei, cu opțiunea de a selecta ce cod să fie procesat pentru opțiuni ulterioare este o soluție perfectă pentru verificarea conținutului digital pe bază de coduri QR.

În cadrul unei astfel de aplicații fiecare din codurile QR detectate și decodificate vor fi acoperite de un pătrat alb semitransparent deasupra căruia va fi amplasat conținutul text.

Meniul conceptual al aplicației, figura 4.18, are trei opțiuni: “Flash toggle”, “Change focus mode” și “Database toggle”.

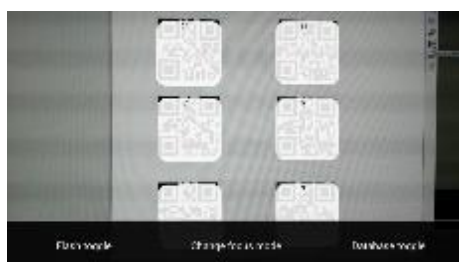


Figura 4.18 – Meniul contextual al aplicației

Primul buton din stânga apare doar dacă dispozitivul are instalată o cameră flash și este disponibilă atunci când utilizatorul scanează codul care este corespunzător luminat. Următorul buton oferă utilizatorilor posibilitatea de a schimba modul focus al camerei: în versiunea implicită, modul focus este setat la modul continuu; în cazul în care codul este prea mic, aproape orice cameră poate fi setată la modul macro. Opțiunile, precum sunt prezentate în figura 4.19, sunt populate funcție de capacitatea dispozitivului.



Figura 4.19 – Opțiunile camerei fotografice

O ultimă opțiune este comutarea din modul bazei de date “Database mode” care schimbă modul în care fluxul video este modificat: dacă textul conținut în cadrul codului QR are o imagine alocată în baza de date internă, aplicația va adăuga fluxului video o imagine plasată deasupra codului QR corespunzător, figura 4.20; dacă textul nu deține o imagine asignată, aplicația va arăta codul QR precum este descris în modul normal, adăugând următorul text: „Imaginea nu a fost identificată!”.



Figura 4.20 – Exemplu de realitate extinsă

În timpul rulării aplicației, utilizatorul poate interacționa cu codurile QR vizibile pe ecran, la o atingere asupra unui cod specific apar următoarele opțiuni:

- revenirea la ecranul principal folosind butonul “Ok” sau butonul dispozitivului “Back”;
- pentru a naviga la adresa dorită folosind un browser instalat în cazul în care codul conține o astfel de adresă; dacă sunt detectate mai multe adrese web, utilizatorul poate alege ce adresă să deschidă;
- pentru a adăuga un nou contact în lista de adrese ale dispozitivului în cazul în care codul conține un vCard, mai multa câmpuri fiind autocompleteate cu informațiile din vCard;
- pentru a deschide un client de email cu adresa indicată de către codul QR în cazul în care aceasta are forma: “mailto:sampleandress@host.suffixes”

Ca răspuns la operația de selecție a utilizatorului, punctul exact este indicat pe ecran cu o serie de cercuri concentrice mici precum și prin o notificare audio scurtă. În cazul inexistenței unui cod QR în jurul punctului selectat, o scurtă notificare apare, indicând: “Niciun cod QR detectat.”

O aplicație mobilă care combină codurile QR cu tehnici AR este chiar mai nesigură, întrucât aceste instrumente pot fi exploatate în funcție de tipul de manipulare urmat de aplicație, o manipulare automată sau orientată spre utilizator. În (Kieseberg, et. al., 2010) sunt prezentate probleme ale codurilor QR privind manipularea precum și atacurile specifice.

Manipularea QR se bazează pe principiul în care dacă cineva poate altera vreo regiune QR, figura 4.21, care conține date sau descriere codul QR, prin schimbarea unui

pixel alb în negru, atunci codul QR nu va mai conține informația inițială care a fost stocată la nivelul acestuia.

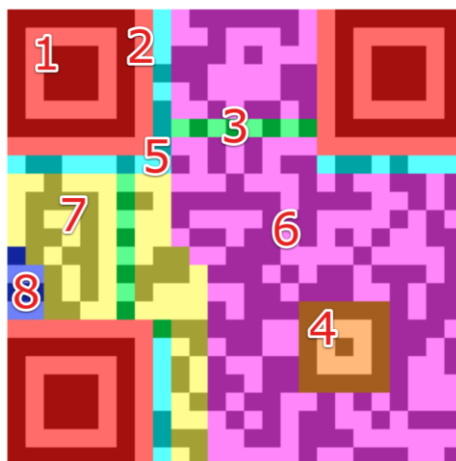


Figura 4.21 – Secțiunile unui cod QR versiunea 2

Regiunile unui cod QR, versiunea 2, sunt:

- regiunea 1 – cele trei colțuri sunt denumite modele de recunoaștere a codurilor QR, aceasta ajută scanarelor să detecteze regiunea codului QR și să-i determine orientarea;
- regiunea 2 – separator cu rolul de a distinge zonele din cadrul codului QR;
- regiunea 3 – model care ajută decodificării de a determina lățimea modulelor;
- regiunea 4 – model de aliniere care ajută decodificării de a corecta deformarea în situația în care codul este îndoit sau curbat;
- regiunea 5 – stochează formatul informației pentru codul QR;
- regiunea 6 – conține datele actuale din cadrul codului QR;
- regiunea 7 – cod de corectare a erorii generat pentru fiecare cuvânt al codului;
- regiunea 8 – biți rămași umpluți cu restul împărțirii datelor la o corecție a erorii în 8 biți de cod.

Un alt exemplu de atac care implică codurile QR este cel prin exploatarea caracteristicilor de a fi necitibile de om ci doar de mașini. Acest lucru permite un atac denumit atac QR de tip vector. Datorită faptului că aceste coduri QR nu sunt interpretabile de oameni și sunt, în general, folosite în cadrul sistemelor automate în care alte produse software procesează datele stocate în codurile QR într-un mod automat, acestea pot induce breșe de securitate pe baza conținutului transmis.

Precum este prezentat în (Kieseberg, et. al., 2010), dacă o aplicație de realitate extinsă sau îmbunătățită se bazează pe adrese web furnizate automat de codurile QR, atunci atacurile care urmăresc schimbarea acestei realități pot fi implementate cu ușurință. În

cazul în care aplicația folosește date senzitive, (Gao, et. al., 2009; Kao, et. al., 2011), acest lucru este și mai periculos deoarece siguranța adreselor decodate nu este garantată.

Datorită faptului că aceste coduri QR furnizează date de intrare pentru sistemele de realitate extinsă, (Gia, et. al., 2012), oricare atac asupra codurilor QR poate afecta aplicațiile de realitate extinsă în modalități care pot altera întregul proces.

Însă o modalitate de protecție a conținutului digital pe bază de coduri QR permite verificarea caracteristicii de autenticitate prin marcarea acestuia cu date care să specifice autorul. Astfel orice material digital care are la bază un cod QR poate fi urmărit până la adevăratul său autor.



Proprietar: RAL, 2015

Autor: Popa Maria

Figura 4.22 – Exemplul unui cod QR și a informației stocată

Aplicațiile de realitate extinsă se pot alimenta prin intermediul link-urilor care determină conținutul virtual ce trebuie afișat. Aceste aplicații își pot modifica ordinea de prezentare a conținutului de o manieră dinamică printr-o cale predefinită citită din cadrul codurilor QR. Dacă aceste coduri QR sunt alterate, atunci și ordinea în care sunt afișate imaginile este modificată, eventual sunt interpus imagini cu un alt conținut afișat în locul celui original, astfel conducând la ideea că autenticitatea este obligatorie atunci când se discută de codurile QR.

O abordare a autenticității unui cod QR, ceea ce include și o verificare a autenticității conținutului digital pe care îl însoțește este aceea de a include o semnătură digitală în cadrul acestuia, (Lorenzi, et. al., 2012), și de a o verifica de fiecare dată când un cod este scanat, prin aceeași modalitate prin care se identifică paginile web false.

În figura 4.23 este prezentată o imagine, conținut digital al unei biblioteci virtuale, care este marcată de un astfel de cod cu răspuns rapid, pe baza căruia se decodifică datele cu privire la proprietarul imaginii. Informația este afișată în timp real cu ajutorul unei aplicații de realitate extinsă, astfel validând sursa de proveniență a imaginii prezentate.

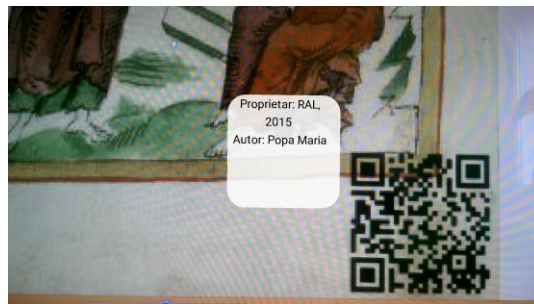


Figura 4.23 – Verificarea autenticității unei imagini

Includerea unei informații în cadrul codurilor cu răspuns rapid care să realizeze în același timp atât o verificare a autenticității cât și o legătură indisolubilă între proprietar și conținutul digital, non-repudiare, reprezintă un mecanism sigur specific semnăturilor digitale. Provocările țin de capacitatea de stocare a codurilor cu răspuns rapid în ceea ce privește semnătura digitală.

5. DISEMINARE

5.1. Soluții originale avansate în lucrarea de cercetare

În vederea inițierii acestui demers științific, cercetătorul se orientează spre o abordare deductivă a problematicii, identificând cadrul general, urmând să propună soluții concrete care să ofere rezultatele scontate pentru îndeplinirea obiectivului propus.

Soluțiile originale expuse în prezenta lucrare de cercetare contribuie la realizarea obiectivului general declarat, cel de a identifica, analiza și implementa soluții de securitate pentru protecția conținutului digital în contextul colaborativ al sistemelor online de biblioteci virtuale din spațiul european.

Contribuțiile autorului sunt prezentate ca parte integrată a fiecărui capitol al lucrării, după cum urmează:

- **SOLUȚII PRIVIND MANAGEMENTUL FLUXULUI DE DATE ÎN SISTEMELE DE BIBLIOTECI VIRTUALE**
 - SO₁ – Analiză asupra modalității de partajare a datelor, în manieră colaborativă, la nivelul sistemelor de biblioteci virtuale;
 - SO₂ – Propunerea unei metodologii de protecție a datelor în cadrul sistemelor integrate de bibliotecă folosind ontologii;
- **SOLUȚII PRIVIND PROTEJAREA CONȚINUTUL DIGITAL LA NIVELUL SISTEMELOR INTEGRATE DE BIBLIOTECĂ**
 - SO₃ – Taxonomii de securitate pentru sistemele integrate de biblioteci virtuale;
 - SO₄ – Soluție pentru clasificarea și ierarhizarea conținutului digital în bibliotecile virtuale;
- **SOLUȚII PRIVIND PROTEJAREA CONȚINUTUL DIGITAL LA NIVELUL SISTEMELOR INTEGRATE DE BIBLIOTECĂ**
 - SO₅ – Soluție pentru asigurarea integrității și confidențialității conținutului digital;
 - SO₆ – Soluție pentru asigurarea autenticității conținutului digital în sistemele integrate de bibliotecă.

În tabelul 5.1 sunt detaliate soluțiile originale dezbătute în capitolele lucrării de cercetare, după cum urmează:

Tabel 5.1 – Descrierea soluțiilor originale

Soluția originală	Definirea soluției	Descrierea
SO₁	Analiză asupra modalității de partajare a datelor, în manieră colaborativă, la nivelul sistemelor de biblioteci virtuale	Partajarea resurselor într-un sistem integrat de bibliotecă este una din principalele funcții pe care acesta le îndeplinește. În cadrul unui sistem integrat de bibliotecă obiectele digitale sunt partajate între mai mulți utilizatori. Sunt prezentate caracteristicile colaborării ca proces de partajare a resurselor precum și principalele aspecte care sunt luate în considerare în cadrul unei biblioteci virtuale.
SO₂	Propunerea unei metodologii de protecție a datelor în cadrul sistemelor integrate de bibliotecă folosind ontologii	Este propusă o metodologie pentru protecția obiectelor digitale din cadrul unei biblioteci virtuale. Metodologia descrie etapele, instrumentele și metodele prin care conținutul digital dintr-un sistem integrat de bibliotecă este protejat de utilizările neautorizate. Aceasta integrează aspecte ale tehnologiei informației prin utilizarea de algoritmi de clasificare supervizată în conjuncție cu analiza semantică pe bază de ontologii și mecanisme de securitate din sfera managementului drepturilor digitale și a steganografiei.
SO₃	Taxonomii de securitate pentru sistemele integrate de biblioteci virtuale	Sunt descrise criteriile pe baza cărora este abordat conținutul digital existent în sistemele integrate de biblioteci virtuale. Pe baza acestor criterii sunt construite nivelurile de importanță ce determină modul în care aparatul de securitate este implementat.
SO₄	Soluție pentru clasificarea și ierarhizarea conținutului digital în bibliotecile virtuale	Se prezintă modalitatea prin care conținutul digital este clasificat și ierarhizat folosind algoritmi de clasificare supervizată precum algoritmul kNN și ontologii lexicale precum, ontologia WordNet.

Soluția originală	Definirea soluției	Descrierea
SO ₅	Soluție pentru asigurarea integrității și confidențialității conținutului digital	Se descrie conceptul de integritate a conținutului digital dintr-o bibliotecă virtuală. Este prezentată o soluție de generare a unor coduri de verificare a integrității și integrare la nivelul obiectelor digitale prin utilizarea unui instrument bazat de înveliș digital.
SO ₆	Soluție pentru asigurarea autenticității conținutului digital în sistemele integrate de bibliotecă	Soluția presupune generarea de cod QR cu elementele bibliografice ale proprietarului și marcarea obiectelor digitale cu acesta prin tehnici de watermarking.

Efortul publicistic al autorului se justifică prin necesitatea de a prezenta soluțiile originale din domeniul securității informatice în strânsă legătură cu domeniul tehnologiei informației, în cadrul revistelor de specialitate și conferințelor internaționale pe teme abordate. Reacțiile primite la prezentările susținute în cadrul conferințelor precum și cele obținute pe baza procesului de recenzie al materialelor publicate în cadrul revistelor de specialitate reflectă în întregime activitatea de publicare.

5.2. Diseminarea în reviste de specialitate și conferințe internaționale

Procesul de diseminare a presupus elaborarea de materiale, în care au fost prezentate soluțiile originale, și susținerea acestora în cadrul unor conferințe internaționale din domeniul cercetării.

După prezentarea direcțiilor de cercetare comunității științifice și primirea unor reacții pozitive în ceea ce privește originalitatea și fezabilitatea obiectivelor declarate, materialele au fost amplu documentate și dezvoltate pentru publicarea în reviste de specialitate cu referenți științifici, cotate în baze de date internaționale.

Publicarea cercetării în reviste de talie internațională reprezintă o etapă importantă în diseminarea rezultatelor.

Activitatea publicistică este reflectată de un număr de 3 articole publicate în reviste cu referenți științifici și 4 materiale diseminate în cadrul conferințelor de specialitate, după cum urmează:

- articole publicate în reviste internaționale cu referenți științifici:
 - Revista de Informatică Economică – (Doinea și Pocatilu, 2014; Doinea, et. al., 2015);
 - Revista Română de Informatică și Automatică – (Doinea, 2015);
- materiale prezentate în cadrul unor conferințe internaționale:
 - Atelierul tematic *Colecțiile de patrimoniu în era digitală* – (Doinea și Dumitrescu, 2014);
 - The 8th International Conference on Security for Information Technology and Communications, SECITC 2015 – (Turcu și Doinea, 2015);
 - The 3rd International Conference on Information Technology and Quantitative Management, ITQM 2015 – (Ivan, et. al., 2015)
 - The 14th International Conference on Informatics in Economy, IE 2015 – (Bătăgan, Boja și Doinea, 2015).

Conferințele internaționale reprezintă cel mai bun mediu de testare a obiectivelor propuse, cu atât mai mult când, cercetarea avansează soluții originale care necesită validare din partea cercetătorilor din domeniu înainte de a fi trimisă spre publicare jurnalelor de referință.

5.3. Stagiul de mobilitate – Universitatea din Catania

În perioada de mobilitate, desfășurată la Universitatea din Catania, Sicilia, Italia, între 25 Mai și 24 Iulie 2015, cercetarea a avut în vedere abordarea tehnicilor de securitate implementate la nivelul managementului drepturilor digitale, concretizându-se în aplicații informatice.

Cercetarea a fost supervizată de profesor universitar Giampaolo Bella, din cadrul departamentului de Matematică și Informatică, sub a cărui îndrumare au fost clarificate aspecte care nu aveau fundament din punct de vedere al necesității.

De asemenea, în cadrul mobilității, s-a avut în vedere diseminarea rezultatelor obținute pe parcursul întregii perioade de proiect în reviste de specialitate și conferințe internaționale pe domeniul proiectului.

Pe parcursul perioadei de mobilitate au fost analizate sistemele de bibliotecă implementate și mecanismele prin care acestea furnizează informații către alte entități externe care rulează în modul de culegere de informații.

În cele două luni de mobilitate au fost publicate două materiale în domeniul securității informatice și analizei datelor de mari dimensiuni, după cum urmează:

- analiza datelor de mari dimensiuni presupune determinarea principalelor caracteristici ale conceptului de Big Data precum și a necesarului de resurse pentru a gestiona un asemenea volum de date; în cadrul acestui concept se ridică problematica fundamentării deciziilor, având la bază un astfel de volum mare de date;
- propunerea unui sistem bazat pe coduri cu răspuns imediat pentru a valida rapid obiectele digitale din punct de vedere al autenticității; se propune o aplicație mobilă cu extensii pe bază de realitate adăugată și coduri cu răspuns rapid care să verifice dacă resursele provin de la un anumit furnizor de conținut digital.

Interacțiunea cu spațiul european a condus la actualizarea temei de cercetare pe baza tendințelor trasate de cercetători cu experiență din universitatea gazdă, în spiritul bunelor practici europene în materie de sisteme de biblioteci virtuale și securitate informatică.

5.4. Direcții viitoare de cercetare

Abordarea orientată pe sisteme integrate de bibliotecă la nivelul cărora este analizat conținutul digital din perspectiva securității drepturilor digitale pe care utilizatorii le dețin, oferă o multitudine de direcții cu impact pozitiv asupra modului în care sunt gestionate resursele. Astfel, în cadrul unui sistem integrat de bibliotecă se pune problema automatizării unor procese care țin de:

- analiza resurselor digitale din punct de vedere al accesului electronic; instrumentele specifice analizei de tip data mining și sistemelor suport de decizie au capacitatea de a analiza modalitatea în care resursele digitale sunt accesate și de a fundamenta decizii; aceste decizii pot reprezenta propuneri de achiziție a unor materiale solicitate și inexistente sau propuneri de arhivare a unora care nu prezintă interes din partea grupului țintă căruia se adresează, în felul acesta optimizând timpul de răspuns la cererile de regăsire lansate de către utilizatori;
- managementul drepturilor digitale la nivelul resurselor electronice cu care sistemele integrate de bibliotecă operează; domeniul managementului drepturilor digitale rămâne unul care oferă soluții pentru protecția conținutului digital, însă acestea necesită o adaptare pentru a se putea implementa în mediile de lucru reale; în zona sistemelor integrate de bibliotecă există un deficit pe partea de securitate la nivelul resurselor digitale expuse utilizatorilor sau partajate prin mecanisme colaborative între diferiți actanți.

Sistemele integrate de bibliotecă se dezvoltă constant și integrează din ce în ce mai multe tehnologii în arhitectura proprie pentru a spori satisfacția utilizatorilor și pentru a optimiza fluxul resurselor în cadrul unei biblioteci. Automatizarea a cât mai multe procese într-un astfel de sistem reprezintă un avantaj major, minimizând resursele implicate și maximizând satisfacția utilizatorilor.

6. CONCLUZII

Demersul privind protecția conținutului digital în cadrul unei biblioteci virtuale al cărui principal rol este de a publica și partaja resursele digitale cu mai multe instituții culturale, precum proiectul Europeana, are în vedere facilitarea accesului la informații pentru toți cetățenii interesați, de o manieră sigură și lipsită de vulnerabilități.

Partajarea resurselor între mai multe instituții culturale de tip bibliotecă virtuale ale spațiului european își găsește justificarea în necesitatea de a crea un patrimoniu durabil care să cuprindă forme digitale de cultură europeană, ușor accesibilă cititorilor acestei zone. Modelul propus se bazează pe standarde actuale bine definite care permit partajarea într-un singur sens a informației digitale.

Într-o astfel de bibliotecă europeană, informații de o importanță culturală deosebită sunt agregate prin coeziunea mai multor sisteme informatice de bibliotecă sub o singură entitate virtuală care să promoveze sloganul Uniunii Europene, „Unitate în diversitate”. Complexitatea rezultată în urma acestui proces reprezintă o premisă pentru a modela mecanisme de securitate existente cu scopul de a fi integrate în contextul sistemelor online de bibliotecă virtuale pentru a proteja patrimoniul cultural european.

Bibliotecile virtuale sunt entități care joacă un important rol atât în viața cotidiană a utilizatorilor uzuali cât și în domeniile de cercetare și mediul academic unde, prin intermediul numeroaselor instrumente care gestionează conținutul digital, oferă suport prin materiale digitale expuse, crescând calitatea proceselor de documentare.

Din acest considerent, toate obiectele digitale gestionate de un sistem integrat de bibliotecă are nevoie de mecanisme de protecție pentru a asigura calitatea sursele și pentru a păstra intactă moștenirea culturală pe care aceste sisteme o promovează. Sistemele integrate de bibliotecă au evoluat, ajungând în acest moment să-și desfășoare procesele în cloud, mediu care favorizează alte tipuri de vulnerabilități în materie de conținut digital. La acest tip de sistem, conținutul digital este cel care necesită toată atenția, în jurul lui desfășurându-se toate procesele. Protecția resurselor digitale reprezintă obiectivul sistemelor de management al drepturilor digitale, însă acestea necesită particularizări majore pentru a putea fi integrate în cadrul unei biblioteci virtuale.

Datorită faptului că sistemele integrate de bibliotecă sunt construite din foarte multe componente și utilizează instrumente multiple pentru a asigura gestiunea corectă a conținutului digital, un nou sistem de securitate necesar protejării resurselor expuse ar îngreuna și mai mult procesele de bază. În felul acesta este surprinsă necesitatea unui proces de optimizare care să permită implementarea tehnicilor de securitate fără însă a scădea eficiența sistemului.

Cercetarea propune o abordare a securității sistemelor de biblioteci virtuale din perspectiva drepturilor digitale asupra conținutului publicat și partajat online. În acest sens este propus un sistem care analizează documentele digitale expuse utilizatorilor, identifică domeniile cele mai expuse din punct de vedere al importanței, definită în termeni de număr de accesări, descărcări și referințe, analiză semantică și sunt identificate metode de combatere a preluărilor și utilizării neautorizate de către un utilizator rău intenționat.

Soluțiile expuse în prezenta lucrare care vizează securitatea conținutului digital gestionat la nivelul bibliotecilor virtuale se bazează pe instrumente de securitate dezvoltate pentru a garanta integritatea, confidențialitatea și autenticitatea resurselor publicate online.

Această abordare este menită să sprijine idea persistenței culturale prin utilizarea sistemelor informaționale și a sistemelor de securitate. Patrimoniul cultural în variantă digitală, expus prin intermediul bibliotecilor virtuale, trebuie păstrat intact, fără a fi denaturat de manipulări neautorizate din partea utilizatorilor rău intenționați.

BIBLIOGRAFIE

Agosti, M., Conlan, O., Ferro, N., Hampson, C., Munnelly, G., Ponchia, C. și Silvello, G. 2014. Enriching digital cultural heritage collections via annotations: The CULTURA approach, *Conference Proceedings of The 22nd Italian Symposium on Advanced Database Systems*, SEBD 2014, 16 – 18 Iunie, pp. 319-326

Agusta, G.M., Hulliyah, K., Arini, R. și Bahaweres, B. 2011. QR code augmented reality tracking with merging on conventional marker based backpropagation neural network, *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 245–248

AWS Products and Services (2015) - Global Compute, Storage, Database, Analytics, Mobile, Application and Deployment Services, Disponibil la: <http://aws.amazon.com/products/?sc_icampaign=ha_en_WhatIsAWS>

Azuma, R.T. 1997. A survey of augmented reality, *Presence*, 6(4), pp. 355-385

Bătăgan L. 2014, The Role of Open Government Data in Urban Areas Development, *Informatica Economică*, 18(2), Disponibil la: <<http://www.revistaie.ase.ro/content/70/08%20-%20Batagan.pdf>>

Batagan, L., Boja, C. și Doinea, M. 2015. Consumer rights in the context of Big and open data society, *Proceedings of the 14th International Conference on Informatics in Economy, IE 2014*, Aprilie 30 – Mai 03, 2015, București, Romania, ASE Printing House

Bechtold, S. 2004. Digital Rights Management in the United States and Europe, *The American Journal of Comparative Law*, 52(2), pp. 323-382

Blanchard, E, Harzallah, M, Briand, H și Kuntz, P. 2005. A typology of ontology-based semantic measures, *Proceeding of EMOI-INTEROP 05*, Portugal

Borda, M. 2011. Fundamentals in Information Theory and Coding, *Springer Publisher*, p. 491

Boyd-Graber, J, Fellbaum, C, Osherson, D și Schapire, R. 2005. Adding Dense, Weighted Connections to WordNet, Disponibil la: <<https://wordnet.princeton.edu/wordnet/publications/jbj-jejufellbaum.pdf>>

Boyd-Graber, J., Fellbaum, C., Osherson, D. și Schapire, R. 2005. „Adding Dense, Weighted Connections to WordNet”, Disponibil la: <<https://wordnet.princeton.edu/wordnet/publications/jbj-jejufellbaum.pdf>>

Breeding, M. 2009. Open Source Library Automation: Overview and Perspective, *Library Technology Reports*, 44 (8), pp. 5–10

Buhanitzky, A și Hirst, G. 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness, *Journal Computational Linguistics*, 32(1), pp. 13-47

Burileanu, D., Dan, C., Sima, M. și Burileanu, C. 1999. A Parser-Based Text Preprocessing For Romanian Language TTS Synthesis, *Proc. EUROSPEECH-99*, pg. 2063-2066, Budapest, Hungary

Burileanu, D., Dan, C., Sima, M. și Burileanu, C.A. 1999. Parser-Based Text Preprocessing For Romanian Language TTS Synthesis, *Proc. EUROSPEECH-99*, Budapest, Hungary, pg. 2063-2066

Carminati, B. și Ferrari, E. 2009. Trust-based information sharing in collaborative communities: Issues and challenges. s.l., *Vieweg+Teubner*, pp. 83-92

Chen, Y.S., Hung, Y.P., Yen, T.F. și Fuh, C.S. 2007. Fast and versatile algorithm for nearest neighbor search based on a lower bound tree, *Pattern Recognition*, pp. 360-375

Cohen, J.E. 2003. DRM and privacy, *Communications of the ACM*, April 2003, 46(4), pp. 46-49

Commission of the European Communities COM(2009) 278 final of 18 June 2009 on Internet of Things — An action plan for Europe. [online]. Disponibil la: <<http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0278:FIN:EN:PDF>>

Constantinescu, R., 2013. Interoperability Solutions for E-Government Services, In INFOREC, *Proceedings of the 12th International Conference on INFORMATICS in ECONOMY (IE 2013)*, Bucharest, Romania, 25-28 April 2013, Bucharest, ASE Publishing House

Cooper, A. și Martin, A. 2006. Towards an open, trusted digital rights management platform In Proceedings of the ACM Workshop On Digital Rights Management, DRM'06. Co-located with the *13th ACM Conference on Computer and Communications Security*, CCS'06, pp. 79-88

Damjanović, B. și Simić, D. 2013. Performance evaluation of AES algorithm under linux operating system, *Proceedings of the Romanian Academy Series A - Mathematics Physics Technical Sciences Information Science*, 14(2), pp. 177-183.

Davies, T., Farhan, H., Alonso, J., Rao, B. și Iglesias, C., 2013. Open Data Barometer, 2013 Global Report – ODI(Open Data Institute) conceptual framework , Disponibil la: <<http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>>

Davies, T., Perini, F. și Alonso, J.M. 2013. Researching the emerging impacts of open data, ODDC (Open Data in Developing Countries) conceptual framework, [online]. Disponibil la: <<http://www.opendataresearch.org/sites/default/files/posts/>>

Demurjian, S., Ren, H., Berhe, S., Devineni, M., Vegad, S. și Polineni, K. 2009. Improving the information security of collaborative Web portals via fine-grained role-based access control, *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*. IGI Global, pp. 430-448

Dilpreet, S. și Chandan, R. 2014. A survey on platforms for big data analytics, *Journal of Big Data*, vol. 2, 20 pg.

Dinu, V., Schileru, I. și Atanase, A. 2012. Attitude of Romanian consumers related to products' ecological labelling, *Amfiteatru Economic*, XIV (31), pp. 8-24.

Doinea, M. și Dumitrescu, F. 2014. Tehnici de securitate pe baza de ontologii în sistemele de biblioteci virtuale, *Workshopul Colecțiile de patrimoniu în era digitală*, 24-25 noiembrie 2014

Doinea, M. și Pocatilu, P. 2014. Security of Heterogeneous Content in Cloud Based Library Information Systems Using an Ontology Based Approach, *Revista Informatica Economica*, 18(4), pp. 101-110

Doinea, M. și Van Osch, W. 2010. Collaborative Systems: Defining and Measuring Quality Characteristics. *Journal of Applied Collaborative Systems*, 2(1), pp. 50-61

Doinea, M., 2015. Securitatea sistemelor de biblioteci virtuale în contextul colaborativ al spațiului european, *Revista Română de Informatică și Automatică*, 25(3), pp. 37-42

Doinea, M., Boja, C., Batagan, L., Toma, C. și Popa, M. 2015 Internet of Things Based Systems for Food Safety Management, *Revista Informatica Economica*, 19(1), pp. 87-97

Don, E. 2004. *Research Methodology in Applied Economics*, 2nd ed., Blackwell Publishing, USA, 2004, p. 248

Elberichi, Z., Rahmoun, A și Bentaalah, M.A. 2008. Using WordNet for Text Categorization, *The International Arab Journal of Information Technology*, 5(1), pp. 16-24

Enderton, H.B. 2001. A mathematical Introduction to Logic, CA: *Academic Press*, p. 295

English, Larry P. 2009. *Information Quality Applied*, Wiley Publishing, 2009, p. 840

EU, 2013. EU implementation of G8 Open Data Charter, [pdf]. Brussels: EU. Disponibil la: <<http://ec.europa.eu/digital-agenda/en/news/eu-implementation>>

Filip, F. G. 2012. Decision-Making Perspective for Designing and Building Information Systems, *International Journal of Comput Commun*, 7(2), pp. 264-272

- Filip, F.G. 2001. A cybernetic model of computerization of the cultural heritage, *Computer Science Journal of Moldova*, 9(2), pp. 101-112
- Filip, F.G. Economia culturii si bibliotecile digitale, *iTOP Business*, 6 (595), XIV/2006, p. 10
- Filip, F.G. și Cojocaru, I. 2010. Economy of Culture in the Information Society Based on Knowledge, *Science and Education Policies in Central and Eastern Europe, Balkans, Caucasus and Baltic Countries*, UNESCO, Italy, vol. 7, pp. 110-120
- Filip, F.G. și Herrera-Viedma, E. 2014. Big Data in the European Union, *The Bridge*, 44(4), pp. 33-37
- G8UK, 2013. G8 Open Data Charter, [pdf]. Disponibil la: <http://www.diplomatie.gouv.fr/fr/IMG/pdf/Open_Data_Charter_FINAL_10_June_2013_cle4a3a4b.pdf>
- Gao, J., Kulkarni, V., Ranavat, H. și Hsing Mei, L.C. 2009. A 2D barcode-based mobile payment system, *Third International Conference on Multimedia and Ubiquitous Engineering*, pp. 320–329
- Garrison, G., Kim, S. și Wakefield, R. 2012. Success Factors for Deploying Cloud Computing, *Communications of the ACM*, 55(9), pp. 62-68
- Goebert, S., Harriehausen-Mühlbauer, B. și Furnell, S., 2014. Towards a unified OAI-PMH registry. s.l., *Society for Imaging Science and Technology*, pp. 97-100
- Gollub, T., Hagen, M., Michel, M. și Stein, B. 2013. From keywords to keyqueries: Content descriptors for the Web, *SIGIR 2013, Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 981-984
- Gonzalez, A, Rigau, G și Castillo, M. 2012. A graph-based method to improve WordNet Domains, *Proceeding CICLing'12 Proceedings of the 13th international conference on Computational Linguistics and Intelligent Text Processing*, vol 1, pp. 17-28
- Google Cloud Platform – Google Developers, Disponibil la: <<https://developers.google.com/cloud/>>
- Hessami, E., Mahmoudi, F. și Jadidinejad, H. 2011. Unsupervised Graph-based Word Sense Disambiguation Using lexical relation of WordNet, *International Journal of Computer Science Issues*, 8(3), pp. 225-230
- Houssos, N., Stamatis, K., Koutsourakis, P., Kapidakis, S., Garoufallou, E. și Koulouris, A. 2014. Enhanced oai-pmh services for metadata sharing in heterogeneous environments, *Library Review*, 63(6-7), pp. 465-489
- Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*, [online]. Disponibil la: <<http://www.loc.gov/z3950/agency/>>

Ivan, I., Zamfiroiu, A., Doinea, M. și Despa, M. 2015. Assigning Weights for Quality Software Metrics Aggregation, *Procedia Computer Science*, pp. 586-592, Information Technology and Quantitative Management Conference (ITQM 2015), Brazil, 2015

Janssen, M., Charalabidis, Y. și Zuiderwijk, A. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, [ejournal], 29(4), pp. 258-268, Disponibil la: <<http://www.tandfonline.com/doi/full/10.1080/10580530.2012.716740>>

Jones, A. 2007. A framework for the management of information security risks, *BT Technology Journal*, 25(1), pp. 30-36

Jonker, W. și Linnartz, J.-P. 2004. Digital rights management in consumer electronics products, *IEEE Signal Processing Magazine*, 21(2), pp. 82 - 91

Jrme, E.și Pavel, S. 2010. Ontology matching, *Springer Publishing*, pp. 334

Kalker T, Samtani R și Wang X. 2012 UltraViolet: Redefining the movie Industry?, *IEEE Multimedia*, pp. 7–11

Kao, Y., Luo, G., Lin, H., Huang, Y. și Yuani, S. 2011. Physical access control based on QR code, *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 285–288

Kieseberg, P., Leithner, M., Mulazzani, M., Munroe, L., Schrittwieser, S., Sinha, M. și Weippl. E. 2010. Qr code security, *Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia*, MoMM '10, pp. 430–435

Kim, S.N. și Baldwin, T. 2005. Automatic Interpretation of Noun Compounds using WordNet Similarity, *Natural Language Processing – IJCNLP*, vol. 3651, pp. 945-956

Kolbe, D., Zhu, Q. și Pramanik, S. 2010. Reducing non-determinism of k-NN searching in non-ordered discrete data space, *Information Processing Letters*, pp. 420-423

Kubesch, A.S. și Wicker, S. 2015. Digital rights management: The cost to consumers, *Proceedings of the IEEE*, 103(5), 1 May 2015, pp. 726-733

Kumar, R. 2005. *Research methodology*, SAGE publications, p. 332

Kundra, V. 2011. Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect, [pdf] Harvard: Harvard University Press, Disponibil la: <http://www.hks.harvard.edu/presspol/publications/papers/discussion_papers/d70_kundra.pdf>

Lagoze, C., Van De Sompel, H., Nelson, M. și Warner, S., 2015. Open Archives. [Interactiv], Disponibil la: <<https://www.openarchives.org/OAI/openarchivesprotocol.html>>

Lee, H., Park, S., Seo, C. și Shin, S.U. 2015. DRM cloud framework to support heterogeneous digital rights management systems, *Multimedia Tools and Applications*, 16 May 2015, p. 21

Lee, H., Seo, C. și Shin, S.U. 2013. DRM Cloud Architecture and Service Scenario for Content Protection, *Journal of Internet Services and Information Security*, 3(34), pp. 94-105

Lin, F și Sandkuhl, K. 2008. A Survey of Exploiting WordNet in Ontology Matching, *IFIP International Federation for Information Processing*, Vol. 276, Artificial Intelligence and Practice II, Boston, Springer, pp. 341-350

Lorenzi, D., Shafiq, B., Vaidya, J., Nabi, G., Chun, S. și Atluri, V. 2012. Using QR codes for enhancing the scope of digital government services, *Proceedings of the 13th Annual International Conference on Digital Government Research*, pp. 21-29

McCarty, D, Koeling, R, Weeds, J. și Carroll, J. 2004. Finding Predominant Word Senses in Untagged Text, *Proceeding ACL '04 Proceedings of the 42th Annual Meeting on Association for Computational Linguistics*, Disponibil la: <http://www2.denizyuret.com/ref/mccarthy_d/95_pdf_2-col.pdf>

Mehmood, K., Afzal, M., Mukaram Khan, M. și Waseemqbal, M.M. 2015. A practical approach to impede key recovery and piracy in Digital Rights Management System, *Proceedings of 2015 12th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2015*, National Centre for Physics Islamabad; Pakistan; 13 – 17 January 2015, pp. 349-353

Mell, P. și Grance, T. 2011. The NIST definition of cloud computing, *NIST*

Mican, D., Tomai, N. și Cocos, R. 2009. Web Content Management Systems, a Collaborative Environment in the Information Society, *Revista Informatică Economică*, 13(2), pp. 20-31.

Moga, L.M., Constantin, D.L. și Antohi, V.M. 2012. A Regional Approach of the Information Technology Adoption in the Romanian Agricultural Farms, *Informatica Economică*, 16(4), pp. 29-36

Morar, G.A., Muntean, C.I. și Silaghi, G.C. 2011. Implementing and Running a Workflow Application on Cloud Resources, *Informatica Economica*, 15(3), pp. 15-27

Open Archives Initiative Protocol for Metadata Harvesting, Protocol de comunicare open source, [online]. Disponibil la: <<http://www.openarchives.org/pmh/>>

Open Knowledge Foundation, 2006. Open Knowledge Definition. [online]. Disponibil la: <<http://opendefinition.org>>

Organization for Economic Co-operation and Development, 2007. OECD

Principles and Guidelines for Access to Research Data from Public Funding, Retrieved January 25, 2014 [online]. Disponibil la: <<http://www.oecd.org/sti/sci-tech/38500813.pdf>>

Passos, A. și Wainer, J. 2009. Wordnet-based metrics do not seem to help document clustering, Disponibil la: <http://www.ic.unicamp.br/~tachard/docs/wnccluster.pdf>

Pedersen, T., Patwardhan, S. și Michelizzi, J. 2004. WordNet::Similarity – Measuring the Relatednes of Concepts, *Proceedings HLT-NAACL – Demonstrations*, pp. 38-41

Pocatilu, P., Alecu, F. și Vetrici, M. 2010. Measuring the Efficiency of Cloud Computing for E-learning Systems, *WSEAS Transactions on Computers*, 9(1), pp. 42-51

Popa, S.C., Avornicului, M.C. și Besfelean, V. P. 2013. Using AMDD method for Database Design in Mobile Cloud Computing Systems, *Informatica Economica*, 17(1), pp. 27-39

Rouillard, J. și Laroussi, M. 2008. Perzoovasive: contextual pervasive QR codes as tool to provide an adaptive learning support, *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, CSTST '08*, pp. 542–548

Samuelson, P. 2003. DRM {and, or, vs.} the law, *Communications of the ACM - Digital rights management*, 46(4), April 2003, pp. 41-45

Sánchez, M., Jiménez, B., Gutiérrez, F.L., Paderewski, P. și Isla, J.L. 2009. Access control model for collaborative business processes, *Engineering the User Interface: From Research to Practice*. Springer London, pp. 117-132

Search/Retrieve via URL, Disponibil la: <http://www.loc.gov/standards/sru/>

Shavitt, Y. și Tankel, T. 2004. Big-Bang Simulation for Embedding Network Distances in Euclidean Space, *IEEE/ACM Transactions on Networking*, 12(6), pp. 993-1006

Standard ISO, 1998/2014. Information and documentation -- Information retrieval (Z39.50) -- Application service definition and protocol specification. s.l.:ISO/TC 46/SC

Stefanoiu, D., Borne, P., Popescu, D., Filip, F.G., Abdelkader, E.K. 2014. *Optimization in Engineering Sciences; Approximate and Metaheuristic Methods*, J. Wiley, London, p. 446

Stone, J. și Partridge, C. 2000. When the CRC and TCP checksum disagree. Stockholm

Synergy Research Group, Microsoft Cloud Revenues Leap; Amazon is Still Way Out in Front [online]. Disponibil la: <<https://www.srgresearch.com/articles/microsoft->

cloud-revenues-leap-amazon-still-way-out-front≥, 29 October 2014

Tao, J., Ma, J., Keranen, M., Mayo, J., Shene, C. și Wang, C. 2014. RSAvisual: A visualization tool for the RSA cipher. s.l., *Association for Computing Machinery*, pp. 635-640.

Tene, O. și Polonetsky, J. 2012. Big data for all: Privacy and user control in the age of analytics, *Northwestern Journal of Intellectual Property*, 11(5), pp. 239–273.

Tinholt, D. 2013., The Open Data Economy Unlocking Economic Value by Opening Government and Public Data, [online]. Capgemini Consulting, Disponibil la: <http://www.capgemini-consulting.com/resource-fileaccess/resource/pdf/opendata_pov_6feb.pdf>

Trausan-Matu, S. 2004. Inteligenta artificiala, Disponibil la: <<http://www.racai.ro/~trausan/ia.pdf>>

Turcu, P. și Doinea, M. 2015. Security Concerns of QR Codes Powered By AR Techniques, *Proceedings of the 8th International Conference on Security for Information Technology and Communications (SECITC'15)*, Iunie 11-12, 2015, Bucharest, Romania, ASE Printing House

UltraViolet Website, 2015, [online], Disponibil la: <<http://www.uvvu.com>>

Van Krevelen, D.W.F. și Poelman, R. 2010. A survey of augmented reality technologies, applications and limitations, *International Journal of Virtual Reality*, 9(2)

Van Tassel, J.M. 2006. Digital Rights Management, *Focal Publisher*, p. 263

Vickery, B. 2009. Information history - an introduction: exploring an emergent field, *Journal of Librarianship and Information Science*, 41(4), pp. 245-246

Welcome to the UltraViolet Wiki System Specification, 2014. Disponibil la: <http://www.uvvuwiki.com/images/3/3f/System-1.1r1.pdf> (2014)

Whitman, M.E. și Mattord, H.J. 2011. Principles of Information Security 4 Edition, *Course Technology*, p. 617

Willett, P. 2006. The Porter stemming algorithm: then and now, *Electronic Library and Information Systems*, 40(3), pp. 219-223

Witten, I.H., Frank, E. și Hall, M.A. 2011. Data Mining: Practical Machine Learning Tools and Techniques – 3rd edition, *Elsevier Publishing*, p. 629

WordNet. A lexical database for English, [online], 2015, Disponibil la: <<http://wordnet.princeton.edu/wordnet/related-projects/>>

Yeluri, R. și Castro-Leon, E. 2014. Building the Infrastructure for Cloud Security, *Apress*

Zhou, F., Duh, H.B.L. și Billinghamurst, M. 2008. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR, *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, pp. 193-202

Exlibris Primo, Prezentarea produsului Primo, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/PrimoOverview>

Exlibris MetaLib, Prezentarea produsului MetaLib, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/MetaLibOverview>

Exlibris Rosetta, Prezentarea produsului Rosetta, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/RosettaOverview>

Exlibris Aleph, Prezentarea produsului ILS Aleph, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/ILSOverview>

Exlibris DigiTool, Prezentarea produsului DigiTool, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/DigiToolOverview>

Exlibris ALMA, Prezentarea produsului ALMA, 2015, [online] Disponibil la: <http://www.exlibrisgroup.com/category/AlmaOverview>

LIBISnet, 2015, [online] Disponibil la: <http://www.libisnet.be/>

Apache Hadoop, 2015, [online] Disponibil la: <https://hadoop.apache.org/>

Anexa 1 – Lista de acronime

Acronim	Descriere
AR	Augmented Reality
CMS	Content Management Systems
CPU	Central Processing Unit
DM	Data mining
DRM	Digital Rights Management
GPU	Graphical Processing Unit
IaaS	Infrastructure as a Service
ILS	Integrated Library Systems
IoT	Internet of Things
kNN	k Nearest Neighbour
MIF	Metadata Interoperability Framework
NER	Name Entity Recognition
NLP	Natural Language Processing
OAI	Open Archives Initiative
OCR	Optical Character Recognition
ODI	Open Data Institute
OECD	Organization for Economic Cooperation and Development
OKF	Open Knowledge Foundation
OPAC	Online Public Access Catalog
PaaS	Platform as a Service
PMH	Protocol for Metadata Harvesting
QR	Quick Response
RELS	Rights Expression Languages
REST	Representational State Transfer
SaaS	Software as a Service
SWOT	Strengths, Weaknesses, Opportunities & Threats
URDD	Unified Resource Discovery and Delivery
URL	Uniform Resource Locator
URM	Unified Resource Management
VM	Virtual Machine
XML	Extensible Markup Language
XrML	Extensible rights Markup Language

Anexa 2 – Lista de figuri

FIGURA 2.1 – CONTRIBUȚIA ILS ÎN FORMAREA DE CUNOȘȚINȚE-----	12
FIGURA 2.2 – ACTANȚII ȘI OPERAȚIILE LA NIVELUL UNUI ILS-----	14
FIGURA 2.3 – ARHITECTURA UNUI SISTEM DE BIBLIOTECĂ -----	16
FIGURA 2.4 – STRUCTURA DEPOZITULUI DE DATE DIN PERSPECTIVA OAI-PMH-----	20
FIGURA 2.5 – COMPONENTELE MODELULUI COLABORATIV ȘI CARACTERISTICILE FIECĂRUI GRUP -----	21
FIGURA 2.6 – ETAPELE PROCESULUI DE DIGITIZARE A CONȚINUTULUI DIGITAL -----	22
FIGURA 2.7 – DIAGRAMA SWOT A MODELULUI COLABORATIV LA NIVELUL BIBLIOTECILOR VIRTUALE -----	26
FIGURA 2.8 – ETAPELE METODOLOGIE DE PROTECȚIE A CONȚINUTULUI DIGITAL -----	28
FIGURA 2.9 – COMPONENTELE PROCESULUI DE CLASIFICARE -----	29
FIGURA 3.1 – IMPACTUL DATELOR CU ACCES LIBER -----	36
FIGURA 3.2 – CICLUL DATELOR CU ACCES LIBER -----	36
FIGURA 3.3 – DOMENII IMPORTANTE ALE DATELOR CU ACCES LIBER -----	37
FIGURA 3.4 – COMPROMISUL ÎNTRE TIMP ȘI CALITATEA REZULTATELOR -----	40
FIGURA 3.5 – UTILIZAREA ONTOLOGIILOR ÎN SISTEMELE ILS-----	41
FIGURA 3.6 – CLASIFICAREA OBIECTELOR DIGITALE -----	42
FIGURA 3.7 – NIVELURILE DE SECURITATE LA NIVELUL CONȚINUTULUI DIGITAL -----	44
FIGURA 3.8 – SCHEMA GENERALĂ A ETAPEI DE PROCESARE LINGVISTICĂ-----	46
FIGURA 3.9 – REPREZENTAREA ARBORESCENTĂ DIN ONTOLOGIA WORDNET SPRE NODUL COMPACT -----	48
FIGURA 3.10 – ELEMENTELE COMPONENT ALE UNUI CONCEPT DIN CADRUL ONTOLOGIEI WORDNET-----	49
FIGURA 3.11 – SENSURI CONCEPT „COUNTRY” ÎN ONTOLOGIA WORDNET-----	50
FIGURA 3.12 – EXEMPLU RULARE ALGORITM PORTER ȘI WORDNET LEMMATIZER-----	51
FIGURA 3.13 – CALCULUL GRADULUI DE SIMILITUDINE, DISTANȚĂ, DINTRE DOUĂ FRAGMENTE TEXT -----	56
FIGURA 3.14 – MATRICEA DISTANȚELOR DINTRE CONCEPTELE REZULTATE ÎN URMA PREPROCESĂRII -----	57
FIGURA 4.1 – METODE DE PROTECȚIE A CONȚINUTULUI DIGITAL UTILIZÂND DRM-----	62
FIGURA 4.2 – ROLUL CONTROALELOR DE SECURITATE DIN CADRUL UNUI SISTEM-----	64
FIGURA 4.3 – ETAPELE PROCESULUI DE PROTECȚIE A CONȚINUTULUI DIGITAL -----	64
FIGURA 4.4 – MODEL DE PROTECȚIE A RESURSELOR BAZAT PE DRM ÎN BIBLIOTECI VIRTUALE -----	67
FIGURA 4.5 – MECANISME DE ASCUNDE A DATELOR -----	68
FIGURA 4.6 – (A) MODELUL RGB DE REPREZENTARE (B) IMAGINE A 4 PUNCTE -----	69
FIGURA 4.7 – DESCRIEREA STRUCTURII UNUI FIȘIER BMP -----	70
FIGURA 4.8 – STRUCTURA UNUI FIȘIER BMP ÎN FORMAT HEXAZECIMAL-----	70
FIGURA 4.9 – SCRIEREA CODULUI DE CONTROL LA NIVELUL UNUI OBIECT DIGITAL DE TIP IMAGINE-----	71
FIGURA 4.10 – EXTRAGEREA CODULUI DE CONTROL DINTR-O IMAGINE MODIFICATĂ-----	72
FIGURA 4.11 – DECODIFICAREA UNEI IMAGINI CU UN ÎNVELIȘ DIGITAL-----	72
FIGURA 4.12 – COD DE BARE UNIVERSAL-----	74
FIGURA 4.13 – EXEMPLU DE QR CODE-----	75
FIGURA 4.14 – STRUCTURA UNUI COD QR -----	76

FIGURA 4.15 – RECLAME CU CODURI QR-----	76
FIGURA 4.16 – ANUNȚ DE VÂNZARE CU CODURI QR -----	77
FIGURA 4.17 – MODUL DE SCANARE AL UNEI APLICAȚII QR-----	78
FIGURA 4.18 – MENIUL CONTEXTUAL AL APLICAȚIEI-----	78
FIGURA 4.19 – OPȚIUNILE CAMEREI FOTOGRAFICE -----	78
FIGURA 4.20 – EXEMPLU DE REALITATE EXTINSĂ-----	79
FIGURA 4.21 – SECȚIUNILE UNUI COD QR VERSIUNEA 2-----	80
FIGURA 4.22 – EXEMPLUL UNUI COD QR ȘI A INFORMAȚIEI STOCATĂ -----	81
FIGURA 4.23 – VERIFICAREA AUTENTICITĂȚII UNEI IMAGINI -----	82

Anexa 3 – Lista de tabele

TABEL 3.1 – BENEFICIILE ECONOMICE ALE DATELOR CU ACCES LIBER _____	35
TABEL 3.2 – CLASIFICARE ÎN FUNCȚIE DE UTILIZAREA DATELOR CU ACCES LIBER _____	37
TABEL 3.3 – VARIABILELE FOLOSITE ÎN CADRUL CLASIFICĂRII SUPERVIZATE KNN ASUPRA CARACTERISTICILOR NUMERICE _____	52
TABEL 3.4 – CALCULUL DISTANȚEI DINTRE DOUĂ PUNCTE REPREZENTATE ÎNTR-UN SPAȚIU ORTOGONAL N-DIMENSIONAL _____	55
TABEL 3.5 – VARIABILELE CLASIFICĂRII SUPERVIZATE KNN ASUPRA CARACTERISTICILOR TEXT _____	57
TABEL 4.1 – CAPACITATEA MAXIMĂ A UNUI COD QR _____	75
TABEL 5.1 – DESCRIEREA SOLUȚIILOR ORIGINALE _____	84