

ŠIAULIŲ UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS KATEDRA

Daiva Lapukaitė

Informatikos (mokytojo) specialybės II magistro kurso dieninio skyriaus studentė

**VIRTUALIOS NUOTOLINIO MOKYMO APLINKOS
DUOMENŲ GAVYBA**

DATA MINING IN VIRTUAL LEARNING ENVIRONMENT

MAGISTRO DARBAS

Darbo vadovas:
Prof. habil. dr. Leonidas Sakalauskas

Recenzentas:
Doc. dr. Sigita Turskienė

Šiauliai, 2009

Tvirtinu, jog darbe pateikta medžiaga nėra plagijuota ir paruošta naudojant literatūros sąrašę pateiktus informacijos šaltinius bei savo tyrimų duomenis.

Daiva Lapukaitė

.....
(parašas)

Darbo tikslas

Darbo tikslas yra sukurti ir ištirti metodologiją bei programinę įrangą, skirtą duomenų, gautų iš virtualios nuotolinio mokymo aplinkos Moodle, pirminiam apdorojimui ir duomenų gavybai. Pasinaudojus sukurtomis priemonėmis pritaikyti StatSoft STATISTICA 7 programinės įrangos paketo duomenų gavybos algoritmus besimokančiųjų mokymosi veiklų duomenų analizei bei parengti išvadas ir rekomendacijas virtualios mokymosi aplinkos tobulinimui.

Uždaviniai

Tikslo įgyvendinimui keliami tokie uždaviniai:

1. Išanalizuoti duomenų gavybos technologijas ir jų pritaikymo virtualioms nuotolinio mokymo aplinkoms tirti galimybes.
2. Atlikti virtualios nuotolinio mokymo aplinkos duomenų gavybos proceso analizę.
3. Sudaryti PHP programinės kalbos priemonėmis vartotojų veiklos duomenų, gautų iš nuotolinio studijų centro Moodle aplinkoje, pradinio apdorojimo programinę įrangą.
4. Ištirti duomenų gavybos metodų, realizuotų su statistinės analizės paketu StatSoft STATISTICA 7, taikymą nuotolinių studijų duomenų analizei besimokančiųjų intensyvumo klasterinės analizės pavyzdžiu.
5. Parengti rekomendacijas nuotolinių studijų veiklų duomenų gavybai.

Darbo vadovas: L. Sakalauskas

.....
(parašas)

Turinys

1. Įvadas.....	5
2. Duomenų gavybos technologijos.....	7
3. Virtualios nuotolinio mokymo aplinkos duomenų gavybos proceso analizė	9
4. Virtualios nuotolinio mokymo aplinkos duomenų gavybos sistema.....	12
4.1. Duomenų rinkimas	13
4.2. Išankstinis duomenų paruošimas.....	13
4.3. Duomenų gavybos pritaikymas	20
4.4. Rezultatai ir rekomendacijos	24
5. Išvados	25
Literatūros sąrašas	26
Anotacija.....	27
Summary.....	27
Priedai.....	28
Kompaktinio disko turinys	28

1. Įvadas

Šiandienos gyvenimas be interneto jau nebeįsivaizduojamas. 2008 metų kovo mėnesį Internet World Stats duomenimis Lietuvoje buvo 2 103 471 interneto vartotojas [5]. Esant tokiam interneto populiarumui nenuostabu, jog internetas tapo ir mokymo pagalbininkas. Tam, kad būtų galima mokytis internetu, yra kuriamos virtualios nuotolinio mokymo aplinkos. Jos tapo labai populiaros ir šiuo metu jau niekam nekyla klausimas: „kas yra virtuali nuotolinio mokymo(si) aplinka?“. Remiantis „Naudojimosi Virtualiomis mokymosi aplinkomis ir Nuotoliniais mokymo kursais tvarka“¹ „Virtuali mokymosi aplinka – tai kompiuterių tinklais ir kitomis informacinėmis ir komunikacinėmis technologijomis pagrįsta informacinė ugdymo sistema, kurioje vyksta mokymosi procesas ir besimokančiųjų bei mokytojų sąveika. Virtuali mokymosi aplinka leidžia kurti ir naudoti įvairius mokymosi scenarijus ir metodus.“ [9] Esant galimybei mokymo procese naudoti virtualias nuotolinio mokymo aplinkas atsirado galimybė į šį procesą įtraukti daugiau mokinių, nei buvo galima tradicinėje mokymo aplinkoje. Padidėjus studijuojančiųjų skaičiui tolygiai padidėja ir duomenų srauto dydis. Nuotolinių aplinkų duomenų bazėse kaupiama informacija apie studentus ir jų veiksmus nuotolinėje aplinkoje: prisijungimo laikas, spęsti testai, darytos užduotys, skaitytos temos ir kt. Kadangi net vienos akademinės grupės šių duomenų skaičius siekia dešimtimis tūkstančių įrašų, tai juos analizuoti rankiniu būdu beveik neįmanoma. Šiai analizei atlikti naudojami duomenų gavybos įrankiai, duomenų gavybos sistemos. Šių dviejų sistemų – virtualios nuotolinio mokymo(si) aplinkos ir duomenų gavybos – sujungimas leidžia virtualios nuotolinio mokymo sistemos vadovui (dėstytojui) paprasčiau analizuoti studentų elgesį sistemoje, atsižvelgiant į gautus rezultatus koreguoti sistemos struktūrą, pateikiamą medžiagą ar studentų darbo grupių sudėtį tam, kad būtų pasiekti kuo geresni mokymo(si) rezultatai. Šiame darbe nagrinėjami Šiaulių nuotolinio studijų centro pateikti duomenys iš virtualios nuotolinio mokymo aplinkos Moodle siekiant sukurti virtualaus nuotolinio mokymo aplinkos duomenų paruošimo duomenų gavybai programinę įrangą, iširti ir pasiūlyti šių duomenų gavybos algoritmus.

Darbo tikslas

Darbo tikslas yra sukurti ir iširti metodologiją bei programinę įrangą, skirtą duomenų, gautų iš virtualios nuotolinio mokymo aplinkos Moodle, pirminiam apdorojimui ir duomenų gavybai. Pasinaudojus sukurtomis priemonėmis pritaikyti StatSoft STATISTICA 7 programinės įrangos paketo duomenų gavybos algoritmus besimokančiųjų mokymosi veiklų duomenų analizei bei parengti išvadas ir rekomendacijas virtualios mokymosi aplinkos tobulinimui.

¹ PATVIRTINTA Švietimo informacinių technologijų direktoriaus įsakymu 2006-12-11 Nr. V1 - 154

Uždaviniai

Tikslo įgyvendinimui keliami tokie uždaviniai:

1. Išanalizuoti duomenų gavybos technologijas ir jų pritaikymo virtualioms nuotolinio mokymo aplinkoms tirti galimybes.
2. Atlikti virtualios nuotolinio mokymo aplinkos duomenų gavybos proceso analizę.
3. Sudaryti PHP programinės kalbos priemonėmis vartotojų veiklos duomenų, gautų iš nuotolinio studijų centro Moodle aplinkoje, pradinio apdorojimo programinę įrangą.
4. Ištirti duomenų gavybos metodų, realizuotų su statistinės analizės paketu StatSoft STATISTICA 7, taikymą nuotolinių studijų duomenų analizei besimokančiųjų intensyvumo klasterinės analizės pavyzdžiu.
5. Parengti rekomendacijas nuotolinių studijų veiklų duomenų gavybai.

Aktualumas

Duomenų gavybos metodologija žinoma jau seniai. Pats terminas apibrėžtas 1990 metais. Ši technologija plačiai taikoma verslo kompanijose duomenų analizavimui. Tačiau duomenų gavybos technologija virtualių nuotolinio mokymosi aplinkų duomenų analizavimui pradėta taikyti labai neseniai. Šių dviejų sričių apjungimas vis dar yra naujas dalykas.

Šio darbo metu sudaryta virtualios nuotolinio mokymo aplinkos duomenų paruošimo duomenų gavybai programinė įranga, panaudojant PHP (angl. PHP - hypertext preprocessor) [6] programavimo kalbą, paruošti duomenys ištirti StatSoft STATISTICA 7 programinės įrangos paketu atliekant klasterinę analizę ir pateikti tyrimo rezultatai. Tokiu būdu gaunama informacija bus naudinga dėstytojams, rengiantiems kursus ir įvertinantiems mokymosi proceso efektyvumą, padės klasifikuoti studentus, atsižvelgiant į jiems reikalingą konsultaciją ir pagalbą mokantis. Tokia analizė, atlikta semestro viduryje, leistų dėstytojui pastebėti nepažangius studentus ir laiku juos paskatinti.

Tyrimo objektas

Šiaulių nuotolinio studijų centro terpė, virtuali nuotolinio mokymo aplinka Moodle, statistinės analizės paketas StatSoft STATISTICA 7, duomenų gavybos algoritmai.

Tyrimo metodai

Mokslinės ir techninės literatūros analizė, praktinis darbas, duomenų analizė.

2. Duomenų gavybos technologijos

Duomenų gavyba – didelių duomenų kiekių tyrinėjimas ir analizė automatizuotu arba pusiau automatizuotu būdu, siekiant rasti naudingus modelius ir taisykles.

„Tobulėjant šiuolaikinėms technologijoms, didėjant duomenų kiekiui darosi vis sudėtingiau juos išanalizuoti ir daryti greitus, efektyvius ir teisingus sprendimus. Duomenų bazės jau peržengė terabaitines ribas ir žmogus jau tampa nepajėgus išanalizuoti visą duomenų gausą. Tokiame milžiniškame kiekyje informacijos gali slėptis ir strategiškai svarbi, ir niekinė informacija. Tokios problemos paskatino atsirasti aukštos kokybės taikomiesiems paketams, programavimo įrankiams, duomenų analizės priemonėms, kurios padeda nepasimesti informacijos gausoje. Savo ruožtu tai kartu padidino ir vartotojų prieinamumą prie pažangiausių technologijų, atvėrė elektroninės komercijos, internetinės analizės ir kitas galimybes.“ [4] Viena iš tokių technologijų yra duomenų gavyba (angl. data-mining). Tai procesas, kuris naudoja įvairius duomenų analizės įrankius, padedančius atrasti tokias duomenų struktūras ir ryšius, kurie būtų panaudojami realioms išvadoms ir sistemos rezultatams apibrėžti. Ši technologija taikoma tose srityse, kur reikia apdoroti labai didelius informacijos kiekius - versle, medicinoje ir kt.

Duomenų gavybos tikslas – iš didelių duomenų kiekių išgauti tam tikrą naują informaciją. Šios srities mokslininkai ir tyrinėtojai duomenų gavybos sąvoką apibrėžia skirtingai, tačiau jos suvokimas yra toks pats.

Pagrindiniai šeši duomenų gavybos veiklos procesai:

1. Klasifikacija (prie vienos iš specifiкуotų grupių ar klasių priskiriamas naujas objektas pagal tam tikrą savybę);
2. Įvertinimas (pateikus tam tikrą kiekį įvedamų duomenų, grąžinama kokia nors nežinoma reikšmė, pvz. pajamos, pelnas ir pan.);
3. Prognozė (tai tas pats procesas kaip klasifikacija ir įvertinimas, tačiau čia įrašai klasifikuojami pagal kažkokį prognozuojamą elgesį ar apytikrę vertę ateityje);
4. Grupavimas pagal bendrus bruožus ar ryšių taisykles (priklausomybių modeliavimas, t.y. nusprendžiama, kurie subjektai dera kartu, pvz. prekių krepšelio analizė);
5. Klasterizavimas (populiacijos dalinimas į tam tikrą kiekį pogrupių ar klasterių);
6. Apibūdinimas ir vizualizacija (vizuali arba tiriamoji duomenų gavyba).

Iš metodinės pusės duomenų gavyba apima:

- Neuroninius tinklus;
- Taisyklių indukciją ir susietumų paiešką (angl. rule induction, association rules);
- K-vidurkių klasterizavimą (angl. k-means clustering);

- Atraminių vektorių regresiją ir klasifikavimą (angl. SVM regression and classification);
- Sprendimų medžius (angl. decision trees);
- Daugiamatį skaliavimą (angl. multidimensional scalling);
- Artimiausio “kaimyno” metodą (angl. the nearest neighbour classification).

Žinių išgavimo iš didelio duomenų kiekio problemos:

- Išskirtų žinių pateikimas;
- Sudėtinga paieška;
- Preliminarių žinių naudojimas išskyrimo procesui pagerinti;
- Išskyrimo operacijų kontrolė;
- Tinkamiausių duomenų gavybos metodų parinkimas tam tikrai duomenų visumai.

Duomenų gavybos programinė įranga

Duomenų gavybos metodu dirbančios programinės priemonės yra vadinamos duomenų gavybos produktais. Išskiriami 3 pagrindiniai duomenų gavybos programinės įrangos tipai:

- Pirmoji grupė: interaktyvios duomenų analizės įrankiai – OLAP (angl. OnLine Analytical Processing [10]) duomenų bazės bei taikomieji įrankiai, kurie yra pagalbinės priemonės OLAP analizei. OLAP duomenų bazės - tai programinis produktas, kuris leidžia visapusiškai analizuoti informaciją realiuoju laiku. Taikomieji įrankiai padeda besinaudojantiems OLAP identifikuoti daugelį svarbių dimensijų ir segmentų, o jų rinkos lyderiais yra laikomi „Business Miner“ and „Cognos Scenario“.

- Antrąją kategoriją sudaro vadinamieji tiesioginiai duomenų gavybos produktai. Jie yra skirti duomenų gavybos specialistams/analitikams kurie atlieka pagrindinį vaidmenį - sprendžia, ar gaunamas modelis, taisyklė arba funkcija yra tikslūs, teisingi ir naudingi. Šių programinių įrankių lyderiais pripažįstami – „IBM DB Intelligent Miner“, „Oracle Darwin“, „SAS Enterprise Miner“, „SGI MineSet“ ir „SPSS Clementine“ .

- Ir trečiasis tipas yra programiniai duomenų gavybos įrankiai, kurie yra skirti tik konkrečioms specifiniams analitiniams procesams nagrinėti, kuriuose duomenų gavybos fazė yra integruota analitinė dalis. Tokios programinės įrangos yra „SAS“ („SAS Enterprise Miner“), „SPSS“ („SPSS Clementine“), „Insightful“ („Insightful Miner“), „IBM“, „Unica“.

Duomenų gavyba padeda analitikams geriau panaudoti turimus duomenis ir greičiau bei tiksliau priimti sprendimus, taip minimizuojant klaidos tikimybę. Svarbiausia duomenų gavybos funkcija yra išrinkti svarbiausius duomenis iš begalės kitų, mažinant laiko bei technologinius resursus ir optimizuojant gaunamus rezultatus.

3. Virtualios nuotolinio mokymo aplinkos duomenų gavybos proceso analizė

Virtualios nuotolinio mokymo aplinkos kaupia didelį kiekį informacijos, kuri yra labai vertinga studentų elgesio analizei atlikti panaudojant duomenų gavybą. Jos įrašo studentų atliktus veiksmus, tokius kaip skaitymas, rašymas, testų sprendimas, įvairių užduočių atlikimas ir net bendravimas su kolegomis – registruojamas kiekvienas pelės paspaudimas, kurį atlieka studentas virtualioje aplinkoje. Šios sistemos dažniausiai taip pat turi duomenų bazę, kurioje kaupiama visos sistemos informacija: asmeniniai duomenys apie vartotojus (profilis), akademiniai rezultatai ir vartotojų sąveikos duomenys. Dėstytojams ir kurso autoriams reikia įrankių, kurie padėtų jiems analizuoti gautus duomenis. Nors kai kurios platformos siūlo savo ataskaitų generavimo įrankius, tačiau dėstytojui tampa sunku gauti naudingą informaciją, kai yra didelis studentų skaičius. Jose nėra specifinių įrankių, leidžiančių pedagogams kruopščiai sekti ir įvertinti visų studentų veiksmus, atsižvelgiant į kurso struktūrą ir turinį. Labai daug žadanti sritis šiam tikslui pasiekti yra duomenų gavybos panaudojimas analizuojant nuotolinio mokymo aplinkų duomenis.

Per paskutinius keletą metų tyrėjai pradėjo tyrinėti įvairius duomenų gavybos metodus, kad padėtų dėstytojams ir administratoriams pagerinti elektroninio studijavimo sistemas. Kai kurios iš naudingiausių duomenų gavybos užduočių ir metodų yra statistika, vizualizacija, klasterizavimas, klasifikacija ir asociacijos taisyklės gavyba. Šitie metodai atskleidžia naujas, įdomias ir naudingas žinias, pagrįstas studentų veiksmų duomenimis.

Dauguma dabartinių duomenų gavybos įrankių yra per sudėtingi naudoti pedagogams ir jų savybės prasilenkia su pedagogų poreikiais.

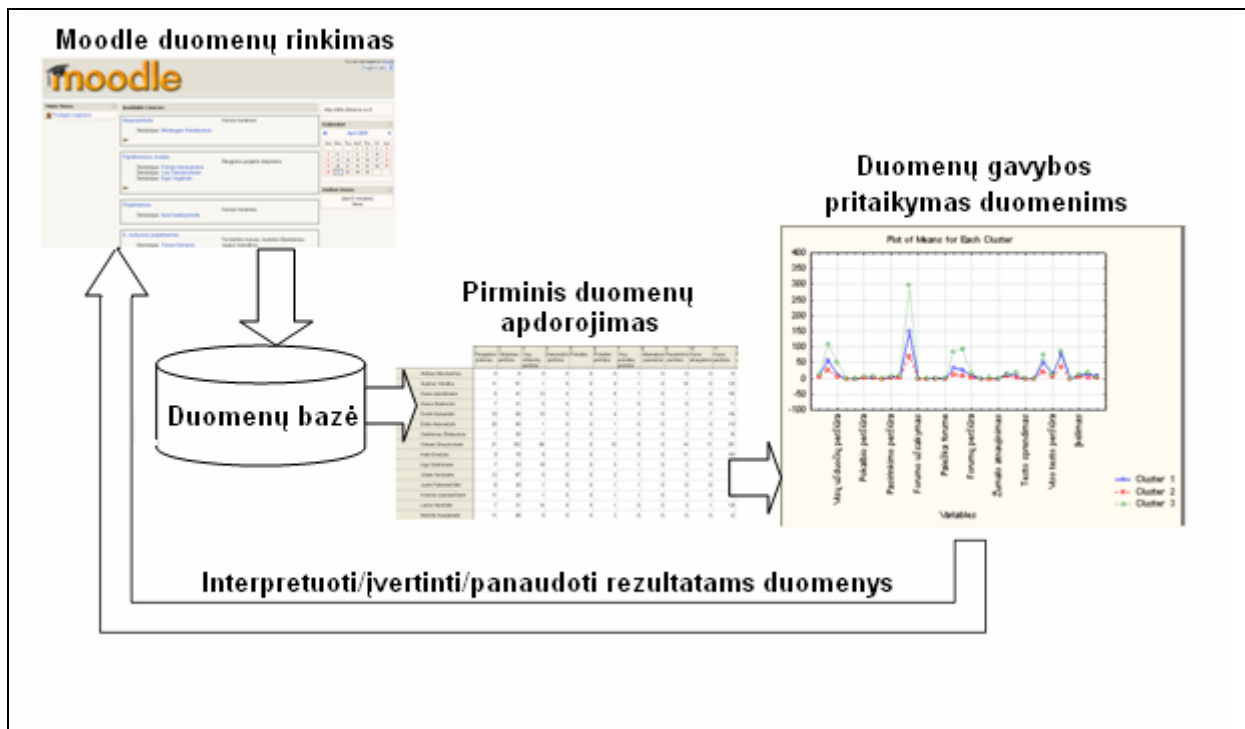
Duomenų gavyba yra pritaikoma duomenims, gaunamiems iš skirtingų švietimo sistemų tipų. Vienokie duomenys yra duomenys, gauti iš tradicinės mokymo aplinkos (klasės). Kitokie duomenys yra duomenys, gauti iš kompiuterine technika ir internetu pagrįsto mokymo, tokio kaip virtualios nuotolinio mokymo aplinkos. Tradicinėse mokymo aplinkose galima gauti informaciją tik apie studentų lankomumą, informaciją apie kursą, mokymo programos tikslus ir individualius plano duomenis. Kompiuterine technika ir interneto tinklu pagrįstas mokymas turi daug daugiau informacijos, nes šitos sistemos gali įrašyti visą informaciją apie studentų veiksmus į įvykių žurnalus arba duomenų bazes.

Duomenų gavyba padeda ne vien tik paverčiant duomenis į žinias, bet taip pat ir gautas žinias panaudojant sprendimų priėmimui. Naudojant virtualios nuotolinio mokymo aplinkos sukauptus studentų veiklos duomenis bei pritaikius duomenų gavybos metodus (grupavimą, klasterizavimą, klasifikavimą), galima identifikuoti studentus pagal mokymosi būdą, stilių, skirti

juos prie atitinkamos grupės ir pateikti pasiūlymus kurso kuratoriui, kaip parinkti kiekvienam studentui tinkamiausią mokymo metodą bei pateikti pritaikytą konkrečiam asmeniui mokymo medžiagą. Atlikus duomenų gilinimą (ang. drill-down) bei pritaikius klasterizavimo metodą, galima išskirti aktyvius studentus, kurie dažnai dalyvauja diskusijose, pradeda naujas temas, rašo laiškus bei domisi visais virtualios nuotolinio mokymo aplinkos įrankiais ir pavesti jiems vadovauti sudarytoms darbo grupėms, skirti papildomus klausimus, siūlymus ir pan. Pasinaudojus šiuo metodu taip pat galima išskirti ir neaktyvius studentus. Šiems studentams galėtų būti teikiama išsamesnė ir specifinė informacija apie naujoves virtualioje nuotolinio mokymo aplinkoje, siūlomi moduliai ar temos, kurių jis dar nestudijavo, atsižvelgus į vertinimo rezultatus besimokantysis galėtų būti nukreiptas patobulinti atitinkamų temų žinias. Norint padaryti tokius studentus aktyvesnius, būtų galima įtraukti juos į diskusijas ar darbo grupes, kuriose jam būtų paskirtas vadovo vaidmuo. Dar viena sritis, kurioje duomenų gavyba gali labai praversti, yra potencialių plagiatorių nustatymas. Tai tokie studentai, kurie tik imituoja mokymąsi prisijungdami ir neatlikdami veiklos, reikalingos mokymuisi. Tokius studentų mokymosi sukčiavimo atvejus labai sunku nustatyti nesinaudojant jokiais duomenų analizės ar duomenų gavybos programinėmis įrangomis. Pasinaudojus daugiamačių išskirčių radimo algoritmais šiuos atvejus galima pastebėti. Šio metodo esmė – rasti veiksmų dėsningumus tarp didelio kiekio atsitiktinių veiksmų įrašų [12].

Elektroninio studijavimo sistemų duomenų gavybos procesas susideda iš tų pačių keturių žingsnių kaip ir bendrame duomenų gavybos procese (žr. paveikslą nr. 1):

- **Duomenų rinkimas.** Studentai naudojami virtualia nuotolinio mokymo aplinka ir veiksmų informacija yra kaupiama duomenų bazėje.
- **Duomenų paruošimas.** Duomenys yra išgryninami ir transformuojami į tinkamą apdorojimui formatą.
- **Duomenų gavybos pritaikymas.** Duomenų gavybos algoritmai yra pritaikomi kuriant modelį, kuris atvaizduoja ir apibendrina žinias pagal vartotojo (dėstytojo, studento, administratoriaus) pageidavimą. Kad tai pasiekti gali būti naudojami pagrindiniai ir specialūs, komerciniai ir nemokami duomenų gavybos įrankiai.
- **Rezultatų interpretavimas, įvertinimas ir pateikimas.** Rezultatai ar gautas modelis yra jau interpretuoti ir jais remdamasis dėstytojas gali atlikti tolimesnius veiksmus. Dėstytojas gali panaudoti gautą informaciją darant sprendimus apie studentų ir nuotolinio kurso sąveiką tam, kad būtų pagerintas studentų mokymasis.



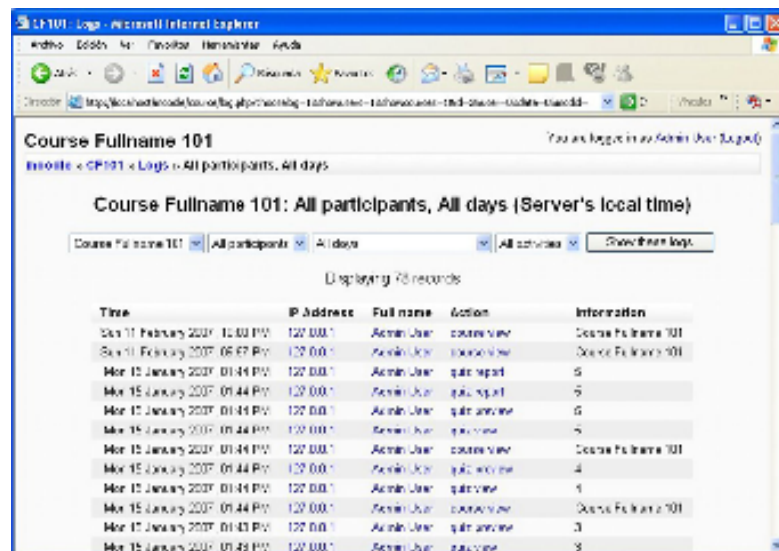
1 pav. Duomenų apdorojimo etapų schema

Šiuo metu viena iš dažniausiai naudojamų virtualaus nuotolinio mokymo aplinkų yra Moodle (Modular Object Oriented Developmental Learning Environment) - atviro kodo studijų valdymo aplinka, kurioje galima kurti lanksčius ir patrauklius internetinius kursus. Ši aplinka taip pat naudojama ir Šiaulių universiteto nuotolinių studijų centre. Nuotolinio mokymo aplinka Moodle nekaupia įvykių kaip tekstinių rinkmenų. Įvykių duomenys kaupiami reliacinėje duomenų bazėje (MySQL ir PostgreSQL yra labiausiai suderinamos, tačiau taip pat gali būti naudojamos Oracle, Access, Interbase ir kitos duomenų bazės). Moodle duomenų bazė turi apytiksliai 145 susijusias tarpusavyje lenteles (lentelių skaičius priklauso nuo aplinkoje naudojamų modulių). Tačiau visa informacija, kurią galima gauti iš šių lentelių, mums nėra reikalinga. Pirmiausia ją reikia paversti tinkamu formatu, kuris yra naudojamas duomenų gavybos algoritmu. Dėl to reikalingas pirminis Moodle sistemos duomenų apdorojimas. Pastebima, kad išankstinį apdorojimą dažniausiai atlieka sistemos administratoriai, o ne dėstytojai tuo atveju, jei tai yra skirtingi asmenys.

4. Virtualios nuotolinio mokymo aplinkos duomenų gavybos sistema

Virtuali nuotolinio mokymo aplinka Moodle saugo detalius visus studentų veiksmų duomenis. Moodle turi integruotą įvykių (angl. log) peržiūros sistemą, tačiau ji nėra išsami. Įvykių duomenys gali būti rūšiuojami pagal kursą, dalyvį (studentą), datą ir veiklą. Dėstytojas šiuos įvykių duomenis gali panaudoti norėdamas sužinoti studentų aktyvumą, veiksmus ir laiką, kada buvo atliekami veiksmai. Jis gali gauti ataskaitas apie konkrečių studentų veiksmus arba studentų sąrašą, kurie atliko kurį nors veiksmą, tačiau detalesnės analizės Moodle aplinkoje atlikti negalima. Detalesnė analizė padėtų dėstytojui interpretuoti, kaip susiję kurso studentų atlikti veiksmai ir pažangumas, nustatyti nesąžiningo elgesio ir plagiato atvejus, iš didelio kiekio įrašų gauti išvadas, kaip geriau parengti kursų programas bei modulius arba kaip formuoti studentų grupes ir t.t., kad būtų pasiekti kiek įmanoma geresni rezultatai.

Pasirinkto kurso visų dalyvių visų veiksmų peržiūra, kuri galima Moodle aplinkoje, nėra labai išsami (žr. paveikslą nr. 2).



Time	IP Address	Full name	Action	Information
Sat 11 February 2006 11:00 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Sat 11 February 2006 06:57 PM	127.0.0.1	Admin User	course edit	Course Fullname 101
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz report	6
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz report	6
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz online	6
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz online	6
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz online	4
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	quiz view	4
Mon 12 January 2006 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 12 January 2006 01:43 PM	127.0.0.1	Admin User	quiz online	3
Mon 12 January 2006 01:43 PM	127.0.0.1	Admin User	quiz view	3

2 pav. Moodle įvykių peržiūros langas

Nuotolinio mokymo aplinkos Moodle duomenų gavybos procesas susideda iš šių etapų:

1. Duomenų rinkimas;
2. Išankstinis duomenų paruošimas;
3. Duomenų gavybos pritaikymas;
4. Rezultatų interpretavimas.

4.1. Duomenų rinkimas

Duomenų rinkimo etapo metu gaunami pirminiai neapdoroti arba labai mažai apdoroti duomenys. Reikia nuspręsti, kurių kursų duomenų gavyba gali būti naudinga. Kursams, kurių aktyvumas yra nedidelis, nėra prasmės taikyti duomenų gavybos metodus, kad gauti naudingos informacijos, nes vien peržiūrėjus veiksmų įrašus tai padaryti yra gan paprasta. Šiame darbe naudojami didelį aktyvumą turėjusių grupių studentų veiksmų įrašų žurnalų duomenys (įrašai buvo eksportuoti iš virtualios nuotolinio mokymo aplinkos Moodle į *.xls failus, nes tik tokiu būdu sistemos administratorius turi galimybę gauti šiuos duomenis). Pasirinkus aktyvias grupes taip pat reikia pasirinkti kokių stebimų veiksmų analizė bus atliekama. Pavyzdžiui: Kordobos universitete yra pasiekiamą informacija apie apytiksliai 4223 studentų 192 kursuose veiklą ir dėstytojai dažniausiai atlieka užduočių pateikimo (angl. assignments) analizę. Atsižvelgdami į šios analizės rezultatus dėstytojai grupuoja studentus į grupes taip, kad būtų pasiekti kiek galima geresni mokymosi rezultatai – į vieną grupę patenka ir pažangūs, ir nepažangūs studentai [11].

Iš Šiaulių universiteto nuotolinių studijų centro gauti duomenys: 6 grupių studentų veiklos Moodle aplinkoje įrašų žurnalai *.xls dokumentuose. Viso 188 studentai, įrašų ~90 000.

4.2. Išankstinis duomenų paruošimas

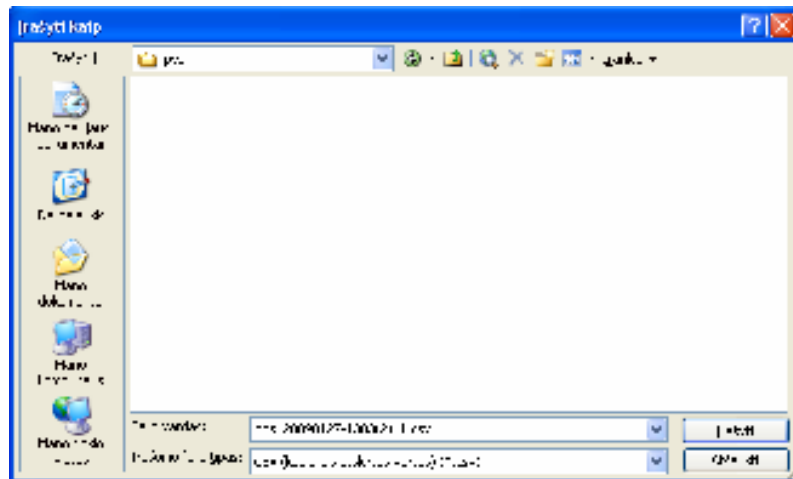
Parengiant duomenis duomenų gavybai reikia sukurti naują dokumentą, kuriame apibendrinama reikiama informacija. Pradiniam virtualios nuotolinio mokymo aplinkos Moodle duomenų apdorojimui galima pasinaudoti duomenų bazės administratoriaus įrankiu arba kokiu nors specifiniu išankstinio duomenų apdorojimo įrankiu, tačiau tai bus tik ribotos ataskaitos, kadangi Moodle išsamių ataskaitų nepateikia. Todėl gauti duomenys iš Šiaulių universiteto nuotolinių studijų centro yra neapibendrinti, o pateikti tik kaip visų veiksmų sąrašas, išsaugotas *.xls formatu.

Išeksportavus duomenis iš Moodle aplinkos duomenų bazės į MS Exel failą turime duomenų lentelę (žr. paveikslą nr. 3). Šiuos duomenis kelsime į MySQL duomenų bazėje sukurtas lenteles pagal grupės pavadinimą. Dėl duomenų konfidencialumo darbe nenaudojamas tikrasis grupės pavadinimas.

	1	2	3	4	5	6	7
3	Pašvaldys	1	20090127 104117	14.10.2009	Tiška pasveikimo	Vakara	FF masė
4	valdyb. L		20090127 104117	14.10.2009	Jaunūjų	course + em	valdyb. J. Karson
5	valdyb. F		20090127 104117	2009 1 85	Vilniaus Tax	vaigumet + em	valdyb. J. Karson
6	valdyb. L		20090127 104117	2009 1 105	Vilniaus Tax	course + em	valdyb. J. Karson
7	valdyb. F		20090127 104117	2009 1 85	Vilniaus Tax	vaigumet + em	PašT. DaP=FF įtaka pašai
8	valdyb. L		20090127 104117	2009 1 105	Vilniaus Tax	course + em	valdyb. J. Karson
9	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	valdyb. J. Karson
10	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
11	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
12	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	valdyb. J. Karson
13	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
14	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	valdyb. J. Karson
15	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	FF įtaka pašai
16	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
17	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
18	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
19	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
20	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
21	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
22	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
23	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
24	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
25	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
26	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
27	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
28	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai
29	valdyb. F		20090127 104117	2009 09 29 14	įtaka pasveikimo	vaigumet + em	PašT. DaP=FF įtaka pašai
30	valdyb. L		20090127 104117	2009 09 29 14	įtaka pasveikimo	course + em	PašT. DaP=FF įtaka pašai

3 pav. Duomenys Exel lentelėje

Tam, kad būtų galima importuoti duomenis į MySQL duomenų bazę pirma reikia *.xls formato failus pakeisti į *.csv (angl. comma-separated values) [3] formatą (žr. paveikslą nr. 4).



4 pav. Duomenų failo saugojimas *.csv formatu

Duomenų sukėlimui į duomenų bazę pasinaudota duomenų įkėlimo iš *.csv failų įrankiu, esančiu MySQL duomenų bazėje (žr. paveikslą nr. 5). Kėlimo metu iškilusios problemos: saugant iš *.xls į *.csv formatą kai kuriuose įrašuose trūko laukų pabaigos simbolių.

Struktūra | Paržiūrėti | SQL | Paimėkta | Inpti | Flopartuoti | Veikimai | Analgti | Sulinti

Textas (SQL) mēk x, fēis	<input type="text"/>	En → (Cifra x= dēlx " 148/R)
Laikotarpis lentelės turinys	<input type="checkbox"/> - laikas	Laikotarpis lentelės turinys (laikas) bus rodomas tik tada, kai bus įrašyti duomenys.
Laukų pobūdis	<input type="text"/>	Laukų pobūdis
Laukų reikšmės apibrėžimas	<input type="text"/> <input type="checkbox"/> Prorintanti	Užrašyti lentelės reikšmės apibrėžimus, kad būtų galima atpažinti duomenis, kurie yra uždėję klaidas.
Laukų reikšmės tagai	<input type="text"/>	Prorintanti veiksmai, kurie apibrėžiami reikšmės apibrėžimuose.
Laukų reikšmės žymė	<input type="text"/>	Užrašyti reikšmės žymę (pvz., "148/R")
Sluoksnio pavadinimas	<input type="text"/>	Leidžia apibrėžti lentelės sluoksnio pavadinimą, kuris bus rodomas laukų reikšmėse.
Laukų matymas	<input type="radio"/> JŲ A <input checked="" type="radio"/> JŲ A ir JŲ B	Užrašyti, kurie laukai bus matomi, kai matomas lentelės turinys. Jei yra įrašyti duomenys, matomas bus visų laukų turinys.

i

5 pav. Duomenų įkėlimas į duomenų bazę

Įkėlus duomenis duomenų bazėje turime po lentelę kiekvienai grupei (žr. paveikslą nr. 6). Lentelės struktūra aprašyta 1 lentelėje. Paveiksle nr. 7 matome naudojamų duomenų laukų tipus. Buvo atlikti duomenų koregavimo veiksmai (vardų ir pavardžių koregavimas, nes kai kurie įrašai buvo be lietuviškų raidžių – tokius vardus į aplinką suvedė besiregistruojantys studentai). Šiam koregavimui įvykdytos laukų atnaujinimo užklauskos. Pavyzdžiui:

```
UPDATE `logai6` SET `Studentas` = 'Artūras Gavedas' WHERE `Studentas` = 'Arturas Gavedas';
UPDATE `logai6` SET `Studentas` = 'Arūnas Svidrius' WHERE `Studentas` = 'Arunas Svidrius';
UPDATE `logai6` SET `Studentas` = 'Diana Sevelioviene' WHERE `Studentas` = 'diana sevelioviene';
```

←T→	ID	Laikas ▲	IP	Studentas	Veiksmas	Veikimas
<input type="checkbox"/>	2	2008.09.18 21:27	87.247.69.168	Vaidas Kiškėnas	course view	VADYBA 8(I kursas)
<input type="checkbox"/>	3	2008.09.18 21:27	87.247.69.168	Vaidas Kiškėnas	quiz view all	
<input type="checkbox"/>	4	2008.09.18 21:28	87.247.69.168	Vaidas Kiškėnas	course view	VADYBA 8(I kursas)
<input type="checkbox"/>	5	2008.09.18 21:28	87.247.69.168	Vaidas Kiškėnas	forum view forum	
<input type="checkbox"/>	6	2008.09.18 21:39	87.247.69.168	Vaidas Kiškėnas	course view	VADYBA 8(I kursas)
<input type="checkbox"/>	7	2008.09.18 21:39	87.247.69.168	Vaidas Kiškėnas	forum view forum	Naujien
<input type="checkbox"/>	8	2008.09.18 21:39	87.247.69.168	Vaidas Kiškėnas	course view	VADYBA 8(I kursas)
<input type="checkbox"/>	9	2008.09.18 21:39	87.247.69.168	Vaidas Kiškėnas	resource view	
<input type="checkbox"/>	10	2008.09.18 21:39	87.247.69.168	Vaidas Kiškėnas	resource view	Vertinimo formul
<input type="checkbox"/>	11	2008.09.18 21:40	87.247.69.168	Vaidas Kiškėnas	resource view	
<input type="checkbox"/>	12	2008.09.18 21:40	87.247.69.168	Vaidas Kiškėnas	resource view	PLANAS VADYBOS EGZAMINUI

6 pav. Lentelė duomenų bazėje

Lauko pavadinimas	Apibūdinimas
ID	Irašo identifikavimo numeris
Laikas	Veiksmo atlikimo data ir laikas (sekundžių tikslumu)
IP	IP adresas
Studentas	Studento vardas ir pavardė
Veiksmas	Veiksmas, atliktas Moodle sistemoje
Veikimas	Platesnė atlikto veiksmo informacija

	Laukas	Tipas	Palyginimas	Atributai	Null	Nutylint	Papildomai
<input type="checkbox"/>	ID	int(10)			Ne		auto_increment
<input type="checkbox"/>	Laikas	varchar(50)	utf8_lithuanian_ci		Taip	NULL	
<input type="checkbox"/>	IP	varchar(50)	utf8_lithuanian_ci		Taip	NULL	
<input type="checkbox"/>	Studentas	varchar(50)	utf8_lithuanian_ci		Taip	NULL	
<input type="checkbox"/>	Veiksmas	varchar(50)	utf8_lithuanian_ci		Taip	NULL	
<input type="checkbox"/>	Veikimas	varchar(50)	utf8_lithuanian_ci		Taip	NULL	

7 pav. Duomenų bazės lentelių laukų aprašymas

Duomenų bazėje kaupiamų duomenų apie studento veiklą sąrašas su vertimu į lietuvių kalbą pateikiamas 2 lentelėje.

2 lentelė. Kaupiamų duomenų pavadinimų vertimas

Eil. Nr.	Angliškas (originalo) pavadinimas	Vertimas
1.	assignment upload	Prisegtuko įkėlimas
2.	assignment view	Užduoties peržiūra
3.	assignment view all	Visų užduočių peržiūra
4.	blog view	Dienoraščio peržiūra
5.	chat talk	Pokalbis
6.	chat view	Pokalbio peržiūra
7.	chat view all	Visų pokalbių peržiūra
8.	choice choose	Alternatyvos pasirinkimas
9.	choice view	Pasirinkimo peržiūra
10.	course recent	Kurso atnaujinimas
11.	course view	Kurso peržiūra
12.	forum subscribe	Forumo užsakymas
13.	forum unsubscribe	Forumo užsakymo atšaukimas
14.	forum user report	Forumo vartotojo pranešimas
15.	forum search	Paieška forume
16.	forum view discussion	Diskusijos forume peržiūra
17.	forum view forum	Forumo peržiūra
18.	forum view forums	Forumų peržiūra
19.	glossary view	Žodyno peržiūra
20.	glossary view all	Viso žodyno peržiūra
21.	journal update entry	Žurnalo atnaujinimas
22.	journal view	Žurnalo peržiūra
23.	journal view all	Viso žurnalo peržiūra
24.	quiz attempt	Testo sprendimas
25.	quiz close attempt	Bandytas baigti testą
26.	quiz view	Testo peržiūra

27.	quiz view all	Viso testo peržiūra
28.	resource view	Resursų peržiūra
29.	resource view all	Visų resursų peržiūra
30.	upload upload	Įkėlimas
31.	user view	Vartotojo informacijos peržiūra
32.	user view all	Visų vartotojų informacijos peržiūra

Kadangi tokie neapdoroti duomenys yra netinkami tolimesnei analizei, todėl buvo sukurta suvestinė lentelė šiam tyrimui reikalingos informacijos apjungimui (žr. paveikslą nr. 13). Šioje lentelėje pavaizduota visų veiksmų, padarytų kiekvieno studento kurso metu, kiekio suvestinė. Ši suvestinė eksportuojama į *.csv formato failą. Toks duomenų formatas jau yra tinkamas tolimesnei duomenų analizei StatSoft STATISTICA 7 programinės įrangos paketu arba bet kuria kita duomenų analizės programa.

Darbo realizaciją galima rasti internete adresu <http://ik.su.lt/~dailap>. Tam, kad būtų galima atlikti duomenų analizę, reikia pasirinkti kurios grupės duomenys bus analizuojami (žr. paveikslą nr. 8). Pasirinktos grupės analizę galima atlikti dvejopai: pasirinkus „Vieno studento veiklos atvaizdavimas“ parodomas tos grupės studentų sąrašas (žr. paveikslą nr. 9), kuriame galima parinkti vieno konkretaus studento veiklos peržiūrą (žr. paveikslą nr.10). Tai naudinga norint pamatyti konkretaus studento veiklą. Pasirinkus „Visos grupės studentų veiklos atvaizdavimas“ gauname apibendrintus visų grupės studentų visų veiksmų duomenis (kiek kartų kiekvienas studentas atliko konkrečius veiksmus) (žr. paveikslą nr. 11).

8 pav. Grupių sąrašas, kurių duomenis galima analizuoti

Grupės Grupel studentų sąrašas:

1.	Renkuosi	Artūras Mockevičius
2.	Renkuosi	Audrius Volodka
3.	Renkuosi	Daiva Janelūnienė
4.	Renkuosi	Darius Budreckis
5.	Renkuosi	Dovilė Nutautaitė
6.	Renkuosi	Edita Adomaitytė
7.	Renkuosi	Gediminas Šidlauskas
8.	Renkuosi	Gintarė straznickaitė
9.	Renkuosi	Indrė Driežytė
10.	Renkuosi	Inga Gedrimienė
11.	Renkuosi	Jūratė Gerybaite
12.	Renkuosi	Justė Paltanavičiūtė
13.	Renkuosi	Kristina Laurinavičienė
14.	Renkuosi	Laima Norvilaitė
15.	Renkuosi	Mažvilė Ananjevaite

9 pav. Grupės studentų sąrašas

Studentas Povilas Motuza yra iš Grupe2 grupės.
Jo veiksmų sąrašas:

assignment upload	2
assignment view	7
assignment view all	2
choice view	1
course view	43
forum user report	3
forum view discussion	3
forum view forum	2
quiz attempt	6
quiz close attempt	6
quiz view	24
quiz view all	6
resource view	17
upload upload	2
user view	1

10 pav. Studento veiksmų sąrašas

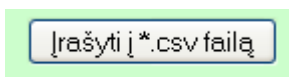
[rašyti į *.csv failą]

Grupės Grupe5 studentų veikla:

Vardas Pavardė	Prisegtuko ikėlimas	Užduoties peržiūra	Visų užduočių peržiūra	Dienoraščio peržiūra	Pokalbis	Pokalbio peržiūra	Visų pokalbių peržiūra	Alternatyvos pasirinkimas	Pasirinkimo peržiūra
Agnė Zlabytė	1	15	0	0	0	0	0	0	1
Andrius Regelskis	1	3	1	0	0	0	1	0	0
Asta Aidukonytė	10	39	1	0	0	0	1	0	0
Asta Jašinskienė	0	24	6	0	0	0	1	0	11
Aušra Kvedienė	12	64	1	0	0	0	0	0	4
Daiva Jokubauskienė	14	32	29	0	0	0	5	0	1
Danutė Pociuvienė	5	30	3	0	0	0	0	0	5
Deimantė Liubinienė	3	11	11	0	0	0	0	0	1
Edita Kavaliauskienė	8	33	0	0	0	0	0	0	2
Egidija Danilevičiūtė	2	35	2	0	0	0	7	0	11

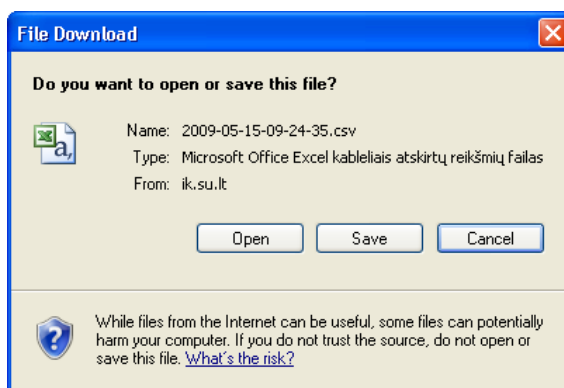
11 pav. Apibendrinti vienos grupės studentų duomenys, gauti iš duomenų bazės

Tolimesnei duomenų analizei matomą apibendrintų duomenų lentelę įsirašome į *.csv failą paspaudus mygtuką „[rašyti į *.csv failą“ (žr. paveikslą nr. 12).



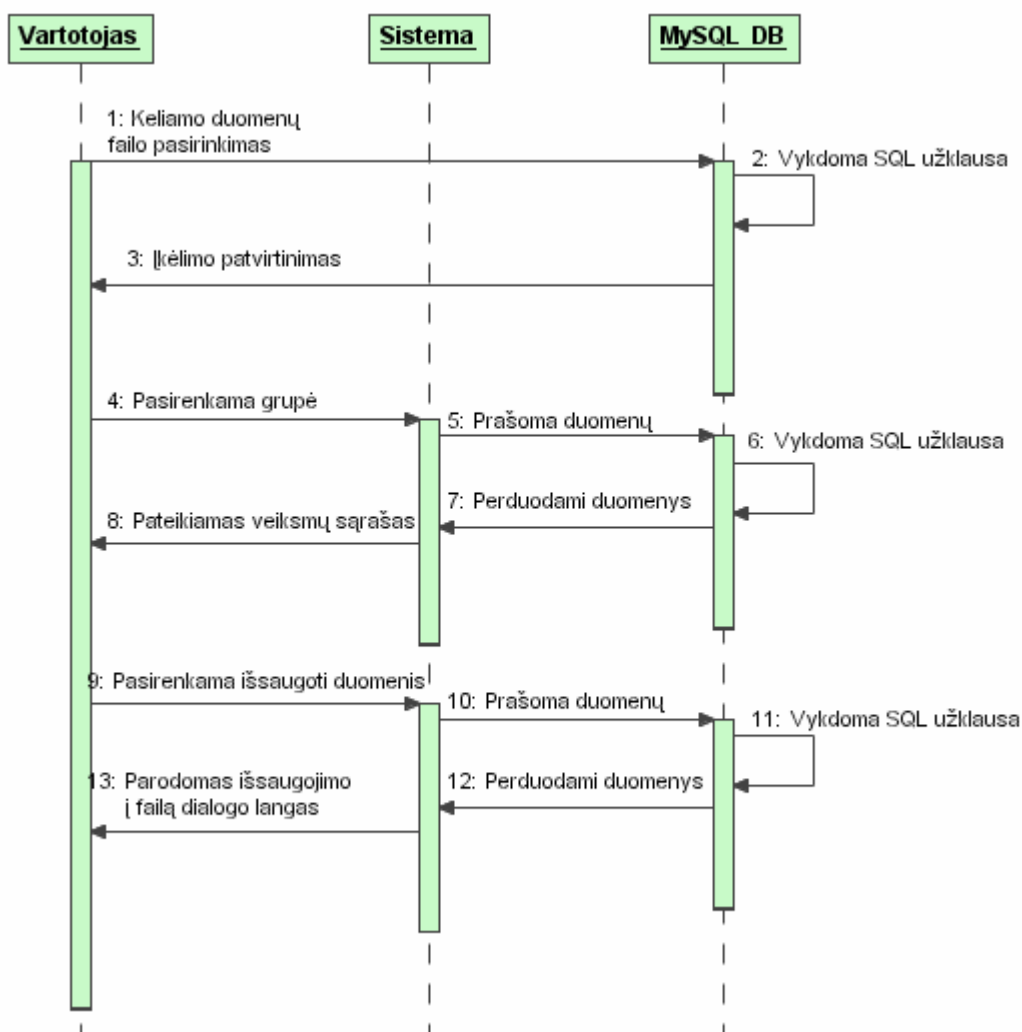
12 pav. Duomenų įrašymo į *.csv failą mygtukas

Norint įsitikinti, ar pateikiami reikalingi duomenys, galima failą tik atidaryti, tačiau patariama jį pirma išsisaugoti (žr. paveikslą nr. 13).



13 pav. Failo siuntimo langas

Paveiksle nr. 14 pavaizduota sekų diagrama, kurioje atvaizduotas etapas nuo duomenų failo įkėlimui pasirinkimo iki *.csv failo su tos grupės duomenimis gavimo.



14 pav. Sekų diagrama duomenų failo gavimui

4.3. Duomenų gavybos pritaikymas

Panagrinėkime virtualios nuotolinių studijų aplinkos duomenų gavybą besimokančiųjų intensyvumą klasterinės analizės pavyzdžiu.

Prieš atliekant analizę reikėtų apgalvoti kokie veiksmai, įrašyti į įvykių žurnalą, yra svarbūs, o kurių galima visai neįtraukti. Lentelėje nr. 3 matome visų veiksmų sąrašą ir sprendimą – naudoti jį analizėje ar ne. Veiksmų pasirinkimą lėmė jų įvykimo dažnumas – nesirenkami neįvykę ar tik kelis kartus įvykę veiksmai. Taip pat buvo nenagrinėjami veiksmai, nesusiję su intensyviu studijų procesu.

3 lentelė. Stebėtų veiksmų naudojimas analizėje

Eil. nr.	Veiksmo pavadinimas	Naudojamas analizėje
1.	Prisegtuko įkėlimas	Taip
2.	Užduoties peržiūra	Taip
3.	Visų užduočių peržiūra	Taip
4.	Dienoraščio peržiūra	Ne
5.	Pokalbis	Ne
6.	Pokalbio peržiūra	Ne
7.	Visų pokalbių peržiūra	Ne
8.	Alternatyvos pasirinkimas	Ne
9.	Pasirinkimo peržiūra	Ne
10.	Kurso atnaujinimas	Ne
11.	Kurso peržiūra	Taip
12.	Forumo užsakymas	Ne
13.	Forumo užsakymo atšaukimas	Ne
14.	Forumo vartotojo pranešimas	Ne
15.	Paieška forume	Ne
16.	Diskusijos forume peržiūra	Taip
17.	Forumo peržiūra	Taip
18.	Forumų peržiūra	Taip
19.	Žodyno peržiūra	Ne
20.	Viso žodyno peržiūra	Ne
21.	Žurnalo atnaujinimas	Ne
22.	Žurnalo peržiūra	Ne
23.	Viso žurnalo peržiūra	Ne
24.	Testo sprendimas	Taip
25.	Bandymas baigti testą	Ne
26.	Testo peržiūra	Taip
27.	Viso testo peržiūra	Taip
28.	Resursų peržiūra	Taip
29.	Visų resursų peržiūra	Ne
30.	Įkėlimas	Taip
31.	Vartotojo informacijos peržiūra	Ne
32.	Visų vartotojų informacijos peržiūra	Ne

Gautus po išankstinio apdorojimo duomenis patogiu pateikti įvairiais formatais, tinkamais pritaikyti įvairius statistinės analizės ir duomenų gavybos įrankius (SPSS, STATISTICA 7, IBM DB2 Intelligent Miner ir t.t.).

Gauti duomenys buvo apdoroti StatSoft STATISTICA 7 programinės įrangos paketu. STATISTICA 7 – duomenų analizės ir duomenų gavybos sistema, apimanti STATISTICA 7 duomenų gavybos programinę įrangą, siūlo labiausiai visapusišką ir efektyvią sistemą su vartotojui paprastai naudotinais įrankiais, skirtais visam duomenų gavybos procesui – nuo duomenų bazių užklausų iki sugeneruotų galutinių ataskaitų. StatSoft duomenų gavybos ir prognozių modeliavimo programinė įranga skirta naudoti vienoje kompiuterizuotoje darbo vietoje, keliose darbo vietose (duodama kelių vartotojų licencija) arba galimas įmonei skirtas programos leidimas [13]. StatSoft STATISTICA 7 yra įdiegta ir plačiai taikoma analizei Šiaulių universitete.

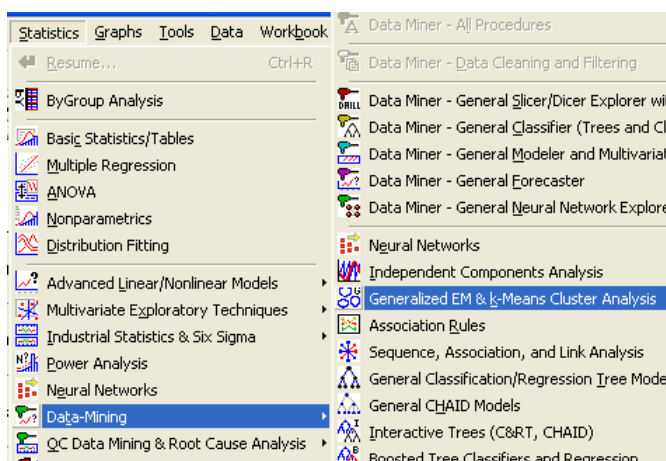
Darbai su STATISTICA 7 turime apibendrintus vienos grupės studentų duomenis, gautus iš duomenų bazės (žr. paveikslą nr. 15). Šiems duomenims apdoroti pritaikyta klasterinė analizė. Būtent šis metodas pasirinktas tam, kad būtų galima suskirstyti studentus į grupes pagal panašius mokymosi ypatumus ir pagal tai juos skirstyti į tokias grupes, kuriose būtų pasiekti geriausi mokymosi rezultatai. Klasterinė analizė arba klasterizavimas yra objektų skirstymas į grupes (vadinamas klasteriais) taip, kad tos pačios grupės objektai būtų kiek įmanoma panašesni tarpusavyje nei su objektais iš kitų klasterių. [2] Gautas jungimo sprendinys paprastai atvaizduojamas grafiku. Pagal šį grafiką matomas klasterių reikšmių pasiskirstymas. [1]

	1	2	3	4	5	6	7	8	9	10	11	12
1. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
2. Anoniminiai	11	11	11	11	11	11	11	11	11	11	11	11
3. Anoniminiai	8	8	8	8	8	8	8	8	8	8	8	8
4. Anoniminiai	7	7	7	7	7	7	7	7	7	7	7	7
5. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
6. Anoniminiai	23	23	23	23	23	23	23	23	23	23	23	23
7. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
8. Anoniminiai	21	21	21	21	21	21	21	21	21	21	21	21
9. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
10. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
11. Anoniminiai	2	3	5	7	7	7	7	7	7	7	7	7
12. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
13. Anoniminiai	1	24	1	7	7	7	7	7	7	7	7	7
14. Anoniminiai	7	7	7	7	7	7	7	7	7	7	7	7
15. Anoniminiai	1	8	7	7	7	7	7	7	7	7	7	7
16. Anoniminiai	1	10	7	7	7	7	7	7	7	7	7	7
17. Anoniminiai	5	14	13	7	7	7	7	7	7	7	7	7
18. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7
19. Anoniminiai	7	8	7	7	7	7	7	7	7	7	7	7

15 pav. Duomenys, gauti iš pradinio duomenų paruošimo

Klasterizavimui pasirinktas duomenų gavybos (angl. data mining) k-vidurkių (angl. k-means) metodas (žr. paveikslą nr. 16). K-vidurkių klasterizavimo metodas yra klasterinės

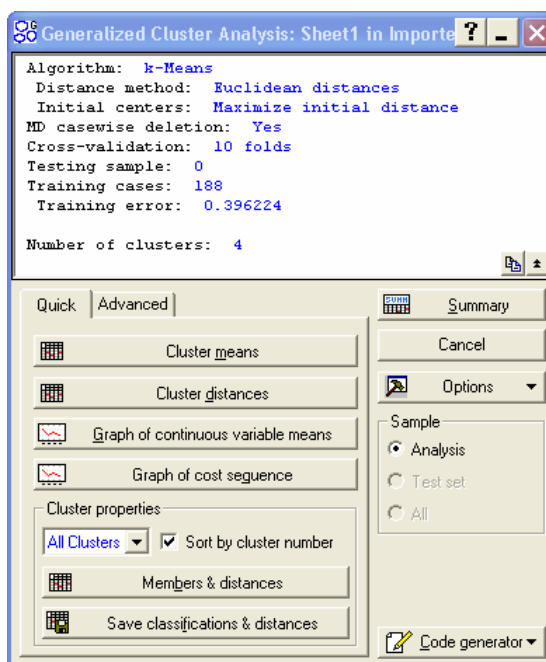
analizės metodas, kurio metu n stebimųjų yra paskirstomi į k kieki grupių (klasterių). Klasteris kiekvienam stebimam objektui parenkamas pagal jo ir klasterio vidurkio panašumą.



16 pav. Klasterizavimo metodo parinkimas

Atlikus viso kurso studentų klasterinę analizę paaiškėjo, jog studentus geriausia skirstyti į 4 klasterius (tiek grupių ir turėjome) (žr. paveikslą nr. 17). Klasterių sudėtis:

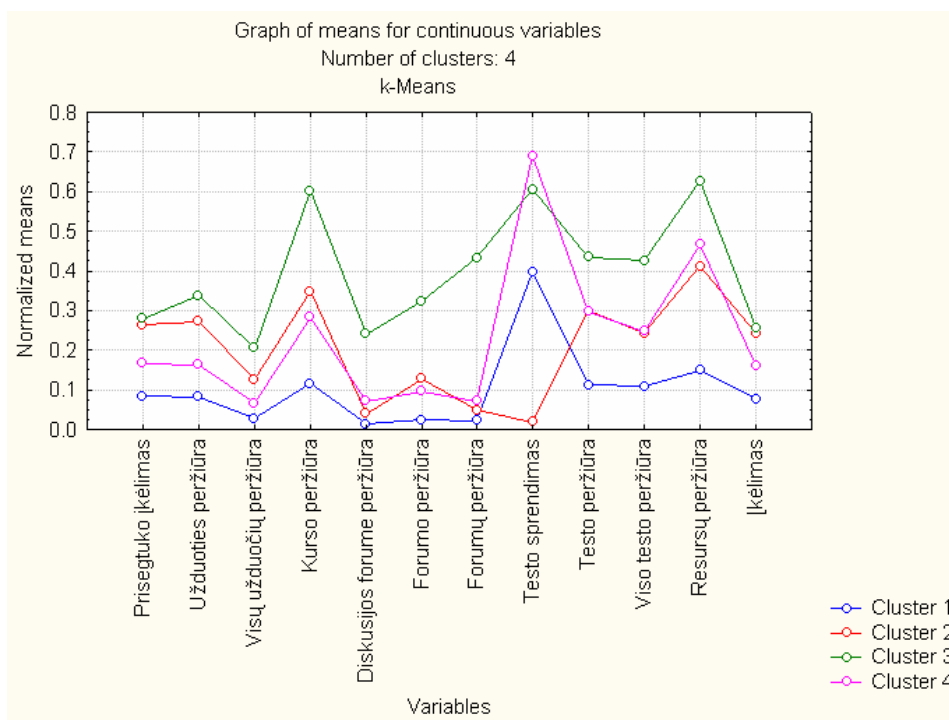
- 1 klasteris – 60 narių;
- 2 klasteris – 36 nariai;
- 3 klasteris – 21 narys;
- 4 klasteris – 71 narys.



17 pav. Klasterinės analizės rezultatas

Nustačius duomenų klasterius labai svarbu juos interpretuoti arba paaiškinti tiriamos dalykinės srities požiūriu. Klasterių vidurkių pasiskirstymo grafiką matome paveiksle nr. 18. Į

pirmą klasterį pateko studentai, kurie vidutiniškai prisegė iki trijų darbų, užduotis peržiūrėjo iki 15 kartų, kursą žiūrėjo iki 52 kartų ir t.t. (žr. paveikslą nr. 19). Atsižvelgiant į kitų klasterių studentų aktyvumą galima teigti, kad į pirmąjį klasterį pateko mažiausiai aktyvūs studentai. Taip pat lyginant kitų klasterių rezultatus galima teigti, kad į 3 klasterį pateko aktyviausi studentai (žr. lentelę nr. 4).



18 pav. Klasterinės analizės grafikas

Centroids for k-means clustering (Sheet1 in Imported from C:\Documents and Settings\Administrator\Desktop\Knyga1.xls)							
Number of clusters: 4							
Total number of training cases: 188							
Cluster	Prisegtuko įkėlimas	Užduoties peržiūra	Visų užduočių peržiūra	Kurso peržiūra	Diskusijos forume peržiūra	Forumo peržiūra	Forumų peržiūra
1	2.98333	14.85000	2.21667	51.8333	5.6833	6.08333	1.63333
2	9.47222	49.66667	10.52778	146.8056	17.0278	31.86111	3.52778
3	10.04762	61.57143	17.19048	250.2857	103.3333	80.85714	31.61905
4	5.98592	29.61972	5.50704	120.2958	30.3803	24.00000	5.22535

Cluster	Testo sprendimas	Testo peržiūra	Viso testo peržiūra	Resursų peržiūra	Įkėlimas	Number of cases	Percentage(%)
1	4.366667	22.31667	4.43333	22.55000	2.733333	60	31.91489
2	0.194444	47.61111	9.97222	62.55556	8.750000	36	19.14894
3	6.666667	66.00000	17.42857	95.33333	9.238095	21	11.17021
4	7.577465	47.38028	10.18310	70.85915	5.788732	71	37.76596

19 pav. Klasterių vidurkiai

Eil. Nr.	Veiksmų pavadinimai	1 klasteris	2 klasteris	3 klasteris	4 klasteris
1.	Prisegtuko įkėlimas	2,9833	9,47222	10,04762	5,98592
2.	Užduoties peržiūra	14,85000	49,66667	61,57143	29,61972
3.	Visų užduočių peržiūra	2,21667	10,52778	17,19048	5,50704
4.	Kurso peržiūra	51,8333	146,8056	250,2857	120,2958
5.	Diskusijos forume peržiūra	5,6833	17,0278	103,3333	30,3803
6.	Forumo peržiūra	6,08333	31,86111	80,85714	24,00000
7.	Forumų peržiūra	1,63333	3,52778	31,61905	5,22535
8.	Testo sprendimas	4,366667	0,194444	6,666667	7,577465
9.	Testo peržiūra	22,31667	47,61111	66,00000	47,38028
10.	Viso testo peržiūra	4,43333	9,97222	17,42857	10,18310
11.	Resursų peržiūra	22,55000	62,55556	95,33333	70,85915
12.	Įkėlimas	2,733333	8,750000	9,238095	5,788732

4.4. Rezultatai ir rekomendacijos

Darbo metu sudaryta sistema, skirta duomenų, gautų iš virtualios nuotolinio mokymo aplinkos Moodle, pirminiam apdorojimui ir paruošimui duomenų gavybai. Gauti duomenys ištirti STATISTICA 7 programa, pritaikant klasterizavimo k-vidurkiu metodu. Šis metodas parodė, jog geriausia studentus skirstyti į 4 klasterius pagal v-fold kriterijų.

Tolimesniam naudojimui sukurta programinė įranga galima pritaikyti tiesioginiam duomenų gavimui iš virtualios nuotolinio mokymo aplinkos Moodle, kad būtų išvengta papildomo duomenų eksportavimo iš pradžių iš Moodle aplinkos, o po to importavimo į programinės įrangos duomenų bazę.

Viena iš duomenų pirminio apdorojimo užduočių taip pat galėtų būti duomenų diskretizavimas. Skaitmeninių verčių diskretizavimas būtinas tam, kad būtų galima duomenis interpretuoti. Diskretizacija skirsto skaitmeninius duomenis į kategorines klases, kurios lengviau suprantamos dėstytojui (kategorinės vertės yra patogesnės dėstytojui negu tikslūs dydžiai ar diapazonas). Galima diskretizuoti visas skaitmenines apibendrinimo lentelės vertes išskyrus kurso identifikavimo numerį.

Taip pat galima duomenis pritaikyti analizei atsižvelgiant ne tik į atliktus veiksmus aplinkoje, bet ir į prisijungimų skaičių bei laiką, praleistą prisijungus nuotolinėje aplinkoje, laiką, skirtą intensyviai mokymuisi, nustatant plagiato arba nesąžiningo elgesio faktus, tobulinant mokymosi programas ir kursus arba grupuojant studentus į besimokančiųjų grupes. Tam būtų galima pritaikyti klasifikavimo, susietumo sąryšių, sprendimo medžių ir kitus duomenų gavybos metodus.

5. Išvados

1. Daugelis plačiai taikomų virtualių nuotolinio mokymo aplinkų neturi patogių duomenų pradinio apdorojimo sąsajų, leidžiančių pasinaudoti žinomomis duomenų analizės technologijomis.
2. Ištirtas virtualios nuotolinio mokymo aplinkos duomenų gavybos procesas ir pagrindiniai etapai.
3. Pasinaudojus PHP programavimo kalbos priemone darbe sudaryta programinė įranga pirminiam virtualios nuotolinio mokymo aplinkos duomenų apdorojimui formatais, tinkamais plačiai naudojamuose statistinės analizės ir duomenų gavybos įrankiuose (Statistica 7, SPSS, IBM DB2 Intelligent Miner).
6. Ištirtos duomenų gavybos metodų, realizuotų su statistinės analizės paketu StatSoft STATISTICA 7, taikymo galimybės nuotolinių studijų duomenų analizei besimokančiųjų intensyvumo klasterinės analizės pavyzdžiu.
7. Parengtos rekomendacijos nuotolinių studijų veiklų duomenų gavybai.

Literatūros sąrašas

1. Butkevičienė E., Rutkauskienė D., Daukilas S., Gudonienė D., Mušankovienė R. E. mokymosi ypatybių švietimo sektoriuje studija [žiūrėta 2009-04-21]. Prieiga per internetą: http://www.lvu.lt/cms/files/liedm/resources/3010_E._mokymosi_ypatybiu_svietimo_sektoriuose_studija.pdf
2. Cluster analysis apibrėžimas [žiūrėta 2009-03-11]. Prieiga per internetą: http://en.wikipedia.org/wiki/Data_clustering
3. CSV apibrėžimas [žiūrėta 2009-04-23]. Prieiga per internetą: <http://lt.wikipedia.org/wiki/CSV>
4. Duomenų gavyba [žiūrėta 2008-12-02]. Prieiga per internetą: <http://www.mokslai.lt/referatai/referatas/duomenu-gavyba.html>
5. Internet World Stats [žiūrėta 2009-04-20]. Prieiga per internetą: <http://www.internetworldstats.com/europa.htm#lt>
6. Kas yra PHP? [žiūrėta 2009-04-13]. Prieiga per internetą: <http://www.php.lt/render/Articles;aid,27>
7. Kurasova O. Daugiamačių duomenų vizuali analizė taikant savireguliuojančius neuroninius tinklus (SOM). Daktaro disertacija. Vilnius, 2005
8. Moodle oficiali svetainė [žiūrėta 2009-02-05]. Prieiga per internetą: <http://www.moodle.org>
9. Naudojimosi Virtualiomis mokymosi aplinkomis ir Nuotoliniais mokymo kursais tvarka [žiūrėta 2009-01-30]. Prieiga per internetą: http://www.emokykla.lt/doc/VMA/files/vma_ir_nmk_tvarka.doc
10. OLAP duomenų bazės [žiūrėta 2009-03-17]. Prieiga per internetą: http://itekspertas.projektas.lt/index.php?option=com_content&task=view&id=94&Itemid=66
11. Romero C., Ventura S., Garsia E. Data Mining in course management systems: Moodle case study and tutorial. [žiūrėta 2008-09-27]. Prieiga per internetą: <http://sci2s.ugr.es/docencia/doctoM6/Romero-Ventura-Garcia-CE.pdf>
12. Sakalauskas L., Predys S. Nuotolinių studijų vartotojų poreikių analizė
13. *STATISTICA* Data Mining and Predictive Modeling Solutions [žiūrėta 2009-03-10]. Prieiga per internetą: <http://www.statsoft.com/datamining/dataminingsolutions.html>

Anotacija

Daiva Lapukaitė. Informatikos magistro darbas. Virtualios nuotolinio mokymo aplinkos duomenų gavyba. Darbo vadovas: prof. habil. dr. Leonidas Sakalauskas. Šiaulių universitetas. Šiauliai, 2009. 27 puslapiai.

Nuotolinių aplinkų duomenų bazėse kaupiamas didelis kiekis informacijos apie studentus ir jų veiksmus nuotolinėje aplinkoje. Kad būtų paprasčiau analizuoti šiuos duomenis į pagalbą pasitelkiama duomenų gavyba.

Darbo tikslas - sudaryti sistemą, skirtą duomenų, gautų iš virtualios nuotolinio mokymo aplinkos Moodle, pirminiam apdorojimui ir duomenų gavybai. Gautus duomenis iširti pritaikant programinės įrangos paketo StatSoft STATISTICA 7 duomenų gavybos algoritmus besimokančiųjų mokymosi intensyvumo duomenų analizei. Įvertinus gautus rezultatus parengtos rekomendacijos tolimesnei duomenų analizei. Duomenų analizei pritaikyta klasterinė k-vidurkių analizė.

Summary

Daiva Lapukaite. The final work for master's degree in informatics: Data Mining in virtual learning environment supervised by professor habilitated dr. L. Sakalauskas. Siauliai University. Siauliai, 2009. 27 p.

The databases of virtual learning environments store large quantity of information about students and their activity. The data mining is usable to easier analysis of these information.

The object of work is to make a system for preprocessing and data mining of data, obtained from virtual learning environment Moodle. The historical learning data can be analysed after preprocessing to study learners learning intensity with data mining algorithms by StatSoft STATISTICA 7 software. k-means cluster analysis was applied as example of data mining of learning data. Recommendations to further application of data mining of learning activities are given, too.

Priedai

Kompaktinio disko turinys

- Katalogas „Realizacija“ – sukurtos programinės įrangos realizacijos failai ir į tekstinį failą „DB.txt“ eksportuota duomenų bazė.
- Katalogas „Duomenys“ – duomenų, gautų iš Šiaulių universiteto nuotolinių studijų centro, *.xls failai.
- Katalogas „Aprašymas“ – darbo ataskaita „Ataskaita_D.Lapukaite.doc“ ir „Ataskaita_D.Lapukaite.pdf“.