



UNIVERSITATEA DIN CRAIOVA
FACULTATEA DE AUTOMATICĂ, CALCULATOARE ȘI
ELECTRONICĂ
DEPARTAMENTUL DE CALCULATOARE ȘI TEHNOLOGIA
INFORMAȚIEI



PROIECT DE DIPLOMĂ

Vladu Octavian-Mihai

COORDONATOR ȘTIINȚIFIC

Conf. dr. ing. Mihăescu Cristian

iulie 2015

CRAIOVA



UNIVERSITATEA DIN CRAIOVA
FACULTATEA DE AUTOMATICĂ, CALCULATOARE ȘI
ELECTRONICĂ
DEPARTAMENTUL DE CALCULATOARE ȘI TEHNOLOGIA
INFORMAȚIEI



Dezvoltarea unui Educational Data Benchmark

Vladu Octavian-Mihai

COORDONATOR ȘTIINȚIFIC

Conf. dr. ing. Mihăescu Cristian

iulie 2015

CRAIOVA

„Există doar 3 culori, 10 cifre și 7 note. Ceea ce facem cu ele este important.”

Jim Rohn

DECLARAȚIE DE ORIGINALITATE

Subsemnatul Vladu Octavian-Mihai, student la specializarea Calculatoare Română din cadrul Facultății de Automatică, Calculatoare și Electronică a Universității din Craiova, certific prin prezenta că am luat la cunoștință de cele prezentate mai jos și că îmi asum, în acest context, originalitatea proiectului meu de licență:

- cu titlul Dezvoltarea unui Educational Data Benchmark,
- coordonată de Conf. dr. ing. Mihăescu Cristian,
- prezentată în sesiunea iulie 2015.

La elaborarea proiectului de licență, se consideră plagiat una dintre următoarele acțiuni:

- reproducerea exactă a cuvintelor unui alt autor, dintr-o altă lucrare, în limba română sau prin traducere dintr-o altă limbă, dacă se omit ghilimele și referința precisă,
- redarea cu alte cuvinte, reformularea prin cuvinte proprii sau rezumarea ideilor din alte lucrări, dacă nu se indică sursa bibliografică,
- prezentarea unor date experimentale obținute sau a unor aplicații realizate de alți autori fără menționarea corectă a acestor surse,
- însușirea totală sau parțială a unei lucrări în care regulile de mai sus sunt respectate, dar care are alt autor.

Pentru evitarea acestor situații neplăcute se recomandă:

- plasarea între ghilimele a citatelor directe și indicarea referinței într-o listă corespunzătoare la sfârșitul lucrării,
- indicarea în text a reformulării unei idei, opinii sau teorii și corespunzător în lista de referințe a sursei originale de la care s-a făcut preluarea,
- precizarea sursei de la care s-au preluat date experimentale, descrieri tehnice, figuri, imagini, statistici, tabele et caetera,
- precizarea referințelor poate fi omisă dacă se folosesc informații sau teorii arhicunoscute, a căror paternitate este unanim cunoscută și acceptată.

Data,

Semnătura candidatului,



UNIVERSITATEA DIN CRAIOVA
Facultatea de Automatică, Calculatoare și Electronică
Departamentul de Calculatoare și Tehnologia Informației

Aprobat la data de
.....
Șef de departament,
Prof. dr. ing.
Marius BREZOVAN

PROIECTUL DE DIPLOMĂ

| | |
|--|---|
| Numele și prenumele studentului: | Vladu Octavian-Mihai |
| Enunțul temei: | Dezvoltarea unui Educational Data Benchmark |
| Datele de pornire: | |
| Conținutul proiectului: | |
| Material grafic obligatoriu: | |
| Consultații: | |
| Conducătorul științific (titlul, nume și prenume, semnătura): | Conf. dr. ing. Mihăescu Cristian |
| Data eliberării temei: | |
| Termenul estimat de predare a proiectului: | |
| Data predării proiectului de către student și semnătura acestuia: | |



UNIVERSITATEA DIN CRAIOVA
Facultatea de Automatică, Calculatoare și Electronică

Departamentul de Calculatoare și Tehnologia Informației

REFERATUL CONDUCĂTORULUI ȘTIINȚIFIC

Numele și prenumele candidatului: Vladu Octavian-Mihai
Specializarea: Calculatoare Română
Titlul proiectului: Dezvoltarea unui Educational Data Benchmark

Locația în care s-a realizat practica de documentare (se bifează una sau mai multe din opțiunile din dreapta):
În facultate ☐
În producție ☐
În cercetare ☐
Altă locație:

În urma analizei lucrării candidatului au fost constatate următoarele:

| | | | | | |
|-----------------------------|--|--|--|---|---|
| Nivelul documentării | | Insuficient <input type="checkbox"/> | Satisfăcător <input type="checkbox"/> | Bine <input type="checkbox"/> | Foarte bine <input type="checkbox"/> |
| Tipul proiectului | | Cercetare <input type="checkbox"/> | Proiectare <input type="checkbox"/> | Realizare practică <input type="checkbox"/> | Altul [se detaliază] |
| Aparatul matematic utilizat | | Simplu <input type="checkbox"/> | Mediu <input type="checkbox"/> | Complex <input type="checkbox"/> | Absent <input type="checkbox"/> |
| Utilitate | | Contract de cercetare <input type="checkbox"/> | Cercetare internă <input type="checkbox"/> | Utilare <input type="checkbox"/> | Altul [se detaliază] |
| Redactarea lucrării | | Insuficient <input type="checkbox"/> | Satisfăcător <input type="checkbox"/> | Bine <input type="checkbox"/> | Foarte bine <input type="checkbox"/> |
| Partea grafică, desene | | Insuficientă <input type="checkbox"/> | Satisfăcătoare <input type="checkbox"/> | Bună <input type="checkbox"/> | Foarte bună <input type="checkbox"/> |
| Realizarea practică | Contribuția autorului | Insuficientă <input type="checkbox"/> | Satisfăcătoare <input type="checkbox"/> | Mare <input type="checkbox"/> | Foarte mare <input type="checkbox"/> |
| | Complexitatea temei | Simplă <input type="checkbox"/> | Medie <input type="checkbox"/> | Mare <input type="checkbox"/> | Complexă <input type="checkbox"/> |
| | Analiza cerințelor | Insuficient <input type="checkbox"/> | Satisfăcător <input type="checkbox"/> | Bine <input type="checkbox"/> | Foarte bine <input type="checkbox"/> |
| | Arhitectura | Simplă <input type="checkbox"/> | Medie <input type="checkbox"/> | Mare <input type="checkbox"/> | Complexă <input type="checkbox"/> |
| | Întocmirea specificațiilor funcționale | Insuficientă <input type="checkbox"/> | Satisfăcătoare <input type="checkbox"/> | Bună <input type="checkbox"/> | Foarte bună <input type="checkbox"/> |

| | | | | | |
|--------------------------------|---------------|--|--|---|---|
| | Implementarea | Insuficientă <input type="checkbox"/> | Satisfăcătoare <input type="checkbox"/> | Bună <input type="checkbox"/> | Foarte bună <input type="checkbox"/> |
| | Testarea | Insuficientă <input type="checkbox"/> | Satisfăcătoare <input type="checkbox"/> | Bună <input type="checkbox"/> | Foarte bună <input type="checkbox"/> |
| | Funcționarea | Da <input type="checkbox"/> | Parțială <input type="checkbox"/> | Nu <input type="checkbox"/> | |
| Rezultate experimentale | | Experiment propriu <input type="checkbox"/> | | Preluare din bibliografie <input type="checkbox"/> | |
| Bibliografie | | Cărți | Reviste | Articole | Referințe web |
| Comentarii și observații | | | | | |

În concluzie, se propune:

| | |
|---|---|
| ADMITEREA PROIECTULUI <input type="checkbox"/> | RESPINGEREA PROIECTULUI <input type="checkbox"/> |
|---|---|

Data,

Semnătura conducătorului științific,

REZUMATUL PROIECTULUI

Proiectul de față are rolul de crea o funcționalitate care să permită efectuarea unei analize a datelor dintr-o platformă online de e-learning. În cadrul prezentului proiect, platforma pentru care a fost dezvoltată această funcționalitate este platforma TESYS, platformă deținută de Fundația "Constantin Belea" din Craiova. Demararea acestui proiect a venit din necesitatea implementării acestei funcționalități în cadrul platformei TESYS, dar și datorită lipsei unui înmagazinării a datelor din acest domeniu.

Pentru asigurarea mediului de dezvoltare a fost folosit modulul TDM (Tesys Data Manager) din cadrul platformei Tesys. Acesta are rolul de administrare a datelor existente în platforma Tesys. Această aplicație este dezvoltată în Java de către contributorii cu care deținătorii platformei au lucrat de-a lungul timpului.

Noțiunile necesare realizării acestui proiect reprezintă o gamă relativ largă despre programare în limbajul Java și folosirea bazelor de date.

Fiind un proiect bazat atât pe parte de cercetare cât și pe parte de dezvoltare, s-au întâmpinat probleme în găsirea unei mulțimi de caracteristici care să fie relevante în legătură cu contextul platformelor de e-learning. Aceste caracteristici au fost fie studiate în cărțile și lucrările de specialitate, fie dezvoltate împreună cu profesorul coordonator. În urma studiului și a cercetării efectuate s-a ajuns la o mulțime ce conține 28 de caracteristici, din care momentan doar 17 pot fi calculate.

Cu ajutorul acestor caracteristici se poate face analiza datelor într-o platformă de e-learning așa cum va fi prezentat în continuarea acestui document.

Termenii cheie: *analiza, date, java, mysql, educatie, e-learning, tesys, uci, machine, learning, cercetare, dezvoltare.*

CUPRINSUL

| | | |
|----------|--|-----------|
| 1 | INTRODUCERE | 10 |
| 1.1 | SCOPUL..... | 10 |
| 1.2 | MOTIVAȚIA..... | 11 |
| 1.3 | ENUNȚUL TEMEI..... | 12 |
| 1.4 | OBIECTIVE | 13 |
| 2 | MEDIUL DE DEZVOLTARE ȘI TEHNOLOGII FOLOSITE..... | 14 |
| 2.1 | XML | 14 |
| 2.2 | MySQL | 15 |
| 2.3 | XAMPP..... | 16 |
| 2.4 | JAVA..... | 17 |
| 2.5 | ECLIPSE | 18 |
| 2.6 | JDBC | 19 |
| 2.7 | HIBERNATE | 19 |
| 3 | STRUCTURA EDUCATIONAL DATA BENCHMARK | 21 |
| 3.1 | CERCETĂRI SIMILARE | 21 |
| 3.1.1 | <i>UCI Data Benchmark – Performanțe studenți.....</i> | <i>21</i> |
| 3.1.2 | <i>PSLC DataShop.....</i> | <i>22</i> |
| 3.2 | CONTRIBUȚII PERSONALE | 23 |
| 3.2.1 | <i>Sintetizare caracteristici.....</i> | <i>23</i> |
| 3.2.2 | <i>Asociere caracteristici cu datele brute ale platformei Tesys.....</i> | <i>26</i> |
| 3.2.3 | <i>Dezvoltarea aplicației pentru obținerea Educational Data Benchmark-ului</i> | <i>41</i> |
| 4 | DEZVOLTAREA APLICAȚIEI | 44 |
| 4.1 | ARHITECTURA SOFTWARE ÎN DETALIU | 44 |
| 4.1.1 | <i>Clase necesare transferului de date.....</i> | <i>44</i> |
| 4.1.2 | <i>Clase necesare configurării</i> | <i>47</i> |
| 4.1.3 | <i>Clase necesare exportului</i> | <i>51</i> |
| 4.2 | MANUALUL DEZVOLTATORULUI..... | 54 |
| 4.3 | MANUALUL UTILIZATORULUI | 57 |
| 4.4 | REZULTATE OBȚINUTE | 62 |
| 5 | CONCLUZII | 64 |
| 6 | BIBLIOGRAFIE | 65 |

| | | |
|---|--------------------|----|
| 7 | REFERINȚE WEB..... | 66 |
|---|--------------------|----|

1 INTRODUCERE

1.1 Scopul

Fundația "Constantin Belea" din Craiova deține o platformă de e-learning intitulată TESYS care a fost dezvoltată de-a lungul timpului cu ajutorul membrilor Fundației, dar și prin contribuția studenților de la Facultatea de Automatică și Calculatoare a Universității din Craiova.

Această platformă este funcțională în prezent și rulează în beneficiul Universității din Craiova facilitând învățământul la distanță. Platforma rulează în prezent pentru mai multe facultăți ale Universității din Craiova astfel că numărul utilizatorilor care folosește această aplicație este unul ridicat.

În tot acest timp, platforma TESYS nu a beneficiat de o funcționalitate prin care să se realizeze o analiză a datelor asupra utilizatorilor săi care să ajute pe viitor la îmbunătățirea funcționalităților pe care aplicația le oferă în favoarea unei utilizări mai facile din punct de vedere al utilizatorului.

Totodată, în cadrul UCI Irwin Machine Learning Repository s-a constatat că nu există înmagazinate date cu privire la activitățile desfășurate în cadrul platformelor de e-learning. UCI Irwin Machine Learning Repository reprezintă o colecție de date, teorii și generatoare de date care sunt folosite de către comunitatea științifică pentru analiza empirică a îmbunătățirii algoritmilor de machine learning și pattern recognition.

Având în vedere aceste două aspecte s-a urmărit implementarea unei funcționalități care să preia datele aferente platformei Tesys, să le analizeze și să obțină niște caracteristici individuale ale unui utilizator în legătură cu diferite aspecte legate de utilizarea acestei aplicații. Totodată, datele obținute au menirea de-a fi colectate, documentate și donate sub forma unui set de date către UCI Irwin Machine Learning Repository.

Cu alte cuvinte, dezvoltarea și implementarea unei astfel de tehnologii are să aducă avantaje atât Fundației "Constantin Belea", deținătoare a platformei Tesys, cât și comunității științifice care dorește îmbunătățirea algorimilor de machine learning și pattern recognition.

În cadrul platformei Tesys, pe lângă aplicația propriu-zisă care reprezintă platforma online, există dezvoltată o aplicație desktop care are rolul de administrare a datelor aferente aplicației online. Această aplicație este intitulată Tesys Data Manager (TDM) și are rolul de a crea un mediu mai lejer și mai prietenos de administrare a datelor decât ar face-o aplicație online.

Astfel, funcționalitatea care va permite analiza datelor aferente platformei Tesys va fi implementată în cadrul modului amintit anterior având rolul de a extrage anumite caracteristici despre diverși utilizatori ai aplicației online și stocarea acestora sub diferite forme (fișiere .csv, .xml sau .arff).

În urma rulării acestei tehnologii asupra unui context oferit de platforma Tesys, datele obținute trebuie documentate și clasificate conform cu normele UCI Irwin Machine Learning Repository și donate către aceasta comunitate științifică.

În consecință scopul acestui proiect este unul dublu. În primul rând îmbunătățirea platformei de e-learning Tesys prin analiza comportamentului utilizatorilor, iar cel de-al doilea îl reprezintă donarea setului de date obținut către UCI Irwin Machine Learning Repository.

1.2 Motivația

Începând cu toamna anului 2014 am fost cooptat de către îndrumătorul meu în echipa de contributori ai modului TDM. Cu timpul m-am obișnuit cu rolul îndeplinit atât de modulul la a cărui dezvoltare urma să contribui cât și cu informații despre întreaga platformă Tesys.

Odată cu acomodarea în ceea ce privește modulul TDM, împreună cu îndrumătorul meu am început să contribui la această aplicație desktop, reușind într-un final să implementez câteva funcționalități.

Având în vedere informațiile expuse în prezentarea scopului acestui proiect, deci necesitatea unei funcționalități în cadrul platformei Tesys care să realizeze analiza datelor stocate de această aplicație, dar și faptul că modulul care se ocupă cu administrarea datelor acestei platforme este aplicația desktop TDM, iar eu sunt în momentul de față contribuitor al acestei aplicații, domnul Conf. Dr. Ing. Mihăescu Cristian mi-a propus dezvoltarea acestei uneelte în cadrul curentului proiect de licență.

Din dorința de a realiza un proiect de licență care să fie util în producție am acceptat propunerea domnului profesor, astfel luându-mi angajamentul de a implementa funcționalitatea cerută în TDM, dar și buna documentare a muncii făcute pentru a realiza acest lucru.

Pe parcursul dezvoltării acestui proiect am avut satisfacția de a dobîndi lucruri noi datorită documentării pe care trebuia să o realizez pentru a atinge finalizarea proiectului. Totodată am întâmpinat și dificultăți în acest proces datorită faptului că trebuia să întocmesc o listă de caracteristici care odată calculate să ofere date despre un anumit utilizator. Aceasta listă a fost până la urmă definitivată odată cu cercetarea diferitelor cărți și lucrări din domeniu, dar și cu ajutorul îndrumatorului.

În prezent, această mulțime de caracteristici conține 28 de elemente care vor fi expuse și documentate în prezentul proiect.

1.3 Enunțul temei

Tema acestui proiect de licență o reprezintă dezvoltarea unei funcționalități în cadrul unei aplicații desktop care să preia date dintr-o baza de date aferente unei platforme online de e-learning, să calculeze aceste caracteristici pe baza datelor preluate și să expună administratorului rezultatele obținute cu ajutorul anumitor tipuri de fișiere folosite pentru a analiza date.

În urma cercetării făcute s-a definitivat o mulțime de caracteristici relevante în legătură cu modul de utilizare a platformei din punct de vedere al unui utilizator, mulțime ce conține 28

de elemente. Toate cele 28 de caracteristici sunt documentate în prezentul document, iar o parte dintre ele sunt calculabile având la dispoziție actuala baza de date, respectiv logică a platformei Tesys.

Rolul acestor caracteristici este de a aduna date despre un utilizator pentru a fi analizate din dorința de a îmbunătăți funcționalitățile aplicației.

1.4 Obiective

Principalul obiectiv al acestui proiect este analiza datelor în cadrul platformei online de e-learning Tesys. Acest lucru ar aduce mari avantaje în îmbunătățirea logicii și funcționalităților oferite de această aplicație.

Acest lucru poate fi făcut atât prin adunarea datelor despre un utilizator și analiza lor cât și prin faptul că în momentul de față unele caracteristici nu se pot calcula având în vedere datele preluate de către aplicație. Acest lucru este benefic în dezvoltarea aplicației datorită faptului ca scoate în evidență logica ce trebuie implementată în continuare în această aplicație astfel încât, pe viitor, aceste caracteristici să poată fi calculate.

Un alt obiectiv important al acestui proiect îl reprezintă dorința de a dona setul de date obținut către comunitatea științifică UCI Irwin Machine Learning Repository pentru a ajuta la îmbunătățirea algoritmilor de machine learning și pattern recognition.

2 MEDIUL DE DEZVOLTARE ȘI TEHNOLOGII FOLOSITE

2.1 XML

Extensible Markup Language (XML) este un meta-limbaj de marcare recomandat de Consorțiul Web pentru crearea de alte limbaje de marcare cum ar fi XHTML, RDF, RSS, MathML, SVG, OWL etc. Aceste limbaje formează familia de limbaje XML. Meta-limbajul XML este o simplificare a limbajului SGML (din care se trage și HTML) și a fost proiectat în scopul transferului de date între aplicații pe internet, descriere structură date ș.a.

XML este acum și un model de stocare a datelor nestructurate și semi-structurate în cadrul bazelor de date native XML.

Datele XML pot fi utilizate în limbajul HTML permițând o identificare rapidă a documentelor cu ajutorul motoarelor de căutare. Cu ajutorul codurilor JavaScript, PHP etc. fișierele XML pot fi înglobate în paginile de internet. Cel mai elocvent exemplu este sitemul RSS care folosește un fișier XML pentru a transporta informațiile dintr-o pagină web către mai multe pagini web.

Avantaje:

- extensibilitate – se pot defini noi indicatori dacă este nevoie.
- validitate – se verifică corectitudinea structurală a datelor.
- oferă utilizatorilor posibilitatea de a-și reprezenta datele într-un mod independent de aplicație

XML este simplu și accesibil (sunt fișiere text create pentru a structura, stoca și a transporta informația), poate fi editat și modificat foarte ușor (necesită doar un editor text simplu precum notepad, wordpad etc.)

2.2 MySQL

MySQL este un sistem de gestiune a bazelor de date relaționale, produs de compania suedeza MySQL AB și distribuit sub Licența Publică Generală GNU. Este cel mai popular SGBD open-source la ora actuală, fiind o componentă cheie a stivei LAMP (Linux, Apache, MySQL, PHP).

Deși este folosit foarte des împreună cu limbajul de programare PHP, cu MySQL se pot construi aplicații în orice limbaj major. Există multe scheme API disponibile pentru MySQL ce permit scrierea aplicațiilor în numeroase limbaje de programare pentru accesarea bazelor de date MySQL cum ar fi: C, C++, C#, Java, Perl, PHP, Python, FreeBasic etc., fiecare dintre acestea folosind un tip specific API. O interfață de tip ODBC denumită MyODBC permite altor limbaje de programare ce folosesc această interfață, să interacționeze cu bazele de date MySQL cum ar fi ASP sau Visual Basic. În sprijinul acestor limbaje de programare, unele companii produc componente de tip COM/COM+ sau .NET (pentru Windows) prin intermediul cărora respectivele limbaje să poată folosi acest SGBD mult mai ușor decât prin intermediul sistemului ODBC. Aceste componente pot fi gratuite (ca de exemplu MyVBQL) sau comerciale.

Licența GNU GPL nu permite încorporarea MySQL în softuri comerciale. Cei care doresc să facă acest lucru pot achiziționa, contra cost, o licență comercială de la compania producătoare, MySQL AB.

MySQL este componentă integrată a platformelor LAMP sau WAMP (Linux/Windows-Apache-MySQL-PHP/Perl/Python). Popularitatea sa ca aplicație web este strâns legată de cea a PHP-ului care este adesea combinat cu MySQL și denumit Duo-ul Dinamic. În multe cărți de specialitate este precizat faptul ca MySQL este mult mai ușor de învățat și folosit decât multe din aplicațiile de gestiune a bazelor de date, ca exemplu comanda de ieșire fiind una simplă și evidentă: „exit” sau „quit”.

Pentru a administra bazele de date MySQL se poate folosi modul linie de comandă sau, prin descărcare de pe internet, o interfață grafică: MySQL Administrator și MySQL Query Browser. Un alt instrument de management al acestor baze de date este aplicația gratuită, scrisă în PHP, phpMyAdmin.

MySQL poate fi rulat pe multe dintre platformele software existente: AIX, FreeBSD, GNU/Linux, Mac OS X, NetBSD, Solaris, SunOS, Windows 9x/NT/2000/XP/Vista.

2.3 XAMPP

XAMPP este un pachet de programe free software, open source și cross-platform web server care constă în Apache HTTP Server, MySQL database și interpretoare pentru scripturile scrise în limbajele de programare PHP și Perl.

Numele XAMPP este un acronim pentru:

- X → cross (din engleză, "cruce") și se referă că este cross-platform
- A → Apache HTTP Server
- M → MySQL
- P → PHP
- P → PERL

Acest program este lansat sub termenii licenței GNU și acționează ca un web server capabil de a servi pagini dinamice. XAMPP este disponibil pentru Microsoft Windows, Linux, Solaris și Mac OS X și este utilizat în principal pentru dezvoltarea proiectelor web. Acest software este util pentru crearea paginilor dinamice utilizând limbaje de programare ca PHP, JSP, Servlets ș.a.

În mod oficial, designerii XAMPP au avut intenția de a îl utiliza numai ca utilitar de dezvoltare, pentru a permite designerilor și programatorilor web să își testeze munca pe calculatoarele proprii, fără a avea nevoie de acces la Internet. Pentru a face posibil acest lucru, multe caracteristici de securitate importante sunt dezactivate în mod implicit. În practică, totuși, XAMPP este uneori utilizat pentru a servi pagini web în World Wide Web.

Un utilitar special este asigurat pentru a proteja prin parolă cele mai importante părți ale pachetului.

XAMPP deasemeni asigură suport pentru crearea și manipularea bazelor de date în MySQL și SQLite între utilizatori.

Odată ce XAMPP este instalat, poți trata adresa de localhost a serverului XAMPP ca pe un server la distanță, prin conectarea utilizând protocolul client FTP. Utilizarea unui program ca FileZilla are multe avantaje atunci când instalezi un sistem de gestiune al conținutului cum ar fi Joomla. Te poți deasemeni conecta la o gazdă locală prin FTP utilizând editorul HTML propriu.

Utilizatorul MySQL implicit este "root", iar acesta nu are o parolă MySQL implicită. Cu programul batch resetroot.bat din subdirectorul mysql al directorului unde este instalat XAMPP, puteți reseta acest utilizator încât să aibă parola vidă.

2.4 Java

Java este un limbaj de programare orientat pe obiecte, puternic tipizat, conceput de către James Gosling la Sun Microsystems (acum filială Oracle) la începutul anilor 1990, fiind lansat în 1995. Cele mai multe aplicații distribuite sunt scrise în Java, iar noile evoluții tehnologice permit utilizarea sa și pe dispozitive mobile gen telefon, agenda electronică etc. În felul acesta se creează o platformă unică, la nivelul programatorului, deasupra unui mediu eterogen extrem de diversificat. Acesta este utilizat în prezent cu succes și pentru programarea aplicațiilor destinate intranet-urilor.

Limbajul împrumută o mare parte din sintaxă de la C și C++, dar are un model al obiectelor mai simplu și prezintă mai puține facilități de nivel jos. Un program Java compilat, corect scris, poate fi rulat fără modificări pe orice platformă care are instalată o mașină virtuală Java (engleză Java Virtual Machine, prescurtat JVM). Acest nivel de portabilitate (inexistent pentru limbaje mai vechi cum ar fi C) este posibil deoarece sursele Java sunt compilate într-

un format standard numit cod de octeți (engleză byte-code) care este intermediar între codul mașină (dependent de tipul calculatorului) și codul sursă.

Mașina virtuală Java este mediul în care se execută programele Java. În prezent, există mai mulți furnizori de JVM, printre care Oracle, IBM, Bea, FSF. În 2006, Sun a anunțat că face disponibilă varianta sa de JVM ca open-source.

Există 4 platforme Java furnizate de Oracle:

1. Java Card – pentru carduri cu cip (smartcard-uri).
2. Java Platform, Micro Edition (Java ME) – pentru hardware cu resurse limitate (PDA sau telefoane mobile).
3. Java Platform, Standard Edition (Java SE) – pentru sisteme cum ar fi stațiile de lucru (workstation-urile). Această platformă se găsește pe PC-uri.
4. Java Platform, Enterprise Edition (Java EE) – pentru sisteme de calcul mari, de multe ori sisteme distribuite.

2.5 Eclipse

Eclipse este un mediu de dezvoltare open-source scris preponderent în Java. Acesta poate fi folosit pentru a dezvolta aplicații Java și, prin intermediul unor plug-in-uri, în alte limbaje cum ar fi C, C++, COBOL, Python, Perl și PHP. De dezvoltarea sa se ocupă Fundația Eclipse.

Este un mediu de dezvoltare integrat (IDE). Acesta conține un spațiu de lucru de bază și un sistem extensibil plug-in pentru personalizarea mediului.

Codul de bază provine de la IBM VisualAge. Kitul de dezvoltare software (SDK), care include instrumente de dezvoltare Java, este destinat pentru dezvoltatorii Java. Utilizatorii pot

extinde abilitățile sale prin instalare de plug-in-uri scrise pentru platforma Eclipse, precum dezvoltarea de toolkit-uri pentru alte limbaje de programare.

Proiectul Eclipse Web Tools Platform (WTP) este o extensie a platformei Eclipse cu instrumente pentru dezvoltarea de aplicații Web și Java EE. Acesta include sursă și editori grafici pentru o varietate de limbi și aplicații predefinite pentru a simplifica dezvoltarea, instrumente și API-uri pentru a sprijini implementarea, rularea și testarea aplicațiilor.

2.6 JDBC

Este o tehnologie Java de conectare la o bază de date oferită de către Corporația Oracle. Această tehnologie este un API pentru limbajul de programare Java care definește cum poate un client să acceseze o bază de date. JDBC oferă metode pentru a interoga și a modifica datele existente într-o bază de date și este orientată către bazele de date relaționale.

Sun Microsystems a lansat JDBC ca parte din JDK 1.1 pe date de 19 februarie 1997. Din acel moment a fost parte integrată din Java SE (Standard Edition). Clasele JDBC sunt conținute de pachetele *java.sql* și *javax.sql*.

2.7 Hibernate

Hibernate ORM (pe scurt, Hibernate) este un context de mapare obiect-relație pentru limbajul Java, oferind un context pentru a mapa un model orientat pe obiect într-o bază de date relațională. Hibernate rezolvă problemele legate de neconcordanța impedanței obiect-relație înlocuind numărul mare de accesări ale bazei relaționale cu funcții pentru manipularea obiectelor de nivel înalt.

Maparea claselor Java în tabele este realizată printr-un fișier XML de configurare sau folosind Java Annotations. Folosind fișiere XML, Hibernate poate genera scheletul codului sursă al claselor obținute din tabele. Acest lucru nu este necesar în cazul folosirii adnotărilor.

Hibernate poate folosi atât fișiere XML cât și Java Annotations pentru a păstra schema bazei de date.

Totodată această tehnologie oferă posibilități de a restitui relațiile 1:m (unul la mai mulți) cât și m:m (mulți la mulți) în cadrul claselor create.

Hibernate oferă un limbaj inspirat din SQL numit Hibernate Query Language (HQL) care permite interogarea datelor deținute de obiecte cu ajutorul interogărilor SQL.

3 STRUCTURA EDUCATIONAL DATA BENCHMARK

3.1 Cercetări similare

3.1.1 UCI Data Benchmark – Performanțe studenți

Acest set de date este donat către UCI Machine Learning Repository de către Paulo Cortez de la Universitatea din Minho, Guimaraes, Portugalia. Rolul acestui set de date este prezicerea performanțelor elevilor din învățământul secundar (liceu).

Pentru acest lucru au fost colecționate date despre rezultatele elevilor din învățământul secundar din două licee din Portugalia. Atributele calculate includ notele elevilor și caracteristici demografice, sociale și în strânsă legătură cu performanțele lor școlare și au fost colecționate cu ajutorul rapoartelor școlare și prin intermediul chestionarelor.

Două seturi de date sunt obținute în ceea ce privește performanțele elevilor la două materii diferite: matematică și limba portugheză.

Fiecare set de date este alcătuit din 33 de caracteristici dintre care 30 sunt comune atât pentru performanțele de la matematică cât și pentru performanțele de la limba portugheză, iar celelalte sunt în strânsă legătură cu materia la care face referire setul de date.

Aceste ultime trei atribute, notate de autor cu G1, G2, respectiv G3 (litera G venind de la cuvântul englezesc "grade" care înseamnă notă), au o strânsă legătură între ele pentru că G1 reprezintă media notele de pe primul semestru, G2 reprezintă media notelor de pe cel de-al doilea semestru, iar G3 reprezintă media anuală (media aritmetică dintre G1 și G2). Astfel, atributului G3 este cel care are cea mai mare importanță, deoarece reprezintă rezultatul final și trebuie găsită o relație între cele 30 de caracteristici și notele obținute de elevi. Toate cele trei caracteristici sunt de tip numeric.

Din celălalte 30 de caracteristici care se regăsesc în acest Data Benchmark aş vrea să le subliniez pe următoarele trei:

→ *sex* – binar – reprezintă genul elevului (bărbat sau femeie).

→ *age* – numeric – reprezintă vârsta elevului.

→ *Pstatus* – binar – reprezintă situația parentală a elevului (împreună sau despărțiți).

Aceste trei caracteristici se regăsesc și în mulțimea de caracteristici alcatuită de mine pentru obținerea unui Data Benchmark din contextul platformei online de e-learning Tesys.

3.1.2 PSLC DataShop

PSLC DataShop – Pittsburg Science of Learning Center DataShop este un o colecție de date din domeniul învățământului care a fost creată pentru a avea caracteristici particulare pentru analiza datelor în sistemul educațional.

În această lucrare se discută metodele prin care membrii comunității utilizează resursele în prezent și cum uneltele DataShop suportă atât analiza datelor cât și extragerea cunoștințelor din datele aferente.

În prezent DataShop este specializată în date obținute din interacțiunea studenților cu software-uri educaționale, incluzând date din cursuri online, sisteme inteligente de tutoring, laboratoare virtuale, sisteme de teme online, medii de învățare colaborative, dar și din simulări. Din punct de vedere istoric, datele educaționale de această natură au fost stocate într-o largă varietate de tipuri de fișiere.

DataShop este o aplicație web care oferă un mediu securizat de stocare a datelor precum și o suită de unelte pentru analiză și vizualizare disponibile cu ajutorul interfeței web. Datele sunt

colectate de la șase cursuri ce se desfășoară în cadrul PSLC: algebră, chimie, limba chineză, limba engleză, limba germană și fizică. Desemenea există surse externe care contribuie în mod regulat la DataShop cum ar fi datele obținute în școala generală de matematică oferite de proiectul ASSISTments (<http://www.assistment.org>), dar și date din cursurile online din licee oferite de Open Learning Initiative (<http://cmu.edu/oli>). Multe alte studii și cercetări folosesc datele DataShop pentru analiza propriilor date.

DataShop poate stoca o gamă largă de tipuri de date provenite din cursurile și studiile computerizate. Aceasta include datele interacțiunii student – software (care pot fi analizate cu ajutorul uneltelor de analiză și vizualizare) precum și publicațiile, fișierele, prezentările sau artefactele electronice pe care cercetătorul dorește să le stocheze.

3.2 Contribuții personale

3.2.1 Sintetizare caracteristici

Procesul de alcătuire a unei mulțimi de caracteristici relevante pentru platforma online de e-learning Tesys a fost unul relativ problematic datorită nivelului mare de informații care au trebuit parcurse în lucrările de specialitate pentru a afla de existența anumitor caracteristici folosite de alții până în acel moment, dar și pentru faptul că multe dintre ele au fost gândite împreună cu profesorul coordonator.

Cărțile și lucrările specializate pe care le-am studiat pentru acest proces sunt "*Data mining in E-Learning*" (editori: C. Romero și S. Ventura) și "*Monitoring and Assessment in Online Collaborative Environments: Emergent Computational Technologies for E-Learning Support*" (scriitori: Angel A. Juan, Thanasis Daradoumis, Fatos Xhafa, Santi Caballe & Javier Faulin).

Datorită acestor lucrări am înțeles logica proiectului pe care trebuia să-l dezvolt și importanța sa în e-learning. Ideea care mi-a rămas întipărită în urma citirii celor două cărți este că pot fi

realizate o gramadă de statistici și analize pe baza datelor provenite dintr-o platformă online de e-learning cu ajutorul cărora să se îmbunătățească funcționalitățile acestor platforma devenind mai pragmatice și mai atractive astfel încât performanțele studenților să crească.

Așa cum am amintit în introducerea acestui document, mulțimea de caracteristici pe care s-a pornit conține 28 de elemente, dintre care, momentan, doar 17 sunt calculabile. Acest lucru se datorează arhitecturii aplicației online, dar acest lucru se va prezenta în capitolele care vor urma.

În continuare o să prezint mulțimea de caracteristici pe care le-am găsit după studiile efectuate cât și o scurtă explicație a fiecărui atribut.

A. Caracteristici referitoare la sesiune

- **TOTALSES** → numărul total de sesiuni începute de un utilizator.
- **MEANINTDUR** → media de timp scursă între două sesiuni consecutive.
- **MEANDUR** → durata medie a fiecărei sesiuni.
- **MEANHITS** → numărul mediu de click-uri efectuate într-o sesiune.
- **W[i]** → numărul total de sesiuni din ziua 'i'.
- **WEEKENDPCT** → oferă detalii despre perioada din săptămână când utilizatorul preferă să înceapă o sesiune (în timpul săptămânii sau în weekend).

B. Caracteristici referitoare la mesaje

i) Caracteristici generale

- **MSG_TOTALSEND** → numărul total de mesaje trimise.

- **MSG_TOTALRECV** → numărul total de mesaje primite.
- **MSG_MEANSEND** → durata medie dintre două mesaje trimise consecutive.
- **MSG_MEANRECV** → durata medie dintre două mesaje primite consecutive.
- **MSG_MEANANSWER** → durata medie dintre un mesaj primit și mesajul trimis ca răspuns.

ii) Caracteristici specifice

- **MSG_TOTALSEND_WDS** → numărul total de cuvinte din mesajele trimise.
- **MSG_TOTALRECV_WDS** → numărul total de cuvinte din mesajele primite.
- **MSG_TOTALSEND_KWS** → numărul total de cuvinte cheie din mesajele trimise.
- **MSG_TOTALRECV_KWS** → numărul total de cuvinte cheie din mesajele primite.
- **MSG_SENDFPERC_KWS** → procentajul de cuvinte cheie din totalul de cuvinte din mesajele trimise.
- **MSG_RECVPERC_KWS** → procentajul de cuvinte cheie din totalul de cuvinte din mesajele primite.

C. Caracteristici demografice

- **GENDER** → reprezintă sexul utilizatorului.
- **AGE** → reprezintă vârsta utilizatorului.
- **HOME** → reprezintă mediul de proveniență al utilizatorului.
- **SIBLING** → reprezintă numărul total de frați și surori pe care utilizatorul îi are.
- **PARENTS** → reprezintă situația familială din care provine utilizatorul.

D. Caracteristici referitoare la activitatea de auto-testare

- ***TOTALQUEST*** → reprezintă numărul total de întrebări la care utilizatorul a răspuns.
- ***GOODQUEST*** → reprezintă numărul total de întrebări la care utilizatorul a răspuns corect.
- ***QUESTPERC*** → reprezintă procentul de întrebări la care utilizatorul a răspuns corect din numărul total de întrebări la care s-a răspuns.
- ***TOTALHW*** → reprezintă numărul total al temelor de casă rezolvate de utilizator.
- ***MEAN_TOTALHW_GRADE*** → reprezintă media notelor primite pentru temele de casă rezolvate.
- ***MEAN_HOMEWORKS_GRADE*** → reprezintă media notelor primite pentru toate temele de casă.

3.2.2 Asociere caracteristici cu datele brute ale platformei Tesys

Procesul de asociere a caracteristicilor cu datele brute ale platformei online de e-learning Tesys a fost unul mult mai ușor față de procesul alcătuirii mulțimii de atribute. Pentru o mai bună înțelegere a asocierii între caracteristici și datele din baza de date, în documentarea caracteristicilor, ca metodă de calculare este trecută direct interogarea conform cu baza de date, iar în cazul în care este necesară o prelucrare a acelor date este menționat acest lucru.

Pentru realizarea cât mai bună a acestui proiect a fost necesară documentarea fiecărei caracteristici astfel încât în viitor să se poată ști fiecare caracteristica în parte pentru ce este folosită și cum se calculează.

Totodată acest lucru este absolut necesar pentru a putea dona setul de date rezultat în urma rulării aplicației dezvoltate în proiectul curent asupra unui context al platformei Tesys către comunitatea științifică UCI Machine Learning Repository.

În continuare o să prezint documentația aferentă fiecărei caracteristici expusă anterior în aceeași ordine, în funcție de clasificarea lor.

A. Caracteristici referitoare la sesiune

TOTALSES

Detalii:

Reprezintă numărul total de sesiuni începute de un utilizator.

Mod de obținere:

- din baza de date

Interogare:

```
select count(id) from `activity` where `userid` = id_user and `actiune` = 'login';
```

Tip: numeric

Numărul de elemente: 1

MEANINTDUR

Detalii:

Reprezintă media de timp dintre două sesiuni consecutive. În momentul de față nu se poate calcula din baza de date pentru că există cazuri în care încheierea unei sesiuni nu este consemnată în baza de date (ex: închiderea browser-ului).

Pentru a putea calcula această caracteristică este nevoie de îmbunătățirea logicii aplicației web astfel încât evenimentele de distrugere a sesiunii să fie înregistrate în baza de date corespunzător.

MEANDUR

Detalii:

Reprezintă durata medie a unei sesiuni. În momentul de față nu se poate calcula din baza de date pentru că există cazuri în care încheierea unei sesiuni nu este consemnată în baza de date (ex: închiderea browser-ului).

Pentru a putea calcula această caracteristică este nevoie de îmbunătățirea logicii aplicației web astfel încât evenimentele de distrugere a sesiunii să fie înregistrate în baza de date corespunzător.

MEANHITS

Detalii:

Reprezintă numărul mediu de click-uri efectuate de utilizator în timpul unei sesiuni. În momentul de față nu se poate calcula din baza de date pentru că aceste evenimente nu se înregistrează în baza de date.

Pentru a putea calcula această caracteristică este nevoie de îmbunătățirea logicii aplicației web astfel încât momentele în care se execută click-uri în timpul unei sesiuni să fie înregistrate în baza de date.

W[i]

Detalii:

Reprezintă numărul total de sesiuni din ziua i , unde i ia valori în intervalul $[0; 6]$ și reprezintă una din zilele săptămânii.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

```
select data from `activity` where `userid` = id_user and `actiune` = 'login';
```

Prelucrarea datelor:

Odată obținute toate datele în care s-au efectuat acțiuni de logare trebuie luată fiecare zi în parte și (cu ajutorul funcției de calendar din limbajul de programare) trebuie stabilită fiecare data ce zi a săptămânii reprezintă.

După acest pas nu rămâne decât să incrementăm numărul de sesiuni asociat zilei respective.

Tip: numeric

Numărul de elemente: 7

WEEKENDPCT***Detalii:***

Reprezintă momentul săptămânii în care un utilizator preferă să înceapă o sesiune. Această caracteristică ne spune dacă utilizatorul preferă să înceapă o sesiune în zilele lucrătoare (luni - vineri) sau la finalul săptămânii (sâmbătă - duminică).

Cu cât valoarea acestei caracteristici este mai mare, cu atât înseamnă că utilizatorul preferă să înceapă o sesiune la finalul săptămânii.

Mod de obținere:

- prelucrarea datelor.

Prelucrarea datelor:

Pentru a calcula această caracteristică trebuie să aplicăm formula următoare pentru caracteristica $W[i]$:

$$\frac{W_5 + W_6}{\sum_{i=0}^{i \leq 6} W_i}$$

Tip: numeric

Numărul de elemente: 1

B. Caracteristici referitoare la mesaje

Caracteristici generale

MSG_TOTALSEND

Detalii:

Reprezintă numărul total de mesaje trimise de utilizator.

Mod de obținere:

- din baza de date;

Interogare:

```
select count(id) from `messages` where `senderid` = id_user;
```

Tip: numeric

Numărul de elemente: 1

MSG_TOTALRECV

Detalii:

Reprezintă numărul total de mesaje primite de utilizator.

Mod de obținere:

- din baza de date.

Interogare:

```
select count(id) from `messages` where `receiverid` = id_user;
```

Tip: numeric

Numărul de elemente: 1

MSG_MEANSEND

Detalii:

Reprezintă media de timp scursă dintre doua mesaje trimise consecutive.

Mod de obținere:

- valori din baza de date;

- prelucrarea datelor;

Interogare:

```
select `data` from `messages` where `senderid` = id_user;
```

Prelucrarea datelor:

Calcularea timpului parcurs între datele a două mesaje consecutive și împărțirea rezultatului la numărul total de mesaje trimise (*MSG_TOTALSEND*).

Tip: numeric

Numărul de elemente: 1

MSG_MEANRECV

Detalii:

Reprezintă media de timp scursă între două mesaje primite consecutive.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

```
select `data` from `messages` where `receiverid` = id_user;
```

Prelucrarea datelor:

Calcularea timpului parcurs între datele a două mesaje consecutive și împărțirea rezultatului la numărul total de mesaje primite (*MSG_TOTALRECV*).

Tip: numeric

Numărul de elemente: 1

MSG_MEANANSWER

Detalii:

Reprezintă media de timp scursă între primirea unui mesaj și trimiterea unui răspuns. În momentul de față nu se poate calcula această caracteristică pentru că în baza de date nu există o evidență clară a răspunsurilor date la mesajele primite.

Caracteristici specifice

MSG_TOTALSEND_WDS

Detalii:

Reprezintă numărul total de cuvinte din mesajele trimise.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

```
select `text` from `messages` where `senderid` = id_user;
```

Prelucrarea datelor:

După obținerea text-ului unui mesaj transmis, acesta se împarte în funcție de anumite caractere care despart cuvintele (ex: spațiul) și se numără câte cuvinte avem. Acest număr se aduna la un numar total de cuvinte din mesajele trimise rezultând în final valoarea acestei caracteristici.

Tip: numeric

Numărul de elemente: 1

MSG_TOTALRECV_WDS

Detalii:

Reprezintă numărul total de cuvinte din mesajele primite.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

```
select `text` from `messages` where `receiverid` = id_user;
```

Prelucrarea datelor:

După obținerea text-ului unui mesaj primit, acesta se împarte în funcție de anumite caractere care despart cuvintele (ex: spațiul) și se numără câte cuvinte avem. Acest număr se adună la un număr total de cuvinte din mesajele primite rezultând în final valoarea acestei caracteristici.

Tip: numeric

Numărul de elemente: 1

MSG_TOTALSEND_KWS***Detalii:***

Reprezintă numărul total de cuvinte cheie din mesajele trimise. În momentul de față nu se poate calcula această caracteristică pentru că nu există o tabelă / o listă cu cuvinte cheie.

Pentru a putea calcula această caracteristică trebuie îmbunătățită arhitectura aplicației web astfel încât să existe o tabelă în baza de date sau un fișier în sistemul de fișiere care să conțină cuvintele cheie.

MSG_TOTALRECV_KWS***Detalii:***

Reprezintă numărul total de cuvinte cheie din mesajele primite. În momentul de față nu se poate calcula această caracteristică pentru că nu există o tabelă / o listă cu cuvinte cheie.

Pentru a putea calcula această caracteristică trebuie îmbunătățită arhitectura aplicației web astfel încât să existe o tabelă în baza de date sau un fișier în sistemul de fișiere care să conțină cuvintele cheie.

MSG_SENDPERC_KWS

Detalii:

Reprezintă procentul de cuvinte cheie din numărul total de cuvinte din mesajele trimise. În momentul de față nu se poate calcula această caracteristică pentru că nu poate fi calculată caracteristica *MSG_TOTALSEND_KWS*.

MSG_RECVPERC_KWS

Detalii:

Reprezintă procentul de cuvinte cheie din numărul total de cuvinte din mesajele primite. În momentul de față nu se poate calcula această caracteristică pentru că nu poate fi calculată caracteristica *MSG_TOTALRECV_KWS*.

C. Caracteristici demografice

GENDER

Detalii:

Reprezintă genul unui utilizator: bărbat sau femeie.

Mod de obținere:

- din baza de date;

Interogare:

```
select `sex` from `iddusers` where `id` = id_user;
```

Tip: numeric

Numărul de elemente: 1

AGE

Detalii:

Reprezintă vârsta unui utilizator.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

```
select `datanasterii` from `iddusers` where `id` = id_user;
```

Prelucrarea datelor:

Având data nașterii trebuie să calculăm cât timp a trecut de atunci până în prezent. Pentru a face acest lucru vom face diferența dintre anul curent și anul nașterii, iar din rezultat vom scădea o unitate în cazul în care utilizatorul nu a împlinit încă vârsta respectivă (comparăm luna, respectiv ziua din data nașterii cu luna, respectiv ziua curentă).

Tip: numeric

Numărul de elemente: 1

HOME

Detalii:

Reprezintă mediul din care provine utilizatorul: rural sau urban. În momentul de față nu se poate calcula această caracteristică pentru că nu există nicio dată care să menționeze mediul din care provine utilizatorul.

Pentru a putea calcula această caracteristică trebuie îmbunătățită arhitectura tablei `iddusers` a bazei de date astfel încât să existe un câmp care să indice mediul de proveniență al utilizatorului.

SIBLING

Detalii:

Reprezintă numărul total de surori și frați pe care utilizatorul îi are. În momentul de față nu se poate calcula această caracteristică pentru că în baza de date nu se găsesc informații legate de familia utilizatorului.

Pentru a putea calcula această caracteristică trebuie îmbunătățită arhitectura tablei `iddusers` a bazei de date astfel încât să existe un câmp care să indice direct valoarea acestei caracteristici sau crearea unui alt tabel care să conțină date despre surorile și frații utilizatorului.

PARENTS

Detalii:

Reprezintă tipul relației dintre părintii utilizatorului: împreună, divorțați, decedați ș.a.). În momentul de față nu se poate calcula această caracteristică pentru că în baza de date nu se găsesc informații legate de familia utilizatorului.

Pentru a putea calcula această caracteristică trebuie îmbunătățită arhitectura tablei `iddusers` a bazei de date astfel încât să existe un câmp care să indice direct valoarea acestei caracteristici.

D. Caracteristici referitoare la activitatea de auto-testare

TOTALQUEST

Detalii:

Reprezintă numărul total de întrebări la care utilizatorul a răspuns.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor.

Interogare:

select count(id) from `intrebariteste` where `userid` = id_user;

select count(id) from `intrebariexamene` where `userid` = id_user;

Prelucrarea datelor:

Valoarea caracteristicii este dată de suma rezultatelor celor două interogări.

Tip: numeric

Numărul de elemente: 1

GOODQUEST**Detalii:**

Reprezintă numărul total de întrebări la care utilizatorul a răspuns corect.

Mod de obținere:

- valori din baza de date;
- prelucrarea datelor

Interogare:

- select count(id) from `intrebariteste` where `userid` = id_user and `raspunsprof` = `raspunsstud`;

- select count(id) from `intrebariexamene` where `userid` = id_user and `raspunsprof` = `raspunsstud`;

Prelucrarea datelor:

Valoarea caracteristicii este dată de suma rezultatelor celor două interogări.

Tip: numeric

Numărul de elemente: 1

QUESTPERC

Detalii:

Reprezintă procentul de răspunsuri corecte din numărul total de răspunsuri.

Mod de obținere:

- prelucrarea datelor.

Prelucrarea datelor:

Valoarea caracteristicii este dată de valoarea:

$$\frac{100 * GOODQUEST}{TOTALQUEST}$$

Tip: numeric

Numărul de elemente: 1

TOTALHW

Detalii:

Reprezintă numărul maxim de teme de casa rezolvate de utilizator.

Mod de obținere:

- din baza de date.

Interogare:

```
select count(id) from `studenthomework` where `studentid` = id_user;
```

Tip: numeric

Numărul de elemente: 1

MEAN_TOTALHW_GRADE

Detalii:

Reprezintă media notelor obținute la temele de casă rezolvate.

Mod de obținere:

- din baza de date.

Interogare:

```
select avg(mark) from `studenthomework` where `studentid` = id_user;
```

Tip: numeric

Numărul de elemente: 1

MEAN_HOMWORKS_GRADE

Detalii:

Reprezintă media notelor obținute pentru toate temele de casă (dacă o temă de casă nu a fost rezolvată, se punctează cu nota 0).

Mod de obținere:

- valori din baza de date;

- prelucrarea datelor.

Interogare:

- select `sectieid` from `sectii` where `userid` = id_user;
- select `id` from `materii` where `sectieid` = id_sectie;
- select count(id) from `homework` where `materieid` = id_materie;
- select sum(mark) from `studenthomework` where `studentid` = id_user;

Prelucrarea datelor:

Primul pas este să obținem id-ul secției la care utilizatorul este înscris. În al doilea rând trebuie să preluăm materiile care se predau la secția respectivă. Pentru fiecare materie trebuie să aflăm numărul total de teme de casă pe care-l vom aduna la un numar total de teme de casă.

Pentru a calcula valoarea acestei caracteristici trebuie să împărțim suma notelor obținute de utilizator la temele de casă rezolvate la numărul total de teme pe care le-a avut.

Tip: numeric

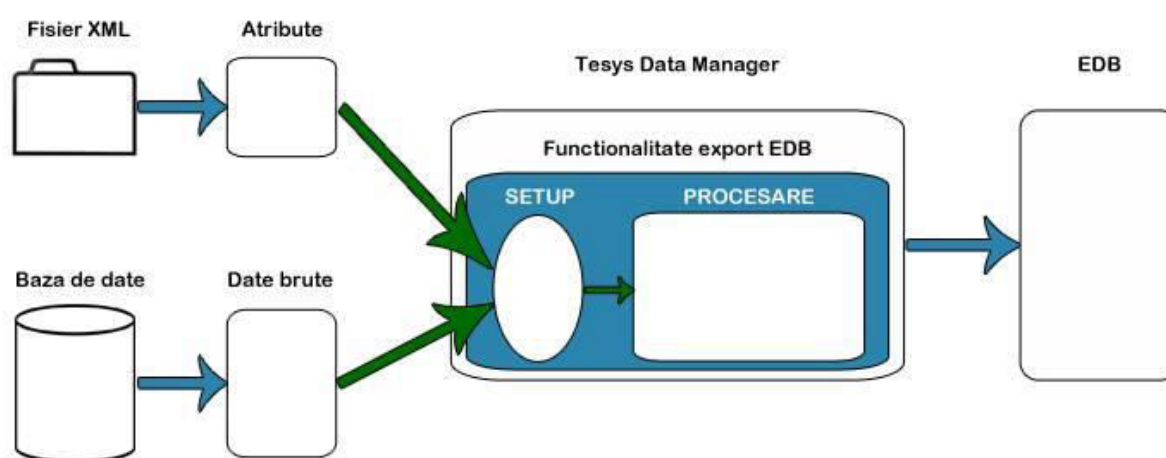
Numărul de elemente: 1

3.2.3 Dezvoltarea aplicației pentru obținerea Educational Data Benchmark-ului

Aplicația în care va fi dezvoltat acest proiect este modulul Tesys Data Manager (TDM) al platformei online de e-learning Tesys. Acest modul are rolul de administrare a datelor aferente aplicației online. Cu ajutorul acestui modul poate fi generat un nou context (ex: la finalul anului și/sau începutul unui nou an), poate fi vizualizat un context sub forma unui pdf (în urma unui export într-un fișier .pdf), poate fi validat un context actual, poate introduce un nou set de studenți etc. Orice operație care se face pe datele bazei de date aferentă platformei online de e-learning Tesys este implementată în modulul TDM.

Având în vedere acest aspect, dar și faptul că proiectul de față are ca scop aplicarea unor operații asupra datelor brute din baza de date aferentă platformei online Tesys, în urma realizării acestui proiect (din punct de vedere al dezvoltării / implementării) modulul TDM va fi îmbunătățit cu o nouă funcționalitate care va avea rolul să exporte un Educational Data Benchmark pentru un context.

Pentru a înțelege mai bine procesul de dezvoltare al acestei funcționalități am să reprezint mai jos arhitectura acestei funcționalități din punct de vedere grafic.



Așa cum se observă din imaginea de mai sus, datele care urmează a fi procesate pentru obținerea unui Educational Data Benchmark provin din baza de date a platformei online de e-learning Tesys sub forma unor date brute, preluate de aplicație în urma activității utilizatorilor săi.

Aceste date împreună cu alte date adiacente care se preia dintr-un fișier .xml sunt preluate de către partea de configurare a funcționalității noastre. Datele care provin din fișierul .xml fac referire la mulțimea de caracteristici și conțin date importante despre statutul fiecărui atribut.

În urma configurării, utilizatorul filtrează datele preluate de către aplicație păstrând astfel doar datele pe care dorește să le folosească în obținerea Educational Data Benchmark-ului.

După configurare, aplicația calculează fiecare caracteristică dorită de utilizator cu datele selectate de utilizator și stochează valorile obținute într-un fișier (.xml, .arff, .csv – în funcție de configurarea făcută de utilizator).

4 DEZVOLTAREA APLICAȚIEI

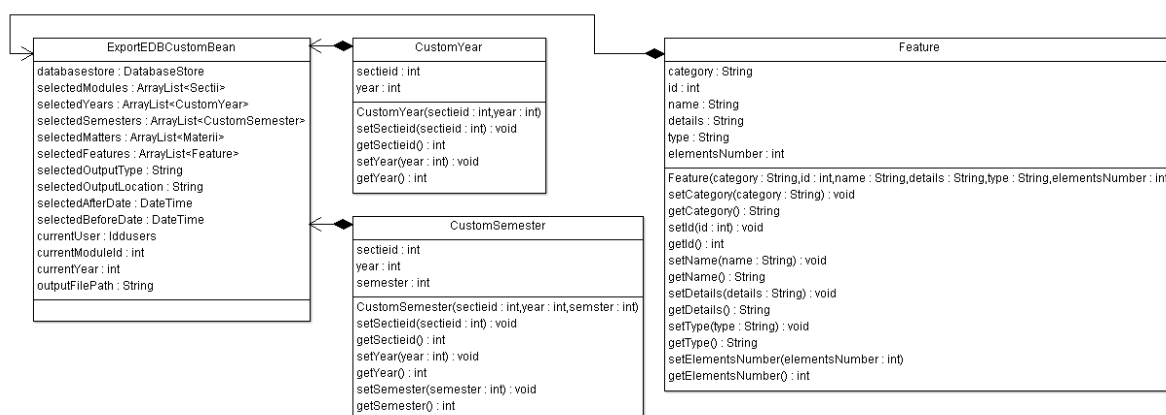
4.1 Arhitectura software în detaliu

Arhitectura software necesară implementării acestei funcționalități însumează un total de 12 clase necesare configurării și calculului caracteristicilor dorite, acestora alăturându-se alte 17 clase care reprezintă caracteristicile calculabile în momentul de față.

Baza de date folosite pentru acest export este baza de date aferentă platformei online de e-learning Tesys, structura sa fiind prezentată în subcapitolele anterioare.

4.1.1 Clase necesare transferului de date

Din totalul de 12 clase necesare exportului Educational Data Benchmark-ului, 4 dintre acestea au rol pur de transmiterea datelor pe parcursul întregului proces de export.



Din această imagine lipsesc clasele "**Sectii**", "**Materii**" și "**Iddusers**", deoarece acestea sunt obținute în cadrul proiectului prin **Reverse Engineering**, astfel că cele două clase sunt în strictă concordanță cu tabelele "Secții", "Materii", respectiv "Iddusers" din baza de date. Atributele acestor două clase reprezintă coloanele tabelurilor din baza de date și sunt de tip **private**, iar metodele reprezintă funcții de acces pentru atributele respective: **setter-i** și **getter-i**.

CustomYear – Această clasă are rolul de a reprezenta obiectele de tip **An** specifice unei anumite secții. Cele doua atribute pe care le conține reprezintă valoarea anului respectiv și id-ul secției din care face parte.

CustomSemester – Această clasă are rolul de a reprezenta obiectele de tip **Semestru** specifice unui anumit an al unei anumite secții. Cele trei atribute pe care le conține reprezintă valoarea semestrului respectiv, anul de care aparține semestrul respectiv și id-ul secției din care face parte.

Feature – Această clasă are rolul de a reprezenta obiectele de tip **Caracteristică**. Conform documentației caracteristicilor ce sunt folosite în acest proiect, fiecare element al mulțimii respective conține același tip de date astfel că aceste obiecte au fost reprezentate în acest proiect cu ajutorul acestei clase. Atributele clasei sunt și reprezintă următoarele:

- *category* – reprezintă numele categoriei din care face parte caracteristica.
- *id* – reprezintă id-ul pe care-l are în fișierul .xml de unde sunt preluate caracteristicile.
- *name* – reprezintă numele caracteristicii.
- *details* – reprezintă detaliile despre caracteristică.
- *type* – reprezintă tipul pe care-l are această caracteristică.
- *elementsNumber* – reprezintă numărul de elemente pe care-l are caracteristica.

ExportEDBCustomBean – Această clasă este cea mai importantă de-a lungul întregului proces de export al Educational Data Benchmark-ului, deoarece culege toate detaliile despre configurarea exportului din pașii de setup, un astfel de obiect fiind transmis de la începutul exportului până la finalul acestuia. Clasa are un total de 14 atribute, fiecare având rolul său în cadrul exportului după cum urmează:

→ *databasestore* – reprezintă datele brute din baza de date preluate și organizate în cadrul aplicației ca **ArrayList**-uri de obiecte identice cu tabelele SQL.

→ *selectedModules* – reprezintă secțiile selectate în cadrul primului pas de setup.

→ *selectedYears* – reprezintă anii selectați în cadrul celui de-al doilea pas de setup.

→ *selectedSemesters* – reprezintă semestrele selectate în cadrul celui de-al doilea pas de setup.

→ *selectedMatters* – reprezintă materiile selectate în cadrul celui de-al treilea pas de setup.

→ *selectedFeatures* – reprezintă caracteristicile selectate în cadrul celui de-al patrulea pas de setup.

→ *selectedOutputType* – reprezintă tipul de fișier care să rezulte în urma exportului. În acest fișier se vor afla valorile caracteristicilor căutate.

→ *selectedOutputLocation* – reprezintă directorul unde va fi salvat fișierul rezultat în urma exportului. În acest fișier se vor afla valorile caracteristicilor căutate.

→ *selectedAfterDate* – reprezintă data minimă pentru care să fie calculate caracteristicile dorite.

→ *selectedBeforeDate* – reprezintă data maximă pentru care să fie calculate caracteristicile dorite.

→ *currentUser* – reprezintă user-ul pentru care se calculează caracteristicile la un moment dat.

→ *currentModuleId* – reprezintă id-ul secției din care face parte user-ul căruia i se calculează caracteristicile la un moment dat.

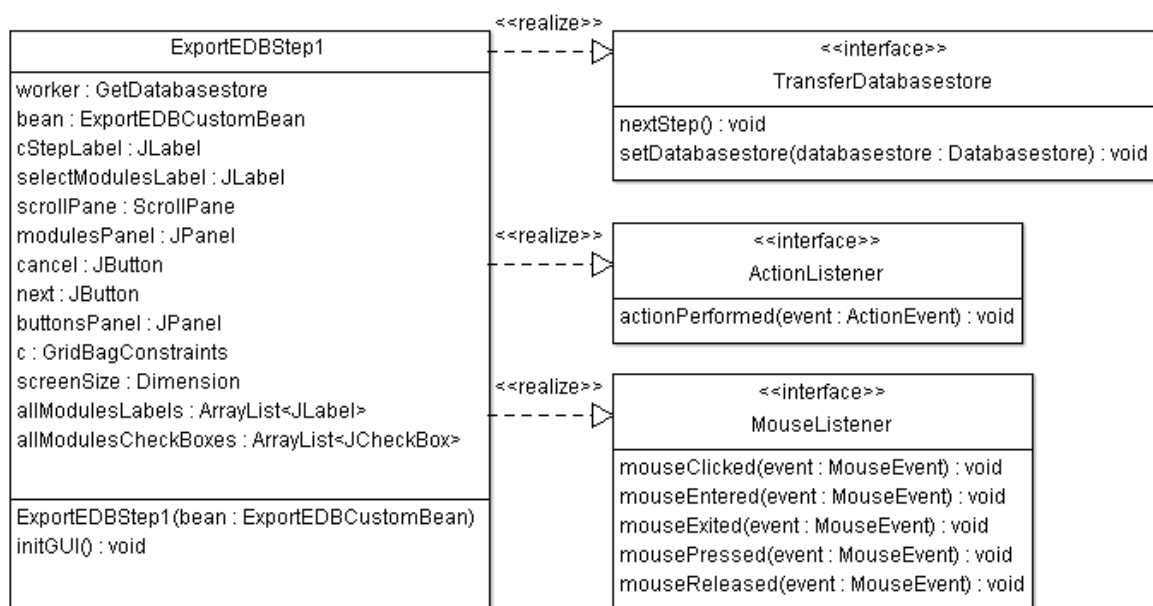
→ *currentYear* – reprezintă anul din care face parte user-ul căruia i se calculează caracteristicile la un moment dat.

→ *outputFilePath* – reprezintă calea absolută către fișierul rezultat în urma exportului. În acest fișier se vor afla valorile caracteristicilor căutate.

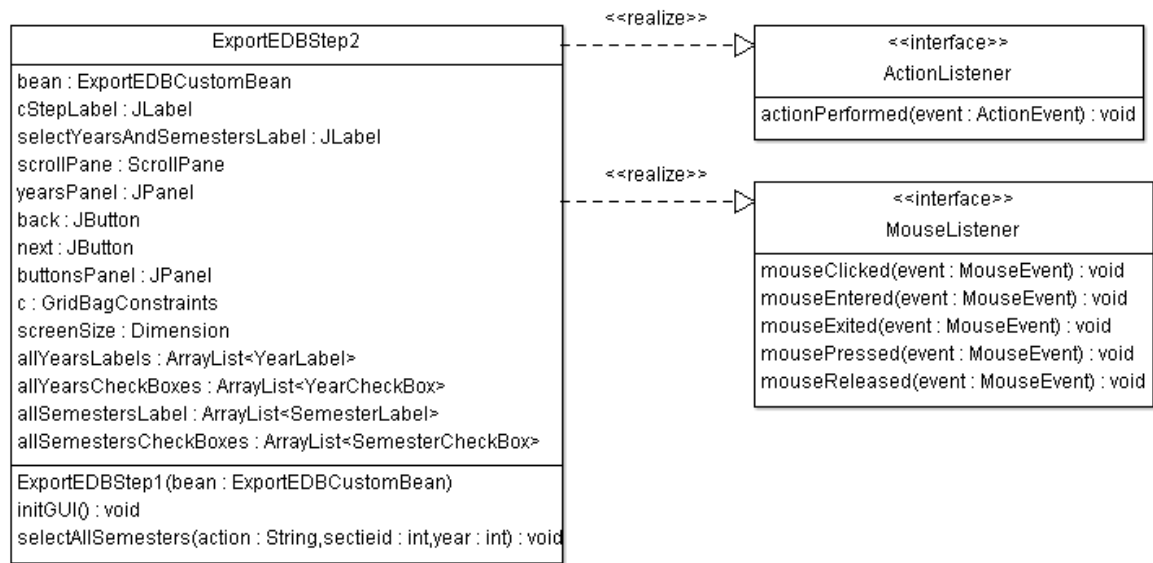
4.1.2 Clase necesare configurării

Configurarea datelor care se doresc a fi obținute este împărțită în această aplicație în patru pași. Primii trei pași au legătură completă unul cu celălalt, fiecare fiind dependent de precedentul, iar cel de-al patrulea pas este independent de primii trei, deoarece face referire la caracteristicile care se doresc a fi calculate, datele între care se dorește calcularea lor și configurarea fișierului de ieșire.

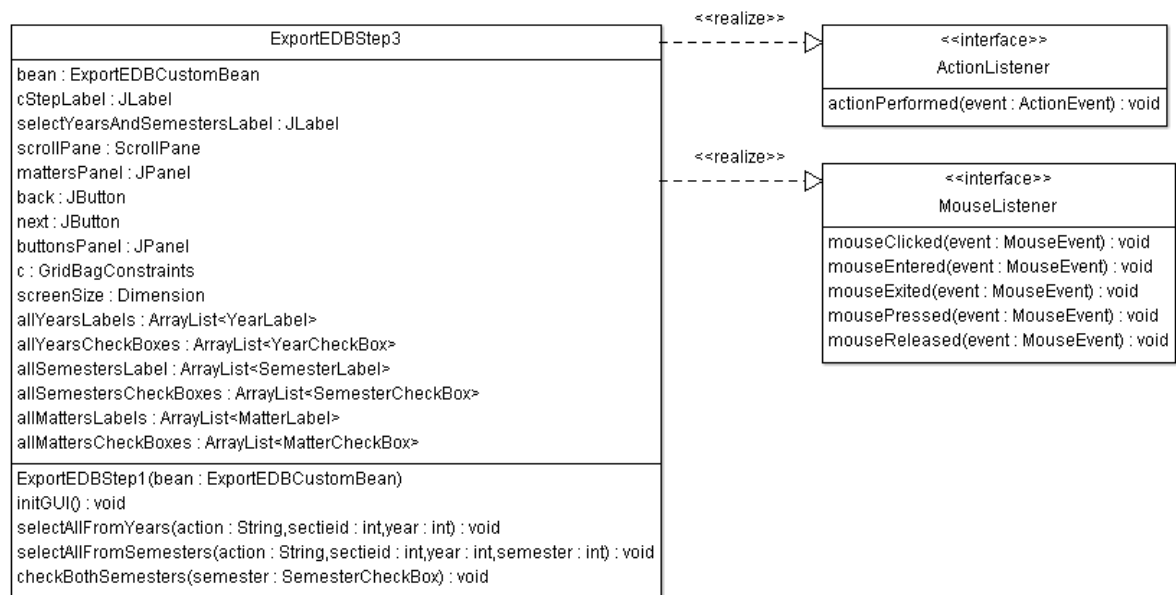
Pasul 1:



Pasul 2:



Pasul 3:



În figura de mai sus sunt prezentate diagramele de clase pentru primii trei pași ai configurării exportului dorit, pași în care se aleg secțiile dorite (pasul 1), anii și semestrele dorite (pasul 2), respectiv materiile dorite (pasul 3). Având în vedere asemănarea celor trei pași, aceștia pot fi tratați împreună datorită structurii asemănătoare a acestora.

TransferDatabaseStore – Această interfață a fost creată de mine la implementarea unei funcționalități anterioare (exportul datelor brute din baza de date în format .pdf). Rolul acestei interfețe este să pună la dispoziție clasei care-o implementează cele două metode pe care le are. Orice clasă care trebuie să preia datele din baza de date și să formeze ArrayList-uri care să conțină obiecte identice cu structura tabelor din baza de date și trebuie să aștepte terminarea acestui proces înainte de a continua, trebuie să implementeze această interfață. După rularea unui fir secundar care realizează această operație, obiectul **DatabaseStore** obținut este transmis către clasa curentă (în cazul acesta clasa **ExportEDBStep1**) prin intermediul metodei *getDatabasestore*, urmând ca apoi să fie apelată metoda *nextStep()* care continuă activitatea firului principal.

ExportEDBStep1 – Această clasă este responsabilă de crearea și gestionarea frame-ului care permite selectarea secțiilor ce se doresc a face parte din obținerea Educational Data Benchmark-ului. Această clasă însumează 13 atribute și o singură metoda, restul metodelor fiind implementate prin intermediul interfețelor. Metoda *initGUI()* are rolul de a crea și a afișa frame-ul pentru selectarea secțiilor.

ExportEDBStep2 – Această clasă este responsabilă de crearea și gestionarea frame-ului care permite selectarea anilor și semestrelor de la secțiile selectate în pasul anterior. Această clasă însumează 14 atribute și 2 metode, restul metodelor fiind implementate prin intermediul interfețelor. Metoda *initGUI()* are rolul de a crea și a afișa frame-ul pentru selectarea secțiilor, iar metoda *selectAllSemesters* are rolul de a selecta sau deselecta cele două semestre ale anului atunci când checkbox-ul anului este selectat, respectiv deselectat.

ExportEDBStep3 – Această clasă este responsabilă de crearea și gestionarea frame-ului care permite selectarea secțiilor ce se doresc a face parte din obținerea Educational Data Benchmark-ului. Această clasă însumează 16 atribute și 4 metode, restul metodelor fiind implementate prin intermediul interfețelor. Metoda *initGUI()* are rolul de a crea și a afișa frame-ul pentru selectarea secțiilor. Metoda *selectAllFromYear* are rolul de a selecta sau deselecta toate semestrele și toate materiile atunci când un an este selectat, respectiv

deselectat. Metoda *selectAllFromSemester* are rolul de a selecta sau deselecta toate materiile atunci când un semestru este selectat, respectiv deselectat. Metoda *checkBothSemester* are rolul de a verifica dacă ambele semestre ale unui an sunt selectate sau deselectate simultan pentru a ști când să selectăm, respectiv deselectăm anul aferent celor două semestre.

În continuare o să fac o scurtă prezentare a atributelor care fac parte din aceste trei clase datorită asemănării structurilor lor.

→ *worker* – reprezintă obiectul prin intermediul căruia se preia datele din baza de date sub forma unor ArrayList-uri de obiecte ce au structura tabelor din SQL.

→ *bean* – reprezintă obiectul care transferă toate datele configurate de-a lungul pașilor de setup.

→ *cStepLabel* – reprezintă label-ul pentru pasul curent al setup-ului.

→ *selectModulesLabel* – reprezintă label-ul pentru denumirea scopului acestui pas de setup.

→ *scrollPane* – reprezintă un panou derulabil pe verticală care urmează a fi creat pe baza unui alt tablou unde se vor afla expuse secțiunile curente.

→ *modulesPanel* – reprezintă panoul în care se vor afla expuse secțiunile curente.

→ *cancel* – reprezintă butonul de anulare a operației de export.

→ *next* – reprezintă butonul de trecere la pasul următor.

→ *buttonsPanel* – reprezintă panoul care va conține butoanele *cancel* și *next*.

→ *c* – reprezintă obiectul cu ajutorul căruia se impun constrângeri la adăugarea elementelor în panourile care au *GridBagLayout* ca *LayoutManager*.

→ *screenSize* – reprezintă dimensiunea monitorului de pe care este rulată aplicația și are rolul de a fixa frame-urile în mijlocul ecranului.

→ *allModulesLabels* – reprezintă label-urile secțiunilor ce urmează a fi selectate. Se regăsește în *ExportEDBStep1*.

→ *allModulesCheckBoxes* – reprezintă checkbox-urile secțiilor ce urmează a fi selectate. Se regăsește în *ExportEDBStep1*.

→ *allYearsLabels* – reprezintă label-urile anilor ce urmează a fi selectați. Se regăsește în *ExportEDBStep2*.

→ *allYearsCheckBoxes* - reprezintă checkbox-urile anilor ce urmează a fi selectați. Se regăsește în *ExportEDBStep2*.

→ *allSemestersLabels* – reprezintă label-urile semestrelor ce urmează a fi selectate. Se regăsește în *ExportEDBStep2*.

→ *allSemestersCheckBoxes* – reprezintă checkbox-urile semestrelor ce urmează a fi selectate. Se regăsește în *ExportEDBStep2*.

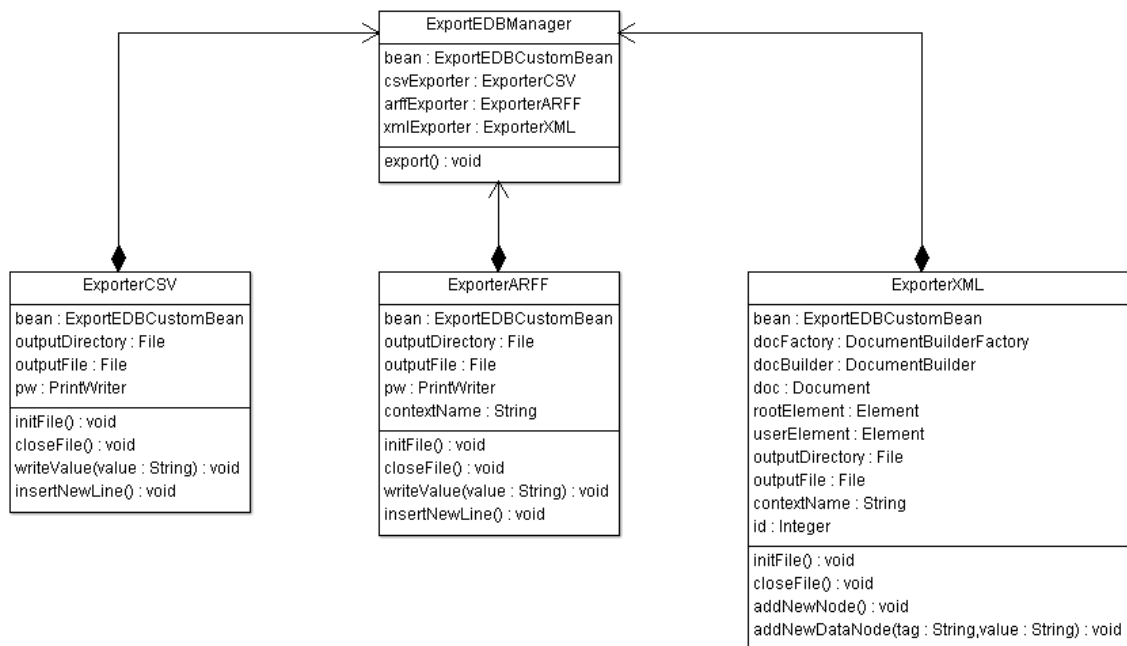
→ *allMattersLabels* – reprezintă label-urile materiilor ce urmează a fi selectate. Se regăsește în *ExportEDBStep3*.

→ *allMattersCheckBoxes* – reprezintă checkbox-urile materiilor ce urmează a fi selectate. Se regăsește în *ExportEDBStep3*.

4.1.3 Clase necesare exportului

Exportul Educational Data Benchmark-ului dorit conform pașilor de configurare prezentați anterior se face în trei formate: .csv, .arff și .xml. Pentru fiecare format avem o clasă care se ocupă de exportul în fișierul respectiv, fiecare clasă fiind coordonată de-o clasă superioară numită *ExportEDBManager*.

Pentru o detaliere mai bună a procesului de export și a componentelor celor 4 clase necesare am să fac o prezentare a acestora pe baza unei diagrame de clase.



ExportEDBManager – În funcție de tipul de fișier selectat pentru output, această clasă instanțiază una din clasele **ExporterCSV** (pentru exportul în fișier .csv), **ExporterARFF** (pentru exportul în fișier .arff) sau **ExportXML** (pentru exportul în fișier .xml). Această clasă însumează 4 atribute și o singură metodă. Metoda *export()* are rolul de a calcula valorile caracteristicilor dorite, urmând ca aceste valori să fie trimise către una din cele trei clase prezentate mai sus pentru a stoca informațiile în fișier.

→ *bean* – reprezintă obiectul care redă toate datele configurate de-a lungul pașilor de setup.

→ *csvExporter* – reprezintă obiectul care face exportul într-un fișier .csv.

→ *arffExporter* – reprezintă obiectul care face exportul într-un fișier .arff.

→ *xmlExporter* – reprezintă obiectul care face exportul într-un fișier .xml.

ExporterCSV / ExporterARFF – Aceste două clase fac exportul datelor obținute într-un fișier .csv (**ExporterCSV**), respectiv într-un fișier .arff (**ExporterARFF**). Aceste clase au fiecare un număr de 4, respectiv 5 atribute și 4 metode fiecare. Toate metodele au același rol și anume: *initFile()* are rolul de a inițializa fișierul, *closeFile()* are rolul de a închide fișierul în urma exportului, *writeValue(String value)* are rolul de a scrie în fișier valoarea lui *value*, iar *insertNewLine()* are rolul de a introduce metacarakterul \n la finalul liniei curente.

→ *bean* – reprezintă obiectul care redă toate datele configurate obținute de-a lungul pașilor de setup.

→ *outputDirectory* – reprezintă directorul unde se va salva fișierul de ieșire.

→ *outputFile* – reprezintă fișierul de ieșire.

→ *pw* – reprezintă stream-ul cu ajutorul căruia se scriu date în fișierul de ieșire.

→ *contextName* – reprezintă numele context-ului din care se exportă Educational Data Benchmark-ul (valabil doar în **ExporterARFF**).

ExporterXML – Această clasă realizează exportul datelor obținute într-un fișier .xml.

Această clasă însumează 10 atribute și 4 metode. Din cele 10 atribute, 4 sunt identice cu cele ale clasei **ExporterARFF** și îndeplinesc același rol (a se vedea mai sus). Metodele *initFile()* și *closeFile()* sunt identice cu cele ale claselor **ExporterCSV** și **ExporterARFF** și îndeplinesc același rol (a se vedea mai sus). Spre deosebire de cele două clase, aici întâlnim două metode noi: *addNewNode()* care are rolul de a crea un nou nod pentru fiecare utilizator în parte, respectiv *addNewDataNode(String tag, String value)* care are rolul de a adăuga un nou nod text pentru fiecare caracteristică.

→ *docFactory* – reprezintă obiectul pe baza căruia se creează un nou document.

→ *docBuilder* – reprezintă obiectul care creează un nou document.

→ *doc* – reprezintă documentul care conține DOM-ul (Data Object Model) fișierului pe care dorim să-l creăm.

→ *rootElement* – reprezintă nodul rădăcină.

→ *userElement* – reprezintă nodul auxiliar pentru introducerea unui nou utilizator în document.

→ *id* – reprezintă id-ul utilizatorului curent.

4.2 Manualul dezvoltatorului

Implementarea funcționalității prezentate în acest proiect s-a făcut în aplicația desktop TDM (Tesys Data Manager) pentru că această aplicație are rolul de a administra datele din baza de date aferentă platformei online de e-learning Tesys. Practic, aplicația TDM are la dispoziție toate datele ce țin de activitatea utilizatorilor platformei Tesys. Având în vedere că pentru a implementa această funcționalitate este nevoie de un context al platformei Tesys, dar și faptul că s-a dorit îmbunătățirea aplicației TDM, implementarea funcționalității a trebuit să depindă de câțiva factori externi datorită modului de funcționare al TDM-ului.

Astfel, datele pentru obținerea Educational Data Benchmark-ului nu sunt preluate direct din baza de date ci din niște obiecte de tip *ArrayList*. Aceste obiecte sunt construite pe baza unor fișiere .xml, iar acele fișiere sunt create preluând datele dintr-o baza de date aferentă unui context al platformei Tesys. În concluzie, aceste obiecte de tip *ArrayList* oferă dezvoltatorului acces direct la datele unui context fără a mai fi nevoie să se lucreze cu baza de date.

Pentru obținerea fișierelor XML respective este nevoie de conectarea la o baza de date aferentă platformei Tesys și preluarea datelor cu ajutorul tehnologiei *reverse-engineering*.

Astfel, de fiecare dată când se dorește exportul unui Educational Data Benchmark trebuie executată operația descrisă în paragraful anterior. Această operație durează în funcție de volumul de date existent în baza de date pe care se lucrează. Cu cât volumul de date este mai mare, cu atât timpul de așteptare crește. După încărcarea datelor în fișierele XML și apoi în

obiectele de tip *ArrayList* a trebuit implementată partea de configurare a datelor ce se doresc a fi obținute.

Primul lucru care a trebuit făcut pentru implementarea acestei funcționalități a fost crearea unei opțiuni în meniul aplicației care atunci când este selectată execută operația descrisă mai sus.

După terminarea acestei operații (avem acces la datele din baza de date prin intermediul obiectelor de tip *ArrayList*) a trebuit implementată logica ce permite configurarea datelor de ieșire. Această operație se întinde pe 4 pași astfel:

→ **Pasul 1** – Permite selectarea secțiilor pentru care se dorește obținerea caracteristicilor.

→ **Pasul 2** – Permite selectarea anilor, respectiv a semestrelor pentru care se dorește obținerea caracteristicilor.

→ **Pasul 3** – Permite selectarea materiilor pentru care se dorește obținerea caracteristicilor.

→ **Pasul 4** – Permite selectarea caracteristicilor ce se doresc a fi calculate, datele între care să fie înregistrate în baza de date (s-a implementat o căsuță care odată selectată permite aplicației să calculeze caracteristicile în funcție de setările celălalte, fără a ține cont de dată), tipul de fișier în care se dorește stocarea datelor de ieșire (csv, arff sau xml) și directorul unde se dorește salvarea fișierului de ieșire.

Pentru primii trei pași datele care se afișează pentru a fi configurate (secțiunile, anii cu semestrele și materiile) sunt preluate cu ajutorul obiectelor de tip *ArrayList* descrise anterior.

Pentru pasul al 4-lea caracteristicile afișate pentru a fi configurate sunt preluat dintr-un fișier XML. Acest fișier a fost creat pentru a standardiza cât mai mult această funcționalitate astfel

încât la apariția a noi caracteristici ce pot fi calculate logica TDM-ului să nu sufere modificări la nivelul pașilor de configurare și a exportului propriu-zis.

Acest fișier conține toate datele necesare despre o anumită caracteristică și anume:

- numele caracteristicii;
- scurtă descriere;
- tipul caracteristicii (*int / int[], double / double[] etc.*);
- numărul de elemente;
- un câmp special care specifică dacă caracteristica poate fi calculată.

Odată configurate datele ce se doresc obținute a fost implementată logica ce realizează exportul propriu-zis. Pentru această operație a fost nevoie de implementarea a 4 clase:

→ ***ExportEDBManager*** – clasă care are rolul de a calcula toate caracteristicile și a transmite valoarea lor către una din celălalte trei clase pentru a stoca valoarea în tipul de fișier dorit de utilizator.

→ ***ExporterCSV*** – clasă care are rolul de a crea un fișier .csv și de a stoca datele primite de la clasa *ExportEDBManager*.

→ ***ExporterARFF*** – clasă care are rolul de a crea un fișier .arff și de a stoca datele primite de la clasa *ExportEDBManager*.

→ ***ExporterXML*** – clasă care are rolul de a crea un fișier .xml și de a stoca datele primite de la clasa *ExportEDBManager*.

Pentru calcularea caracteristicilor s-a ales obțiunea implementării a câte o clasă pentru fiecare caracteristică și folosirea lor în clasa *ExportEDBManager* prin tehnologia *reflection*. Pentru a

realiza acest lucru s-a făcut o convenție ca fiecare clasă ce reprezintă o caracteristică anume să respecte următorul standard:

→ numele clasei este numele caracteristicii pe care o reprezintă așa cum este ea trecută în fișierul XML din care se preia caracteristicile (*case-sensitive*).

→ această clasă are o funcție definită *public* cu numele *getX* (unde *X* reprezintă numele caracteristicii pe care o reprezintă clasa așa cum este ea trecută în fișierul XML din care se preia caracteristicile – *case-sensitive*), iar această funcție primește ca parametru un obiect de tip *ExportEDBCustomManager*, acest obiect fiind cel care a preluat de-a lungul pașilor de configurare toate datele ce se doresc obținute.

În concluzie, indiferent de câte caracteristici vor apărea pe viitor tot ceea ce trebuie făcut pentru a putea fi preluate în TDM și calculate este să se completeze fișierul XML cu datele necesare despre caracteristica respectivă și să se creeze o clasă care să respecte structura declarată anterior. În cadrul acestei clase, în funcția descrisă anterior trebuie implementată logica pentru obținerea caracteristicii respective. Astfel s-a reușit, pe lângă implementarea acestei funcționalități, standardizarea într-o măsură destul de mare a procesului de export al unui Educational Data Benchmark.

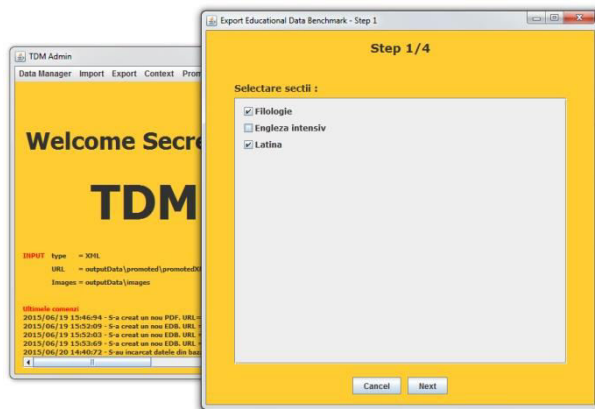
4.3 Manualul utilizatorului

Pentru utilizarea acestei aplicații lucrurile sunt cât se poate de simple. După deschiderea aplicației TDM, pentru obținerea unui Educational Data Benchmark, tot ce trebuie făcut este selectarea acestei opțiuni din meniul aplicației.



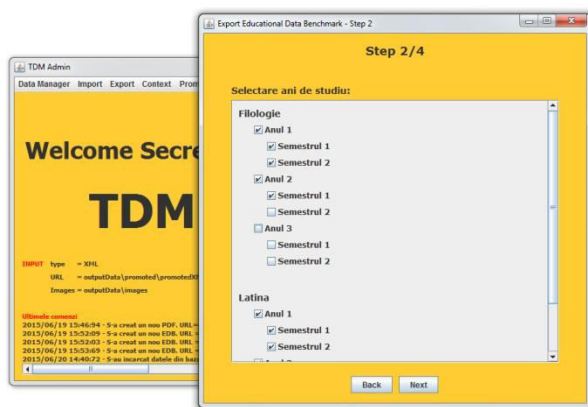
După cum se vede în imagine această opțiune este la îndemână și se poate găsi foarte ușor în meniul *Export*. După selectarea opțiunii trebuie să așteptați o perioadă de timp (în funcție de contextul pe care lucrează aplicația în momentul respectiv – *cu cât contextul este mai bogat în date cu atât timpul de așteptarea crește*), iar apoi o să urmeze patru pași ce permit setarea datelor ce se doresc obținute.

În primul pas se permite selectarea secțiilor pentru care se dorește calcularea caracteristicilor.



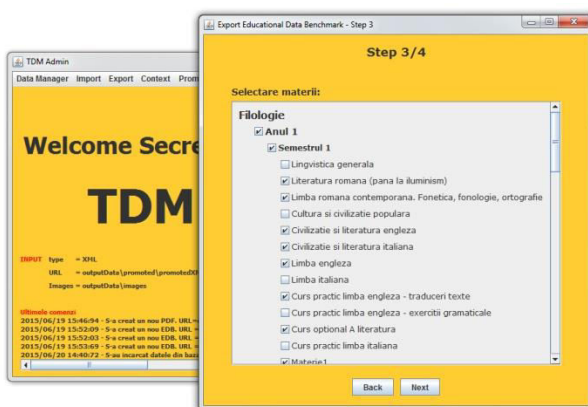
Acest lucru se realizează prin selectarea/deselectarea căsuțelor din fața numelor secțiilor ce se află în context.

În pasul al 2-lea se permite selectarea anilor și a semestrelor fiecărei secții selectate.



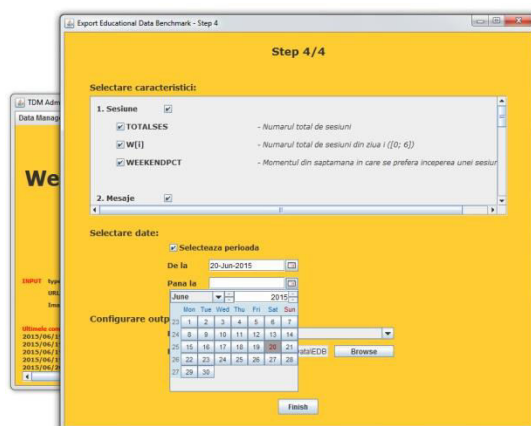
Acest lucru se realizează prin selectarea/deselectarea căsuțelor din fața fiecărui an sau semestru. La selectarea/deselectarea unui an vor fi automat selectate/deselectate semestrele aferente anului respectiv.

În pasul al 3-lea se permite selectarea materiilor din fiecare semestru selectat.



Acest lucru se realizează prin selectarea/deselectarea căsuțelor din fața numelor materiilor. La selectarea/deselectarea unui semestru vor fi automat selectate/deselectate materiile aferente semestrului respectiv. La selectarea/deselectarea unui an vor fi automat selectate/deselectate semestrele aferente, respectiv materiile aferente semestrelor respective.

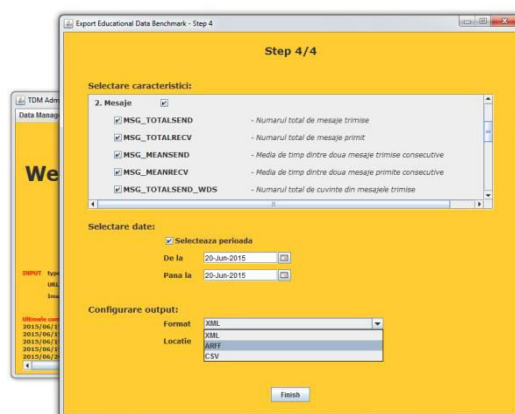
În pasul al 4-lea se permite selectarea caracteristicilor ce se doresc a fi calculate, perioada de timp în care să se încadreze valorile necesare calculării caracteristicilor, tipul de fișier de ieșire și directorul unde să fie salvat.



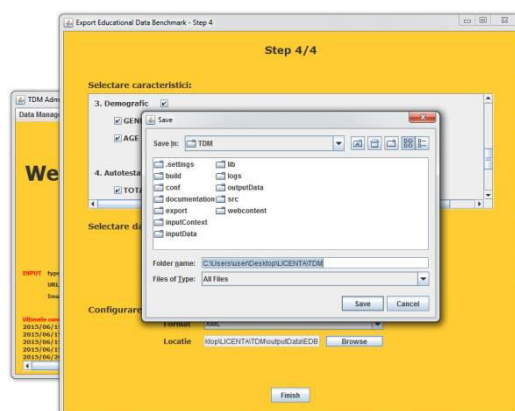
Selectarea caracteristicilor ce se doresc a fi obținute se realizează prin selectarea/deselectarea căsuțelor din fața numelor caracteristicilor. La selectarea/deselectarea unei categorii vor fi automat selectate/deselectate toate caracteristicile aferente categoriei respective.

Configurarea datelor se face în trei pași:

1. Selectarea/deselectarea căsuței din dreptul text-ului "*Selectarea perioada*". Acest lucru permite calcularea caracteristicilor într-o perioadă specificată de utilizator (*atunci când căsuța este bifată*) sau pentru orice dată (*atunci când căsuța este debifată*).
2. Selectarea datei minime pentru care să se preia valorile din baza de date (*înainte de...*).
3. Selectarea datei maxime pentru care să se preia valorile din baza de date (*după...*).



Selectarea tipul de fișier dorit se face cu ajutorul unui *combobox*, iar pentru a vedea toate opțiunile disponibile se face click pe săgeata din dreapta care derulează lista de opțiuni disponibile după cum se poate vedea în imagine. Pentru selectarea unei opțiuni se face click pe câmpul respectiv (*în poză* – va fi selectată opțiunea *arff*).



Pentru selectarea directorului unde să fie salvat fișierul de ieșire se face click pe butonul *Browse* care deschide un *filechooser* ce permite navigarea prin sistemul de fișiere pentru a selecta fișierul dorit.

Odată configurate toate aceste date trebuie apăsat butonul *Finish* care permite aplicației să realizeze exportul Educational Data Benchmark-ului cu specificațiile oferite de utilizator.

4.4 Rezultate obținute

Unul din rolurile acestui proiect, pe lângă obținerea unui Educational Data Benchmark pentru platforma online de e-learning Tesys, a fost donarea datelor obținute către UCI Irwin Machine Learning Repository.

UCI Irwin Machine Learning Repository reprezintă o colecție de date, teorii și generatoare de date care sunt folosite de către comunitatea științifică pentru analiza empirică a îmbunătățirii algoritmilor de machine learning și pattern recognition.

S-a obținut un set de date pentru contextul Facultății de Economie și Administrare a Afacerilor din cadrul Universității din Craiova. Această facultate își desfășoară activitatea de învățământ la distanță prin intermediul platformei online de e-learning Tesys. Acest context are un număr total de 700 de utilizatori dintre care 559 sunt studenți, un număr de 9 secții (*module*) și un număr total de 464 de materii.

Acest set de date a fost documentat conform cerințelor UCI, iar de donarea sa către această comunitate științifică se va ocupa domnul profesor Cristian Mihăescu, deoarece momentan nu s-a obținut acordul Facultății de Economie și Administrare a Afacerilor pentru publicarea datelor.

Pentru obținerea Educational Data Benchmark-ului pe acest context a fost nevoie de un timp de 9 minute pentru transferul datelor din baza de date în fișierele XML și apoi în obiecte de tip *ArrayList* și de alte 3 minute pentru calcularea caracteristicilor și stocarea lor în fișiere (*indiferent de tipul fișierelor*).

Pentru testarea funcționalității în timpul implementării a fost folosită o bază de date de teste care avea 150 de utilizatori, 3 secții (*dintre care doar una conținea date, restul fiind pentru*

testarea încărcării secțiilor). Pe această bază de date încărcarea datelor și obținerea Educational Data Benchmark-ului se realizau aproape instantaneu.

5 CONCLUZII

Un prim lucru bun în implementarea acestei funcționalități este faptul că acest proiect poate continua cu noi abordări cum ar analiza datelor obținute și/sau efectuarea unor statistici pe baza datelor obținute. În momentul de față acest proiect nu permite acest lucru pentru că scopul său a fost altul și anume obținerea unui Educational Data Benchmark care să ofere niște caracteristici relevante la activitatea utilizatorului în cadrul platformei online de e-learning Tesys.

Așa cum am prezentat în capitolul precedent, implementarea funcționalității în cadrul aplicației desktop Tesys Data Manager (TDM) s-a dorit a fi cât mai standard astfel încât apariția a noi caracteristici ce pot fi calculate să nu modifice cu nimic pașii de setup și exportul propriu-zis. În concluzie, proiectul Educational Data Benchmark-ul poate fi îmbunătățit cu alte caracteristici ce pot fi calculate fără a modifica logica de export.

Un alt obiectiv al proiectului care a fost atins a fost donarea unui set de date către comunitatea științifică UCI Irwin Machine Learning Repository, dar momentan nu s-a primit niciun răspuns din partea lor datorită timpului mic scurs între momentul predării Educational Data Benchmark-ului și momentul susținerii acestui proiect.

6 BIBLIOGRAFIE

A data repository for the EDM community: The PSLC DataShop

Data mining in E-Learning – Editors: C. Romero & S. Ventura

Monitoring and Assessment in Online Collaborative Environments: Emergent Computational Technologies for E-Learning Support – Angel A. Juan, Thanasis Daradoumis, Fatos Xhafa, Santi Caballe & Javier Faulin

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROESIS, ISBN 978-9077381-39-7.

7 REFERINȚE WEB

About, UCI Machine Learning Repository

<http://archive.ics.uci.edu/ml/about.html>

Eclipse (software), Wikipedia

[https://ro.wikipedia.org/wiki/Eclipse_\(software\)](https://ro.wikipedia.org/wiki/Eclipse_(software))

Hibernate (Java), Wikipedia

[https://en.wikipedia.org/wiki/Hibernate_\(Java\)](https://en.wikipedia.org/wiki/Hibernate_(Java))

Java Database Connectivity, Wikipedia

https://en.wikipedia.org/wiki/Java_Database_Connectivity

Java (limbaj de programare), Wikipedia

[https://ro.wikipedia.org/wiki/Java_\(limbaj_de_programare\)](https://ro.wikipedia.org/wiki/Java_(limbaj_de_programare))

MySQL, Wikipedia

<https://ro.wikipedia.org/wiki/MySQL>

Student Performance Data Set, UCI Machine Learning Repository

<http://archive.ics.uci.edu/ml/datasets/Student+Performance>

XAMPP, Wikipedia

<https://en.wikipedia.org/wiki/XAMPP>

XML, Wikipedia

<https://ro.wikipedia.org/wiki/XML>