

# Sistem de regăsire a informației muzicale. Noțiuni teoretice și aplicații

Adrian Simion

Universitatea POLITEHNICA București

Catedra de Calculatoare a Facultății de Automatică și Calculatoare

Splaiul Independenței, nr. 313, 060042, București, România

simion.adrian@gmail.com

## REZUMAT

În această lucrare sunt analizate unele probleme specifice domeniului regăsirii informației muzicale. De asemenea sunt prezentate noțiuni introductive precum și câteva aplicații practice din domeniu. Multe din noțiunile discutate precum și unele abordări în rezolvarea problemelor își vor găsi aplicarea într-un program dezvoltat de autor, detaliat spre sfârșitul lucrării. Ultima parte își propune să tragă câteva concluzii despre componentele dificile ale MIR (*Music Information Retrieval*) și despre viitorul acestui domeniu.

## Cuvinte cheie

Regăsirea informației muzicale, MIR, Sistem Informațional Multimedia.

## Clasificare ACM

H.5.1 Multimedia Information Systems

H.5.5 Sound and Music Computing

## INTRODUCERE. REGĂSIREA INFORMAȚIEI MUZICALE

Imaginați-vă o lume în care cineva ar deschide calculatorul și ar cânta fragmentul de melodie ce-i rămăsese în minte de la dejun. Calculatorul acceptă modalitatea de introducere (chiar dacă fragmentul este în altă gamă sau cu note false pe alocuri), corectează cererea și sugerează prompt faptul că „Roadhouse Blues” este cauza iritării. Persoana confirmă sugestia calculatorului de a asculta unul din multele fișiere MP3 pe care le-a găsit și respinge oferta de a asculta toate versiunile cântecului, ce includ un remix rap italian sau o orchestratie formată dintr-un flaut și un cimpoi. Un astfel de sistem nu există în momentul de față. Cercetările în domeniu ne confirmă că în viitor vom avea ceva similar, dar majoritatea cercetătorilor au decis ca acest lucru nu va fi ușor de pus în practică.

Canțitatea de muzică digitală este într-o creștere continuă, promovată de interesul utilizatorilor și de dezvoltarea noilor tehnologii pentru plăcerea de a asculta muzică.

Sunt câteva motive ce explică acest trend, în primul rând, caracteristica esențială a limbajului muzical. Muzica este o formă de artă ce poate fi împărtășită de oameni ce aparțin unor culturi diferite deoarece depășește barierele limbajului național și al fondului cultural. Ca exemplu muzica din vestul Americii are mulți entuziaști în Japonia, și multe persoane din Europa apreciază muzica clasică indiană. Aceștia pot aprecia muzica fără nevoia unei

traduceri, ce este în mod normal necesară pentru accesarea unor lucrări textuale străine.

Un alt motiv este faptul că tehnologia pentru înregistrarea muzicală, transformarea digitală și redare permite utilizatorilor accesul la o informație care este aproape comparabilă cu înregistrarea unei performanțe live, cel puțin la nivelul calității audio, raportul semnal/zgomot fiind mai bun pentru unele formate digitale față de multe formate analogice.

În cele din urmă, muzica este o formă de artă ce poate fi cultă și populară în același timp, iar câteodată este imposibil de a trage o linie între cele două, cum ar fi jazz-ul și muzica tradițională.

Music Information Retrieval (MIR) este un domeniu interdisciplinar ce își are originea în domenii conexe: muzicologie, psihologie cognitivă, lingvistică și nu în ultimul rând știința calculatoarelor. MIR propune un număr mare de metode diferite față de regăsirea informației textuale, scopul principal fiind gestionarea informațiilor muzicale, acces ușor și în cele din urmă plăcerea de a asculta muzică.

Cea mai mare parte a cercetării făcute pe MIR, sau tehnicile propuse, și sistemele dezvoltate se bazează pe conținut. Ideea principală a abordărilor bazate pe conținut este că un document poate fi descris de o serie de caracteristici ce sunt calculate direct din conținutul său.

Tehnicile primare de regăsire a informației (IR), ce se bazează pe statistică și teoria probabilității, pot fi aplicate și în afara cazului textual, deoarece modelele descriu caracteristici fundamentale ce sunt împărtășite de diferite medii, limbaje și domenii de aplicare. Din această cauză, rezultatele obținute în IR, în particular în cazul documentelor text, pot fi folosite ca referințe pentru metodele MIR.

## EXPUNEREA PROBLEMEI

MIR s-a confruntat de la început cu o dificultate primară[3][4]: nu a existat o metodă clară pentru echipele de cercetare de a compara științific metodele și rezultatele.

Acest lucru se întâmpla deoarece nu există :

- o colecție standard de muzică pe care echipele să-și poată testa metodele,
- un set standardizat de sarcini ce vizează performanța,
- un model standardizat pentru metrice de evaluare.

Comunitatea MIR a recunoscut de timpuriu nevoia unei paradigme mai riguroase și comprehensive de evaluare.[2] Pentru regăsirea de informație textuală a fost dezvoltată o paradigmă de testare și evaluare. În această paradigmă fiecare echipă de regăsire de informație textuală are acces la:

- o colecție de text, de o mărime impresionantă, standardizată,
- un set standardizat de interogări pentru teste,
- o evaluare standardizată a rezultatelor pe care fiecare echipă le generează.

Din cauza asemănărilor dintre MIR și IR-ul tradițional, a fost sugerat ca cercetătorii MIR să exploreze modelul TREC [20] și ca acesta să devină o componentă cheie în evaluarea MIR.

Luând ca exemplu competiția MIREX pentru algoritmi aplicabili în acest domeniu, sarcinile pot fi grupate în: clasificare, extragerea caracteristicilor, aliniere și regăsire.

• Clasificare:

– Identificarea audio a artiștilor. Asocierea înregistrărilor audio cu etichete ce identifică artistul ce a creat înregistrarea.

– Clasificare audio/simbolică a genurilor. Determinarea genului muzical pentru date de tip audio sau MIDI.

• Extragerea caracteristicilor:

– Extragerea liniei melodice. Extragerea liniei melodice principale din date audio polifonice. Acest lucru implică două subsarcini: detectarea vocilor (ce decide dacă un anumit cadru de timp conține o “înălțime melodică”) și detectarea înălțimii notelor (decizându-se care este înălțimea cea mai probabilă pentru un anumit cadru de timp).

– Detectarea “atacului”. Listarea notelor și a “atacului” din înregistrările audio.

– Detectarea ritmului (percuție – tobe). Determinarea evenimentelor și corespondența numelor de clase pentru percuție în muzica polifonică; spre deosebire de sarcina precedentă, detectarea doar a tobelor (instrumentelor de percuție) în locul detectării fiecărei note, și de asemenea analizarea clasei din care toba face parte.

– Extragerea tempo-ului. Determinarea tempoului pentru înregistrările audio.

– Determinarea gamei, a tonice și a notelor dominante pentru o înregistrare sau pentru un fișier audio.

– Urmărirea măsurii. Spre deosebire de detecția ritmului sau a tempo-ului, problema este determinarea măsurilor într-o înregistrare audio

– Similaritatea muzicală și regăsirea. Calcularea unei matrice de similaritate pentru o listă dată de fișiere audio.

– Interogare prin prezentarea parțială a temei muzicale (fredonare). Dându-se o interogare, să se caute într-o colecție piesele ce conțin un material muzical similar.

## NOȚIUNI FUNDAMENTALE DESPRE MUZICĂ ȘI DATE AUDIO

### Ce este sunetul?

Sunetul este un fenomen complex ce implică noțiuni de fizică și percepție. Probabil că cel mai simplu mod de a îl defini este prin cel puțin trei lucruri : ceva se mișcă, ceva transmite rezultatele acestei mișcări, ceva (sau cineva) aude rezultatele mișcării.

Toate lucrurile care produc sunete se mișcă, și într-un sens foarte metafizic, toate lucrurile care se mișcă (dacă nu se mișcă foarte încet sau foarte repede) produc sunete.

Odată cu mișcarea, lucrurile “împing și trag” aerul înconjurător (sau apă, sau orice fel de mediu ocupă), producând variații de presiune. Aceste variații de presiune sau unde de sunet sunt de fapt ce numim sunet.[11]

Sunetul este reprezentat vizual astfel:

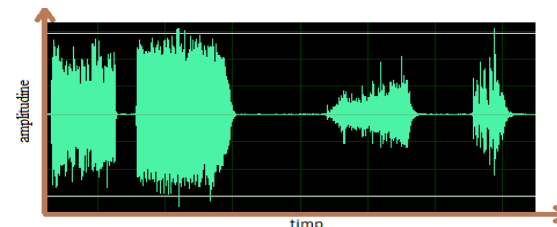


Figura 1. Reprezentare digitală a sunetului

### Cum putem descrie sunetul?

Sunetul poate fi descris în diferite moduri. Avem multe cuvinte pentru sunete și diferite moduri de a vorbi despre acestea. Spunem despre un sunet ca este „macabru”, „întunecat”, „intens” etc. De fapt, limbajul nostru de a vorbi despre sunet este familiar dar din punct de vedere științific, este destul de imprecis.

Pentru a manipula sunetele digitale cu ajutorul unui calculator este util să avem acces la o descriere, astfel trebuie să punem (și să răspundem) următoarele tipuri de întrebări: Cât de tare este, care este înălțimea, care este spectrul, ce frecvențe sunt prezente, cum se modifică sunetul în timp, de unde vine sunetul, care sunt caracteristicile obiectului fizic ce produce sunetul?

## PERCEPȚIA MUZICII ȘI REGĂSIREA ACESTEIA

Aproape toată muzica cunoscută este construită pornind de la note, la fel cum textul este construit din caractere sau cuvinte; notele sunt mult mai apropiate de caractere decât de cuvinte, totuși similaritatea nu este atât de evidentă.

Cei patru parametri ce definesc o notă muzicală sunt următorii:

- înălțime: cât de înalt sau jos este sunetul, analogie perceptuală a frecvenței,
- durată: cât durează o notă,
- intensitate sonoră: analogie perceptuală a amplitudinii,
- timbru sau calitatea tonului.

Oamenii aud muzica într-un mod “non-liniar”: studiile în percepția muzicală și cogniție arată multe aspecte neintuitive și subtile, iar acești parametri nu se pot separa atât de ușor. Pentru a da un exemplu simplu, notele foarte scurte sunt percepute ca având o intensitate sonoră mai

mică decât note identice de o durată mai mare. Atunci când un grup de note se află într-o secvență, o melodie, efectele asupra percepției nu sunt evidente. Spre exemplu, schimbând timbrul se poate transforma o linie melodică simplă în voci multiple și vice-versa. Dacă un sistem MIR ar lucra doar cu o reprezentare înalt-structurată, adică, partituri – aceste efecte n-ar fi o problemă foarte mare. Dar majoritatea sistemelor au nevoie să opereze în alte reprezentări. Mai mult, interogările muzicale se bazează pe amintirile ascultătorului, astfel nu sunt foarte stricte și pot conține erori și efecte cognitive și perceptuale.[1]

### TREC (TEXT RETRIEVAL CONFERENCE)

Importanța delimitării naturii sarcinilor MIR – și a interogărilor asociate – pentru a fi folosite în evaluarea testării este o altă temă semnificativă.

Fiind dat consensul copleșitor pentru stabilirea unei paradigme similare TREC, este totuși interesant de ce nu a fost adoptată o abordare de tip TREC încă. În cadrul conferințelor ISMIR s-au conturat patru tipuri de cerințe ce pot da un răspuns acestei probleme :

1. complexitatea informației muzicale;
2. complexitatea interogărilor;
3. relevanța în contextul MIR și aplicabilitatea preciziei rememorării în metrici de evaluare;
4. lipsa accesului la colecții muzicale / în legătură cu dreptul de proprietate intelectuală practicat de industria muzicală.

Ordinea primelor trei probleme nu este aleatoare. Complexitatea informației muzicale poate fi văzută ca fiind cauza complexității găsite în interogările din lumea reală. Complexitatea interogărilor, la rândul său, contribuie la dificultatea asociată găsirii relevanței.

#### Problema nr.1 : Complexitatea informației muzicale

Complexitatea informației muzicale este mult mai mare decât complexitatea informației textuale. Informația muzicală este un amalgam multilateral format din elemente de înălțime, tempo, ritm, timbru, text, elemente bibliografice și practice.

Muzica poate fi reprezentată prin note, fișiere MIDI și alte codări discrete, și într-un număr mare de formate digitale și analogice (e.g. benzi magnetice, CD-uri, MP3-uri etc). Spre deosebire de majoritatea textelor, muzica este foarte plastică, adică, fiind dată o anumită bucată muzicală, aceasta poate fi transpusă, i se poate schimba ritmul, armonia, orchestrația, versurile etc, și totuși să fie percepută ca “aceeași” piesă muzicală. Interacțiunea complexității muzicale și plasticitatea acesteia fac elementele de regăsire foarte problematice. Se pot sintetiza următoarele probleme:

- până la folosirea unui „repertoriu universal” de muzică, determinarea celor mai „reprezentative” versiuni (și formate) ale obiectelor muzicale pentru folosirea în colecții de test rămâne o problemă deschisă;

- mărimea colecției de test este un motiv de îngrijorare. Deoarece este nevoie de mai multe instanțe ale informației: simbolice, audio, metadata etc, pentru fiecare piesă din colecție, plaja de teste pentru MIR, va ajunge la,

sau chiar depăși, limita centrelor de cercetare. Fișierele audio tind să devină mari, relative la perechile simbolice, acest lucru contribuind în mod semnificativ la această problemă;

- stabilirea și menținerea diferitelor manifestări ale fiecărei lucrări. Trebuie evitat ca o metodă de regăsire să nu fie „privilegiată” față de alta. Acest lucru duce la noțiunea de „neutralitate a regăsirii”;

- interogările muzicale – acestea fiind tot un fel de informație muzicală – sunt de asemenea plastice, complexe și multilaterale. Acest lucru implică faptul că este nevoie de o încapsulare formalizată a interogărilor în „înregistrări de tip interogare” pentru utilizarea într-o paradigmă de tip TREC.

#### Problema nr.2: Complexitatea interogărilor

Pentru complexitatea interogărilor în MIR este un fond de literatură formală destul de mic, iar acesta se preocupă cu analiza informațiilor din lumea reală în legătură cu nevoile și utilizarea MIR/MDL. Această deficiență este cauzată de inexistența sistemelor MIR/MDL care să conțină muzică pe care utilizatorii să o vrea. Totuși, atunci când se încearcă astfel de studii, distanța dintre presupunerile făcute de către cercetătorii MIR despre natura interogărilor muzicale (i.e. fredonarea melodiilor, regăsirea elementelor cunoscute, identificarea cântecelor) și situațiile reale, este remarcabilă. O opinie împărtășită de membrii comunității este că o atenție deosebită trebuie acordată în dezvoltarea unor interogări de tip TREC, deoarece folosirea lor va avea ramificații științifice semnificative, în special în validarea rezultatelor experimentale.

Principiul “neutralității regăsirii” poate fi divizat în două tabere: cea legată de cercetarea în regăsirea simbolică și cea care explorează tehnicile de procesare audio și a semnalelor.

Din perspectiva faptului că nu există date pentru a compara punctele forte și pe cele slabe ale tehnicilor folosite de către cele două tabere, consensul este că paradigma de tip TREC să aducă mijloacele pentru a face evaluări ale meritelor celor două abordări.

Sintetizând sugestiile făcute de către cercetători în legătură cu acest subiect, se pare că o înregistrare pentru o interogare de tip TREC trebuie să includă aceste elemente principale: reprezentări de calitate înaltă, metadata exhaustive (utilizator, nevoi, folosire) și reprezentări simbolice ale muzicii prezentate.

#### Problema nr.3: Relevanță, precizie și rememorare

Comunitatea de IR pentru informația textuală a avut un set standardizat de evaluare a performanței metricilor în ultimele patru decade. De la experimentele lui Cranfield, două metrici au fost proeminente: precizia (i.e. raportul dintre numărul de documente relevante regăsite și numărul de documente regăsite) și rememorarea (i.e. raportul dintre numărul de documente relevante regăsite și numărul de documente relevante prezente în sistem). Aceste metrici stau la baza paradigmei de evaluare TREC. Cheia în folosirea preciziei și rememorării ca metrici este determinarea documentelor ce pot fi “relevante” pentru o anumită interogare. În timp ce au fost dezbateri despre natura „relevanței”, semnificația a fost îndeajuns de

stabilă, astfel încât evaluările TREC să fie posibile. Pentru a simplifica lucrurile, un „document” este „relevant” relativ la o interogare dacă documentul este „despre” același subiect ca și interogarea (i.e. există o intersecție a „înțeleșului” între interogare și document). În contextul evaluării MIR, totuși, această abordare bazată pe „înțeleș” nu este adecvată. Spre exemplu, ce înseamnă de fapt Sonatele pentru Pian ale lui Beethoven, sau solo-urile de chitară ale lui Hendrix? Comunitatea MIR recunoaște această insuficiență. De fapt, definiția „relevanței” în contextul MIR a fost atât de problematică astfel încât precizia și rememorarea se regăsesc foarte rar în literatura MIR. În ciuda acestei probleme, opinia generală spune că cercetările comunității MIR ar trebui să creeze mijloace pentru evaluarea sistemelor MIR în paradigma TREC și astfel să continue examinarea preciziei și a rememorării ca metrici fundamentale. În acest sens, se speră că, dacă se vor face înregistrări pentru interogări cât mai „bogate” în informație, atunci ar putea să apară un standard „persoană rezonabilă” pentru judecarea relevanței elementelor regăsite. Cu alte cuvinte, ar trebui să fie destulă informație conținută în înregistrările interogărilor astfel încât diferite persoane ar putea contribui la aflarea răspunsului dacă un element satisface intenția interogării. Validitatea presupunerii „persoanei rezonabile” ar fi, desigur, subiectul unei verificări empirice.

#### Problema nr.4: Construirea colecțiilor

Construirea colecțiilor și proprietatea intelectuală pot fi costisitoare. În perioada actuală (Post-Napster [13]), cei ce dețin drepturile intelectuale sunt foarte activi pentru a opri distribuția gratuită a muzicii. Schimbările recente în legile de copyright pun sub semnul întrebării chiar existența sursei înregistrărilor „domeniului public”. Acestea sunt motivele ce au oprit dezvoltarea colecțiilor de test, mari, accesibile de către comunitate, ce ar fi cuprins reprezentările audio, simbolice și metadata. Deținătorii de colecții nu le fac accesibile altora din cauza legilor copyright.

În ciuda acestor dificultăți reale, câteva lucruri recente au făcut posibilă începerea construirii bazei de date ce va conține colecția de test. Acest lucru a fost posibil prin convingerea celor ce dețin drepturile de copiere, de faptul ca cercetătorii MIR sunt de încredere și respectă proprietățile intelectuale ale acestora. Asta a însemnat dezvoltarea unor mecanisme prin care proprietatea intelectuală poate fi securizată prin oprirea accesului neautorizat și a distribuției.[4]

#### METODE ȘI SARCINI ALE SISTEMELOR MIR

##### Sisteme MIR

În momentul de față un număr considerabil de unelte și Framework-uri sunt în curs de dezvoltare pentru a combina componente în sisteme ce pot efectua operații cum ar fi extragerea, regăsirea muzicii și evaluarea automată. Multe din aceste unelte sunt Open Source și, astfel, comunitatea ce se ocupa cu cercetarea este încurajată să participe la dezvoltarea lor.

#### Unelte de cercetare în muzică

Această secțiune descrie sumar câteva sisteme de regăsire a informației muzicale. Sistemele sunt împărțite în unelte de cercetare muzicală și unelte ce măsoară performanțele.

Câteva exemple de unelte de cercetare în muzică sunt: MARSYAS [10] - un Framework audio de nivel jos, CLAM - Framework pentru extragerea și sinteza muzicii, OMRAS2 [14] - un Framework distribuit ce permite integrarea componentelor sistemului MIR în unelte software.

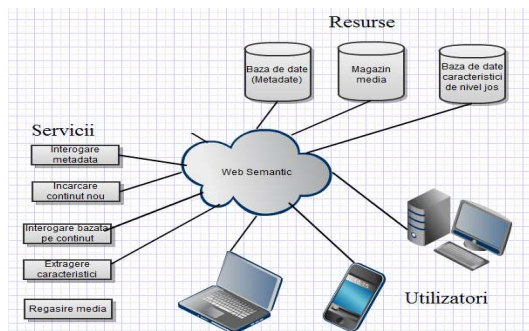


Figura 2. Prezentarea generală a Framework-ului OMRAS2 ce arată folosirea web-ului semantic pentru a face disponibile resurse și servicii cercetătorilor MIR.

#### Metode folosite de sisteme MIR

##### Căutarea datelor simbolice

- Metodele bazate pe șiruri pentru melodiile monofonice.

Muzica monofonică poate fi reprezentată de șiruri unidimensionale, unde fiecare caracter descrie o notă sau o pereche de note consecutive. Șirurile pot reprezenta secvențe de interval, secvențe de înălțimi, și binecunoscuții algoritmi de potrivire cum ar fi algoritmi pentru calcularea distanțelor, găsirii celei mai mari sub-secvențe comune, sau găsirea aparițiilor unui șir în altul. Acești algoritmi au fost aplicați, câteodată cu câteva adaptări pentru a-i face adecvați în găsirea melodiilor.

Următoarea figură ilustrează faptul că există tehnici obișnuite de variație care nu sunt în mod normal suportate de metodele bazate pe șiruri. În mod obișnuit, variația unei melodii este percepută ca fiind similară din punct de vedere muzical, dar conține multe alte note, ceea ce duce la o reprezentare prin șiruri cu multe alte caractere. Construind o metodă de găsire a similarității bazată pe șiruri în cazul anterior fără a produce „false positives” nu este ușor.[2]



Figura 3. Detectarea similarității muzicale folosind metode bazate pe șiruri poate fi dificilă în cazuri cum ar fi acesta

Unele sisteme MIR verifică doar potrivirile exacte sau cazuri în care șirul de căutare este un subșir în intrările din baza de date. Pentru astfel de sarcini, se pot folosi algoritmi standard de căutare cum ar fi Knuth-Morris-Pratt și Boyer-Moore. „Themefinder” [18] caută înregistrări în baza de date folosind expresii regulate. În

acest caz, nu există o noțiune de distanță, dar diferite șiruri pot fi găsite folosind aceeași expresie regulată.

Pentru o potrivire aproximativă, poate fi util să calculăm o distanță de editare cu ajutorul programării dinamice. Musipedia [12] este un exemplu de sistem ce face acest lucru. Simpla calculare a unei distanțe de editare între șirurile de interogare și datele din baza de date nu este de ajuns, totuși, deoarece aceste șiruri pot reprezenta piese muzicale de diferite lungimi. Astfel, poate fi necesar să găsim subșirurile adecvate înaintea calculării distanței de editare.

În indexarea pentru găsirea subșirurilor ce se potrivesc în totalitate se pot folosi metodele standard pentru indexarea textului (spre exemplu, inverted files, B-trees, etc).

- Metode bazate pe colecții pentru muzica polifonică.

Spre deosebire de metodele bazate pe șiruri, metodele bazate pe colecții nu presupun că notele sunt ordonate. Muzica este privită ca un set de evenimente cu proprietăți ca atac, înălțime și durată. Aceste lucruri fac metodele bazate pe colecții adecvate pentru muzica polifonică. „Distance Measures Clausen” propune o metodă de căutare ce vede partiturile și interogările ca o colecție de note.

- Potrivirea probabilistă.

Scopul metodelor de potrivire probabilistă este de a determina proprietățile probabilistice ale pieselor candidat și de a le compara pe acestea cu proprietăți corespundente ale interogării. De exemplu, sistemul GUIDO[6] calculează modele Markov descriind probabilitățile tranzițiilor de stări în piese și apoi comparând matricele ce descriu probabilitățile de tranziție.

Caracteristicile criteriilor de măsurare a distanțelor melodiilor cum ar fi secvențele de interval, de înălțime sau ritm pot fi calculate prin lanțuri Markov. În aceste lanțuri Markov, stările pot corespunde unor caracteristici cum ar fi o anumită înălțime, un anumit interval sau durata unei note, iar probabilitățile de tranziție reflectă numărul de apariții ale diferitelor stări ulterioare.

Similaritățile dintre o interogare și o piesă candidat în baza de date pot fi determinate prin calcularea intersecției probabilităților de tranziție, bazate pe matricea de tranziție a piesei candidat, pentru fiecare pereche de stări consecutive în interogare.

Indexarea matricelor de tranziție ierarhice poate fi organizată ca un arbore. Frunzele sunt matrice de tranziție ale pieselor din baza de date, în timp ce nodurile interioare sunt matricele de tranziție ce descriu concatenarea pieselor din subarbore.

### Căutarea datelor audio

- Extragerea caracteristicilor relevante perceptual.

O metodă naturală de a compara înregistrări audio într-un mod ce are înțeles este de a extrage o descriere abstractă a unui semnal audio ce reflectă aspectele relevante percepțional ale înregistrării, urmate de aplicarea unei funcții de aflare a distanței pentru informația extrasă. O înregistrare audio este în mod normal împărțită în cadre scurte, care se suprapun, iar aceste cadre conțin evenimente singulare. Câteva caracteristici ce sunt extrase

în mod normal din cadre audio cu o durată între 25 și 40 de milisecunde sunt: intensitatea sonoră, înălțimea, nuanța, tonul (claritate și banda de frecvențe) și derivatele. Sistemele de regăsire audio, cum ar fi SuperMBox[7] compară vectorii acestor caracteristici pentru a găsi înregistrări audio ce sună similar unei interogări date.

- Ampretele digitale audio.

În cazul în care scopul nu este de a identifica o anumită lucrare, ci o înregistrare, tehnicile ce au la bază amprente digitale se descurcă foarte bine. Sistemele telefonice pentru identificarea melodiilor cunoscute (e.g. Shazam) folosesc o formă de amprente audio digitale. Aceste amprente audio, de obicei au doar câțiva bites per segment de înregistrare, sunt apoi stocate într-un index al bazei de date, împreună cu referințe către înregistrările unde se regăsesc. Același mod de a extrage caracteristici este folosit pe interogare, și cu ampretele audio ce au fost extrase din interogare. Înregistrările candidat pentru a se potrivi cu înregistrările pot fi extrase foarte ușor. Numărul acestor candidați poate fi redus prin verificarea dacă ampretele au loc în ordinea corectă și la intervalele corecte de timp.

- *Self-Organizing Map*

*Self-Organizing Map* (SOM), un algoritm foarte cunoscut al rețelelor neuronale din categoria învățare nesupervizată, a fost folosit pentru clusterizarea și clasificarea melodiilor similare. O rețea de tip SOM este alcătuită din unități ce sunt ordonate într-un grid de puține dimensiuni (de obicei bidimensional). Un vector model în spațiul înalt-dimensional este asignat fiecărei unități. În timpul antrenării, vectorii model sunt legați de date astfel încât distanțele dintre elementele ce conțin date și vectorii model corespunzătorii sunt minimizezate. Vectorii model pot conține orice caracteristici.[21]

### MOTOR DE CĂUTARE A INFORMAȚIEI MUZICALE ȘI CRAWLER STRUCTURAT

Autorul și-a propus realizarea unui motor de căutare pentru muzică, în acest demers fiind util și un *spider* (*crawler*) pentru a căuta informația necesară indexării din diferite surse (MP3-Blogs, disc dur local, bază de date externă). Pentru extragerea de informație relevantă s-au încercat diferite abordări.

#### Resurse și privire de ansamblu

Aplicația are o interfață web, accesibilă din browser. În cazul unui utilizator simplu, aplicația îi va da posibilitatea de regăsire a melodiilor pe baza unor criterii selectate de acesta: tip, gen/artiști similari, etichetă ID, *bitrate*, an. În cazul în care se dorește administrarea aplicației, se accesează un link separat și se va face o autentificare pe baza de utilizator/parolă. Administratorul are posibilitatea de a schimba diferiții parametri ai aplicației: algoritmi de analiză, reîmprospătarea listei Media RSS, pornire/oprire spider.

Tehnologiile folosite pentru dezvoltare au fost: Java, Javascript, Ajax, Html, Oracle/MySql.

Aplicația conține un *crawler* ce indexează (într-o bază de date MySQL sau Oracle) diferitele informații legate de fișierele găsite prin intermediul *Multimedia Syndication*

RSS. Pe baza unor algoritmi matematici și tehnici de regăsire a modelelor (HMM și PCA), fiecărei melodii i se vor atașa informații în baza de date: tip, un *hash code*, secvențe relevante (pattern-uri).

Pentru aplicarea MVC (*Model View Controller*) în Java s-a folosit o structură de tip *Web-Tier* și componente extinse ale limbajului e.g. J2EE (JSP, JSF, Servlet) precum și diferite biblioteci/Framework-uri.

Partea de *Spider (Crawler)* precum și diferiții parametri ai motorului pot fi accesați de către un utilizator special ce are drepturi depline în aplicație.

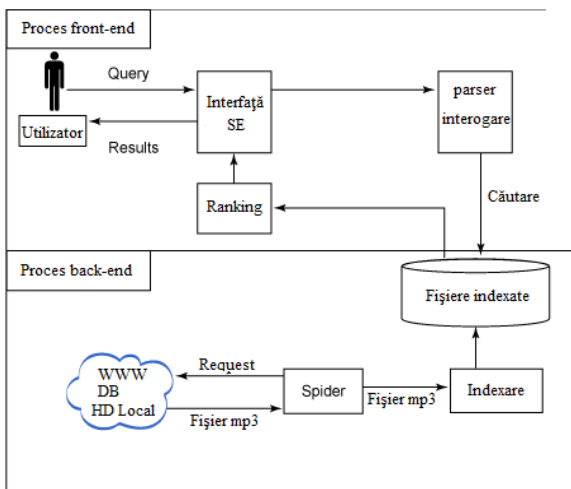


Figura 4. Descrierea sistemului

Sistemul pornește de la o diagramă obișnuită a unui motor de căutare ce este adaptată ulterior la cerințele problemei inițiale.

### Specificații funcționale

#### Informație prelucrată

Spațiul de căutare cuprinde: WEB Logs (MP3-Blogs – Media RSS), disc dur local, Baza de date. Astfel, spider-ul extrage informațiile din sursele anterioare după care aceste informații sunt indexate și persistate. Informațiile extrase sunt: tip, similarități, informație textuală - etichete ID3, cale (web, local).

Tipul fișierelor indexate este mp3 (algoritmul putându-se adapta și altor fișiere audio ulterior – flac, ogg, wav etc).

#### Biblioteci/Framework-uri, clase utilizator și caracteristici

Pentru implementarea aplicației s-a folosit limbajul Java și biblioteci open-source. Astfel, bibliotecile folosite sunt următoarele: JLayer [8] – pentru decodarea MP3-urilor, Encog [5] – o bibliotecă ce folosește diferite tehnici și algoritmi de inteligență artificială, Yahoo MediaRSS [22] – proiectul folosește această bibliotecă pentru a accesa feed-urile de tip Atom/Syndicate din care se pot extrage fișierele audio, Lucene [9] – această bibliotecă de la Apache Foundation pune la dispoziție mecanismele principale folosite de un motor de căutare obișnuit, Tika [18] – este un set de instrumente pentru analiza conținutului de la Apache Software Foundation și împreună cu Lucene pune la dispoziție metode de

investigare al conținutului media, SEAM [17] – un *Framework* dezvoltat de comunitatea JBoss ce combină ultimele tehnologii J2EE.

### Abordări în rezolvarea problemelor MIR

Sistemul dezvoltat încearcă rezolvarea câtorva probleme prezentate anterior.

#### Colecție

Una din problemele principale este existența unei colecții pe care se pot face teste. În rezolvarea acestei probleme am recurs la o extensie a standardului RSS, Media RSS. Cu ajutorul bibliotecilor puse la dispoziție de Yahoo! am reușit, după analizarea paginilor web, găsirea și indexarea unui număr semnificativ de resurse audio.

#### Clasificare audio și similaritate

După identificarea fișierelor mp3 se încearcă decodificarea informației audio cu ajutorul bibliotecii JLayer (Javazoom) după care cu ajutorul acestei informații se încearcă o clusterizare (în principal legată de ritm). Informația audio se transformă în date de training pentru o rețea neuronală de tip SOM (*Self-Organizing Map*) Kohonen. Astfel, fiecare fișier audio va aparține unui anumit cluster, nefiind legat în mod bijectiv de tipul real al melodiei – eventual extras din tag-ul ID3.



Figura 5. Reprezentare grafică a clusterizării audio [16]

În urma analizării clusterului se poate spune că unele melodii aparțin unui gen “comun” putând face chiar propuneri utilizatorului motorului de căutare pentru regăsirea unor artiști “similari”. Pentru completitudinea soluției am folosit și un serviciu web de sugestii ale unor artiști similari (indexat manual pe baza ID3-tag-urilor) EchoNest.

#### Relevanță și regăsire

Factorii ce intervin în algoritmul de calculare a scorului pentru Lucene sunt: *tf (term frequency)* frecvența notei (*pattern-ului*) în melodie, *idf (inverse document frequency)*, *coord* (numărul de termeni din interogare ce au fost folosiți în document), *lengthNorm* (criteriu de măsurare a importanței termenului în funcție de numărul total de termeni existenți), *queryNorm* (factorul de normalizare astfel încât interogările să poată fi comparate), *boost (index)* compensarea câmpului la index-time, *boost (query)* compensarea câmpului la interogare.

Pentru a modifica algoritmul de calcul al scorului este de ajuns să moștenim clasa *DefaultSimilarity* și să facem *override* la metodele pe care vrem să le schimbăm.

## Rezultate

Folosind intrări din surse ca: *East African Tube* am făcut teste legate de varietatea și cantitatea informației muzicale. Astfel... după o zi de indexare automată (conexiune medie internet 1MB/s și luând ca medie 7.5MB pentru un mp3 colecția s-a ridicat la 800 de melodii distincte (cantitate). Varietatea, testată de către trei utilizatori diferiți a demonstrat că cele 800 de melodii cuprind 20 de clase (genuri) printre care Bongo, Hip-hop, Muzică Mauritaneană, Rock, Reggae etc. Prin folosirea MediaRSS am rezolvat principala problemă într-un mod automatizat: lipsa unui set variat și consistent de muzică.

Clusterizarea realizată cu ajutorul rețelelor Kohonen nu a corespuns în proporții acceptabile cu genurile indicate de cei 3 utilizatori sau cu sugestiile EchoNest, acest lucru indicând faptul că genurile muzicale sunt în primul rând subiective și că tehnicile folosite trebuiesc rafinate (Hidden Markov Models și Principal Component Analysis).

Problemele de mai sus se extind și către algoritmul de calcul al scorului pentru Lucene, în cazul folosirii unei metode clasice textuale, sistemul de regăsire a informației textuale dând rezultate bune pentru fișierele ce conțin etichetele ID3 (în acest caz sistemul putând fi interpretat ca o bibliotecă de text – metadata).

Artist	Titlu	Album	Track#	An.	Comentarii	Similar	Cale
Anathema	Alternative 4	Alternative 4	7			My Dying Bride, On Thorns I Lay, Katatonia, Novembers Doom, Paradise Lost	E:\Work\Project\Mp3Repository\Anathema - Alternative 4.mp3
Anathema	Fragile Dreams	Alternative 4	2			My Dying Bride, On Thorns I Lay, Katatonia, Novembers Doom, Paradise Lost	E:\Work\Project\Mp3Repository\Anathema - Fragile Dreams.mp3

Figura 6. Pagina de rezultate

Aplicația răspunde majorității necesităților unui motor de căutare și încearcă o abordare din punct de vedere audio a informației, neuitând de partea textuală conținută de fișierele mp3.

## CONCLUZII

### De ce este dificilă MIR?

#### Provocarea multireprezentatională

Cu excepția aspectului bibliografic, fiecare caracteristică poate fi reprezentată sub formă de simboluri, audio sau sub ambele forme. Reprezentările simbolice includ note imprimate, partituri, text și o miriadă de codări discrete pentru calculator, ce include *Musical Instrument Digital Interface* (MIDI), *Notation Interchange File Format* (NIFF). În figura următoare este prezentat un mod de comunicare a informației muzicale. Deasupra versurilor sunt acordurile ce susțin armonia cântecului, progresia F, C, Bb (Fa major, Do major și Si major bemol).

```

From: XXXXXXXX
Subject: Early 80's - Please identify this song! (it's *very* difficult, though)
Newsgroups: alt.music.lyrics
Date: 2000-12-14 09:42:24 PST

Hi, this is so difficult because I only remember those damn FRAGMENTS of it, which can (in combination with possible errors) make it VERY difficult to identify this song!
OK here I go... (gonna add the chords for you guitarists out there ;)

[verse 1]

F C Bb Bb C
Crazy ..... onto the ..... café

F C Bb
I'm drinking coffee, she came away

F C Bb Bb C
She ordered ..... precious sum of money ???

F C Bb
deedeedeedeedeedeedeede...<remaining text deleted>
    
```

Figura 7. Exemplu de informație muzicală

Reprezentările audio includ concerte live și înregistrări de studio, ambele în format digital sau analog (e.g. LP-uri, MP3-uri, CD-uri și benzi magnetice). Alegerea reprezentărilor, chiar dacă sunt simbolice sau audio, este un amalgam de factori ce includ scopul folosirii sistemelor, resursele de calcul și lățimea de bandă. Reprezentările simbolice tind să necesite mai puține resurse de calcul și lățime de bandă decât reprezentările audio. Spre exemplu, un fragment de 10 secunde de muzică reprezentată în calitate digitală stereofonică de CD are nevoie de aproximativ 14 mega-octeți de date să fie procesată, transmisă și stocată. Sub simplul test al reprezentării simbolice, același eveniment muzical ar putea fi reprezentat prin 8 sau 16 biți. Totuși, din cauză că marea majoritate a ascultătorilor înțeleg muzica doar în formă auditivă, mulți dezvoltatori MIR văd benefica includerea reprezentărilor audio, în ciuda consumării inerente a resurselor, ce sunt absolut necesare. Prezența (sau absența) pragmatică a unor reprezentări influențează deciziile de design. Spre exemplu, mulți cercetători se limitează la utilizarea formatelor MIDI, CD și/sau MP3 deoarece este facilă construirea unor mici colecții de acest tip folosind tehnici de „Web Spidering”. Trebuiesc luate în considerare și dificultățile create de proprietatea intelectuală.

#### Provocarea multiculturală

Informația muzicală este, desigur, multiculturală. Există totuși o înclinare către muzica vestică. Sunt foarte multe stiluri muzicale pentru care codările audio și simbolice nu sunt disponibile, non-standard sau incomplete. Improvizații jazz, muzica electronică, muzica asiatică, sau spectacolele indiene ragas sunt doar câteva exemple. De asemenea, nu avem încă o reprezentare muzicală comprehensivă pentru cântecele tribale africane sau pentru muzica „guturală” a iniuților. Astfel, ar trebui dezvoltate noi modalități de codare. În momentul de față este mult mai eficient să se construiască sisteme bazate pe resurse ce sunt mai ușor de obținut și de manipulat.

#### Provocarea multiexperiențială

Muzica, în cele din urmă, există în mintea celui care o percepe. Astfel, percepția, aprecierea și experiența muzicii va fi diferită nu doar prin prisma multitudinii de minți ce o înțeleg, dar va varia în mintea fiecărui individ în funcție de dispoziție, situație și circumstanțe.

Înregistrările sunt ascultate de mulți ca un *aide-mémoire* pentru a rememora experiențe plăcute. Muzica poate fi experimentată ca o continuare a tradițiilor familiare de a cânta imnuri, colinde, cântece de tabără etc. Unii oameni percep muzica într-un mod religios, ca o formă de expresie sublimă sau extatică. David Huron sugerează că muzica are calitățile unui drog.

Infinita varietate de exprimare prin muzică pune două probleme semnificative pentru dezvoltatorii MIR. În primul rând, ridică problema audienței și a folosirii sistemului. Chiar dacă s-ar putea coda, căuta și regăsi efectele similare drogurilor a unor piese muzicale într-o bază de date ar răspunde un astfel de sistem nevoilor analitice ale unui muzicolog?

În al doilea rând, problema multiexperiențială pune întrebări chiar despre natura relevanței muzicale. În cele mai multe cazuri, noțiunea de similaritate cu scop de regăsire a fost limitată la aspecte melodice, ritmice, armonice și timbrale ale muzicii.

Obiectele muzicale ce au unele intervale, măsură, acorduri și/sau orchestrație în comun sunt considerate „similare” la un anumit nivel, și astfel sunt considerate potențial „relevante” pentru evaluare.

În ce moduri, totuși, putem evalua experiența unui utilizator cu a altuia relative la o piesă din colecție? Care este dispoziția dorită sau efectul psihologic considerat “similar” al unei lucrări muzicale? Poate unele combinații de similarități melodice, ritmice, armonice și timbrale joacă un rol important în similaritatea experiențelor, și cercetările viitoare vor arăta în ce mod.

#### Provocarea multidisciplinară

Viziunea globală eterogenă asupra disciplinelor este problematică. Fiecare disciplină ce contribuie la dezvoltarea sistemelor de regăsire a informației muzicale aduce comunității MIR setul propriu de scopuri, punerea în aplicare a unor metode uzuale, întrebări valide de cercetare și paradigme de evaluare.[15]

#### Viitorul MIR

Metodele ce au fost descrise sunt aproximative și în timp se pot schimba, totuși, acestea furnizează o fundație folositoare. Unii cercetători au sugerat ca procesul de regăsire a muzicii să se facă în două faze. Scopul primei faze este de a regăsi un set larg de potriviri generale, cu precizie mică dar cu regăsire mare. În a doua fază, acest set înseamnă filtrare și potrivirile false sunt eliminate. Sarcina de regăsire este împărțită deoarece caracteristicile folosite pentru a obține o regăsire mare sunt aceleași pentru o precizie mare. Sarcinile de regăsire de mare precizie au de asemenea potențialul de a fi costisitoare din punct de vedere computațional, ceea ce adâncește nevoia de faze intermediare pentru a filtra colecția. Este astfel de mare folos pentru comunitatea de regăsire a informației muzicale (MIR) să aibă în vedere cât mai multe tehnici de extragere posibile și modalități de regăsire.

#### REFERINȚE

1. Byrd, D. și Crawford, T., Problems of Music Information Retrieval in the Real World, Center for Intelligent Information Retrieval, Department of

Computer Science, University of Massachusetts, *Information Processing and Management*, 2002, 249-272

2. Casey, M. A., Veltkamp R., Goto M., Leman M., Rhodes C. și Slaney M., Content Based Information Retrieval: Current Direction and Future Challenges, University of London, *Proceedings of the IEEE*, 2008
3. CLEF <http://clef2011.org/>
4. Downie, J. S., Toward the Scientific Evaluation of Music Information Retrieval Systems, University of Illinois at Urbana-Champaign, *Conference on Music Information Retrieval, (ISMIR)*, 2003
5. ENCOG <http://www.heatonresearch.com/encog>
6. GUIDO [http://en.wikipedia.org/wiki/GUIDO\\_music\\_notation](http://en.wikipedia.org/wiki/GUIDO_music_notation)
7. Jang J. R., Lee H. și Chen J., Super MBox: An Efficient/Effective Content-based Music Retrieval System, Multimedia Information Retrieval Laboratory, C.S. Department, National Tsing Hua University, Taiwan, *Proceedings of the ninth ACM international conference on Multimedia*, 2001, 1-2
8. JLAYER <http://www.javazoom.net/javalayer/sources.html>
9. LUCENE <http://lucene.apache.org/java/docs/index.html>
10. MARSYAS <http://www.marsyas.info/>
11. Muller M., Information Retrieval for Music and Motion, Meinard Müller Institut für Informatik III Universität Bonn Römerstr, Springer (2007)
12. MUSIPEDIA <http://www.musipedia.org/>
13. NAPSTER <http://music.napster.com/>
14. OMRAS2 <http://www.omras2.org/>
15. Pickens J., A Survey of Feature Selection Techniques for Music Information Retrieval, Center for Intelligent Information Retrieval, *Symposium on Music Information Retrieval (ISMIR)*, 2001
16. Rauber A., Pampalk. E și Merkl D., Content-based Music Indexing and Organization, Austrian Research Institute for AI, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 2002
17. SEAM <http://seamframework.org/>
18. THEMETFINDER <http://www.themefinder.org/>
19. TIKA <http://tika.apache.org/>
20. TREC <http://trec.nist.gov/>
21. Typke R., Wiering F. și Veltkamp R. C., A Survey of Music Information Retrieval Systems, Universiteit Utrecht, 2005, 1-8
22. Y!MEDIA-RSS <http://groups.yahoo.com/group/rss-media/>