

Analiza primjena rudarenja podataka u sustavima za e-učenje

Igor Jugo

Odjel za informatiku, Sveučilište u Rijeci
ijugo@inf.uniri.hr

Sažetak – *Educational data mining* je relativno nov naziv za područje istraživanja u kojem se metode rudarenja podataka (temeljene na algoritmima iz područja umjetne inteligencije i strojnog učenja), ali i neke nove metode razvijene u ovom području, primjenjuju na podatke dobivene iz sustava za e-učenje, s ciljem pružanja informacija, razjašnjavanja fenomena, te pružanja osnove za automatsku prilagodbu sustava potrebama, očekivanjima i navikama korisnika. Takve sustave nazivamo inteligentnim i prilagodljivim sustavima za e-učenje. Jedna od najvećih prednosti rudarenja podataka je njegova nenametljivost. Na ovaj način koristimo podatke koje nam je, kroz svoju interakciju sa sustavom, dao korisnik bez da smo to od njega eksplicitno zahtjevali, što je postupak koji u podatke uvijek unosi subjektivnost korisnika. Prednost je i u tome što možemo raditi i vremenske analize ponašanja korisnika i ponoviti sve njegove interakcije sa sustavom. Cilj ovog rada bila je kvalitativna analiza područja s ciljem identifikacije ključnih izvora, radova, autora i smjerova istraživanja, uz poseban osvrt na područja od interesa autora za daljnje istraživanje pri izradi doktorske disertacije - vizualizaciju podataka i dobivanje povratnih informacija za podršku nastavnicima. Na kraju rada dan je i pregled istraživanja u RH i susjednim zemljama.

Ključne riječi: rudarenje podataka, sustavi za e-učenje, vizualizacija podataka, prilagodljivi sustavi za učenje.

I. UVOD

Rudarenje podataka (engl. *Data Mining*, dalje: RP) je naziv za fazu analize velike količine podataka u procesu otkrivanja novog znanja iz tih podataka (engl. *knowledge discovery in databases*). RP se bavi izvlačenjem ili otkrivanjem novih i korisnih informacija iz podataka na temelju otkrivanja ponavljajućih uzoraka. Otkrivanje znanja temelji se na brojnim algoritmima od kojih je većina razvijena u području umjetne inteligencije i strojnog učenja koji tako čine njegovu tehničku osnovu. Strojno učenje opisuje se kao pronalaženje/dohvaćanje strukturalnih opisa iz primjera. Ti opisi se kasnije koriste za predviđanje i, što je najveća prednost ispred statističkih modela, objašnjavanje i razumijevanje fenomena koji se istražuju [1].

II. EDUCATIONAL DATA MINING

Educational data mining (dalje: EDM) je jedno od najnovijih područja istraživanja nastalo primjenom i/ili prilagodbom metoda RP na sustave za e-učenje. Pored toga, jedan od važnih ciljeva je razvoj potpuno novih

metoda za poboljšanje e-učenja odnosno učenja općenito. Iako su brojni autori primjenjivali ove metode na sustave za e-učenje već 90-ih godina, ovo područje istraživanja je danas široko prihvaćeni naziv dobilo 2007. godine izlaskom preglednog rada [2] koji je obuhvatio sve najvažnije radove u periodu od 1995 do 2005. godine.

U tom radu istaknute su ključne specifičnosti tog područja koje zahtjevaju prilagodbu metoda RP: nova domena istraživanja (e-učenje), nove karakteristike i izvori podataka (manje količine podataka u odnosu na transakcijske sustave, više mogućih izvora – log datoteke, baze podataka, analogni izvori), specifični ciljevi (poboljšanje učenja i sustava za e-učenje) i nove tehnike (neke tehnike RP mogu se direktno primjeniti na ovu domenu, dok je druge potrebno prilagoditi, a potrebno je i razviti nove).

Također su istaknuti mogući subjekti primjene:

- Studenti – dobivaju unaprijedene sustave za e-učenje, koji se prilagođavaju potrebama studenta, sugeriraju dobre prakse, nude sugestije o načinima prolaska kroz materijale ili nude dodatne materijale
- Učitelji – dobivaju više povratnih informacija o učenju i efikasnosti procesa učenja, mogućnosti klasifikacija učenika, pronalaženja uzoraka ponašanja i predviđanja posljedica, pronalaženja najčešćih grešaka, sugestije u prilagodbi materijala i strukture znanja
- Administratori – dobivaju bolji uvid u općenitu efikasnost sustava, koji se može koristiti za poboljšanje programa studija.

Iako se metode RP mogu primjenjivati na sve tipove podataka prikupljene u procesu učenja, danas se EDM pretežno koristi dinamički generiranim podacima nastalima prilikom interakcija korisnika sa sustavima za e-učenje, a najčešće s ciljem dodavanja karakteristike prilagodljivosti (engl. *adaptive and intelligent web-based educational systems* - AIWES).

Ovi sustavi su nastali zajedničkim razvojem inteligentnih sustava za poučavanje (engl. *intelligent tutoring systems*) i prilagodljivih hipermedijskih sustava (engl. *adaptive hypermedia systems*) [3]. Razvijen je velik broj ovakvih sustava, često za nastavne i istraživačke potrebe pojedinih sveučilišta.

Tablica 1. Pregled tema istraživanja u području EDM u periodu '95.-'05.

Metoda	1995	1998	2000	2001	2002	2003	2004	2005	UKUPNO
<i>Sequence pattern</i>	1	1		1	1	2	1		7
<i>Prediction</i>			1		1		1	1	4
<i>Classification</i>			2			1	2	1	6
<i>Association</i>			3	1	1		6	1	12
<i>Text mining</i>			1				2	3	6
<i>Visualization</i>					1			2	3
<i>Statistics</i>		1				2	1	3	7
<i>Clustering</i>					1	1	3	1	6
<i>Outlier detection</i>							1	2	3
UKUPNO	1	2	7	2	5	6	17	14	

U svom pregledu autori daju tablicu objavljenih radova koji se bave primjenom metoda RP u sustavima za učenje. Tablica referencira ukupno 48 radova, podjeljenih prema metodi RP koju primjenjuju. Podatci iz spomenute tablice (u skraćenom obliku bez navođenja svih referenci i tipa sustava za učenje na koji su metode primjenjivane) prikazani su u Tablici 1. Napomena: manji broj radova je koristio dvije metode, pa smo umjesto uvođenja takvih kategorija broj radova za obje metode uvećali za 1.

Iako se u tablici navodi 48 radova, cjeloviti rad navodi 81 referencu.

Iste godine objavljen je još jedan pregledni rad pod nazivom „*Applying Data Mining Techniques to e-Learning problems*“ [4] koji je imao za cilj dati cjeloviti pregled područja u periodu od 1999-2006. godine. Rad navodi 99 referenci, a pregled radova je dan kroz pet tablica koje su predstavljale područja primjene EDM:

1. Procjena rada učenika
2. Prilagodba sadržaja i preporuke
3. Evaluacija i testiranje
4. Povratne informacije temeljene na ponašanju/aktivnostima učenika
5. Detekcija atipičnog ponašanja učenika

Sve tablice su imale istu strukturu (referenca, metoda, tehnika, sudionik, vrsta publikacije). U Tablici 2. ispod prikazani su sumirani rezultati iz svih pet tablica.

Pored navedenog, u radu se daju brojni pregledi: trenutnih projekata na području povezivanja metoda RP i sustava za e-učenje, važnih časopisa i konferencija specijaliziranih za e-učenje, ključnih radova i knjiga na području e-učenja, repozitorija objekata za učenje,

Tablica 2. Korištene tehnike RP po smjerovima istraživanja

Tablica	1	2	3	4	5	UKUPNO
Cilj	Referenci					
Statistics	5		1	2		8
Classification	18	14	6	6		44
Clustering	5	1	1	2		9
Visualization	4		1			5
Prediction	8	6		6		20
Outlier detection					3	3
	40	21	9	16	3	89

standarda i pregled postojećih sustava za e-učenje otvorenog koda.

Godine 2008. održana je prva konferencija o EDM-u, a sljedeće godine izdan je prvi broj časopisa *Journal of Educational Data Mining*¹.

U tom časopisu objavljen je sljedeći rad koji je donio pregled razvoja EDM [5]. U tom radu se prvi puta, uz osnovnu klasifikaciju metoda RP koje se koriste u EDM (predviđanje, klasterizacija, otkrivanje pravila, statistika, vizualizacija), spominje otkrivanje (znanja) pomoću modela (engl. *discovery with models*). Model neke pojave razvija se bilo kojim procesom koji se može provjeriti (npr. predviđanje), a zatim se koristi u drugim analizama (npr. otkrivanje pravila). Ovaj pristup daje nove mogućnosti analize pa se tako primjerice koristi za otkrivanje od kojih dijelova materijala za učenje studenti imaju najviše koristi [6], kako tipovi ponašanja na različite načine utječu na učenje [7] i kako varijacije u dizajnu inteligentnih sustava za poučavanja utječu na ponašanje studenata tijekom vremena [8].

Ovaj rad donosi pregled ključnih područja primjene EDM metoda:

- unaprijeđenje modela studenata radi unaprijeđenja učenja
- otkrivanje i unaprijeđivanje strukture domene znanja
- istraživanje pedagoške podrške (npr. suradničko učenje)
- traženje empirijskih dokaza za usavršavanje i proširivanje teorija učenja

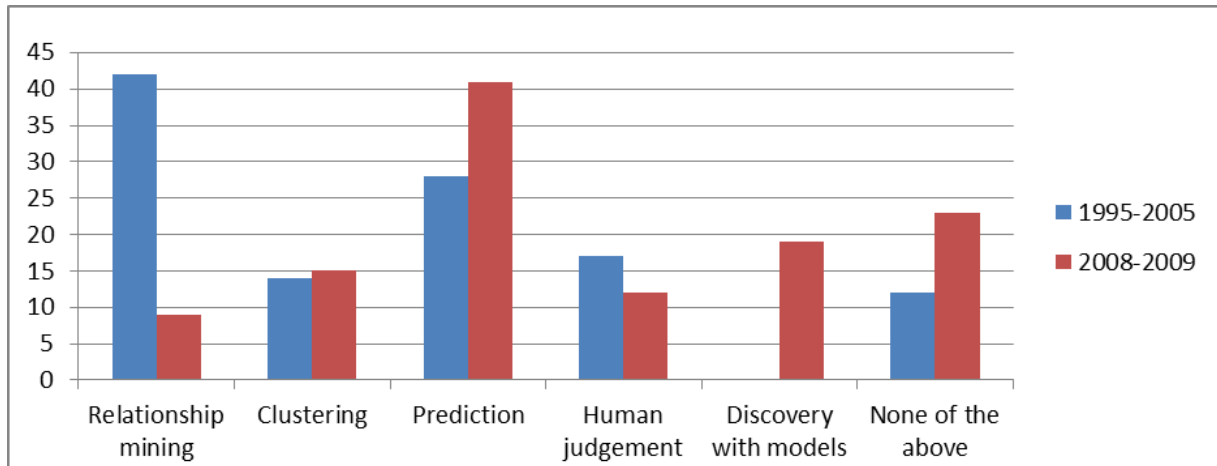
U radu se također daje popis 8 najcitiranih radova spomenutih u preglednom radu Romera [2].

U radu je napravljena i analiza promjena u korištenim metodama opisanima u [2] i metodama korištenima u radovima objavljenima na EDM konferencijama 2008 i 2009. godine².

Usporedni prikaz područja istraživanja iz oba perioda prikazan je na Slici 1. U radu se također ističe važnost

¹ <http://www.educationaldatamining.org/JEDM>

² <http://www.educationaldatamining.org/EDM2008> i <http://www.educationaldatamining.org/EDM2009/>



Slika 1. Usporedba područja istraživanja po periodima

pojave javnih repozitorija podataka o e-učenju kao što je Pittsburgh Science of Learning Center (PSLC) Data Shop³, te sve veća podrška prikupljanju podataka u najraširenijim alatima za učenje poput Moodle-a i WebCT-a.

Sljedeći pregledni rad, ujedno i posljednji koji se pojavio do danas [9], je proširena i ažurirana verzija rada [2] koja obuhvaća period do 2010. godine (rad je objavljen krajem 2010). Rad obuhvaća čak 306 referenci, koje pokazuju potpuno novu klasifikaciju područja istraživanja unutar EDM prikazanu na Slici 2.

Važno je istaknuti buduće smjerove istraživanja u području EDM koje navode autori: a) učiniti alate za RP pristupačnijima učiteljima, b) standardizacija metoda i podataka o e-učenju, c) integracija alata za RP sa sustavima za e-učenje i d) razvoj specifičnih/efikasnijih metoda RP koje integriraju znanje iz domene obrazovanja.

Smjerovi navedeni pod a) i c) poklapaju se s ciljevima našeg istraživanja za potrebe izrade doktorske disertacije. Metode rudarenja podataka u sustavima za e-učenje ili nastavi općenito još su daleko od široke primjene, osobito u osnovnim i srednjim školama. Smatramo da bi se integracijom alata i tehnika rudarenja podataka u sustave za e-učenje, te prilagodbom sučelja značajno povećao broj korisnika takvih sustava, te u konačnici povećala kvaliteta učenja. Naše istraživanje bi pri tome spojilo dva segmenta istraživanja u ovom području – vizualizaciju podataka (u web pregledniku) i prilagodbu nastavnih materijala na temelju primjene metoda RP. U nastavku donosimo pregled radova koji se bave spomenutim problemima.

III. VIZUALIZACIJA PODATAKA

Vizualizacija podataka je naziv za grafičku prezentaciju informacija, s ciljem otkrivanja i razumijevanja uzoraka u velikim skupovima podataka.

Vizualizacija podataka je usko povezana uz područje statističke obrade podataka i rudarenje podataka. Grafičkim prikazom podataka iskorištava se ogroman potencijal čovjeka za uočavanje uzoraka, iznimaka i važnih detalja koji u tabličnom obliku teško mogu biti otkriveni. Rezultati rudarenja podataka vrlo često se

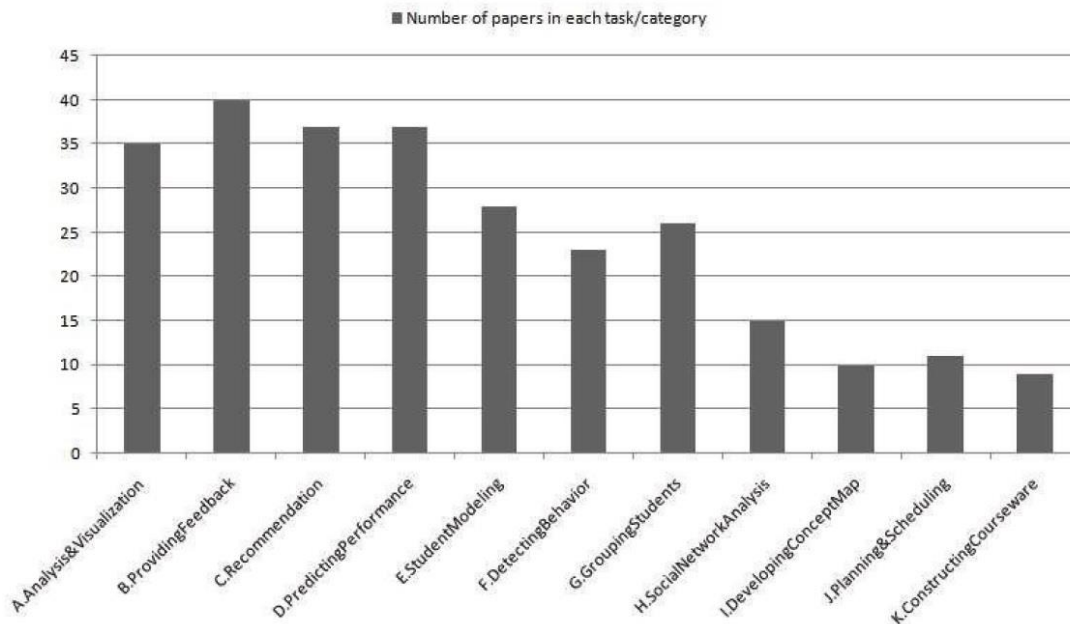
prikazuju u 2D, a sve više u 3D grafikonima, te u animacijama. Kompleksnije vizualizacije često se koriste animacijom kako bi se istaknule promjene podataka u vremenu.

U području e-učenja, raste svijest o potrebi i vrijednosti vizualizacija za sve sudionike tog procesa. Učenici lakše i dublje shvaćaju složenije koncepte putem vizualizacija, nastavnici dobivaju bolji uvid u aktivnosti učenika ili skupine, dok administracija dobiva uvid u rezultate na razini cijelog programa ili institucije. Primjer korištenja vizualizacija u nastavnom radu (uz primjenu EDM-a), dan je u [10], a demonstrira vrijednost vizualizacije rezultata grupe učenika. U radu se opisuje vizualizacija aktivnosti učenika pri timskom rješavanju zadataka iz programiranja. Kao vizualizacija korišten je motiv stabla (grane, cvijetovi) koje vrlo jasno svakom studentu pokazuju razinu njegove aktivnosti u odnosu na grupu. Ovakve vizualizacije mogu se koristiti kao motivacijski element čime se istraživači i nastavnici vrlo često koriste u području učenja putem ozbiljnih igara. U drugom radu [11] demonstrira se korištenje vizualizacija uz pomoć kojih nastavnici mogu dobiti sveobuhvatni ili detaljni pregled aktivnosti, rezultata i uzoraka ponašanja cijele grupe ili jednog učenika, te izvući sugestije za promjenu strukture nastavnih materijala.

Vizualizacija podataka razvila se iz potrebe za prezentacijom (bitnih, skrivenih) informacija temeljenih na ogromnim količinama podataka s kojima se ljudi svakodnevno susreću. Razvoj vizualizacija trenutno ide u mnogim smjerovima s ciljem prezentiranja brojnih vrsta i izvora podataka, korištenjem najrazličitijih tehnologija. Vizualizacije mogu biti 2D ili 3D, animirane ili statične, biti povezane sa “živim” tokom podataka i stalno se osvježavati, te imati karakteristiku interaktivnosti.

Kao najveći izazovi prilikom generiranja vizualizacija ističu se: točnost i sukladnost podataka, isticanje korisnih informacija i novog znanja (tj. izbjegavanje sugestija prema pogrešnim zaključcima), te prilagođenost području.

³ <https://pslclatashop.web.cmu.edu/>



Slika 2. Usporedba tema istraživanja u području EDM [9]

Očekuje se da će glavni pravci razvoja biti u području dinamičkih vizualizacija “živih” tokova podataka, te interaktivnosti, što će korisnicima omogućiti veće mogućnosti manipulacije, te lakše uočavanje novih informacija.

Razvojem alata za vizualizaciju podataka, te standardizacijom brojnih tokova podataka koji nas okružuju biti ćemo u mogućnosti povezivati interdisciplinarnе izvore podataka što će omogućiti nove uvide u veze između pojava koje ti izvori opisuju.

U izvješću HORIZON 2010 [12] New Media Consortiuma (međunarodna zajednica eksperata iz područja obrazovanja) i Educause inicijative (neprofitna organizacija IT stručnjaka i znanstvenika posvećena unaprijeđenju visokog obrazovanja) koje govori o *cutting edge* tehnologijama i pristupima koji će se u budućnosti koristiti u obrazovanju, „vizualna analiza podataka“ smještena je u period “usvajanja u praksi” od sljedećih 4-5 godina. Isti izvještaj spomenuo je više projekata na tom području koji su tada bili u razvoju: Gapminder.com (vizualizacija važnih svjetskih trendova), ManyEyes (IBM, Java, vizualizacija javno dostupnih podataka), FlowingData i Roambi (vizualizacije na tabletima i mobilnim uređajima). U istom izvješću izdanom 2013. godine [13] vizualizacija se spominje kao dio „*data analytics*“, a smještena je u srednjeročni period (sljedeće 2-3 godine) uz bok MOOC⁴ sustavima.

A. Vizualizacija podataka na WWW-u

Vizualizacija podataka na WWW-u pratila je razvoj tehnologija na kojima se WWW temelji. Taj razvoj mogli bi podijeliti u tri faze. Faza 1: Do razvoja skriptnih serverskih jezika vizualizacija podataka bila je ograničena na upload grafika na poslužitelj i njihovo uključivanje na željena mjesta u statičnim web stranicama. Faza 2: Sa

razvojem skriptnih serverskih jezika pojavila se je mogućnost dinamičkog generiranja grafika (najčešće PNG datoteka), te njihovo periodično osvježavanje prilikom generiranja sadržaja web stranice. Faza 3: Razvojem klijentskog skriptnog jezika JavaScript, a posebno platforme JQuery koja je ujedno olakšala pisanje novog koda, kao i standardizirala postupak razvoja novih modula (engl. *plugin*) omogućen je razvoj „živih“ vizualizacija (koje više nisu grafičke datoteke već se iscrtavaju na samoj web stranici), te su u mogućnosti gotovo trenutačno osvježavati svoj sadržaj korištenjem XMLHttpRequest (AJAX) poziva. Razvojem HTML5 specifikacije i sa njome povezanih tehnologija (Canvas, SVG podrška), te implementacijom istih u moderne web preglednike omogućene su dvije važne karakteristike vizualizacije podataka na WWW-u – interaktivnost i animacija.

Time je praktično dosegnut stupanj razvoja koji je postojao na klasičnim (desktop) sustavima (osim 3D animacije) i isključena potreba za instalacijom dodatka (Java, Flash, Silverlight) u web preglednike. Korisnici tako mogu: mijenjati raspone podataka, analizirati pojedine točke (skupove podataka), mijenjati skupove podataka na osima grafikona, dodavati ili uklanjati skupove ili tokove podataka, te pokretati animacije kako bi uočavali promjene u simuliranim ili realnim vremenskim intervalima.

U ovom trenutku postoji relativno malen broj gotovih biblioteka temeljenih na jeziku JavaScript, odnosno JQuery platformi, koje omogućavaju implementaciju brojnih postojećih vizualizacija ili razvoj vlastitih. Od tih biblioteka vrijedi istaknuti sljedeće:

- D3⁵ je JavaScript biblioteka koja se orjentirana na poštivanje standarda (HTML5, CSS3, SVG) s ciljem maksimalne kompatibilnosti sa web preglednicima. D3 izrađuje SVG vizualizacije

⁴ <http://www.educause.edu/library/massive-open-online-course-mooc>

⁵ <http://d3js.org/>

dodavanjem elemenata u strukturu web stranice (engl. *Document Object Model*⁶ – dalje: DOM) čime se korisniku prepušta maksimalna sloboda uređivanja sadržaja korištenjem CSS-a i JavaScripta. D3 omogućuje kreiranje animacija i interaktivnih grafika.

- Processing⁷ je projekt koji se razvija paralelno sa Processing vizualnim programskim jezikom (temeljen na jeziku Java) pisanim za Web. Vizualizacije se pokreću korištenjem HTML5 Canvas elementa,
- Raphael⁸ je JavaScript (jQuery) biblioteka sa naglaskom na jednostavnosti i slobodi kreiranja novih elemenata. Temelji se na SVG standardu., a omogućuje kreiranje grafikona, ali i karata, mreža pojmova, animacija, itd.

Jedan od problema u području e-učenja je taj što vizualizacija podataka nije dovoljno iskorištena. Primjerice, poznati sustavi za e-učenje poput Moodle-a nemaju u svojoj standardnoj instalaciji alate za vizualizaciju podataka. U biblioteci mogućih proširenja Moodle sustava postoji proširenje GISMO⁹ koje generira vizualizacije podataka o interakcijama korisnika sa sustavom, dobivene metodama deskriptivne statistike. GISMO generira vizualizacije korištenjem JQuery platforme i plugina JQPlot¹⁰, ali ne pruža mogućnosti animacije niti interakcije sa podacima.

Od najnovih radova u ovom području vrijedi istaknuti doktorsku disertaciju (u postupku izrade) po naslovom *Applied Visual Analytics for Higher Education Decision Making*⁴ [14]. Autor tog rada opisuje sustav koji će „koristiti tehnike rudarenja podataka i strojnog učenja, te ih usko povezati sa tehnikama vizualizacije (podataka) kako bi se omogućilo dobivanje vrijednih spoznaja i prepoznale ... veze između ... brojnih tokova podataka od

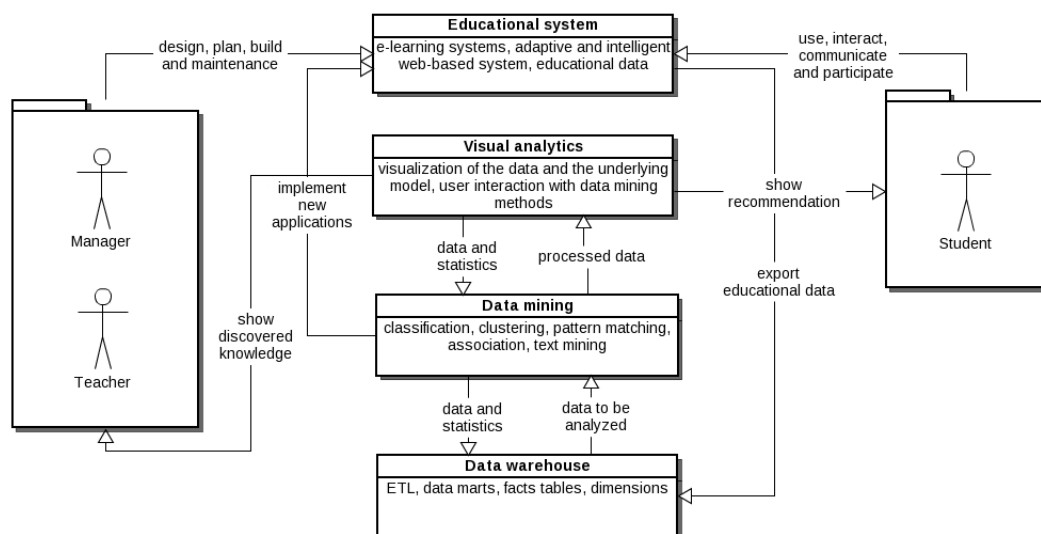
interesa za Sveučilište⁴. Metode analize (RP i SU) biti će prilagodljive vizualnim kontrolama, dok će sučelje biti „prilagođeno stručnjacima i laicima u području RP⁴. Autor također navodi da će se vizualizacije koristiti „kao polu-automatska tehnika analize podataka u fazi pred-procesiranja⁴. Alternacijom između vizualnih i automatskih metoda analize (RP) dobilo bi se trajno unaprjeđivanje i evaluacija privremenih rezultata. Korisnici će moći pristupiti ulaznim podacima i podešavati postavke metoda za rudarenje i vizualizaciju podataka. Predviđeni sustav bi sa ostatkom informacijskog sustava komunicirao putem PMML standarda. Potencijal (primjene) ovakvog sustava autor vidi u:

- vizualizaciji rasprava na forumu,
- vizualizaciji društvenih mreža,
- vizualizaciji sustava preporučenih zadataka ili nastavnih materijala.

Blok-shema predloženog sustava prikazana je na Slici 3.

U radu iz 2012. godine (područje bioinformatike) autori su demonstrirali korištenje Flex (Flash) tehnologije u web pregledniku za izradu kompleksnih vizualizacija nastalih primjenom metoda strojnog učenja [15].

U radu iz 2013. godine demonstriran je alat za RP u obliku „analytics-as-a-service“ kojem se pristupa putem web preglednika [16]. Alat omogućuje upload podataka od strane korisnika, nad kojima može pokrenuti nekoliko algoritama RP, te generirati brojne vizualizacije. Alat se oslanja na Java/Matlab platformu, te, što je vrlo rijetko, nudi prilagodbu predznanju korisnika (prikaz naprednih ili osnovnih rezultata). Sustav je namijenjen širokoj primjeni, a za potrebe ovog rada primjenjen na Sveučilištu Cantabria (Španjolska) kako bi pomogao nastavnicima analizirati uzorke ponašanja studenata i navigacijske putanje prilikom rada sa Moodle sustavom.



Slika 3. Struktura sustava predloženog u [14]

⁶ <http://www.w3.org/DOM/>

⁷ <http://processingjs.org/>

⁸ <http://raphaeljs.com/>

⁹ <http://gismo.sourceforge.net/>

¹⁰ <http://www.jqplot.com/>

Drugi autori bavili su se problemima povezivanja alata za RP i vizualizaciju rezultata, te njihovim prihvaćanjem od strane nastavnika. U [17] autori su opisali razvoj alata LOCO-Analyst kao samostalne Java aplikacije sa brojnim mogućnostima vizualizacije podataka (primjerice društvenih mreža), te njegovu primjenu u sustavu za e-učenje. Kasnije su analizirali stavove nastavnika i njihovo prihvaćanje takvog alata u praksi [18]. Taj rad posebno je zanimljiv jer se smješta u područje „analitike učenja“ (engl. *learning analytics*) koje se definira kao „mjerenje, prikupljanje, analiza i izvještavanje o podacima o učenicima i kontekstu učenja, sa ciljem razumijevanja i optimizacije učenja i okruženja u kojem se događa.“¹¹. *Learning analytics* opisuje se u [19] kao novo, interdisciplinarno područje koje povezuje EDM, sociologiju, računarstvo, statistiku, psihologiju i pedagogiju, a orijentirano je na razumijevanje učenja kao cjeline, te prilagodbu obrazovanja mogućnostima i potrebama svakog učenika.

Još jedan rad objavljen 2012. godine bavi se povezivanjem alata za RP i vizualizaciju interakcije studenata na forumu [20]. Alat se temelji na jezicima JavaScript i Perl, te generira vizualizacije u pregledniku korištenjem SVG tehnologije. Razgovori studenata i nastavnika na forumu prikazuju se kao mapa koncentričnih i povezanih kružnica različitih boja, a omogućeno je i zumiranje kako bi se dobilo više informacija o pojedinom elementu grafikona (temi, autoru (student, nastavnik), broju pogleda i sl.). U radu se također daje opsežan pregled postojećih alata ili modula za vizualizaciju podataka u pregledniku, te daje usporedba njihovih karakteristika.

IV. PRILAGODBA NASTAVNIH MATERIJALA

Veći broj autora koristio je EDM kako bi pokušao poboljšati sustave za e-učenje sugeriranjem potrebnih izmjena u mreži povezanih nastavnih materijala, njihovom sadržaju ili sa njima povezanim provjerama (testovi/kvizovi).

U jednom od ranijih radova, iz 2001. godine, navode se reference na nekolicinu radova u razdoblju od '90. –'99. u kojima su analizirane interakcije korisnika sa sustavima za e-učenje korištenjem metoda RP. Na temelju rezultata identificirani su pojmovi ili nastavni materijali sa kojima su učenici imali najviše problema. U samom radu predloženo je adaptivno okruženje za e-učenje koje može ponuditi različite nastavne materijale za različite studente prema njihovim sposobnostima i rezultatima [21].

Jedan od citiranih radova u ovom području, iz 2004. nastao je na temelju istraživanja u sklopu doktorske disertacije. U radu se opisuje primjena raznih algoritama RP i genetičkog programiranja za otkrivanje veza između pojedinih elemenata sadržaja, te generiranje preporuka za poboljšanje te strukture [22].

Rudarenje pravila nad rezultatima pisanja testova u sustavu za učenje nakon čega se pravila pretvaraju u veze sa preduvjetima za stvaranje mape koncepata opisano je u [23]. Pravilima se dodaju dodatni atributi kako bi se bolje istaknule veze između koncepata, nakon čega se

primjenjuje heuristični algoritam za automatsku izradu mape koncepata. Zanimljivost ovog rada je što izrađene mape prikazuje u web pregledniku (generiranjem statičnih slika). Kada je mapa generirana nalizira se redundantnost i cirkularnost mape.

U radu iz 2008. godine [24] autori koriste Weka alat kroz njegovu aplikativno programsko sučelje (engl. *application programming interface*, API). Koriste se ugrađeni Weka algoritmi za klasifikaciju i rudarenje pravila koji otkrivaju probleme u svladavanju gradiva. Analiza je olakšana primjenom njihovog *Simulog* alata koji prati korisnike i izrađuje log zapise prilagođene Weka sustavu. Izrađen je alat pod nazivom *A²* (kao samostalna Java aplikacija koja poziva Weku) koji na temelju otkrivenih pravila sugerira nastavniku problematične vježbe ili lekcije.

U radu iz 2009. godine autori opisuju sustav koji koristi suradnju nastavnika i eksperata u otkrivanju najboljih preporuka za poboljšanje sadržaja ili ispita [25]. Sustav se sastoji od klijentske (Java) aplikacije za otkrivanje pravila na temelju podataka o korištenju sustava za e-učenje i polaganja *online* testova. Autori su koristili posebnu verziju Apriori algoritma kako bi pojednostavili rad nastavnicima. Sustav ima i web aplikaciju na kojoj se sakupljaju pronađena pravila, gdje im eksperti mogu dati ocjene. Kolaborativnim pristupom odabiru se (i u sustave za e-učenje ugrađuju) najbolja pravila.

Apriori algoritam, ali i broje druge tehnike RP korištene su u radu iz 2012. godine [26]. U radu se opisuje rudarenje asocijativnih pravila kojima se dobiva uvid u ponašanje korisnika, konkretno korištenja nastavnih materijala. Iz pronađenih pravila moguće je uočiti problematične dijelove strukture znanja, pogreške u rasporedu nastavnih materijala, testova i sl. na temelju čega nastavnik može reagirati i ispraviti pogrešku.

V. STANJE U RH I SUSJEDNIM ZEMLJAMA

Pretragom poznatih baza znanstvenih radova i popisa magistarskih radova i doktorskih disertacija objavljenih u RH (vrij. objav. > 2009.), pokušali smo ustvrditi stanje istraživanja u području primjene RP u sferi obrazovanja. Pronađen je vrlo mali broj članaka hrvatskih i slovenskih autora, dok se tim područjem intenzivnije bave u Srbiji. Na Fakultetu elektrotehnike i računarstva su se 2007. godine bavili primjenom metoda RP u obrazovnom okruženju [27]. Godine 2010. u Splitu je objavljen rad „*Student dropout analysis with application of data mining methods*“ [28]. Slovenski autori su se 2002. godine bavili iskoristivošću rudarenja teksta u izdavaštvu i obrazovanju [29]. Skupina autora sa Sveučilišta u Beogradu je 2007. godine, sa kolegama iz Kanade započela rad na razvoju sustava koji je koristio metode RP za potrebe analize podataka iz sustava za e-učenje [30], koji je nastavljen u ranije spomenutim radovima [17] i [18]. Druga skupina autora sa istog Sveučilišta bavila se 2009. godine problemom stvaranja prilagodljivog sustava za e-učenje uz pomoć metoda klasifikacije i izrade modela studenta [31]. Metodama klasifikacije u analizi uzoraka u ponašanju korisnika [32], te razvojem cjelovitog sustava

¹¹ <https://tekri.athabasca.ca/analytics/>

za e-učenje podržanog metodama RP [33] bavili su se na Tehničkom fakultetu u Čačku 2011. i 2013. godine.

U popisima magistarskih radova i doktorskih disertacija u RH pronašli smo jednu doktorsku disertaciju koja bi se mogla povezati uz ovo područje. Disertacija pod naslovom „Inteligentne metode u prilagodbi sustava elektroničkog učenja preferencama i stilovima učenja studenata“ [34] obranjena je na FOI, Varaždin, 2010. Godine. Rad opisuje sustav u kojem je primjenjena neuronska mreža za prepoznavanja preferencija studenata, te prilagodbu prikaza nastavnih materijala.

VI. ZAKLJUČAK I SMJERNICE ZA BUDUĆA ISTRAŽIVANJA

U ovom radu predstavljeno je područje koje se naziva *Educational Data Mining* u kojem se dobro poznate i provjerene metode rudarenja podataka primjenjuju na sustave za e-učenje. U počecima su se spomenute metode koristile sa ciljem davanja uvida u rezultate, načine korištenja sustava od strane učenika i predviđanja uspjeha, da bi se s vremenom počele prilagođavati i koristiti za podršku adaptivnosti takvih sustava. Posljednjih godina razvijaju se novi, originalni pristupi analizi podataka sustava za e-učenje kao što je otkrivanje znanja uz pomoć modela. U radu su istaknuti ključni pregledni radovi, u kojima su naglašeni trendovi u smjerovima razvoja ovog područja. Kroz veći broj radova, kao jedan od glavnih izazova provlači se potreba za približavanjem alata i metoda RP laicima - u ovom slučaju učiteljima na različitim razinama obrazovnog sustava, kao i potreba za integracijom spomenutih alata u sustave za e-učenje. Nakon pregleda područja fokusirali smo se na dva segmenta istraživanja koja smatramo značajnima za ostvarivanje spomenutih izazova - vizualizaciju podataka (dobivenih metodama RP) u web preglednicima, te mogućnosti prilagodbe strukture nastavnih materijala (zasnovane na metodama RP). Dan je prikaz onih radova iz oba segmenta koji se bave temema srodnima našim istraživačkim interesima, sa posebnim naglaskom na najnovim istraživanjima. Iako su spomenuti radovi istraživali brojne srodne teme, smatramo da postoji prostor za istraživanje direktne integracije sustava za e-učenje i alata za RP, kroz sučelje prilagođeno nastavnicima, a uz primjenu najmodernih tehnologija za vizualizaciju podataka. Takva integracija omogućila bi direktnu povratnu vezu između aktivnosti korisnika, analize i intervencije u sustav s ciljem poboljšavanja sustava, a time i rezultata (usvojenosti znanja) korisnika. Posebno interesantnim smatramo područje vizualizacije putanja korisnika kroz strukturu znanja, te povezivanje istih sa pojedinim skupinama korisnika, s ciljem poboljšanja strukture znanja i povezanih nastavnih materijala. S obzirom da se radi o web sustavima, lako su moguće intervencije eksperta kroz prilagođeno napredno sučelje. Na kraju rada, dali smo pregled istraživanja u RH i okolnim zemljama. U RH objavljen je veći broj magistarskih radova koji su primjenjivali metode RP, ali pretežno u ekonomskim i tehničkim znanostima, dok je broj radova u društvenim znanostima, odnosno obrazovanju, vrlo malen.

REFERENCE

- [1] Witten, I. H., Eibe, F., Hall M. A.; Data Mining: practical machine learning tools and techniques, Morgan Kaufmann, USA, 2011
- [2] Romero C., Ventura S.; Educational Data Mining: A survey from 1995 to 2005; Expert Systems with Applications; Vol 33, 2007., pp 135-146.
- [3] Brusilovsky, P., Peylo, C.; Adaptive and intelligent web-based educational systems, International Journal of Artificial Intelligence in Educations, 2005., pp. 153-160.
- [4] Castro F., Vellido A., Nebot A., Mugica F.; Applying Data Mining Techniques to e-Learning Problems; Evolution of Teaching and Learning Paradigms in Intelligent Environment - Studies in Computational Intelligence; Springer Berlin Heidelberg; 2007; pp. 183-221
- [5] Baker R.S.J.d., Yacef K.; The State of Educational Data Mining in 2009: A review and Future Visions; Journal of Educational Data Mining, Vol. 1, Issue 1; 2009., pp. 3-17
- [6] Beck, J.E., Mostow, J.; How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students, Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 2008., pp. 353-362.
- [7] Cocea, M., Hershkovitz, A., Baker, R.S.J.D., The Impact of Off.task and Gaming Behaviors on Learning: Immediate or Aggregate?, Proceedings of the 14th International Conference on Artificial Intelligence in Education, 2009., pp. 507-514.
- [8] Jeong, H., Biswas, G., Mining Student Behavior Models in Learning-by-Teaching Environments, Proceedings of the 1st International Conference on Educational Data Mining, 2008., pp. 127-136.
- [9] Romero, C., Ventura, S.; Educational Data Mining: A Review of the State-of-the-Art, Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews; Vol. 40, Issue: 6, pp. 601-618.
- [10] Romero C., Ventura S., Pechenizkiy M., Baker R.S.J.d.; Handbook of Educational Data Mining, CRC Press, USA, 2010, Chapter 12.
- [11] Liu F. , Shih B.; Application of Data-Mining Technology on E-Learning Material Recommendation, E-learning Experiences and Future, Safeullah Soomro (Ed.), ISBN: 978-953-307-092-6, InTech, DOI: 10.5772/8816. (2010).
- [12] Johnson, L., Levine, A., Smith, R., & Stone, S.; The 2010 Horizon Report., Austin, Texas: The New Media Consortium, 2010.
- [13] Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., Ludgate, H.; NMC Horizon Report: 2013 Higher Education Edition., Austin, Texas: The New Media Consortium, 2013.
- [14] Geryk, J.; Applied Visual Analytics for Higher Education Decision Making, PhD Thesis, Masaryk University, Faculty of Informatics, Czech Republic, 2012
- [15] Koelling, J.; Langenkaemper, D.; Abouna, S.; et al. WHIDE-a web tool for visual data mining colocation patterns in multivariate bioimages, Bioinformatics, Vol. 28, Issue: 8, pp. 1143-1150.

- [16] Zorrilla, M., García-Saiz, D., A service oriented architecture to provide data mining services for non-expert data miners, *Decision Support Systems*, Vol. 55, Issue 1, 2013., pp. 399-411.
- [17] Liaqat, A., Hatala, M., Gašević, D., Jovanović, J.; A qualitative evaluation of evolution of a learning analytics tool, *Computers & Education*, Vol. 58, Issue 1, 2012., pp. 470-489.
- [18] Liaqat, A., Mohsen, A., Gašević, D., Jovanović, J., Hatala, M., Factors influencing beliefs for adoption of a learning analytics tool: An empirical study, *Computers & Education*, Vol. 62, 2013., pp. 130-148.
- [19] U.S. Department of Education; Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief., Center for Technology in Learning SRI International, USA, 2012.
- [20] Jyothi, S., McAvinia, C., Keating, J.; A visualisation tool to aid exploration of students' interactions in asynchronous online communication, *Computers & Education*, Vol. 58, Issue 1, 2012., pp. 30-42.
- [21] Tsai, C.J., Tseng, S.S., Lin, C.Y.; A Two-phase fuzzy mining and learning algorithm for adaptive learning environment. In *International Conference on Computational Science*, San Francisco, 2001., pp. 429-438.
- [22] Romero, C., Ventura, S., De Bra, P.; Knowledge discovery with genetic programming for providing feedback to courseware author. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research* 14, 2004., pp. 425-464.
- [23] Tseng, S.S., Sue, P.C., Su, J.M., Weng, J.F., Tsai, W.N.; A new approach for constructing the concept map. In *Computers & Education Journal*, 49, 2007., pp. 691-707.
- [24] Vialardi, C., Bravo, J., Ortigosa, A.; Improving AEH courses through log analysis, In *Journal of Universal Computers Science*, 14 (17), 2008., pp. 2777-1798.
- [25] Garcia, E., Romero, C., Ventura, S., Castro, C.; Collaborative Data Mining Tool for Education. In *International Conference on Educational Data Mining*, Cordoba, Spain, 2009b., pp. 299-306.
- [26] Kinnebrew J.S., Biswas G.; Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution, *Proceedings of the 5th International Conference on Educational Data Mining*, 2012, pp. 57-64.
- [27] Vranic, M., Pintar, D., Skocir, Z.; The use of data mining in education environment, In *International Conference on Telecommunications*, Zagreb, 2007., pp. 243-250.
- [28] Jadrić, M., Grača, Ž., Čukušić, M.; Student dropout analysis with application of data mining methods, *Business intelligence in South-East Europe, Management*, Split, 15, 2010., pp. 31-46.
- [29] Grobelnik, Mladenec, Jermol, Exploiting text mining in publishing and education, In *Proceedings of the ICML2002 workshop on data mining lessons learned*, 2002., pp. 34-39.
- [30] Jovanovic, J., Gasevic D, Brooks, Devedzic, Hatala; LOCO-Analyst : A tool for raising teacher's awareness in online learning environments, In *European Conference on Technology Enhanced Learning*, Crete, 2007., pp. 112-126.
- [31] Radenković, B., Despotović, M., Bogdanović, Z., Barać, D.; Creating Adaptive Environment for e-Learning Courses, *JIOS*, Vol. 33, No. 1, 2009., pp. 179-189.
- [32] Blagojević Marija, "The application of clustering in the analysis of users' behavior patterns", *Technics Technologies Education Management* Volume 6, Number 4, (2011) pp. 1327-1332.
- [33] Blagojević Marija, Micić Živadina, "A web-based intelligent report e-learning system using data mining techniques", *Computers and Electrical Engineering*, Volume 39, Issue 2, (2013) pp. 465-474.
- [34] Đurđević, I.; *Inteligentne metode u prilagodbi sustava elektroničkog učenja preferencama i stilovima učenja studenata*, Sveučilište u Zagrebu, Fakultet organizacije i informatike Varaždin, 2010.