



TECHNICS AND INFORMATICS IN EDUCATION

6th International Conference, Faculty of Technical Sciences, Čačak, Serbia, 28–29th May 2016

TEHNIKA I INFORMATIKA U OBRAZOVANJU

6. međunarodna konferencija, Fakultet tehničkih nauka, Čačak, Srbija, 28–29. maj 2016.

UDK: 004:316.774]:378.4

Naučni rad u formi prethodnog saopštenja

Neke od primena tekst analitike u institucijama visokoškolskog obrazovanja

Predrag Stolić¹, Snežana Stolić i Aleksandra Milosavljević²

¹ Tehnički fakultet u Boru, Bor, Srbija

² Institut za rudarstvo i metalurgiju Bor, Bor, Srbija

e-mail pstolic@tfbor.bg.ac.rs

Rezime: *Dosadašnja znanja o okruženju u kojima neka institucija vrši rad zasnivala su se na upotrebi strukturiranih podataka. Zadnjih godina prediktivna analiza sve više upotrebljava nestruktuirane podatke. Uticaj nestruktuiranih podataka dominantan je u okviru raznih vrsta tekstova, te je jedan od vodećih pristupa u analizi i dobijanju neophodnih informacija zasnovan na upotrebi tekst analitike. U radu je prikazana realna upotrebna vrednost tekst analitike uz poseban akcenat na neke od mogućih aspekata upotrebe tekst analitike u prostoru rada visokoškolskih institucija.*

Ključne reči: *tekst analitika; otkrivanje podataka iz teksta; visokoškolsko obrazovanje*

1. UVOD

Zajednička karakteristika svih visokoškolskih institucija u svetu jeste realizacija studijskih programa u vrlo dinamičnom okruženju. Spomenuta dinamičnost nastaje kao rezultat kombinacije raznih parametara, počev od socioloških, preko tehničkih i tehnoloških pa sve do ekonomskih. Za razliku od visokoškolskog obrazovanja u nekom tradicionalnom smislu gde su sva unapređenja vršena periodično i sa određenim kašnjenjem usled tadašnjih tendencija koje je iskazivalo društvo, današnje, savremeno, visokoškolsko obrazovanje podrazumeva kontinuirano unapređenje u svim segmentima svog procesa u cilju zadržavanja konkurentnosti.

Savremeno visokoškolsko obrazovanje pred sebe postavlja cilj modernizacije studijskih programa kao kontinuiranog procesa kojim se obezbeđuje izvrsnost u ostvarivanju očekivanih ishoda učenja i potrebnih kompetencija budućih svršenih studenata na svim nivoima akademskih i strukovnih studija [1]. Kako se školovanje studenata izvodi u cilju osposobljavanja studenata za buduće poslove koji će biti obavljani u okvirima stručne osposobljenosti, moderni studijski programi moraju biti koncipirani tako da odgovore zahtevima koje pred njih postavlja aktuelno tržište rada u ekonomskom, tehničko-tehnološkom, socijalnom i naučnom smislu. Kako bi se postigli zadati ciljevi u pogledu usklađenosti studijskih programa sa trenutnim tržištem, u procese kreiranja, oblikovanja, reforme, verifikacije i validacije studijskih programa moraju biti uključeni hijerarhijski svi relevantni državni organi, naučno-istraživačka zajednica, zaposleni u okviru visokoškolskih institucija, ali i studenti i kao jedna jako važna komponenta oni koji će apsorbovati krajnji produkt visokoškolskog rada, a to su krajnji korisnici, odnosno poslodavci.

Izvori podataka koji se mogu iskoristiti u prethodnim analizama u cilju ostvarenja konkurentnih studijskih programa na svim nivoima studija nalaze se svuda u okviru eko sistema u kojem visokoškolske institucije ostvaruju svoj rad. Postavlja se pitanje kako izdvojiti relevantne i upotrebljive podatke iz velike količine podataka koji se generišu svakodnevno velikom brzinom i kako te podatke obraditi u najkraćem mogućem vremenskom intervalu kako bi bila izvršena pravovremena analiza u cilju obezbeđenja instrumenata koji bi definisali nove pravce razvoja i unapređenja studijskih programa. Tradicionalne analize i aparati koji su do skora korišćeni davali su određene rezultate ali se ipak primećuje izvesno kašnjenje u adekvatnim reakcijama na sve promenljivije tržište rada. Zadnjih godina u okviru razvoja informaciono-komunikacionih tehnologija pojavio se skup tehnika poznat pod imenom Big Data koji omogućava obradu velike količine informacija u razumnom vremenskom intervalu uz adekvatno prepoznavanje relevantnih izvora i podataka, pravovremenu analizu i odgovarajuće prezentovanje rezultata izvršenih analitičkih radnji. U poglavljima koja slede biće predstavljena jedna od njih pod imenom tekst analitika (text analytics) i mogućnosti njene primene na unapređenje rada institucija u okviru visokoškolskog obrazovanja.

2. NESTRUKTURIRANI PODACI

Godinama unazad svet informacionih sistema i informacionih tehnologija oslanjao se na strogo strukturane podatke pohranjene u bazama podataka. Gotovo sve od podataka sa kojima su se računarski sistemi susretali, počev od tzv. "sirovih podataka" koji su bili rezultat neke od vrsti akvizicije podataka, pa sve do rezultata obrade nad tim podacima, bez obzira na vrstu, upotrebljivost i količinu, bilo je podređeno nekom od sistema za upravljanje bazama podataka. Pri tome, najzastupljeniji vid baza podataka bile su relacione baze podataka, zasnovane na dobro poznatim matematičkim postulatima relacija i skupova. Napretkom informacionih tehnologija, pristup zasnovan na relacionim bazama podataka pokazao se kao vrlo ograničavajući i u pojedinim momentima neadekvatan za savremene izazove koji svet novih informacionih tehnologija donosi pred korisnike u bližoj i daljoj budućnosti [2].

U skladu sa napred navedenim nasuprot strogo strukturanim podacima, započet je i proces korišćenja nestrukturiranih i polustrukturiranih podataka. Može se reći da su za ekspanziju upotrebe nestrukturiranih podataka, umesto tradicionalnih strukturiranih, zaslužna tri, međusobno izuzetno spregnuta, bitna faktora: masovne računarske komunikacije, eksplozija upotrebe socijalnih mreža i težnja čoveka da se u svetu ova prethodna dva izražava na način blizak svakodnevnom izražavanju, odnosno svom prirodnom govoru. Podaci se generišu ogromnom brzinom uz javljanje ekstremnog broja izvora podataka (blogovi, komentari na sajtovima, socijalne mreže, razni dokumenti u cloudu i slično), ali se, zbog nestrukturiranog načina predstavljanja, većina podataka čine nerelevantnim ili neadekvatnim za dalji proces obrade nad njima.

Detaljnijom analizom može se utvrditi da odgovarajući skupovi nestrukturiranih podataka, za koje se smatra da nose irelevantne podatke, mogu dovesti do značajnih, relevantnih, otkrića u zavisnosti od načina na koji se podaci tretiraju. Stoga je neophodno izvršiti promenu pristupa prilikom analize podataka. Tradicionalne tehnike i tehnologije korišćene u analizi i obradi podataka moraju se napustiti u celosti ili izvršiti njihova modifikacija kako bi se omogućilo dejstvo i na domen nestrukturiranih podataka.

3. OTKRIVANJE PODATAKA IZ TEKSTA (TEXT DATA MINING) I TEKST ANALITIKA (TEXT ANALYTICS)

Ekspanzijom upotrebe informaciono-komunikacionih tehnologija u obavljanju zadataka namenjenih modernom poslovanju, tekstovi raznih vrsta, oblika i dužina počinju sve više da dobijaju na značaju kao nosioci informacija u vidu raznih izveštaja, dopisa, memoranduma, ali i čitavih brošura, uputstava. Zato je prepoznata esencijalna vrednost mogućnosti dobijanja podataka iz raznih vrsta tekstova i počelo je da se traga za metodama za realizovanje pomenutog.

U početku ekstrakcija podataka iz teksta sastojala se u prostom brojanju reči, odnosno pojave zadate reči u nekom tekstu. Vremenom su se zadaci usložnjavali i izvodile su se kompleksnije analitičke operacije nad samim tekstem što je dovelo do stvaranja nove računarske discipline označene kao text data mining (često se sreće i termin text mining). Text data mining označava otkrivanje podataka iz teksta odnosno procesiranje nestruktuiranih tekstualnih informacija [3] u cilju mogućnosti upotrebe informacija iz teksta u daljim tokovima obrade podataka poput upisa odgovarajućih podataka u baze podataka, obrade u statističkim i machine learning algoritmima i sličnom [4]. Tipični zadaci koji se ostvaruju putem text data mininga jesu razne vrste marketinških analiza kroz obradu anketa i formulara, automatska klasifikacija teksta poput automatske detekcije neželjenih poruka (spam messages) [5], analiza ponašanja korisnika na društvenim mrežama putem sentimenata [6], analiza informacija objavljenih na sajtovima, blogovima, stranicama na nekoj od društvenih mreža i slično [7].

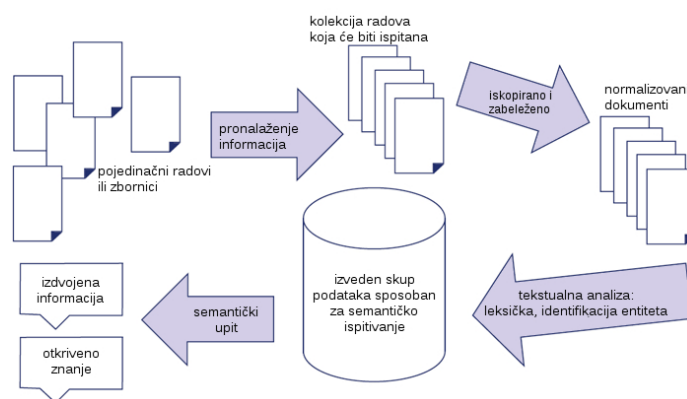
Tekst analitika označava objedinjen skup lingvističkih, statističkih metoda i metoda namenjenih mašinskom učenju koje imaju za cilj ekstrakciju informacija iz tekstualnih izvora. Dobijene informacije, upotrebom tekst analitike, namenjene su korišćenju u okviru Big Data sistema, poslovnoj inteligenciji, raznim statističkim analizama i obradama, predikciji ponašanja i budućih dešavanja [8]. Danas u upotrebi prevladava termin tekst analitika (text analytics) kao sinonim za text data mining iako postoje određene razlike u upotrebi ova dva termina, međutim te razlike su sve manje, a zajedničke osobine su sve više izražene, pa se tekst analitika u svetu modernog poslovanja sve više ističe kao dominantan pojam.

4. TEKST ANALITIKA ZA POTREBE VISOKOŠKOLSKIH INSTITUCIJA

4.1. Detekcija potencijalnih falsifikata

Upotreba tekst analitike u okvirima visokoškolskih ustanova u svetu i kod nas zasniva se prevashodno na analizi raznih vrsta dokumenata (naučni radovi, diplomski radovi, doktorske teze i slično) u cilju otkrivanja potencijalnih falsifikata i njihovih autora. Ako uzmemo u obzir da je, nekoliko godina unazad, prosek publikovanja naučnih radova godišnje nekoliko miliona na svetskom nivou i da je većina tih radova danas dostupno na neki od načina u tekstualnom formatu, upotreba tekst analitike u ovom segmentu predstavlja obaveznu kariku u celom procesu [9,10]. Na taj način jača se integritet institucija kroz podizanje kvaliteta naučnih radova i smanjenje pojave mogućih falsifikata.

Na sl. 1 dat je šematski prikaz rada sistema koji koristi tekst analitiku u za analizu naučnih sadržaja u svrhu detekcije mogućih falsifikata. Ovakav sistem bi se mogao u celosti primeniti i prilikom analize sadržaja udžbenika i ostalih pratećih materijala koji se koriste u okviru realizacije studijskih programa. Međutim, zbog zakona koji važe u oblasti autorskog i srodnog prava sistem bi bio limitiran u svom radu jer bi mogao da analizira samo ona izdanja koja nisu zaštićena autorskim pravima, ili bi mogao da analizira samo delove izdanja koja je izdavač učinio javno dostupnim. Time bi funkcionalnost sistema bila dovedena u pitanje, te se ovakvi sistemi za sada još uvek retko susreću u domenu analize udžbenika i ostale slične literature.



Slika 1. Šematski prikaz procesa uključених u tekst analitiku naučnih sadržaja [9]

4.2. Analiza tržišnog okruženja

Današnje visokoškolske institucije ostvaruju svoju delatnost u uslovima vrlo promenljivog tržišnog okruženja. Spomenute promene su izuzetno dinamične, te je potrebno razviti mehanizme praćenja i identifikovanja istih kako bi se kadrovi, nakon svršetka školovanja na spomenutim institucijama, mogli uklopiti u odgovarajuće tržišne trendove. Tradicionalne tehnike praćenja i segmentisanja tržišta za potrebe projekcija i usklađivanja visokoškolskih programa daju određene rezultate, ali je prisutan veliki stepen inertnosti koji se u krajnjem cilju manifestuje da nakon primene dobijenih rezultata na visokoškolske programe u smislu njihovog osavremenjivanja i usklađivanja, ponovo imamo određenu dozu zastarelosti i neusklađenosti u odnosu na aktuelnu situaciju na tržištu.

Primenom tekst analitike u domenu analize tržišnog okruženja vrši se minimizacija pomenutih problema karakterističnih za tradicionalni pristup. Neutralisanje navedenih činilaca vrši se kroz pretvaranje tekstualnih podataka u podatke od izuzetne organizacione i poslovne vrednosti kojima se može automatski pristupati a potom izvršiti analiza i tumačenje, odnosno razumevanje istih i delovati u skladu sa informacijama otkrivenim iz elektronskih tekstova koji su predmet pristupanja i analize [11]. Ovakav pristup čini da se sada, prilikom raznih analiza, u svakodnevnom smislu upotrebe, mogu iskoristiti resursi društvenih mreža, elektronskih upitnika, komunikacije izvršene putem elektronske pošte, ali i skriveni resursi u vidu pohranjenih informacija u portfolijima kompanija, oglasima za posao, berzama rada i sličnim domenima koji u tradicionalnom smislu nisu otkrivali mnogo ili su otkrivali tek u nekom indirektnom kontaktu sa njima [12]. Sada se ovim resursima tekst analitikom direktno

pristupa i naizgled skrivene informacije postaju vredno oruđe u identifikaciji svih prednosti, ali i potencijalnih problema koje promene na tržištu mogu uneti u sferu visokoškolskog obrazovanja.

4.3. Praćenje zadovoljstva studenata

Nekoliko godina unazad postoji izraženo fokusiranje visokoškolvstva u Republici Srbiji na praćenje zadovoljstva studenata, a naročito nakon formiranja Komisije za akreditaciju i proveru kvaliteta. Razne vrste anketa, eksternih evaluacija i samoevaluacija postale su ključni segment i alat unapređenja rada visokoškolskih institucija. Međutim, u praksi su se pojavili i realni problemi u vidu dobijanja podataka čija je pouzdanost i dalja upotrebljivost diskutabilna. Na primer prilikom evaluacije predmetnih nastavnika pojavila se izražena sklonost studenata da će nižu ocenu dobiti nastavnik koji je fokusiran na disciplinu, a nasuprot tome visoku ocenu dobiće nastavnik koji izuzetno toleriše česte izostanke sa predavanja, kašnjenja i slično. Studenti i dalje izražavaju sumnju u anonimnost svakog ispitivanja koje se sprovodi od strane visokoškolskih institucija.

Nasuprot tome, studenti izražavaju sklonost da objektivnije ocenjuju stanje i rad visokoškolske institucije u okviru manje formalne komunikacije koja se vodi na primer po raznim forumima i segmentima društvenih mreža. Smatra se da upravo ovakav vid podataka i informacija mogu dovesti do realnijeg i sveobuhvatnijeg sagledavanja svih potencijalnih problema sa kojima se studenti susreću u okviru visokoškolskih institucija, a samim tim nakon identifikacije mogu se ostvariti i odgovarajući mehanizmi u cilju unapređenja rada same institucije [13-15]. U ovom slučaju tekst analitika predstavlja način za procesiranje ovakvih informacija u cilju izdvajanja većeg skupa podataka i znanja o zadovoljstvu studenata i dobijanja jedne autentičnije i kvalitetnije slike o stanju kako u visokoškolskim institucijama, tako i u visokoškolvstvu u globalu [16].

Na sl. 2 dat je uprošćeni (generalizovani) šematski prikaz faza koje tekst analitika ostvaruje prilikom analize zadovoljstva studenata na osnovu određenog izdvojenog teksta bez obzira na izvor samog teksta (društvene mreže, forumi, email i slično).



Slika 2. Analiza mišljenja studenata korišćenjem tekst analitike

Određeni tekst se analizira i vrši se izdvajanje rečenica, a potom se pristupa daljoj analizi svake rečenice parcijalno kako bi se identifikovali svi delovi svake od dobijenih rečenica [17]. Nad dobijenim delovima vrši se sprovođenje odgovarajućih jasno definisanih analitičkih postupaka u cilju karakterisanja i vrednovanja svakog od delova [18]. Cilj analitičke obrade predstavlja izdvojeno mišljenje iz polaznog teksta. Ovakav postupak poznat je pod imenom analiza sentimenta (sentiment analysis) [12] i predstavlja jedan kompleksan postupak koji pored tekst analitike, kao okosnice samog postupka, obuhvata i procesiranje na prirodnom jeziku (natural language processing - NLP) i računarsku lingvistiku (computational linguistics).

Prednost ovakvog pristupa prezentovaćemo na kratkom primeru ocenjivanja predmeta. U okviru neke evaluacione ankete, predmet student može oceniti ocenom četiri (vrlo dobar). I takva ocena se može iskoristiti u svrhu daljih statističkih obrada i procena. Međutim, daleko

bolji zaključak o samom predmetu možemo dobiti recimo iz teksta o predmetu koji je napisao student na nekom forumu. Na primer, student je nekom drugom studentu objašnjavao na forumu osnove predmeta, šta je naučio, šta je sve probao, kako je predmet inovativan, kako je predmet moderan i slično i na kraju izlaganja napisao tekst slengom "Jednom rečju ovaj predmet je strava!!!". Upotrebom tekst analitike odavde možemo dobiti informaciju na primer da je student oduševljen predmetom, da ga je podstakao, da ga preporučuje drugim kolegama i čitav drugi skup relevantnih informacija. Kao što se iz navedenog vidi dobijamo ceo skup informacija kojima možemo meriti uspeh predmeta kroz zadovoljstvo studenta, nasuprot polaznoj jednoj oceni iz ankete (4 - vrlo dobar).

5. CONCLUSION

Institucije visokoškolskog obrazovanja svoju delatnost ostvaruju u uslovima vrlo promenljivog i dinamičnog okruženja. Tradicionalne metode analize ne mogu u potpunosti ispratiti savremene trendove koji trenutno vladaju u domenu visokoškolskog obrazovanja i neposrednoj okolini koja ga okružuje. Kako bi se obezbedila dalja konkurentnost neophodno je uvesti savremene metode u sve aspekte rada visokoškolskih institucija.

Tekst analitika je metoda koja može dati ogroman doprinos u analizi rada visokoškolskih institucija. U radu je prikazano samo nekoliko aspekata upotrebe tekst analitike u domenu visokoškolskog obrazovanja, dok je primena praktično neograničena. Suštinski, kako čovek gotovo sve procesira na određenom jeziku, odnosno u vidu nekog teksta, tako i tekst analitika se može uobličiti u neophodan alat u raznim segmentima rada.

Upotreba tekst analitike u polju visokoškolskog obrazovanja biće u neprestanom porastu. Takva kretanja ogledaće se prvenstveno u upotrebi tekst analitike u pravovremenim analizama faktora koji utiču na sam rad visokoškolskih institucija, a tom pravovremenom analizom moći će se ostvariti predikcija budućih kretanja i formirati jasan odgovor na ta kretanja. Samim tim tekst analitika će kao alat budućnosti omogućiti bolju segmentaciju, bolje rezultate, bolje pozicioniranje visokoškolskih institucija.

Očekuje se da će se, uvođenjem tekst analitike u domen visokoškolskog obrazovanja, vreme potrebno za vršenje analiza redukovati značajno. Ekonomski posmatrano analize će postati prihvatljivije i pristupačnije različitim visokoškolskim institucijama. Organizaciono, analize će biti primenjive na svim nivoima organizacione strukture, od katedara u okviru fakulteta do senata univerziteta pod gotovo istovetnim uslovima.

LITERATURA

- [1] Ministarstvo prosvete i nauke Vlade Republike Srbije. (2012). *Strategija razvoja obrazovanja u Srbiji do 2020. godine*. Službeni glasnik Republike Srbije broj 107/2012. Belgrade, Serbia: Službeni glasnik.
- [2] Zaki, A. K. (2014). *NoSQL databases: new millennium database for big data, big users, cloud computing and its security challenges*. International Journal of Research in Engineering and Technology, Vol. 3(3), pp 403-409. doi: 10.15623/ijret.2014.0315080
- [3] He, W. (2013). *Examining students' online interaction in a live video streaming environment using data mining and text mining*. Computers in Human Behavior, Vol. 29(1), pp 90-102. doi: 10.1016/j.chb.2012.07.020
- [4] StatSoft, Inc. (2013). *Electronic Statistics Textbook*. Tulsa, Oklahoma, USA: StatSoft.
- [5] Manning, C.D., Raghavan, P. & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England, UK: Cambridge University Press.

- [6] Irfan, R. (2004). A Survey on Text Mining in Social Networks. *The Knowledge Engineering Review*, Volume 30, pp 157-170. doi: 10.1017/S0269888914000277
- [7] Bail, C. (2014). *The cultural environment: measuring culture with big data*. *Theory and Society*, Vol. 43(3-4), pp 465-482. doi: 10.1007/s11186-014-9216-5
- [8] Ellaway, R. H., Pusic, M. V., Galbraith, R. M. & Cameron, T. (2014). *Developing the role of big data and analytics in health professional education*. *Medical Teacher*, Vol. 36(3), pp 216-222. doi: 10.3109/0142159X.2014.874553
- [9] JISC. (2012). *The Value and Benefit of Text Mining to UK Further and Higher Education*. Bristol, England, UK: JISC.
- [10] Chen, N.-S., Kinshuk, Wei C.-W. & Chen H.-J. (2008). *Mining e-Learning domain concept map from academic articles*. *Computers & Education*, Vol. 50(3), pp 1009-1021. doi: 10.1016/j.compedu.2006.10.001
- [11] SAS. (2012). *SAS Text Analytics - Derive more value from unstructured content*. Cary, North Carolina, USA: SAS Institute Inc.
- [12] Chen, H., Chiang, R. H. L. & Storey V. C. (2012). *Business intelligence and analytics: from big data to big impact*. *MIS Quarterly*, Vol. 36(4), pp 1165-1188.
- [13] Picciano, A. G. (2012). *The Evolution of Big Data and Learning Analytics in American Higher Education*. *Journal of Asynchronous Learning Networks*, Vol. 16(3), pp 9-20.
- [14] Eynon, R. (2013). *The rise of Big Data: what does it mean for education, technology, and media research?*. *Learning, Media and Technology*, Vol. 38(3), pp 237-240, doi: 10.1080/17439884.2013.771783
- [15] Romero, C. & Ventura, S. (2013). *Data Mining in Education*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 3(1), pp 12-27. doi: 10.1002/widm.1075
- [16] Daniel, B. (2015). *Big Datas and analytics in higher education: Opportunities and challenges*. *British Journal of Educational Technology*, Vol. 46(5), pp 904-920. doi: 10.1111/bjet.12230
- [17] Abdous, M. & He, W. (2011). *Using text mining to uncover students' technology-related problems in live video streaming*. *British Journal of Educational Technology*, Vol. 42(1), pp 40-49. doi:10.1111/j.1467-8535.2009.00980.x
- [18] Xu, Y. & Reynolds, N. (2012). *Using Text Mining Techniques to Analyze Students' Written Responses to a Teacher Leadership Dilemma*. *International Journal of Computer Theory and Engineering*, Vol. 4(4), pp 575-578. doi: 10.7763/IJCTE.2012.V4.535